Performance characteristics of five triage tools for major incidents involving traumatic injuries to children

C L Price[1], S J Brace-McDonnell[2, 3], N Stallard[2], A Bleetman[2], I Maconochie[4,5], G D Perkins[2,3]

[1]Operational Research and Management Sciences Group (ORMS), Warwick Business School, University of Warwick, Coventry, CV4 7AL

C L Price, Principal Teaching Fellow Statistician

[2]Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, CV4 7AL, UK

S J Brace-McDonnell, NIHR Clinical Doctoral Research Fellow, N Stallard, Professor of Medical Statistics, A Bleetman, Consultant in Emergency Medicine, G D Perkins, Professor in Critical Care Medicine

[3]Heart of England NHS Foundation Trust, Bordesley Green East, Birmingham, B9 5SS, UK

[4] Imperial College NHS Healthcare Trust, London, W2 1NY, UK

I Maconochie, Consultant in Paediatric Emergency Medicine

[5]The Trauma Audit & Research Network, Manchester Medical Academic Health Sciences Centre, University of Manchester 3rd Floor, the Mayo Building, Salford Royal NHS Foundation Trust, Stott Lane, Salford, M6 8HD

Correspondence to:

Dr Charlotte Price
Warwick Business School
University of Warwick
Coventry
CV4 7AL
charlotte.price@wbs.ac.uk

Keywords: trauma, mass casualty, triage, paediatric, accuracy, prediction

Abstract

Context: Triage tools are an essential component of the emergency response to a major incident. Although fortunately rare, mass casualty incidents involving children are possible which mandate reliable triage tools to determine the priority of treatment.

Objective: To determine the performance characteristics of five major incident triage tools amongst paediatric casualties who have sustained traumatic injuries.

Design, Setting, Participants: Retrospective observational cohort study using data from 31,292 patients aged less than 16 years who sustained a traumatic injury. Data were obtained from the UK Trauma Audit and Research Network (TARN) database.

Interventions: Statistical evaluation of five triage tools (JumpSTART, START, CareFlight, Paediatric Triage Tape/Sieve and Triage Sort) to predict death or severe traumatic injury (injury severity score >15).

Main outcome measures: Performance characteristics of triage tools (sensitivity, specificity and level of agreement between triage tools) to identify patients at high risk of death or severe injury.

Results: Of the 31,292 cases, 1,029 died (3.3%), 6,842 (21.9%) had major trauma (defined by an injury severity score > 15) and 14,711 (47%) were aged eight years or younger. There was variation in the performance accuracy of the tools to predict major trauma or death (sensitivities ranging between 36.4% and 96.2%; specificities 66.0% to 89.8%). Performance characteristics varied with the age of the child. CareFlight had the best overall performance at predicting death, with the following sensitivity and specificity (95% CI) respectively: 95.3% (93.8 to 96.8) and 80.4% (80.0 to 80.9). JumpSTART was superior for the triaging of children under 8 years; sensitivity and specificity (95% CI) respectively: 86.3% (83.1 to 89.5) and 84.8% (84.2 to 85.5). The triage tools were generally better at identifying patients who would die than those with non-fatal severe injury.

Conclusion: This statistical evaluation has demonstrated variability in the accuracy of triage tools at predicting outcomes for children who sustain traumatic injuries. No single tool performed consistently well across all evaluated scenarios.

Introduction

The term 'big bang' major incident is used to describe a major incident caused by sudden catastrophic events with little or no warning, where the number of casualties is relatively constant from the time of the incident but has the potential to outstrip resources.[1,2] Such incidents test the response of emergency medical services and hospitals and it is essential that resources are used in an optimal way to target those with greatest need.[3] In order to achieve this, one of the first priorities is to undertake rapid and accurate triage to prioritise and provide care to as many casualties as possible with the intention of minimising loss of life and suffering, moderated by the available resources. However, there is uncertainty around the efficacy of commonly-used triage systems, particularly in children,[4] and a recent systematic review of the literature concluded that there is limited evidence of the validity of triage tools in major incidents of this nature.[1]

This study aims to assess the performance accuracy of five manual / paper based triage tools when assessing paediatric casualties and to compare the level of agreement between them. The tools assessed are: JumpSTART (age ≤ 8 years),[5] START (age > 8 years),[6] CareFlight,[7] Paediatric Triage Tape/Sieve,[8] Triage Sort[9].

Methods

Study design and data collection

A retrospective observational cohort study was undertaken. Approval was obtained from the Trauma Audit and Research Network (TARN; www.tarn.ac.uk) to analyse data from the TARN database. TARN collects and records data from hospitals across England and Wales for patients who sustain injury resulting in hospital admissions for > 3 days, critical care admission or death. A dataset was obtained in August 2009 containing 31,560 paediatric trauma patient records for patients aged less than 16 years, and included respiratory rate (breaths per minute), systolic blood pressure (mmHg), cardiac arrest (yes/no), intubated (yes/no), age (years), capillary refill time (> 2 seconds / < 2 seconds), heart rate (beats per minute), Glasgow Coma Scale (GCS) score and Injury Severity Score (ISS). No personal identifiable information was provided. The GCS is used to assess the conscious state of a person and is a 13-point scale ranging between 3 and 15, where 3 indicates a state of deep unconsciousness [10]. The ISS assesses trauma severity and ranges between 1 and 75 (worst) [11]. Patient survival was recorded in the dataset as alive or dead. [10] The use of anonymised data from a research database does not require specific ethical approval in the UK (Governance Arrangements for Research Ethics Committees 2012).

Triage tools

Five triage tools were evaluated: JumpSTART [5] (age ≤ 8 years), START [6] (age> 8 years), CareFlight [7], Paediatric Triage Tape/Sieve [8] and Triage Sort [9]. Each triage tool leads to one of three priority outcomes, named slightly differently depending on the tool used: deceased (dead/unsalvageable), immediate (priority 1 or 2) or delayed (urgent/priority 3). For this research, the deceased and immediate outcomes were combined into a single 'immediate priority' outcome. The performance of each tool was assessed according to its ability to accurately distinguish between 'immediate priority' and 'delayed priority' patients.

The following assumptions informed the mapping of TARN data to the various triage tools: An open airway, or the ability to breathe, was indicated by a respiratory rate > 0. The patient was assumed to have palpable pulse if the systolic blood pressure was > 60 mmHg and no palpable pulse if systolic blood pressure was ≤ 60 mmHg. The ability to obey commands was indicated by a GCS score ≥ 14 and a patient was assumed to be unable to obey commands if the GCS score was < 14 or if this was missing and the patient had been intubated.

Statistical analysis

31,292 patient records (99.2%) were eligible for analysis, 268 patients (0.8%) were excluded due to missing vital sign information or ISS. Multiple imputation[12,13] was used to replace missing values for the following variables (proportion missing): respiratory rate (27.4%), heart rate (15.2%), systolic blood pressure (25.6%) and intubated (10.3%) using a model with 29 variables to give five imputed datasets with results combined as proposed by Rubin [13]. If a patient had a missing GCS score but had been intubated, it would have been impossible to obtain the score and this was indicated as a separate category in the dataset.

The primary outcome of interest was patient survival (alive or dead). However, for comparison, the tools were also assessed against injury severity (ISS ≤ 15 or ISS > 15). Descriptive statistics were used to explore the data by age (≤ 8 years and > 8 years) and survival (alive or dead). These included means (medians) with standard deviations (interquartile ranges) and frequencies with percentages. Sensitivities and specificities with 95% confidence intervals were calculated for each triage tool against both survival and injury severity. Sensitivity indicates the proportion of patients who died / had ISS > 15 who were correctly assigned to the immediate priority group and specificity indicates the proportion of patients who did not die / had ISS ≤ 15 who were correctly assigned to the delayed priority group.

Since JumpSTART and START apply to different ages (≤ 8 years and > 8 years respectively), analyses for all triage tools were conducted separately on these two age groups. Acknowledging that the PTT is weight and length based, the weight for each child was calculated as (age + 4) × 2 and the appropriate PTT algorithm was used. [14] A further analysis was undertaken to compare the tools using all cases (regardless of age) by combining JumpSTART and START into a single tool. A complete case analysis was also undertaken to compare the results with and without using multiple imputation. Patient records were assumed to be independent.

Agreement between each pair of triage tools for the two age groups (≤ 8 years and > 8 years) was estimated using the kappa statistic. A value of 1 indicates perfect agreement and a value of 0 indicates no agreement.[15,16]

Data cleaning was undertaken using SPSS v.17. All other analyses were performed in the R statistical software (www.r-project.org; downloaded in UK). In particular, a bootstrapping approach was adopted to undertake the multiple imputation using the aregImpute function in the Hmisc package (http://cran.r-project.org/web/packages/Hmisc).

Results

A total of 31,292 patients aged less than 16 years were included in the study; 10,048 females (32.1%) and 21,244 males (67.9%), with mean ages 7.9 years (standard deviation 4.9 years) and 8.7 years (standard deviation 4.8 years) respectively. A total of 1,029 patients (3.3%) died and the median ISS was 9 (IQR 5 to 13), with 6,842 (21.9%) having an ISS > 15. Within the group of patients who survived, 19.4% (5878/30263) had an ISS > 15 compared to 93.7% (964/1029) of those in the non-survivor group. Splitting the data by age, 14,711 patients (47%) were aged less than or equal to 8 years and 16,581 patients (53%) were aged over 8 years. Patient characteristics in the two age groups by survival (alive or dead) are shown in Table 1.

Table 1: Patient characteristics by age group (≤ 8 yrs or > 8 yrs) and survival (alive or dead)

Characteristic	Age ≤ 8 years (n = 14,711) Alive (n = 14,235)	Age ≤ 8 years (n = 14,711) Dead (n = 476)	Age > 8 years (n = 16,581) Alive (n = 16,028)	Age > 8 years (n = 16,581) Dead (n = 553)	Total (n = 31,292)
Age (years); mean (SD)	3.9 (2.6)	3.6 (2.7)	12.4 (1.9)	12.5 (2.0)	8.4 (4.8)
median (IQR)	4 (2 to 6)	3 (1 to 6)	13 (11 to 14)	13 (11 to 14)	9 (4 to 13)
Gender (male); n (%)	9295 (65.3)	294 (61.8)	11294 (70.5)	361 (65.3)	21244 (67.9)
Respiratory rate (breaths pm); mean (SD)	24.5 (8.1)	21.0 (9.1)	20.3 (5.8)	17.8 (8.3)	22.2 (7.4)
Heart rate (beats pm); mean (SD)	112.4 (26.2)	113.1 (36.8)	91.8 (19.5)	102.1 (33.4)	101.7 (25.6)
Systolic BP (mmHg); mean (SD)	115.0 (21.5)	102.0 (34.1)	123.2 (20.1)	111.1 (33.0)	119.1 (21.9)
ISS:					
median (IQR)	9 (9 to 10)	26 (25 to 38)	9 (5 to 10)	34 (25 to 45)	9 (5 to 13)
> 15; n (%)	2747 (19.3)	438 (92.0)	3131 (19.5)	526 (95.1)	6842 (21.9)
GCS (< 14); n (%)	2492 (17.5)	387 (81.3)	2636 (16.4)	488 (88.2)	6003 (19.1)
Intubated; n (%)	400 (2.8)	63 (13.2)	249 (1.6)	40 (7.2)	752 (2.4)

SD standard deviation; IQR interquartile range

Paediatric triage tool accuracy

Sensitivities and specificities with 95% confidence intervals, calculated separately against survival (alive or dead) and injury severity (ISS ≤ 15 or ISS > 15), are given in Tables 2 to 4.

Table 2: Sensitivity and specificity for each triage tool against survival (alive or dead) and ISS (≤ 15 or > 15), with 95% confidence intervals, Age ≤ 8 years (n = 14,711)

Outcome	Triage tool	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Survival (dead or alive)	PTT	25.0 (20.7, 29.3)	88.4 (87.8, 89.0)
	Triage Sort	60.2 (59.3, 61.1)	60.2 (59.3, 61.1)
	JumpSTART	86.3 (83.1, 89.5)	84.8 (84.2, 85.5)
	CareFlight	95.0 (93.0, 97.0)	78.9 (78.2, 79.6)
ISS (≤ 15 or > 15)	PTT	17.6 (16.2, 19.0)	89.5 (88.9, 90.1)
	Triage Sort	76.6 (75.1, 78.2)	68.0 (67.1, 69.0)
	JumpSTART	43.9 (42.1, 45.8)	89.8 (89.3, 90.4)
	CareFlight	65.4 (63.6, 67.2)	88.1 (87.5, 88.7)

Table 3: Sensitivity and specificity for each triage tool against survival (alive or dead) and ISS (≤ 15 or > 15), with 95% confidence intervals, Age > 8 years (n = 16,581)

Outcome	Triage tool	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Survival (dead or alive)	PTT	48.8 (43.7, 53.9)	46.2 (45.3, 47.0)
	Triage Sort	96.5 (94.7, 98.2)	77.9 (77.3, 78.6)
	START	96.5 (94.7, 98.3)	57.8 (57.0, 58.6)
	CareFlight	95.6 (93.6, 97.5)	81.8 (81.2, 82.4)
ISS (≤ 15 or > 15)	PTT	52.8 (51.1, 54.5)	46.1 (45.1, 47.0)
	Triage Sort	65.4 (63.8, 67.0)	87.0 (86.4, 87.6)
	START	73.2 (71.7, 74.7)	64.2 (63.4, 65.1)
	CareFlight	63.8 (62.0, 65.5)	91.4 (90.9, 91.9)

Table 4: Sensitivity and specificity for each triage tool against survival (alive or dead) and ISS (≤ 15 or > 15), with 95% confidence intervals, all ages (n = 31,292)

Outcome	Triage tool	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Survival (dead or alive)	PTT	37.8 (34.1, 41.4)	66.0 (65.4, 66.6)
	Triage Sort	96.2 (94.9, 97.4)	69.6 (69.0, 70.1)
	JumpSTART/START	91.8 (90.0, 93.5)	70.5 (70.0, 71.1)
	CareFlight	95.3 (93.8, 96.8)	80.4 (80.0, 80.9)
ISS (≤ 15 or > 15)	PTT	36.4 (35.2, 37.7)	66.5 (65.9, 67.2)
	Triage Sort	70.6 (69.5, 71.7)	78.1 (77.5, 78.6)
	JumpSTART/START	59.6 (58.3, 60.8)	76.3 (75.8, 76.9)
	CareFlight	64.5 (63.3, 65.7)	89.8 (89.4, 90.2)

Figure 1 summarises the performance accuracy of the tools, particularly in their ability to correctly triage patients who died and those with an ISS > 15, as indicated by the sensitivity values.

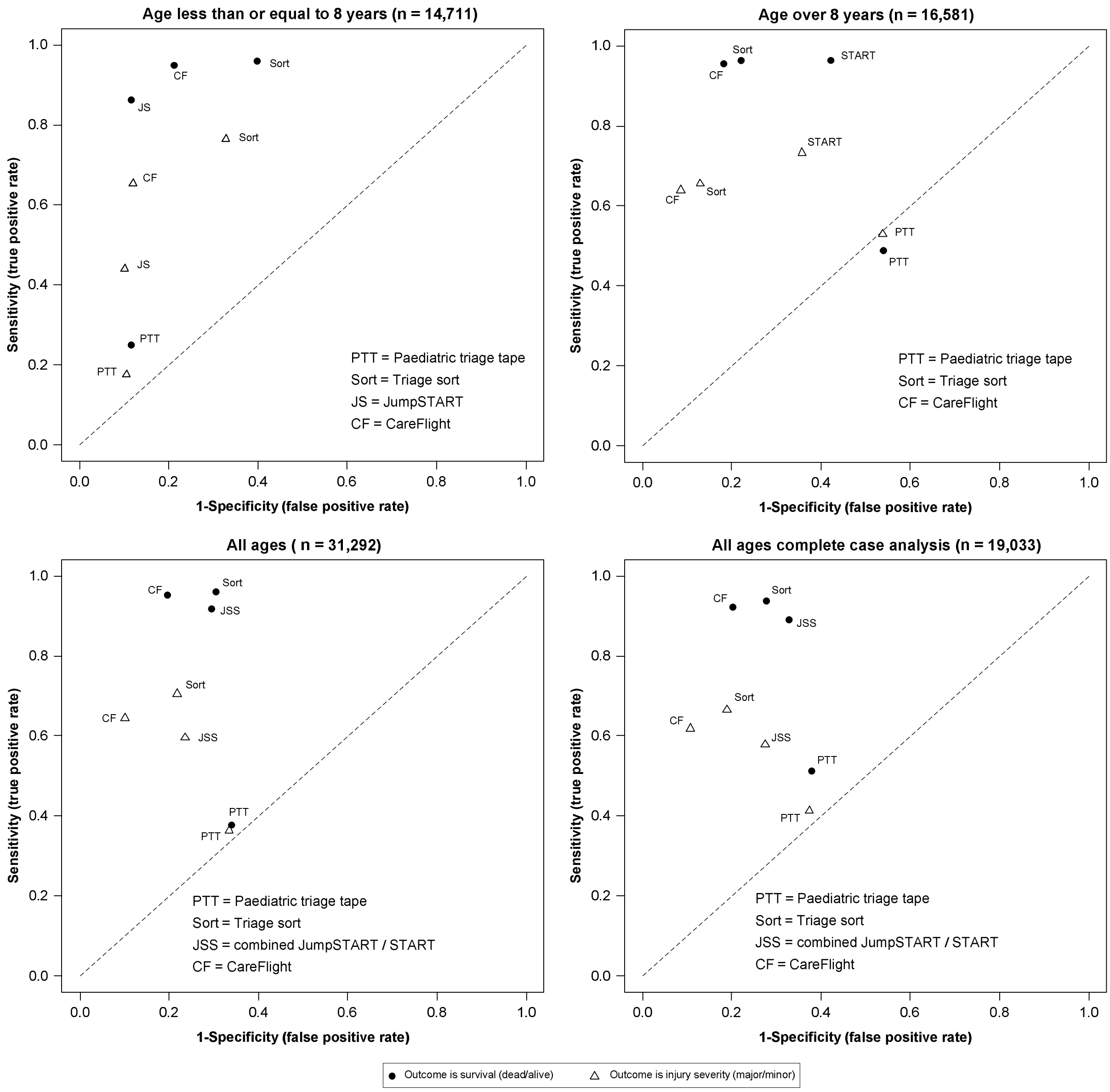

Figure 1: Triage tool performance

Performance for each triage tool is shown as the false positive rate (1 – specificity) against the true positive rate (sensitivity). Solid circles show tool performance at predicting death. Open triangles show tool performance at predicting severe injury (injury severity score > 15). The best performing tools are those closest to the top left hand section of each plot. A completely random guess would give a point along the diagonal dashed line.

The results of the complete case analysis were very similar to those of the analysis using imputed data. This provides reassurance that the imputed analysis has produced accurate estimates whilst enabling the use of the full dataset.

Level of agreement between triage tools

Table 5 shows the kappa statistics for pairs of triage tools. For children under 8, agreement between pairs of tools ranges between poor and moderate with the highest level of agreement between Triage Sort and CareFlight (kappa 0.54) and JumpSTART and CareFlight (kappa 0.54). For children aged 8 years and over, this pattern continues with only a 'good' level of agreement between Triage Sort and CareFlight (kappa 0.79).

Table 5: Level of agreement between pairs of triage tools (kappa statistic)

Age ≤ 8 years (n = 14,711)			
Triage tool	Triage Sort	JumpSTART	CareFlight
PTT*	**Poor**	**Fair**	**Poor**
	0.11	0.31	0.10
Triage Sort		**Fair**	**Moderate**
		0.31	0.54
JumpSTART			**Moderate**
			0.54
Age > 8 years (n = 16,581)			
Triage tool	Triage Sort	START	CareFlight
PTT*	**Poor**	**Poor**	**Poor**
	0.03	0.05	0.00
Triage Sort		**Moderate**	**Good**
		0.48	0.79
START			**Moderate**
			0.49

* Paediatric Triage Tape

Discussion

There are two key issues in assessing the performance of a triage tool in paediatric major incidents. The tool must be sensitive enough to identify patients at greatest need, but at the same time must ensure the best use of available resources by delaying treatment for patients who do not require immediate attention. [17] This study has used a large dataset to evaluate the performance of five paediatric triage tools against two separate outcomes, survival (dead or alive) and injury severity (ISS ≤ 15 or ISS > 15), and has revealed clear differences in performance accuracy.

Paediatric Triage Tape has low sensitivity and low specificity regardless of the outcome used for assessment, as well as poor agreement with the other triage tools.[8] In fact, its performance is barely superior to randomly assigning patients to triage priority categories. Not only does this delay treatment for those who need it most, it assigns immediate treatment to patients who do not require it, therefore wasting resources. For the remaining triage tools, sensitivities were very high for children aged eight years and over when calculated against survival, ranging from 95.6% to 96.5%. In contrast, the sensitivities were more variable for children under eight, ranging from 60.2% (Triage Sort) to 95.0% (CareFlight). In both age groups, the tools showed reduced accuracy in assigning patients to the immediate priority group when calculated against injury severity. This means that the triage systems tend to assign those patients who will ultimately die from their injuries to the immediate priority category, but are less accurate at assigning immediate treatment to patients with major injuries, many of whom will survive. This is perhaps not surprising since patients who die are likely to have suffered extreme trauma and are therefore easier to triage. In fact, within the TARN dataset, of the 1,029 patients who died (3.3%), 835 patients

(81.1%) suffered extreme trauma with an ISS ≥ 25. Children's physiology differs from adults, including major differences in airway, stable intracranial pressure in infants with open fontanelles, compliant chest walls and the ability to compensate for cardiovascular compromise.[18]

This study had some limitations. In order to apply the triage tools to the obtained dataset, a number of clinical assumptions were made to enable the available variables to be used. However, the assumptions were based on clinical expertise and the resulting methodologies matched closely to those used in practice. The variables included in the analysis were the first physiological observations recorded for each patient. Since it was not possible to determine if treatment was started before the information was recorded, we were unable to control for this in the analysis. In a mass casualty incident the ability of clinicians to record observations in potentially dark and contaminated scenes is also a circumstance we were unable to control for and is an acknowledged limitation of attempting pragmatic analyses of existing databases. The data used in the study will thus have variable input reliability. The TARN dataset also had missing data on important variables. However, this was dealt with by using multiple imputation to replace the missing values. This not only allowed the analysis to be carried out on the complete dataset, but avoided potential problems of unreliability that are generally associated with a complete case analysis where any records with a missing value are removed.[19,20] Patients who died at the scene were not considered in this study as they are not included in the TARN database and this may have influenced the performance of the tools. The TARN dataset contains only trauma patients and so these tools may perform differently if applied to mass casualty incidents that contain chemical, biological, radiological or nuclear components. The practical usefulness of the tools is paramount to their performance and testing this was beyond the remit of this analysis.

Conclusion

There is variation in the performance of existing triage tools used for predicting patient outcome and treatment priority in children. No single tool performed consistently well across all evaluated scenarios.

References

1. Kilner TM, Brace SJ, Cooke MW, Stallard N, Bleetman A, Perkins GD. In 'big bang' major incidents do triage tools accurately predict clinical priority?: A systematic review of the literature. Injury 2010.
2. Health Do. NHS Emergency Planning Guidance 20052005.
3. Carley S, Mackway-Jones K. Major incident medical management and support : the practical approach in the hospital. London ;: BMJ, Oxford : Blackwell; 2005.
4. Lyle K, Thompson T, Graham J. Pediatric Mass Casualty: triage and planning for the prehospital provider. Clinical Pediatric Emergency Medicine 2009;10:173-85.
5. Romig LE. Pediatric triage. A system to JumpSTART your triage of young patients at MCIs. JEMS 2002;27:52-8, 60-3.
6. Benson M, Koenig KL, Schultz CH. Disaster triage: START, then SAVE--a new method of dynamic triage for victims of a catastrophic earthquake. Prehosp Disaster Med 1996;11:117-24.
7. Nocera A, Garner A. An Australian mass casualty incident triage system for the future based upon triage mistakes of the past: the Homebush Triage Standard. Aust N Z J Surg 1999;69:603-8.
8. Hodgetts T.J. HJ, Maconochie I., et al. Paediatric triage tape. Pre-Hospital Immed Care 1998;2:155-9.
9. Advanced Life Support Group. Major incident medical management and support: the practical approach. 2nd Ed ed. London: BMJ Books; 2005.
10. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. Lancet 1974;2:81-4.
11. Baker SP, O'Neill B, Haddon W, Jr., Long WB. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. J Trauma 1974;14:187-96.
12. Steyerberg EW. Clinical prediction models : a practical approach to development, validation, and updating. New York ; London: Springer; 2009.
13. Little RJA, Rubin DB. Statistical analysis with missing data. 2nd ed. ed. Hoboken, N.J. ; [Chichester]: Wiley; 2002.
14. Advanced Life Support Group. Advanced Paediatric Life Support:The practical Approach. 5th Edition ed. Oxford: Wiley Blackwell; 2011.
15. Landis JR, Koch GG. Measurement of Observer Agreement for Categorical Data. Biometrics 1977;33:159-74.
16. Altman DG. Practical statistics for medical research: Chapman & Hall; 1991.
17. Frykberg ER. Medical management of disasters and mass casualties from terrorist bombings: How can we cope? Journal of Trauma-Injury Infection and Critical Care 2002;53:201-12.
18. Upperman J, Pierce J. Paediatric disaster preparedness: essential concepts for paramedics. Journal of Paramedic Practice 2010;2:461-5.
19. Clark TG, Altman DG. Developing a prognostic model in the presence of missing data: an ovarian cancer case study. J Clin Epidemiol 2003;56:28-37.
20. van Dijk MR, Steyerberg EW, Stenning SP, Habbema JD. Survival estimates of a prognostic classification depended more on year of treatment than on imputation of missing values. J Clin Epidemiol 2006;59:246-53.