

Anaesthesist 2019 · 68 (Suppl 1):S15–S24
 DOI 10.1007/s00101-017-0352-y
 Received: 8 January 2017
 Revised: 30 March 2017
 Accepted: 7 June 2017
 Published online: 10 August 2017
 © Springer Medizin Verlag GmbH 2017



A. R. Heller¹ · N. Salvador¹ · M. Frank³ · J. Schiffner⁴ · R. Kipke⁵ · C. Kleber²

¹ Department of Anesthesiology and Critical Care Medicine, Medical Faculty Carl Gustav Carus, TU-Dresden, Dresden, Germany

² University Center of Orthopedics and Trauma Surgery, TU-Dresden, Dresden, Germany

³ DRF Air Rescue Base “Christoph 38”, Dresden, Germany

⁴ Red Cross Saxonia, German Society of Disaster Medicine (DGKM) Division Mitteldeutschland, Dresden, Germany

⁵ EMS Teaching Unit, City of Dresden Fire Department and Disaster Preparedness, Dresden, Germany

Diagnostic precision of triage algorithms for mass casualty incidents. English version

Electronic supplementary material

The online version of this article (doi: [10.1007/s00101-017-0352-y](https://doi.org/10.1007/s00101-017-0352-y)) additionally contains Tables 6 and 7 on computer simulation of algorithms. The complete contribution and the supplementary material are available from the electronic full text archive at <http://www.springermedizin.de/der-anaesthesist>. You will find the supplementary material at the end of the contribution as “Supplementary Material”.

Background

Recent mass casualty incidents (MCI) [1–3] underline the key issue of correct assignment of triage categories according to the urgency of treatment and transportation (■ **Table 1**) for survival and patient quality of life [4]. The principle of triage of casualties emerged over several hundred years and is widely accepted [5, 6]. In preparing for the soccer world championship in 2006 the German authorities and medical societies in particular were involved with the issue of triage beginning in 2002 [7, 8].

This work contains data from the doctoral thesis of Nicholas Salvador, TU-Dresden.

English version of: Heller AR, Salvador N, Frank M et al. (2017) Diagnostische Güte von Vorsichtungsalgorithmien für den Massenansturz von Verletzten und Erkrankten. Anaesthesist 10.1007/s00101-017-0336-y.

Due to the shortage of emergency physicians on scene preliminary triage by non-physicians is widely accepted. Several consensus conferences were held on the issue and results were confirmed by the German Medical Association [7–10]. Preliminary triage is defined as [11]:

Immediate identification and indications of the most severely injured. It is a provisional conditional assessment, carried out by physicians and non-physicians, followed by a triage by a physician.

Preliminary triage may not in any case replace triage by a physician. Regarding treatment and transportation priorities time delay until definitive triage by a physician, however, may conditionally imply the first physician based triage to take place in hospital [8, 10]. In these cases preclinical decisions on treatment priorities will *de facto* be taken merely by non-physician personnel. For that purpose suitable algorithms have to be readily available [8, 11].

In recent decades various algorithms were developed to avoid deficits in treatment and to enable an early assessment at the scene [11–14]. Both over-triage and under-triage are well known limitations of all triage algorithms [7, 15–18]. While over-triage causes untargeted use of resources, under-triage prevents patient access to individually necessary treatment resources (■ **Table 2**). Both types of misclassification may be followed by acute

threats to the patients. Due to wastage of treatment resources over-triage is accompanied with a linear increase of mortality [22].

As a result of the 5th German Triage Consensus Conference a new preliminary triage algorithm “Primary Ranking for Initial Orientation in Emergency Medical Services” (PRIOR) was established under the leadership of the German Association of Disaster Medicine (DGKM). It was a particular goal to consider both trauma and non-trauma patients [12], of which the latter only were poorly represented in present algorithms. Furthermore, recent work on patient distribution in MCI suggested a higher proportion of non-trauma patients [19], and a lower percentage of severely injured than previously assumed. To date, triage algorithms are only poorly validated in a relevant cohort of patients in terms of quality and quantity [8, 11, 20]. The present study was designed and carried out to close this gap.

Methods

After institutional review board approval (EK DD 270 06 2015) 500 consecutive helicopter emergency medical service (HEMS) missions (“Christoph 38”, DRF Luftrettung Dresden) between 1 August 2014 and 31 December 2014 were included and were evaluated retrospectively. For this purpose both the

Table 1 Indications, description, and consequences of triage categories according to 6th Triage Consensus Conference 2015 [8]

Category	Description	Consequence
T1 (red)	Vital threat	Immediate treatment
T2 (yellow)	Severely injured/ill	Urgent treatment
T3 (green)	Slightly injured/ill	Delayed treatment
T4 (blue)	Without chance of survival	Palliative treatment
T0 (black)	Dead	–
<i>Additional indications on applying color tag</i>		
TP (red/yellow)	Transportation priority	Transportation priority
B (white)	Non-injured affected	Support
K	Contamination	Protective measures/decontamination

Table 2 Quality criteria of the triage for the patient allocation in the correct triage category T1

	T1 patient		Consequence for the patient		
	Yes	No			
Triage result T1	Yes	Inclusion correct positive	Over-triage false positive	Concurrence	Resources individually undersized
	No	Under-triage false negative	Exclusion correct negative	No admission	
Statistical problem	–	Sensitivity	Specificity	–	

Sensitivity proportion of T1 patients correctly classified as T1 by the test method (correct positive);

Specificity proportion of non-T1 patients correctly recognized as not T1 (correct negative)

electronic emergency documentation (MEDAT, DRF *mission record system*) and the hand-written records (DIVI-DOK Version EPRO 4.2) of each patient were consulted. Recent definitions of triage categories by the 6th Triage Consensus Conference of the German Federal Office of Civil Protection and Disaster Assistance (Table 1) served for patient classification [8].

To serve as a basis for comparison of the triage algorithms, appraisal and classification of patient records to triage categories was accomplished by 19 independent emergency physicians (anesthetists $n = 16$, trauma surgeons $n = 2$ and 1 general practitioner), including 6 onsite commanding physicians, 6 HEMS, and 7 Emergency Medical Service (EMS) physicians. After exclusion of 8 patients who had already passed away on HEMS arrival, 492 mission records were distributed over 3 triaging physicians each. Accordingly, every mission record was independently judged in triplicate without the use of any triage algorithm. Thus, 1476 triage processes were carried out. The level of experience of the triaging

physicians was regarded within the analysis: judgements of onsite commanding physicians received threefold weighting, HEMS physicians and EMS physicians with experience of at least 3 years twofold, and all others were single weighted. Such weighted means of the category allocation were rounded to a joint category when necessary, serving as a reference for calculating the precision measures of the triage algorithms at hand.

The triage algorithms PRIOR [12], modified simple triage and rapid treatment (mSTaRT) [14], field triage score (FTS) [21], Amberg-Schwandorf algorithm for pre-triage (ASAV), simple triage and rapid treatment (STaRT), Care Flight, and Triage Sieve [11] were programmed as query sequences over the established database (Excel, MS-Office Standard 2010, version 14.0, Microsoft, Redmond, WA), which delivered the individual triage category T1–T4 as well as the number of necessary steps until decision within the according algorithms for all patients (see online supplement). Exceptions were the algorithms PRIOR and FTS, which do not consider T4;

however, comparison of all algorithm-related precision measures in identifying the severely injured was enabled by classifying T1 and T4 together. The unified classification of T1 and T4 is particularly relevant for the practice of preliminary triage by non-physicians, because in Germany the T4 classification may not be used by non-physicians [8]. Moreover, it avoids systematic underestimation of the precision of those algorithms which support a possible T4 result [22].

Besides the categorical measures sensitivity (SE) and specificity (SP) the deviation of the cohorts at hand was analyzed with regard to over-triage and under-triage also concerning triage categories. In addition to the measures SE and SP as well as in analogy to the Bland-Altman depiction [23] the presentation of the mean difference from the reference triage category (ΔT) gives a lucid description of the deviations direction and magnitude (possible range of values -2 to $+2$). In this respect decimals of ΔT may not be interpreted as a precision of tenths in the determination of the triage category of a single patient, but with respect to the whole cohort at hand. For example, a ΔT of $+0.5$ means that every second patient is classified too high by one triage category, or every fourth patient is classified by two triage categories too high.

Table 2 shows the systematics and the consequences of the different error estimation possibilities. The classification of emergency medical management problems was performed according to the ABCDE rule [15]: Airway (A), breathing (B), circulation (C), neurological deficits (D), and exposure/environment (E). On the existence of several simultaneous problems, the guiding problem was always defined as the higher order problem, regarding “treat first what kills first” (i.e. A before B before C etc.) [15].

Statistical procedures

Descriptive statistics were created using the pivot table function of MS Excel. The quality of the triage procedures was determined by SE (true positives of all the corresponding T) and SP (true negatives

A. R. Heller · N. Salvador · M. Frank · J. Schiffner · R. Kipke · C. Kleber

Diagnostic precision of triage algorithms for mass casualty incidents. English version**Abstract**

Background. Regarding survival and quality of life recent mass casualty incidents again emphasize the importance of early identification of the correct degree of injury/illness to enable prioritization of treatment amongst patients and their transportation to an appropriate hospital. The present study investigated existing triage algorithms in terms of sensitivity (SE) and specificity (SP) as well as its process duration in a relevant emergency patient cohort.

Methods. In this study 500 consecutive air rescue missions were evaluated by means of standardized patient records. Classification of patients was accomplished by 19 emergency physicians. Every case was independently classified by at least 3 physicians without considering any triage algorithm. Existing triage algorithms Primary Ranking for Initial Orientation in Emergency Medical Services

(PRIOR), modified Simple Triage and Rapid Treatment (mSTaRT), Field Triage Score (FTS), Amberg-Schwandorf Algorithm for Triage (ASAV), Simple Triage and Rapid Treatment (STaRT), Care Flight, and Triage Sieve were additionally carried out computer based on each case, to enable calculation of quality criteria.

Results. The analyzed cohort had an age of (mean ± SD) 59 ± 25 years, a NACA score of 3.5 ± 1.1 and consisted of 57% men. On arrival 8 patients were deceased. Consequently, 492 patients were included in the analysis. The distribution of triage categories T1/T2/T3 were 10%/47%/43%, respectively. The highest diagnostic quality was achieved with START, mSTaRT, and ASAV yielding a SE of 78% and a SP ranging from 80–83%. The subgroup of surgical patients reached a SE of 95% and a SP between 85–91%. The newly established

algorithm PRIOR exerted a SE of 90% but merely a SP of 54% in the overall cohort thereby consuming the longest time for overall decision.

Conclusion. Triage procedures with acceptable diagnostic quality exist to identify the most severely injured. Due to its high rate of false positive results (over-triage) the recently developed PRIOR algorithm will cause overload of available resources for the severely injured within mass casualty incident missions. Non-surgical patients still are poorly identified by the available algorithms.

Keywords

Mass casualty Incident · Disaster management · Emergency preparedness · Triage

Diagnostische Güte von Vorsichtungs- und Triagealgorithmen für den Massenansturm von Verletzten und Erkrankten. Englische Version**Zusammenfassung**

Hintergrund. Jüngste Großschadenslagen verdeutlichen, wie entscheidend die korrekte Einstufung von Patienten entsprechend ihrer Behandlungs- und Transportdringlichkeit für das Überleben und die Lebensqualität ist. Diese Studie prüft internationale Vorsichtungsverfahren hinsichtlich Sensitivität (SE) und Spezifität (SP) sowie ihres Zeitbedarfs in einem Notfallmedizinisch relevanten Patientengut.

Methodik. Fünfhundert konsekutive Luftrettungseinsätze wurden anhand der Einsatzdokumentation ausgewertet. Die Eingruppierung der Patienten erfolgte interdisziplinär durch 19 Notfallmediziner. Jedes Protokoll wurde durch mindestens 3 Sichter unabhängig voneinander und ohne Verwendung von Vorsichtungs- und Triagealgorithmen nach Sichtungskategorie (SK) klassifiziert. Die Algorithmen der Vorsichtungsverfahren PRIOR (Primäres Ranking zur Initialen Orientierung im Rettungsdienst), mSTaRT („modified

Simple Triage and Rapid Treatment“), FTS („Field Triage Score“), ASAV (Amberg-Schwandorf Algorithmus für die Vorsichtung), STaRT („Simple Triage and Rapid Treatment“), Care Flight, Triage Sieve wurden für jeden Patienten computergestützt durchlaufen, um Testgütekriterien für alle Verfahren zu erheben.

Ergebnisse. Die Kohorte hatte ein Alter (MW ± SD) von 59 ± 25 Jahren, einen National Advisory Committee for Aeronautics (NACA) Score von 3,5 ± 1,1 und bestand zu 57 % aus Männern. Bei Eintreffen waren 8 Patienten bereits verstorben, sodass 492 Patienten in die weitere Analyse eingeschlossen wurden. Die prozentuale Verteilung der SK I/II/III war 10%/47%/43 %. Die höchste diagnostische Güte zeigten die Verfahren START, mSTaRT und ASAV mit 78%iger SE und einer SP von 80 bis 83 %. Dabei erreichten chirurgische Patienten eine SE von 95 % und eine SP von 85 bis 91 %. Das neu etablierte Verfahren

PRIOR hatte insgesamt bei einer SE von 90 % lediglich eine SP von 54 % und lässt den größten Gesamtaufwand erwarten.

Schlussfolgerung. Zur Identifikation der Schwerverletzten stehen etablierte Verfahren mit akzeptabler diagnostischer Güte zur Verfügung. Das PRIOR-Verfahren zeigte in unserer Studie eine hohe Rate falsch-positiver Ergebnisse (Übertriage), was im Einsatz zu einer Überbeanspruchung der Ressourcen für die Schwerverletzten führen könnte und damit zur Unterversorgung der richtig-positiven SK-I-Fälle. Internistische Krankheitsbilder werden nach wie vor als qualitativ ungenügend eingestuft.

Schlüsselwörter

Massenanfall von Verletzten · Katastrophenmedizin · Notfallvorsorge · Vorsichtungsverfahren

of all who do not belong to the respective T) (Table 2).

Normal distributed data are given as mean values and the standard error of means (SEM), which, if there is no overlapping, roughly indicate statistical sig-

nificance. Due to the multiple comparative possibilities and the resulting uncontrollable alpha error inflation, further inferential statistical analyses were omitted.

Results

Table 3 shows the demographic distribution of the analyzed cohort of 492 HEMS patients.

Table 3 Demographic data of the investigated cohort. Percentage of the total cohort $n = 492$ in brackets

	Surgical patients ($n = 194$)	Internal patients ($n = 298$)
Age \pm SEM (years)	49 \pm 2	65 \pm 1
Female	77 (15.7%)	135 (27.4%)
Male	117 (23.8%)	163 (33.1%)
T1 (red)	19 (3.9%)	31 (6.3%)
T2 (yellow)	93 (18.9%)	138 (28.0%)
T3 (green)	82 (16.7%)	129 (26.2%)
TP	35 (7.1%)	58 (11.8%)
NACA 1	4 (0.8%)	9 (1.8%)
NACA 2	13 (2.6%)	33 (6.7%)
NACA 3	110 (22.4%)	124 (25.2%)
NACA 4	30 (6.1%)	68 (13.8%)
NACA 5	35 (7.1%)	54 (11.0%)
NACA 6	1 (0.2%)	4 (0.8%)
NACA 7	1 (0.2%)	6 (1.2%)

NACA-Score National Advisory Committee for Aeronautics

Triage category and guiding emergency medical care problem (ABCDE)

The distribution of the reference triage category according to the individual emergency medical care problem (ABCDE) and the surgical or internal cause is depicted in **Fig. 1** and gives an impression of the weight of the individual key problems in the overall view. While every ABCDE problem [15] may justify the classification into the T1 in individual cases, **Fig. 1** clearly shows that a T1 assignment with a total of 10% of the cases analogous to Brüne [19] occurs comparatively rarely. Additionally, **Fig. 1** shows that the triage category correlates with the ABCDE key problem, that there are significant differences between surgical and internal patients and that the presence of B-E causes [15] itself does not certainly justify the classification into T1.

Number of discriminants until definitive classification into one triage category

The aim of triage procedures is to recognize the severely affected and injured as early as possible and correctly. Without consideration of the grouping accuracy by the respective algorithm, **Fig. 2** shows the number of process steps until

definitive classification into T1–T4. An assignment of patients to the T1 category by the PRIOR algorithm is therefore at least one step faster than with mSTaRT. It is particularly striking that the PRIOR algorithm requires the most algorithm steps for the identification of the most uncritical T3 (green) patients.

Test quality of triage algorithms

Depending on surgical or internal patients **Fig. 3** shows the decision-making quality of the triage algorithms for the identification of the severely injured or ill. The further a procedure is shown towards the bottom of **Fig. 3**, the more under-triage is associated with it (missing recognition of the corresponding patients, false negative). The further to the right a method is represented, the more it over-triages (wrong positive assignment to T1 (red)). This means that the most suitable methods are shown in the upper left. **Fig. 3** shows the merging of the algorithm results T1 (red) and T4 (blue), which are not distinguished in the triage by non-physicians in Germany. All comparison-relevant test-quality criteria of the methods are shown in **Table 4** regarding triage categories and patient cohorts.

In the same way as a patient's classification into the T4 category is not possible for a non-physician in Germany,

the death of a patient can also not be ascertained by a non-physician. Therefore, combined assessment of the severely injured and deceased (T1 + T4 + T0) is necessary for the critical appraisal of the suitability and quality of triage algorithms for use by non-physicians. Both the sensitivity of the identification of the severely injured and the specificity in the detection of T2 and T3 showed marginal deviations with or without consideration of the 8 of the 500 patients who were already dead on arrival. In that analysis the sensitivity in PRIOR increases from 90% to 91% in the T1 cases with unchanged specificity. The quality comparisons between the algorithms are ultimately not affected by the consideration of the deceased patients.

Fig. 4 shows the degree of false assignments based on the average deviation (ΔT) of the triage categories assigned by the algorithms from the reference categories. The clearest over-triage is shown in T3 patients with the PRIOR method, in particular in the case of internal patients (**Fig. 4**, right). A ΔT of + 1.1 for internal T3 patients can be interpreted in the way that each T3 patient is declared as a T2 patient by the PRIOR algorithm and in addition every tenth T3 patient even gets a classification into the T1. Conversely, the FTS with a ΔT of + 0.1 is best able to identify surgical T3 patients, since only one tenth of T3 patients are falsely classified into the T2; however, the problems arise in the FTS with its significant tendency to under-triage, in which FTS classifies 90% of the T1 surgical patients as T2.

Level of decision makers experience when triaging without the help of algorithms

Table 5 shows that both novice and advanced EMS physicians who are not active onsite commanding physicians are subject to an under-triage of 0.4 categories for T1 patients, and of 0.2 for T2 cases. In this context the slight underestimation in T2 patients and the marginal overestimation by 0.1 categories in T3 must be regarded as of subordinate clinical importance.

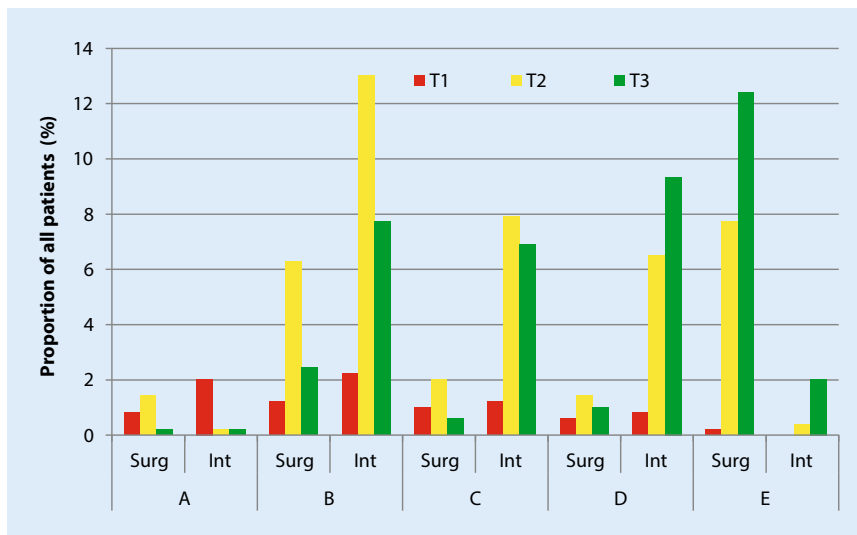


Fig. 1 ▲ Triage categories in surgical and in internal medical patients with respect to the leading problem: airway (A), breathing (B), circulation (C), neurological deficits (D), exposure/environment (E). Data given in percent of all patients

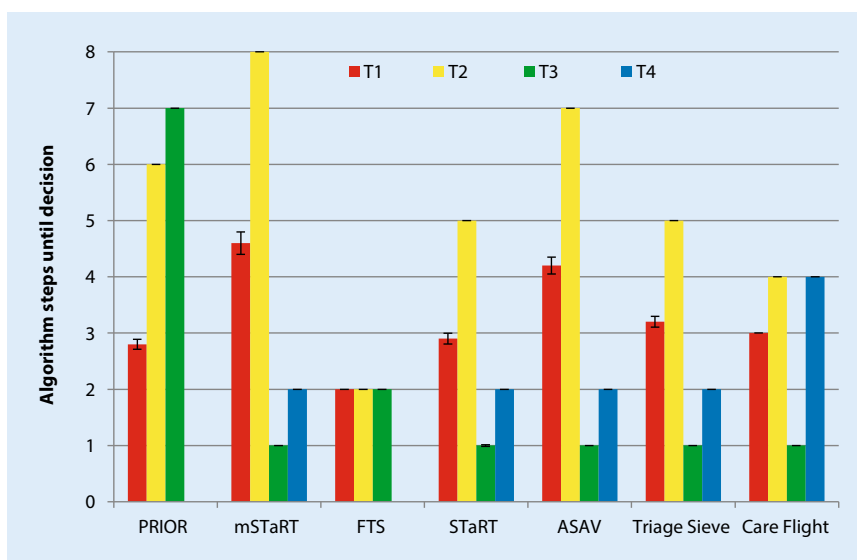


Fig. 2 ▲ Necessary steps (mean ± SEM) for each algorithm until the assignment to a triage category is established. *PRIOR* Primary Ranking for Initial Orientation in emergency medical services, *mSTaRT* modified Simple Triage and Rapid Treatment, *FTS* Field Triage Score, *ASAV* Amberg-Schwandorf algorithm for triage, *STaRT* Simple Triage and Rapid Treatment

Discussion

The successful management of MCI necessitates a prioritization of the treatment measures with the earliest possible return to individual medical treatment principles [15, 24]. In this respect, the additional set-up of treatment capacities is an ethical moral discussion that evolves around what is medically desirable, what is logistically feasible, and what is politically preferred [6]. For a long time, the

insufficient identification of the severely injured (under-triage) with the consequence of an under-supply of these patients was the focus of the improvement and training activities [7, 15–17]. This also led to projects for the improvement and/or new development of triage algorithms in order to enable priority-based treatment in MCI regardless of the type of injury/disease (surgical/internal) [12].

Even though the investigated cohort of patients did not originate from a MCI,

the frequency and the higher degrees of disease and injury in HEMS missions (Table 3) offer the opportunity to investigate the quality of triage algorithms. For the focus of the work, namely the standardized comparison of triage algorithms, the distribution of the patient severity is of secondary importance, as long as a sufficient number of patients of all severity classes are present. This condition is met with 67 surgical and 132 internal patients with an NACA score ≥ 4 . Because of the variability in real life, such a cohort is also more suitable for testing the field suitability of triage methods than actor-patients or paper-based simulation patients [18], which are made easier to classify for methodological/didactic reasons [25]. The quality of the triage procedures, obtained from regular HEMS patients, is consistent with other studies [14, 26], even from MCI [22]. It can be assumed that the observed strengths and weaknesses of the algorithms from the present computer simulation will also become apparent in real MCI when they are applied by EMS forces. The strength of the computer simulation in this context is, in particular the fact that the inadequacies of the algorithms themselves cannot be compensated for by the clinical experience of EMS forces and, thus, the pure logic of the algorithms is compared.

Congruent with Fig. 2, an evaluation study of the *PRIOR* algorithm [26] showed that the time duration for the triage of surgical T3- patients (green) taking 42 s was the longest as compared with both the other triage categories, and with the *mSTaRT* algorithm. The time requirement for the *PRIOR* triage process for T1/T2/T3 was indicated with 27/28/42 s and for *mSTaRT* with 35/20/10 s, respectively. When comparing time requirements, the distribution of triage categories in the cohort considered must be taken into account. A simple model calculation shows that in a patient distribution T1/T2/T3/T0 of 15%/20%/60%/5%, respectively, in 100 patients being triaged by the *PRIOR* algorithm, time requirements of 42 min emerge solely for triage of T3 casualties. In comparison, the time consumption by *mSTaRT* for this category is 10 min. Shifting the patient distribution in favor

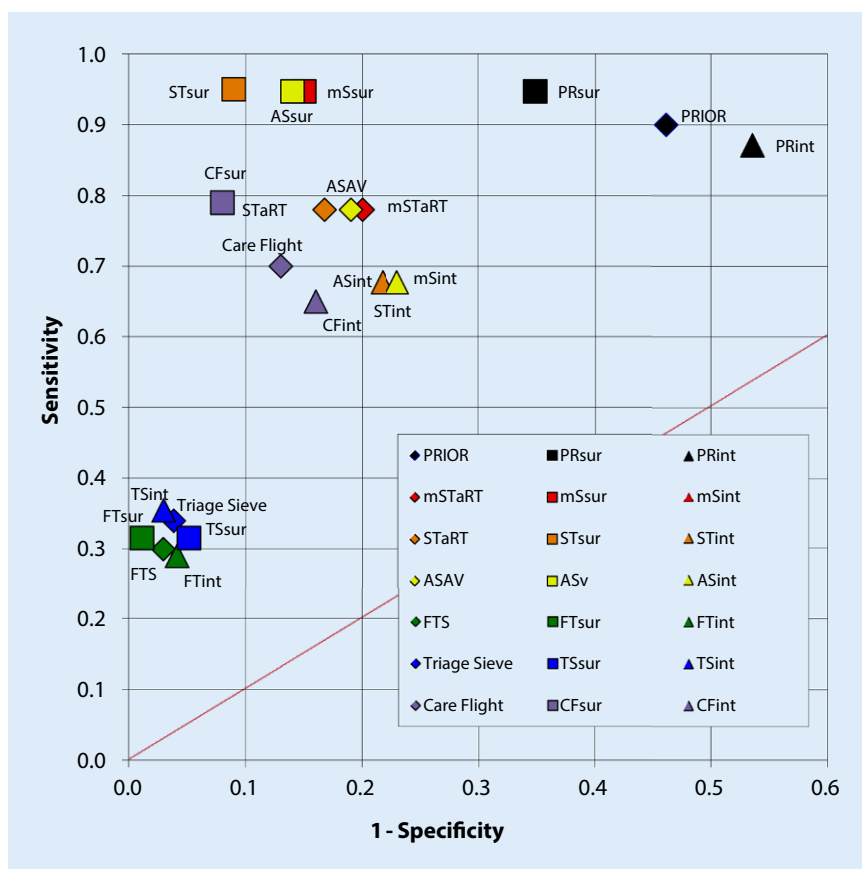


Fig. 3 ▲ Test quality of triage procedures for the detection of severely injured/ill patients (combined triage category T1+T4). Total cohort (*diamonds*), surgical patients (*squares (+ "sur")*), internal patients (*triangles (+ "int")*). Sensitivity = proportion of correct inclusion, specificity = proportion of correct exclusion. *PRIOR* Primary Ranking for Initial Orientation in emergency medical services, *mSTaRT* modified Simple Triage and Rapid Treatment, *FTS* Field Triage Score, *ASAV* Amberg-Schwandorf Algorithm for Triage, *STaRT* Simple Triage and Rapid Treatment

of T3, as found by Brüne [19] in MCI (T1/T2/T3 of 7%/19%/74%, respectively) the *PRIOR* method uses 81% of triage time on slightly Injured.

Binding forces with triage activities delays medical treatment within the casualty collecting points. Thus, when the precision is comparable, a method should be preferred which is quicker, for example by limiting the discriminants. In addition to the effects that result solely from the construction of the algorithm, experience and routine in dealing the algorithms must also be taken into account. In a mass casualty incident exercise of the Federal 24th Medical Task Force (MTF), 4 × 25 dynamic simulation patients [27] with a T1/T2/T3 distribution of 40/28/32, respectively were standardized triaged according to *PRIOR* or *mSTaRT* by physicians with different emergency medical experience [28]. The average time re-

quirement for a *PRIOR* triage procedure independent of the triage category was 23 ± 13 s (*mSTaRT* 31 ± 23 s). In the experienced physician, the time requirement independent of the algorithm was 19 ± 11 s per triage (novice 36 ± 22 s).

Compared to the model calculations with the expected patient distributions, *PRIOR* can show its strengths in terms of speed mainly when a high proportion of T1 (red) patients are present and the algorithm is exited earlier with the respective result. The level of experience, however, decisively determines the length of time. Regarding typical patient distributions in MCI as described by Brüne [19], however, the overwhelming expenditure of time with *PRIOR* remains for T3 (green) patients.

In the management of MCI, one can recognize the tendency that preclinically assessed patients in doubt are classified

into a higher triage category (over-triage) in order to avoid an under-treatment and hence a potential threat to life. On the example of a bus accident with ultimately 11 deceased and 69 injured in 2014 in Dresden, the triage category proved to be less critical on admission to the university hospital by 0.4 categories ($T2.5 \pm 0.3$) as it was handed over by the rescue service (2.1 ± 0.2). Out of 10 patients, four were preclinically rated higher by one category than they actually were [29]. In this respect, the NACA classification of the missions by the emergency physician must be critically questioned and cannot be transferred 1:1 into triage categories. For the validation of both the preclinical assessment as well as the grouping according to triage categories, the downstream scoring in the hospital according to SAPS II or ISS (only trauma patients) would have been interesting, but was not completely present due to the varying destination hospitals within the HEMS operational area. Independent of this, the actual goal of the study is hardly influenced because both the reference triaging physicians and the algorithms worked under the same conditions and, thus, comparability was obtained.

Mortality in a MCI largely depends on the degree of over-triage: data from Frykberg [30] show that the mortality for each percentage of over-triage increases by almost 0.5%. Over-triage, elicits wasting of treatment resources on non-urgent cases (T2/T3) which are thus no longer available to T1 patients [30]. Conversely, due to insufficient sensitivity, under-triage leads to a lack of assignment to the correct, more severe category, and, thus, prevents access of the patient to the necessary treatment resources of this category (Table 2; [18]).

An evaluation study in surgical actor and simulation card patients [27] revealed a sensitivity of 99% with a specificity of 45% for the T1 category using *PRIOR*. The *mSTaRT* exerted a T1 sensitivity of 55% with a specificity of 93% [26]. In surgical EMS patients Paul et al. showed a sensitivity of 50% and a specificity of 97% [14] for *mSTaRT*. These data are in agreement with the results found here in Fig. 3 and Table 4. The *PRIOR* algorithm, thus, captures almost all T1

Table 4 Test quality of the triage method in percent

Patients	Method	Triage category 1+4			Triage category 2			Triage category 3		
		AUC	SE (%)	SP (%)	AUC	SE (%)	SP (%)	AUC	SE (%)	SP (%)
ALL	PRIOR	0.48	90	54	0.15	17	87	0.33	45	73
	mSTaRT	0.62	78	80	0.18	21	84	0.36	67	53
	STaRT	0.65	78	83	0.21	26	82	0.36	67	53
	ASAV	0.63	78	81	0.17	21	83	0.36	67	53
	FTS	0.29	30	97	0.20	25	81	0.30	83	36
	Triage Sieve	0.33	34	96	0.29	43	67	0.36	67	53
	Care Flight	0.61	70	87	0.24	31	78	0.36	67	53
SURG	PRIOR	0.62	95	65	0.20	25	79	0.41	54	76
	mSTaRT	0.81	95	85	0.24	33	74	0.39	62	63
	STaRT	0.86	95	91	0.31	43	72	0.39	62	63
	ASAV	0.81	95	86	0.24	33	72	0.39	62	63
	FTS	0.31	32	99	0.16	19	83	0.27	89	30
	Triage Sieve	0.30	32	95	0.31	51	60	0.39	62	63
	Care Flight	0.73	79	92	0.31	45	69	0.39	62	63
INT	PRIOR	0.40	87	46	0.11	12	91	0.28	39	72
	mSTaRT	0.52	68	77	0.12	13	90	0.33	71	47
	STaRT	0.53	68	78	0.12	14	88	0.33	71	47
	ASAV	0.52	68	77	0.12	13	89	0.33	71	47
	FTS	0.28	29	96	0.23	29	80	0.31	80	39
	Triage Sieve	0.34	35	97	0.28	38	72	0.33	71	47
	Care Flight	0.55	65	84	0.18	21	84	0.33	71	47

AUC product of sensitivity (SE) and specificity (SP); highlighted: highest quality results. The higher the SE the lower the under-triage, the higher the SP the less the over-triage

PRIOR Primary Ranking for Initial Orientation in emergency medical services, mSTaRT modified Simple Triage and Rapid Treatment, FTS Field Triage Score, ASAV Amberg-Schwandorf Algorithm for Triage, STaRT Simple Triage and Rapid Treatment, ALL all patients, SURG surgical patients, INT internal patients

patients (SE 90%), however, at the price of a high over-triage (SP 50%) with the aforementioned problems of treatment resource waste and presumed mortality increase (Table 2, [18, 30]). In contrast to prior studies which neglected the T4 patients as severely injured, the mSTaRT algorithm showed significantly better values in this study with a sensitivity of 95% for surgical patients. The methodological advantage of this combined assessment of the classification of severely injured persons was already explained in the method section. The quality of the classification of the slightly injured T3 (green) was described for the PRIOR algorithm on surgical actor patients and simulation cards [27] with a sensitivity of 97% and a specificity of 96%, while mSTaRT had a sensitivity of 97% and a specificity of 82% [26]. In contrast, the present study shows a specificity

of 65% for PRIOR with a comparative sensitivity in the surgical cohort. This could be due to the fact that actor patients and simulation cards for didactic reasons have a clear predefined triage category. The greater variability of symptoms given by real life patients in this study places higher demands on the separation accuracy of the algorithms and thus allows a more realistic assessment of their quality [31].

Methods which use the “ability to walk” as a criterion for the categorization of only slightly injured and disabled persons (T3) consistently show a sensitivity of 67% and a specificity of 53%. Surgical patients are lower in sensitivity (62%), but higher in specificity (63%). An inverse picture is found in the internal patients with a higher sensitivity (71%) and a lower specificity (47%). This may be caused by the walking-disabled (surgical) patients with slight injuries to

the lower extremities, which are over-triaged because of their lack of walking ability [14]. In summary, “able to walk” is not a good discriminant but is excellently suited to performing a rapid, but coarse division of patients in tangled scenarios. Patients lacking walking ability must be immediately rescued from the injury zone. For subsequent exact triage and allocation of treatment priority, the discriminate “able to walk” is unsuitable.

The use of the motor component of the Glasgow Coma Scale (GCS) within the FTS as the decision-making criterion seems to include the greater part of the actually only slightly injured (SE 83%), but the seriously injured and disabled are not safely excluded from the T3 (SP 36%). If the area under the curve (AUC) is used as a quality criterion, the procedures which regard the ability to walk are at best poor with 0.36, where values between 0.9 and 1 should be striven for. Overall, such an AUC value for diagnostic tests is unsatisfactory. Thus, the discriminant “capable of walking” in our view is in itself not suitable for a valid assignment into one of the triage categories. For the remaining triage category T2 (yellow), test quality values which hardly exceed that of a coin toss are shown for almost all procedures (Table 4).

In the appraisal of the algorithm-independent reference triage category classification by the 19 physicians, the dimensions of affiliation to a medical subspecialty and experience level are of interest. Since the few non-anesthesiologists are members of the group of onsite commanding physicians, the analysis by subspecialty is cross-correlated with the level of experience and is thus not valid. While the group of onsite commanding physicians, as expected, has a high sensitivity and specificity in the determination of T1 patients (Table 5), which also is superior to all triage methods, the sensitivity in the two non-commanding groups is lower than that for STaRT, mSTaRT and ASAV as depicted in Fig. 3. While in the latter two groups the exclusion of non-severely injured persons seems to work well and at the same level as the commanding physicians, a relevant level of under-triage (false negatives) can be seen. An interesting aspect results from

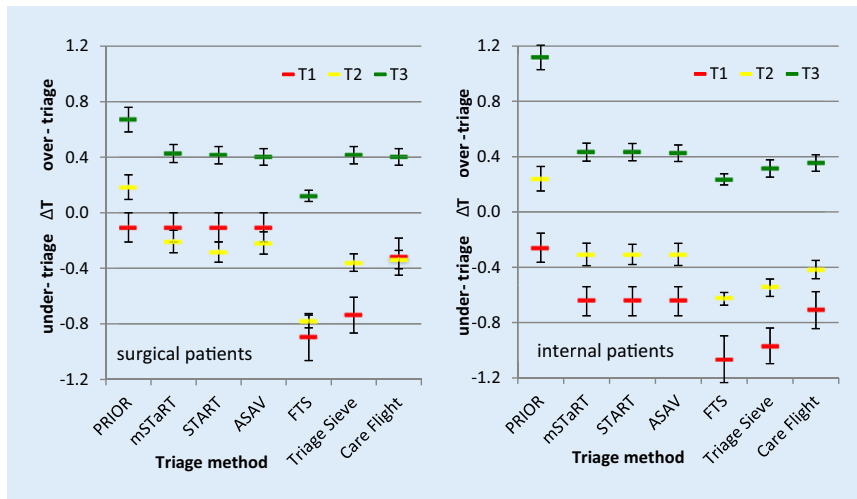


Fig. 4 ▲ Over-triage or under-triage by triage methods regarding triage categories and patient cohorts. The differences (ΔT mean values \pm SEM) of the triage category determined by the triage algorithm from the physician reference triage are given. *PRIOR* Primary Ranking for Initial Orientation in Emergency Medical Services, *mSTaRT* modified Simple Triage and Rapid Treatment, *FTS* Field Triage Score, *ASAV* Amberg-Schwandorf Algorithm for Triage, *STaRT* Simple Triage and Rapid Treatment

the slightly higher test quality for the determination of the T1 patients among the beginners against the advanced EMS physicians without command qualification. The reasons and the actual clinical significance of this observation can only be speculated: caused by their justified uncertainty and a compensatory accuracy, novice EMS physicians might make a higher classification (correctly positive), without, however, achieving false positive results (correct exclusion succeeds).

Summarizing **■ Figs. 3 and 4**, and **■ Table 5**, and considering the limitations of the group size of the classifying physicians, it is to be concluded that emergency physicians who have not dealt intensively with the topic of triage, should better apply the aforementioned triage algorithms for their inclusion decisions in the T1 category to avoid under-triage.

In diagnostic tests in medicine, two-step procedures are quite common in order to bundle the strengths of test procedures and thus to eliminate weaknesses. Transferred to the correct grouping of slightly injured casualties without missing comorbidities, an algorithm initially based on the walking ability could be useful for gaining a quick overview over the scene; however, due to the lack of test quality, an algorithm with detection of

physiological parameters must be carried out downstream (2nd triage) [18].

As this study empirically shows, surgical and internal disease patterns differ with respect to the achievable separation quality to the triage categories. One reason for this could be that the cut-off values of the physiological parameters defined in the algorithms are too rigid for internal patients. For example, acute hypertension or a stroke of many procedures is classified into the T3 category. Likewise, algorithms in particular using anatomical discriminants (critical bleeding, unstable pelvis, etc.) are ineffective in the case of internal diseases. Prioritization algorithms used in medical civil protection still need to provide usability for non-surgical patients [8]. A Federal Medical Task Force order to go into action could read to temporarily supersede a hospital, and prioritization decisions must be taken with the available basic medical resources, which also include stroke, of which 22 were included in this study. This supports the demand for an adjustment of the triage discriminants with appropriate consideration of internal diseases. This problem, which is still unresolved, is also subject to the procedures described with best test quality in **■ Table 4**. The leading methods *START*, *ASAV* and *mSTaRT*, with an AUC of ≤ 0.65 for the total cohort, are far below the values which must be

expected from a diagnostic test despite all mentioned challenges; however, even with the diagnostic abilities of a large emergency department, there is a much better match between surgical admission diagnoses and discharge diagnoses than in internal patients [32]. This observation may be an indication that the differences in the diagnostic precision of surgical and non-surgical patients insolvably lie in the nature of these diseases, in particular when regarding the limited pre-clinical medical capabilities.

Multiparameter Early Warning Scores (MEWS), which take into account multi-stage measurement deviations, both upwards and downwards [33], could contribute to the resolution of the sensitivity problem, particularly in internal patients. The effort associated with the manual acquisition makes these MEWS useless at least for the first triage process in medical civil protection. The same applies to WLAN-based commercially available automated MEWS systems [34], particularly in spatially extended scenes. Similar to the *PRIOR* algorithm, MEWS might result in a high number of false positive ratings. The suitability of MEWS for a second triage in a T3 casualty collection point formed according to the ability to walk in order to avoid an under-triage must be examined in follow-up studies.

Conclusion

According to the current data, the goal to be acknowledged is that T1 (red) and in particular non-surgical patients have to be detected more reliably by means of optimization of triage algorithms; however, the lower diagnostic quality in non-surgical patients may be due to the higher complexity of these patients and may not be resolved preclinically, especially in MCI. In general, further developments of algorithms must follow the rules, which also apply to other diagnostic tests in medicine. The desire to derive a MCI prioritization algorithm from a method, which is well-established in individual trauma treatment, with high sensitivity in detection and treatment of injuries [15] but disregarding the necessary specificity may not be the key driver. As the present study shows, a method such as

Table 5 Deviations of the individual triage procedures from reference (ΔT) and test quality criteria of the physician raters as a percentage regarding the degree of experience (mean \pm SEM)

		Professional experience EMS physicians		
		<3 years n = 7 (2.1 \pm 0.4 years)	>3 years n = 6 (8.8 \pm 0.9 years)	LNA n = 6 (15.3 \pm 2.9 years)
T1	ΔT	-0.4 \pm 0.2	-0.4 \pm 0.1	-0.1 \pm 0.1
	Sensitivity (%)	70	65	92
	Specificity (%)	95	95	94
T2	ΔT	-0.2 \pm 0.1	-0.2 \pm 0.0	0.0 \pm 0.0
	Sensitivity (%)	61	56	80
	Specificity (%)	87	86	88
T3	ΔT	0.1 \pm 0.0	0.1 \pm 0.0	0.1 \pm 0.0
	Sensitivity (%)	89	89	86
	Specificity (%)	74	71	92
Total	ΔT	-0.1 \pm 0.0	-0.1 \pm 0.0	0.0 \pm 0.0

T triage category. Reference: weighted total result of category assignment. In brackets: professional experience in EMS in years. LNA onsite commanding physicians

PRIOR is highly sensitive for the assignment to the T1 category (red) because of the many possibilities of inclusion, but categorizes many false positive decisions (over-triage), so that the suitability for fast and reliable identification of the severely injured in MCI must be doubted. Rather, there is reason to worry about this degree of over-triage, with an increased mortality of the T1 patients, when applied as a triage method in casualty collection points. This is particularly true when taking into account a scenario-specific share of T1 patients between 3–14% as described by Brüne [19] but consuming the greatest time proportion for the triage of the T3 patients.

Given the high availability of the forces and resources of both EMS and downstream medical facilities (e.g. in a metropolitan area), a very sensitive algorithm with a restricted specificity, such as PRIOR [12] will be less negative in the sense of increasing mortality [30] than in a rural area. In the view of the incident command and the downstream care facilities, it is nevertheless more advantageous to prioritize and distribute patients with clearly defined knowledge of the actual triage category than with a high degree of uncertainty in the categorization.

In particular, taking into account a scenario in which a triage by physicians cannot be carried out until reaching the hospital and the decisions are made

on the basis of algorithms by non-physicians, the T4 category (blue) may not play a role in triage on scene. Established triage procedures have good sensitivities and specificities for trauma patients and may be applied. Although functioning triage concepts exist, it is ultimately decisive for success to keep this knowledge available to the EMS personnel of all professions at all times and to keep them ready for application.

Corresponding address

Prof. Dr. A. R. Heller, DEAA, MBA

Department of Anesthesiology and Critical Care Medicine, Medical Faculty Carl Gustav Carus, TU-Dresden
Fetscherstrasse 74, 01307 Dresden, Germany
axel.heller@uniklinikum-dresden.de

Acknowledgements. The authors thank the participating reference triage physicians, T. Bluth, A. Bundy, V. Dalicho, C. Eisold, A. Friedrich, K. Kirschneck, T. Kiss, W. Lorenz, M. Mirus, B. Maschek, S. Mittelstädt, T. Müller, H. Pich, R. Paul, S. Rüder, S. Shmygalev, O. Vicent, as well as K. Juncken for database maintenance. We also thank the colleagues of the DRF HEMS "Christoph 38" in Dresden, and R. Nail for proofreading the English manuscript.

Compliance with ethical guidelines

Conflict of interests. A.R. Heller: membership DGAI, DGKM, Med Course Director PHTLS, Instructor ERC, ATLS, ETC, GIC, member of the Triage Consensus Conference of the Federal Office of Civil Protection and Disaster Assistance. J. Schiffner: membership DGKM.

R. Kipke: membership DGKM. C. Kleber: member of the DGU, DGKM and Triage Consensus Conference of the Federal Office of Civil Protection and Disaster Assistance. N. Salvador and M. Frank declare that they have no competing interests.

The use of existing HEMS patient records for this analysis was approved by the IRB of the Medical Faculty, TU Dresden (EKDD270062015).

Literatur

- Carles M, Levraut J, Gonzalez JF, Valli F, Barnard L (2016) Mass casualty events and health organisation: terrorist attack in Nice. *Lancet* 388(10058):2349–2350
- Hirsch M, Carli P, Nizard R, Riou B, Baroudjian B, Baubet T et al (2015) The medical response to multisite terrorist attacks in Paris. *Lancet* 386(10012):2535–2538
- Biermann K, Faigle P, Geisler A, Müller D, Musharbash Y, Polke-Majewski K et al (2017) Was wir über den Anschlag in Berlin wissen
- Heller AR (2011) Be prepared!: hospital planning for major public events. *Dtsch Arztebl Int* 108(28–29):481–482
- Genfer Abkommen zur Verbesserung des Loses der Verwundeten und Kranken der bewaffneten Kräfte im Felde, Diplomatische Konferenz zur Revision des Genfer Abkommens vom 27. Juli 1929 (1945)
- Frank M, Heller AR (2006) Sichtung durch Infrastruktur ersetzen – Wunsch und Wirklichkeit. *Dtsch Arztebl Int* 103(48):A3250
- Sefrin P (2013) Vorsichtung notwendig – Bericht von der Nachfolge Sichtungskonferenz. *Notarzt* 2014(30):85
- Bundesamt für Bevölkerungsschutz und Katastrophenhilfe (2015) Protokoll der 6. Sichtungskonsensuskonferenz. 6 Sichtungskonsensuskonferenz 2015 November 24
- Bundesärztekammer (2009) Stellungnahme der Bundesärztekammer zur ärztlichen Sichtung Verletzter/Erkrankter
- Kowalzik B (2013) Vorsichtung – Workshop zur Evaluierung von medizinischen Vorsichtungsverfahren bei MANV. *Bevölkerungsschutz* 4:11–13
- Streckbein S, Kohlmann T, Luxen J, Birkholz T, Prückner S (2016) Sichtungskonzepte bei Massenunfällen von Verletzten und Erkrankten. *Unfallchirurg* 119(8):620–631
- Bubser F, Callies A, Schreiber J, Grüneisen U (2014) PRIOR: Vorsichtungssystem für Rettungssassistenten und Notfallsanitäter. *Rettungsdienst* 37(8):730–734
- Hiereth K, Hornburger P, Eyer F, Gerstenhöfer S, Schmöller G, Prückner S et al (2013) mStART Trauma & Intox. *Notf Rettungsmid* 16:627–636
- Paul AO, Kay MV, Huppertz T, Mair F, Dierking Y, Hornburger P et al (2009) Validierung der Vorsichtung nach dem mStART-Algorithmus: Pilotstudie zur Entwicklung einer multizentrischen Evaluation. *Unfallchirurg* 112(1):23–30 (32)
- National Association of Emergency Medical Technicians (NAEMT) (2016) Präklinisches Trauma-management: Prehospital Trauma Life Support (PHTLS), 3 edn. Urban & Fischer Elsevier, München
- Bartels F (2001) Katastrophenmedizin – Wir müssen uns schnell auf die neue Lage einstellen. *Dtsch Arztebl* 98(43):A2770–A2772
- Szasz G (1968) Disaster medicine. *Can Fam Physician* 14(11):26–31
- Kleber C, Cwojdzinski D, Strehl M, Poloczek S, Haas NP (2013) Results of in-hospital triage in 17

- mass casualty trainings: underestimation of life-threatening injuries and need for re-triage. *Am J Disaster Med* 8(1):5–11
19. Brüne F (2013) Reale Verteilung von Sichtungskategorien bei MANV Einsätzen – Auswirkungen auf die Schutzziele. Mathematisch – Naturwissenschaftliche Fakultät der Rheinischen Friedrich-Wilhelms-Universität, Masterstudengang Katastrophenvorsorge-Katastrophenmanagement, Bonn
 20. Garner A, Lee A, Harrison K, Schultz CH (2001) Comparative analysis of multiple-casualty incident triage algorithms. *Ann Emerg Med* 38(5):541–548
 21. Eastridge BJ, Butler F, Wade CE, Holcomb JB, Salinas J, Champion HR et al (2010) Field triage score (FTS) in battlefield casualties: validation of a novel triage technique in a combat environment. *Am J Surg* 200(6):724–727
 22. Kahn CA, Schultz CH, Miller KT, Anderson CL (2009) Does START triage work? An outcomes assessment after a disaster. *Ann Emerg Med* 54(3):424–430
 23. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1(8476):307–310
 24. Sefrin P, Weidringer JW, Weiss W (2003) Sichtungskategorien und deren Dokumentation. *Dtsch Arztebl* 100(31–32):A2057–A2058
 25. Sautter J, Schneider F, Kippnich U, Wirth S, Habermann M, Siebel C et al (2016) Durchführung und Auswertung von MANV Übungen, 1 edn. Deutsches Rotes Kreuz, Berlin
 26. Zahn T (2015) PRIOR Evaluierung – Design und Ergebnisse. In: Schreiber J, Grüneisen U (eds) Symposium 21.11.2015: Medizinisches Management im MANV unter Verwendung der PRIOR-Vorsichtungssystems. Deutsche Gesellschaft für Katastrophenmedizin (DGKM), Leipzig
 27. Brüne F (2012) Kurzbeschreibung Dynamische Patientensimulation
 28. Püsche K, Heller A (2017) Implementierung und Evaluierung zielgruppenspezifischer Simulationstechniken für Großschadenslagen. Med. Dissertation, Professur für Notfallmedizin, Medizinische Fakultät Carl Gustav Carus an der TU Dresden
 29. Heller A (2014) Interner Einsatzbericht Uniklinikum Dresden Busunfall A4 190200-190800jul14. Universitätsklinikum Carl Gustav Carus, Dresden
 30. Frykberg ER (2002) Medical management of disasters and mass casualties from terrorist bombings: how can we cope? *J Trauma* 53(2):201–212
 31. Salvador N, Heller A (2017) Evaluation der Entscheidungsgüte von Vorsichtungsverfahren beim Massenanfall von Verletzten und Erkrankten. Dissertation, Professur für Notfallmedizin, Medizinische Fakultät Carl Gustav Carus an der TU Dresden
 32. Wilk F, Dormann H (2016) Qualitätsindikatoren einer Notaufnahme. DIVI 2016 Abstractbuch, Deutscher Ärzteverlag
 33. Alam N, Hobbelenk EL, van Tienhoven AJ, van d V, Jansma EP, Nanayakkara PW (2014) The impact of the use of the Early Warning Score (EWS) on patient outcomes: a systematic review. *Resuscitation* 85(5):587–594
 34. Bellomo R, Ackerman M, Bailey M, Beale R, Clancy G, Danesh V et al (2012) A controlled trial of electronic automated advisory vital signs monitoring in general hospital wards. *Crit Care Med* 40(8):2349–2361