

Does START Triage Work? An Outcomes Assessment After a Disaster

Christopher A. Kahn, MD, MPH From the Department of Emergency Medicine, University of California, Irvine, Orange, CA (Kahn, Carl H. Schultz, MD Schultz, Miller, Anderson); and the Orange County Fire Authority, Irvine, CA (Miller).
 Ken T. Miller, MD, PhD
 Craig L. Anderson, PhD

Study objective: The mass casualty triage system known as simple triage and rapid treatment (START) has been widely used in the United States since the 1980s. However, no outcomes assessment has been conducted after a disaster to determine whether assigned triage levels match patients' actual clinical status. Researchers hypothesize that START achieves at least 90% sensitivity and specificity for each triage level and ensures that the most critical patients are transported first to area hospitals.

Methods: The performance of START was evaluated at a train crash disaster in 2003. Patient field triage categories and scene times were obtained from county reports. Patient medical records were then reviewed at all receiving hospitals. Victim arrival times were obtained and correct triage categories determined a priori using a combination of the modified Baxt criteria and hospital admission. Field and outcomes-based triage categories were compared, defining the appropriateness of each triage assignment.

Results: Investigators reviewed 148 records at 14 receiving hospitals. Field triage designations comprised 22 red (immediate), 68 yellow (delayed), and 58 green (minor) patients. Outcomes-based designations found 2 red, 26 yellow, and 120 green patients. Seventy-nine patients were overtriaged, 3 were undertriaged, and 66 patients' outcomes matched their triage level. No triage level met both the 90% sensitivity and 90% specificity requirement set forth in the hypothesis, although red was 100% sensitive (95% confidence interval [CI] 16% to 100%) and green was 89.3% specific (95% CI 72% to 98%). The Obuchowski statistic was 0.81, meaning that victims from a higher-acuity outcome group had an 81% chance of assignment to a higher-acuity triage category. The median arrival time for red patients was more than 1 hour earlier than the other patients.

Conclusion: This analysis demonstrates poor agreement between triage levels assigned by START at a train crash and a priori outcomes criteria for each level. START ensured acceptable levels of undertriage (100% red sensitivity and 89% green specificity) but incorporated a substantial amount of overtriage. START proved useful in prioritizing transport of the most critical patients to area hospitals first. [Ann Emerg Med. 2009;54:424-430.]

Provide [feedback](http://www.annemergmed.com) on this article at the journal's Web site, www.annemergmed.com.

0196-0644/\$-see front matter

Copyright © 2009 by the American College of Emergency Physicians.

doi:10.1016/j.annemergmed.2008.12.035

SEE EDITORIAL, P. 431.

INTRODUCTION

Background

More than 255 million people are affected annually by disasters, and providing medical care to the victims of such events is a daunting task.¹ The recent cyclone in Myanmar and the earthquake that struck China's Sichuan Province illustrate the magnitude of the challenge. One tool that can help optimize the initial management of mass casualties generated by these events is triage.

In the 1980s, one of the first civilian triage systems was developed in Orange County, CA.^{2,3} This system, known as

simple triage and rapid treatment (START) (Figure 1), was rapidly adopted across the United States and in some international settings as well. It was the triage standard for the Domestic Preparedness Program created by the Department of Defense. This program trained personnel in the 120 most populous cities in the United States on the management of nuclear, biological, and chemical terrorism.⁴ It now serves as the de facto national triage standard for mass casualty incidents.

Importance

Although START is nearly ubiquitous within the United States, surprisingly little research exists to support its use. Triage disaster victims with the START methodology could

Editor’s Capsule Summary

What is already known on this topic

Disaster triage is inexact. Various systems and methods exist, but there has been almost no field validation.

What question this study addressed

The study analyzed whether simple triage and rapid treatment (START), a commonly used method, was accurate in assigning acuity levels to victims of a real train crash.

What this study adds to our knowledge

START triage overtriaged 79 of the 148 victims to higher acuity levels than needed. Only 3 were undertriaged.

How this might change clinical practice

Triage levels had poor agreement with actual outcomes, and many patients were mistriaged. However, this study points the way to the development of more accurate methods.

significantly increase mortality by inappropriately assigning a low acuity status to victims with critical injuries (undertriage), thus delaying vital treatment. Conversely, assigning high acuity status to stable patients (overtriage) may result in an inundation of noncritical patients at area hospitals, consuming resources needed for the more seriously injured. The effect of overtriage on mortality is less clear but may affect overall survival.⁵⁻⁷ The magnitude of this mistriage threat remains unknown to the large number of communities in the United States that use START for disaster triage. In addition, no data exist demonstrating that START triage influences decisions about which patients should be transported to hospitals first.

A 2001 study by Garner et al,⁸ comparing START with 2 other triage systems, had significant limitations. First, the study subjects were not actually victims of a disaster. The participants involved were designated trauma patients injured individually or in small numbers, rather than representing a true mass casualty population. Second, triage categories were assigned by investigators retrospectively, solely according to the objective criteria contained within each algorithm, instead of assigned at the scene by paramedics who are actually tasked with using these systems. Last, investigators measured only the sensitivity and specificity of the immediate (red) START triage category. The rest of the algorithm was not examined.

Goals of This Investigation

It remains unclear whether START can sort disaster victims accurately. To our knowledge, this study is the first investigation to examine the effectiveness of START triage—or any mass

casualty triage algorithm—using patient outcomes for all victims assessed after an actual disaster. The authors hypothesize that START can achieve a 90% sensitivity and specificity for each triage category in sorting disaster victims and is effective in controlling scene evacuation so that the most critical patients are transported first to area hospitals.

MATERIALS AND METHODS

Study Design and Setting

This study is a retrospective analysis of a crash involving a commuter train carrying 262 persons that crashed head-on with a freight train carrying 2 persons on April 23, 2002. Paramedics dispatched to the scene used START triage to categorize victim acuity per their usual fire department protocol.

Approvals were obtained from the institutional review boards for the University of California, Irvine and the Orange County Health Care Agency to examine these patients’ records. Waivers of informed consent and Health Insurance Portability and Accountability Act authorization were also granted. Institutional review board approval or other authorization was additionally obtained at each receiving hospital. For patients who were transferred within the first 6 hours of arrival, similar approvals were obtained from the next receiving facility.

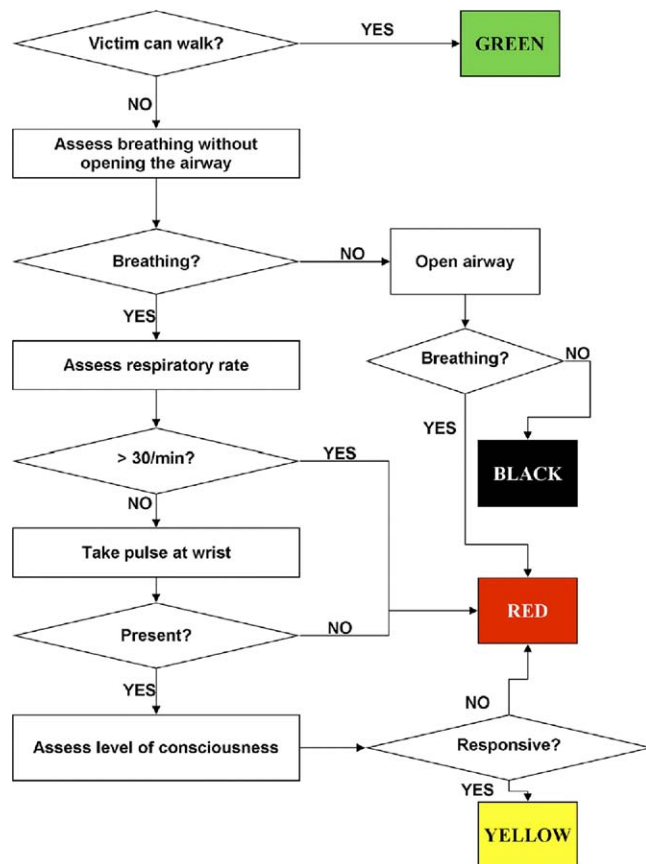


Figure 1. START algorithm.

Chest decompression (needle or tube thoracostomy)

Intravenous fluids for a systolic blood pressure <90 mm Hg, or absence of radial pulse

Blood transfusion

Assisted ventilation or airway procedure

Invasive central nervous system monitoring with brain imaging or other evidence of increased intracranial pressure

Nonorthopedic operation (except pelvic stabilization) with positive findings

Figure 2. Modified Baxt criteria.

Data Collection and Processing

Investigators used several methods to identify victims, their paramedic-assigned field triage level, transport times, and the hospitals to which they were sent. The National Transportation Safety Board reviewed this event and recorded the names and hospital destinations of victims. This information was given to the study team. Researchers examined records from the Orange County Emergency Medical Services Agency, which listed victim triage status, scene departure times, and hospital destinations. Additionally, emergency department (ED) triage logs from the date of the crash were reviewed for further verification.

Data on patient outcomes were obtained by abstracting hospital medical records. Each receiving hospital (or, in the case of one subsequently closed facility, the custodian of records) was contacted to locate charts. Charts were reviewed by the same 2 investigators (C.A.K., C.H.S.) at all hospitals, with the exception of one facility that required that 5 charts be reviewed only by the county EMS medical director and assistant medical director, because of HIPAA concerns. One of these 2 individuals was also a coinvestigator (K.T.M.). Because all but 5 charts were abstracted by the same reviewers simultaneously, investigators did not calculate a κ statistic. Data were abstracted with a standardized data collection instrument (Appendix E1, available at <http://www.annemergmed.com>). Data on arrival and discharge times were obtained from the ED medical record or inpatient nursing notes. Triage acuity was determined by the presence of a triage tag in the medical record, the scene departure and hospital arrival times compared with the EMS transport data identifying patient triage status with such times, and initial nursing notes identifying triage status. Points of ambiguity during data abstraction were clarified by consensus among the reviewing investigators. For patients who were transferred within the first 6 hours of arrival, charts were also obtained from the next receiving facility for review.

Outcome Measures

The modified Baxt criteria were defined, a priori, as the outcomes criteria for this study (Figure 2).^{8,9} These criteria,

when met in the field or within 6 hours of hospital arrival, signify the presence of immediately life-threatening conditions. Accordingly, patients meeting these criteria were considered to fall within the red, or “immediate,” outcome category.

Patients who did not meet the modified Baxt criteria, but were admitted to the hospital for at least 24 hours, were considered to fall within the yellow, or “delayed,” outcome category. Patients who did not meet either of the above criteria were considered to fall within the green, or “minor,” outcome category.

Primary Data Analysis

The data were analyzed using Stata (version 9.2; StataCorp, College Station, TX). Sensitivity, specificity, positive and negative predictive values, and likelihood ratios were calculated for each category with a stepwise progression consistent with the application of START triage (C.L.A.). Patients were first examined within the “green/not green” pair and were then considered within the “red/yellow” pair. This grouping most accurately reflects the application of START triage, in which patients are first assigned to the green group or the “not green” group and are then further stratified into black, red, or yellow. The descriptive statistics reported for the green triage level reflect this application and are reported with “red/yellow” being considered the negative outcome and green considered the positive outcome. The black, or “deceased,” category was not examined. Only 1 patient was tagged black on the scene, providing an insufficient sample size for any meaningful comparisons. Although predictive values depend heavily on prior probabilities, which vary from incident to incident, they are included to assist in the description of instrument performance at this specific incident.

Two summary statistics, overall accuracy and the Obuchowski statistic, were also calculated. The latter statistic was calculated assigning a loss function of 1 for all cases without agreement, with a routine written in Mata, the Stata matrix language (C.L.A.). We verified that the results of our routine agreed with a published example.¹⁰ The Obuchowski statistic is interpreted as the probability that, in any randomly selected pair of subjects with different outcome classes, the subject with the more severe outcome had a more severe field triage score.¹⁰ This statistic is used analogously to the area under a receiver operating characteristic curve, in which 0.5 indicates random allocation and 1.0 indicates perfect coding. Unlike a receiver operating characteristic curve, however, the Obuchowski statistic is designed for use with nonbinary data. Bias-corrected bootstrap confidence intervals (CIs) for the Obuchowski statistic were calculated by sampling with replacement from the study data, with 20,000 repetitions.¹¹

The target of 90% sensitivity and specificity was selected because it is thought to represent the best overall compromise between an ideal triage system (with greater accuracy) and one that is simple enough to apply in an actual disaster. Systems yielding higher sensitivity and specificity are overly complex and impractical for field use under disaster conditions.¹²

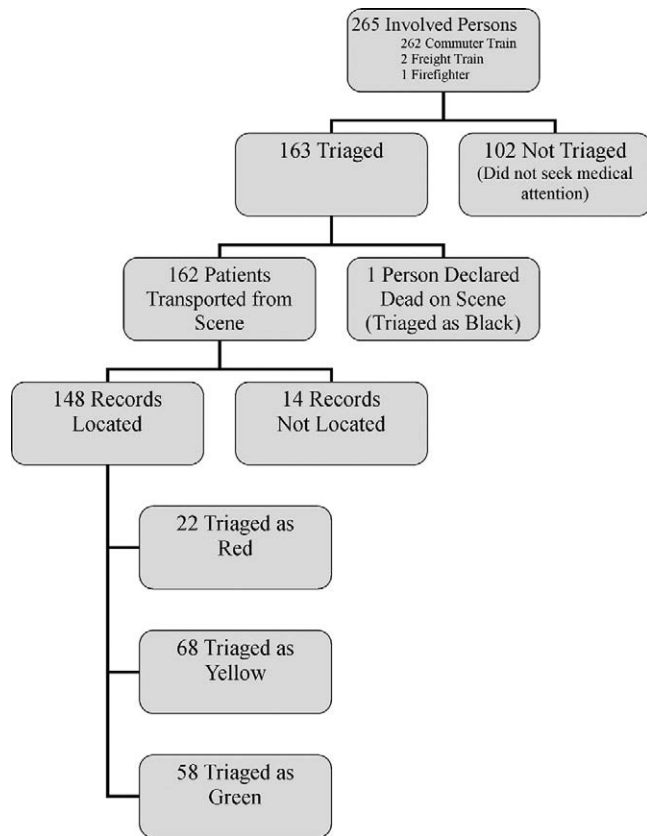


Figure 3. Flow diagram.

For each of the 3 triage groups, investigators calculated the median time from the train crash to the patients’ arrival at the first hospital. These means were compared by using the Kruskal-Wallis rank test.

A nonproprietary index to the Social Security Death Master File was searched to determine whether any known victims of the train crash died in the subsequent 30 days.¹³

RESULTS

A total of 163 persons were triaged and reported to the Orange County communications center, which at that time was responsible for overall coordination of mass casualty incidents (Figure 3). With the exception of one on-scene fatality, these patients were transported to 13 receiving hospitals.

The authors reviewed 148 patient records at 13 receiving hospitals. One patient’s records were reviewed at a 14th hospital, to which he was transferred within 6 hours of arrival. Of these patients, 22 were triaged as red, 68 were triaged as yellow, and 58 were triaged as green. With the a priori outcomes criteria as the determinant, 2 patients were truly red, 26 patients were truly yellow, and 120 patients were truly green. (Table 1), which represents 3 patients who were undertriaged by 1 level (ie, had a yellow outcome but were triaged as green), 79 patients who were overtriaged by 1 or 2 levels, and 66 patients whose outcomes matched their triage levels.

Table 1. Frequencies of triage and outcome levels.

Triage Level	Outcome Level			Total
	Red	Yellow	Green	
Red	2	14	6	22
Yellow	0	9	59	68
Green	0	3	55	58
Total	2	26	120	148

The overall accuracy of START was 44.6%. The Obuchowski statistic is 0.81 (95% CI 0.71 to 0.89), meaning that in any randomly selected pair of subjects with different outcomes, the subject who met the higher outcome criteria had an 81% probability of receiving a higher field triage score.¹⁰

Statistics describing the performance of each triage level are listed in Table 2. For one patient, it was not possible to determine whether the assigned triage level was red or yellow. However, a sensitivity analysis did not reveal any significant change in these statistics. Likelihood ratios describing the function of each triage level are also listed in Table 2.

The median elapsed time from the moment of the train crash to patient arrival at receiving hospitals was 1.29 hours (95% CI 1.17 to 1.67 hours) for patients triaged red, 2.35 hours (95% CI 2.25 to 2.5 hours) for patients triaged yellow, and 2.33 hours (95% CI 2.33 to 2.33 hours) for those triaged green (Figure 4). (The CI for green is a single point because 18 patients were recorded as having the same arrival time.) The distribution of the times to arrival was different for the 3 groups ($P=.0001$; Kruskal-Wallis rank test), but there was no difference between the yellow and green groups ($P=.10$).

The Social Security Death Master File search did not reveal any deaths among victims of this train crash in the subsequent 30 days that were not already revealed by hospital records from the initial admission.¹³

LIMITATIONS

This investigation has several limitations. Most notably, the study methodology could not discern whether errors in assignment of triage categories resulted from failure of the triage algorithm as a tool or failure of emergency personnel to apply it correctly. Researchers did observe that some of the assigned triage levels differed from what strict application of the START algorithm would have mandated. This is evident from discrepancies in the out-of-hospital and hospital care records, such as documentation of patients “walking on scene” for 6 persons triaged as yellow and 1 person triaged as red, all of whom met the green outcomes criteria. However, this intention-to-treat analysis was not designed to identify why each victim received the triage category assigned, so it is not possible to determine where the errors occurred except for the 7 individuals previously mentioned. These types of errors probably contributed to an overtriage bias, decreasing the apparent specificity of the system. Also, there are a small number of lost records. Investigators could not review 14 charts

Table 2. Descriptive statistics by triage level.

Triage Level	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Positive Predictive Value, % (95% CI)	Negative Predictive Value, % (95% CI)	Positive Likelihood Ratio (95% CI)	Negative Likelihood Ratio (95% CI)
Red	2/2 100 (15.8–100)	68/88 77.3 (67.1–85.5)	2/22 9.1 (1.1–29.2)	68/68 100 (94.7–100)	4.4 (3.0–6.5)	0*
Yellow	9/23 39.1 (19.7–61.5)	8/67 11.9 (5.3–22.2)	9/68 13.2 (6.2–23.6)	8/22 36.4 (17.2–59.3)	0.44 (0.26–0.75)	5.1 (2.5–10.6)
Green	55/120 45.8 (36.7–55.2)	25/28 89.3 (71.8–97.7)	55/58 94.8 (85.6–98.9)	25/90 27.8 (18.9–38.2)	4.3 (1.4–12.7)	0.61 (0.49–0.75)

*Unable to calculate a negative likelihood ratio CI for a value of zero.

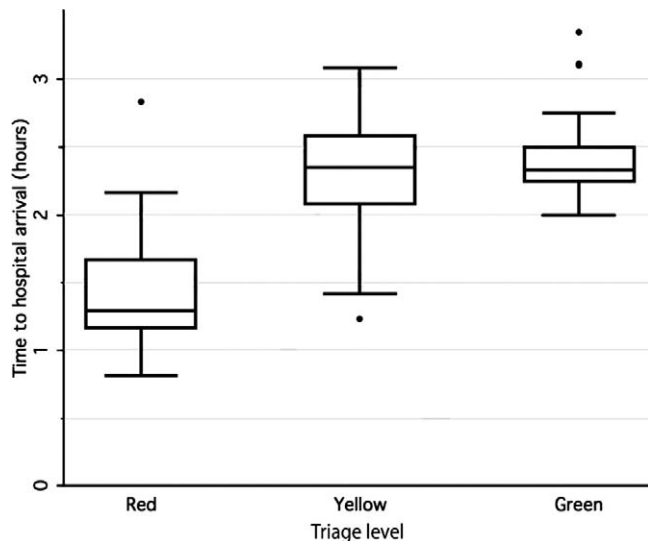


Figure 4. Time analysis. The box indicates the 25th percentile, the median, and the 75th percentile. The height of the box indicates the interquartile range (IQR). The whiskers indicate the most extreme values that are within 1.5 IQR of the box. The dots indicate more extreme values.

because they were missing or contained no data. This is consistent with the National Transportation Safety Board's finding that 20 records could not be located for patients who were identified as having been transported; the National Transportation Safety Board's presumption is that these patients were uninjured.¹⁴ The lack of children in the study population did not affect the analysis because START is not intended for use in triaging children. Because of the small number of victims who died on scene, any potential analysis of the black category is not statistically meaningful.

DISCUSSION

Disasters represent a significant threat to populations of all nations regardless of economic status. Hurricanes, tsunamis, earthquakes, pandemics, and acts of terrorism have the potential to overwhelm the medical resources of even highly developed nations. If such a disaster occurs in the United States, it is likely

that first responders will use START triage for the initial assessment of these victims.

Overall, this investigation demonstrated poor agreement between START triage categories assigned at the scene of the train crash and the a priori outcomes criteria for each triage category. START adequately identifies many patients with minor injuries but poorly discriminates between those with immediate life threats and those with significant but more stable injuries. No triage level met both the 90% sensitivity and 90% specificity requirement set forth in the hypothesis. The use of START did ensure that almost all patients received at least as much care as was needed but incorporated a significant amount of overtriage, which may be wasteful of potentially limited resources.

The 3 patients triaged as green who met yellow outcomes criteria were each admitted for more than 24 hours but did not require any resource-intensive intervention. One patient had an anterior chip fracture of a vertebra, one had a central cord syndrome (observed to be walking in the ED), and the last was diagnosed with concussion and a rib fracture. Of the 2 patients meeting red criteria, 1 died on arrival at the receiving hospital; the second, who had serious injuries and a prolonged hospital course, died approximately 6 weeks later.

Of the 6 red-triaged, green-outcomes patients, 3 had possible brief loss of consciousness or lack of recall, 1 had a respiratory rate of 30 breaths/min with labored and shallow respirations (who was observed to be "anxious" in the ED), 1 hit his or her head and had a history of brain surgery but was complaining only of back pain and was without neurologic deficits, and 1 had no indication why he or she was triaged as red.

The 20 red-triaged, nonred outcomes patients did not have any time-dependent immediate interventions, although this is by necessity a subjective account, given the nonstrict definition of "time dependent." One did go to the operating room within 6 hours of the crash, but this was for repair of knee lacerations (not open fractures). Some did have laceration repairs in the ED. The majority of yellow-outcomes (admitted) patients had lacerations, fractures, and contusions.

Two important findings about the potential for undertriage by START emerged. The "walking filter" that defines the green triage level appears to have functioned well in identifying a less-severely injured group of victims, with a specificity of

risks exacerbating disaster morbidity and mortality while suboptimally using resources. Additional outcomes-based assessments of mass casualty triage (both START and other systems) are imperative for further development of triage systems. These analyses must use data from actual disasters because studies to date based on simulations have failed to predict the results found in this investigation. Without such inquiry, it will not be possible to compare systems meaningfully, refine methodology, and possibly standardize currently divergent triage protocols. In addition, these investigations must have significantly fine granularity to distinguish between errors in applying the algorithms and failure of the triage tools themselves.

Supervising editor: Jonathan L. Burstein, MD

Author contributions: CAK and CHS conceived the study and designed the trial. CAK obtained research funding. CAK and CHS supervised the conduct of the trial and data collection. CAK, CHS, and KTM undertook recruitment of participating centers and collected data. CAK, CHS, and CLA managed the data, including quality control. CLA provided statistical advice on study design and analyzed the data, with assistance from CAK and CHS. CAK drafted the article, and all authors contributed substantially to its revision. CAK and CHS take responsibility for the paper as a whole.

Funding and support: By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article that might create any potential conflict of interest. See the Manuscript Submission Agreement in this issue for examples of specific conflicts covered by this statement. This article was supported by an F32 fellowship training grant from the Agency for Healthcare Research and Quality, awarded to Christopher Kahn, #HS-15768.

Publication dates: Received for publication September 21, 2008. Revision received December 5, 2008. Accepted for publication December 19, 2008. Available online February 5, 2009.

Presented as an oral abstract at the Society for Academic Emergency Medicine annual meeting, Chicago, IL, May 2007.

Reprints not available from the authors.

Address for correspondence: Christopher A. Kahn, MD, MPH, Department of Emergency Medicine, University of California, Irvine, 101 The City Drive, South, Route 128-01, Orange, CA 92868; 714-456-5239, fax 714-456-5390; E-mail ckahn@uci.edu.

REFERENCES

1. Guha-Sapir D, Hargitt D, Hoyois P. *Thirty Years of Natural Disasters 1974-2003: The Numbers*. Louvain-la-Neuve, Belgium: Presses Universitaires de Louvain; 2004.
2. Schultz CH, Koenig KL, Noji EK. A medical disaster response to reduce immediate mortality after an earthquake. *N Engl J Med*. 1996;334:438-444.
3. Super G, Groth S, Hook R, et al. *START: Simple Triage and Rapid Treatment Plan*. Newport Beach, CA: Hoag Memorial Presbyterian Hospital; 1994.
4. *Domestic Preparedness Training Program Instructor Guide*. Soldier and Biological Chemical Command, Department of Defense; 1999.
5. Frykberg ER. Medical management of disasters and mass casualties from terrorist bombings: how can we cope? *J Trauma*. 2002;53:201-212.
6. Hirshberg A. Multiple casualty incidents: lessons from the front line. *Ann Surgery*. 2004;239:322-324.
7. Hupert N, Hollingsworth E, Xiong W. Is overtriage associated with increased mortality? insights from a simulation model of mass casualty trauma care. *Disaster Med Public Health Prep*. 2007;1(1 suppl):S14-24.
8. Garner A, Lee A, Harrison K, et al. Comparative analysis of multiple-casualty incident triage algorithms. *Ann Emerg Med*. 2001;38:541-548.
9. Baxt WG, Upenieks V. The lack of full correlation between the Injury Severity Score and the resource needs of injured patients. *Ann Emerg Med*. 1990;19:1396-1400.
10. Obuchowski NA. Estimating and comparing diagnostic tests' accuracy when the gold standard is not binary. *Acad Radiol*. 2005;12:1198-1204.
11. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. *Stat Med*. 2000;19:1141-1164.
12. Sacco WJ, Navin DM, Fiedler KE, et al. Precise formulation and evidence-based application of resource-constrained triage. *Acad Emerg Med*. 2005;12:759-770.
13. Social Security Death Index Interactive Search. [Rootsweb.com](http://ssdi.rootsweb.ancestry.com). Available at: <http://ssdi.rootsweb.ancestry.com>. Accessed April 6, 2008.
14. Engleman EG, Rosenker MV, Goglia JJ, et al. *Collision of Burlington Northern Santa Fe Freight Train With Metrolink Passenger Train, Placentia, California, April 23, 2002*. National Transportation Safety Board; 2003.
15. Risavi BL, Salen PN, Heller MB, et al. A two-hour intervention using START improves prehospital triage of mass casualty incidents. *Prehosp Emerg Care*. 2001;5:197-199.
16. 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;27:570-575.
17. Aylwin CJ, Konig TC, Brennan NW, et al. Reduction in critical mortality in urban mass casualty incidents: analysis of triage, surge, and resource use after the London bombings on July 7, 2005. *Lancet*. 2006;368:2219-2225.
18. National Transportation Safety Review Board. 49CFR §830.2.

PRE-HOSPITAL DATA

- 1) Age:
- 2) Sex: M F
- 3) Initial VS: HR RR BP / Field? ED?
- 4) Triage level: Green Yellow Red Black
- 5) Triage criteria: (Assumed or documented)
- 6) Retriage? Y N [From: G Y R B To: G Y R B]
- 7) If "Y", why?
- 8) Logistics: Scene Depart Time Hosp Arrival Time
- 9) Extricated? Y N

HOSPITAL DATA

- 10) "Red" outcome criteria met: Y N
- 11) If "Y", which "red" criteria?
 - Chest decompress (field/ED)
 - IVF (field/ED) for SBP<90 or for no palpable pulse, or transfused
 - Assist vent. or airway proc (field/ED)
 - Invasive CNS monitor w/+ CTH or evidence of high ICP
 - Non-ortho OR w/in 6hrs (except pelvic stabilization) w/+ findings (Detail procedures/findings/rads in above box)
- 12) "Yellow" outcome criteria met (hosp stay >24 hrs, not "red"): Y N
- 13) "Green" outcome criteria met (not admitted, not "red" or "yellow"): Y N
- 14) Admitted to hospital? Y N
- 15) Length of stay (hours):-----
- 16) Discharge diagnoses:
 - 1)
 - 2)
 - 3)
- 17) Hospital name:
- 18) Distance from scene to hospital (miles):

CALCULATED DATA

- 19) Time on scene (event → departure):
- 20) Transport (departure → hospital):

.....
Triage tag number: Victim ID number:

Appendix E1. Data collection instrument.