

UC Irvine

UC Irvine Previously Published Works

Title

Does START triage work? An outcomes assessment after a disaster.

Permalink

<https://escholarship.org/uc/item/89p3t51j>

Journal

Annals of emergency medicine, 54(3)

ISSN

1097-6760

Authors

Kahn, Christopher A
Schultz, Carl H
Miller, Ken T
[et al.](#)

Publication Date

2009-09-01

DOI

10.1016/j.annemergmed.2008.12.035

Peer reviewed

Does START Triage Work? An Outcomes Assessment after a Disaster

Abstract

Objective

The mass casualty triage system known as START (Simple Triage and Rapid Treatment) has been widely utilized 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 hypothesized 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 over-triaged, three were under-triaged, 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% CI 15.8-100) and green was 89.3% specific (95% CI 71.8-97.7). 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. Red patients arrived at hospitals 0.92 hours (95% CI 0.71-1.1) faster than other patients.

Conclusions

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 under-triage (100% red sensitivity and 89% green specificity) but incorporated a significant amount of over-triage. START proved useful in prioritizing transport of the most critical patients to area hospitals first.

Introduction

Background

Over 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, California.^{2,3} This system, known as START Triage (Simple Triage and Rapid Treatment, 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. It is possible that triaging disaster victims utilizing the START methodology could significantly increase mortality by inappropriately assigning a low acuity status to victims with critical injuries (under-triage), thus delaying vital treatment. Conversely, assigning high acuity status to stable patients (over-triage) may result in an inundation of non-critical patients at area hospitals, consuming resources needed for the more seriously injured. The impact of over-triage on mortality is less clear

but may impact overall survival.⁵⁻⁷ The magnitude of this mistriage threat remains unknown to the large number of communities in the U.S. that utilize START for disaster triage. In addition, no data exist demonstrating that START triage influences decisions regarding which patients should be transported to hospitals first.

A 2001 study by Garner, et al, comparing START with two other triage systems suffered from significant limitations.⁸ 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, based solely on the objective criteria contained within each algorithm, instead of assigned at the scene by paramedics who are actually tasked with using these systems. Lastly, investigators only measured 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. 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 following an actual disaster event. 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.

Methods

Study Design and Setting

This study is a retrospective analysis of a collision involving a commuter train carrying 262 persons that impacted head-on with a freight train carrying two persons on April 23, 2002. Paramedics dispatched to the scene employed 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 HIPAA authorization were also granted. IRB approval or other authorization was additionally obtained at each receiving hospital. For patients who were transferred within the first six hours of arrival, similar approvals were obtained from the next receiving facility.

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 (NTSB) 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 triage logs from the date of the collision 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 two

investigators (CAK, CHS) at all hospitals, with the exception of one facility which required that five charts be reviewed only by the county EMS medical director and assistant medical director, due to HIPAA concerns. One of these two individuals was also a co-investigator (KTM). Since all but five charts were abstracted by the same reviewers simultaneously, investigators did not calculate a kappa statistic. Data were abstracted using a standardized data collection instrument. Data on arrival and discharge times were obtained from the emergency department medical record or in-patient nursing notes. Triage acuity was determined by the presence of a triage tag in the medical record, the scene departure and hospital arrival times when compared to 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 six 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 (Table 1).^{8,9} These criteria, when met in the field or within six 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, Stata Corp., College Station, TX). Sensitivity, specificity, positive and negative predictive values, and likelihood ratios were calculated for each category using a stepwise progression consistent with the application of START triage (CLA). 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 one patient was tagged black on the scene, providing an insufficient sample size for any meaningful comparisons. Although predictive values depend heavily upon 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, using a routine written in Mata, the Stata matrix language (CLA). 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, where 0.5 indicates random allocation and 1.0

indicates perfect coding. Unlike an ROC curve, however, the Obuchowski statistic is designed for use with non-binary data. Bias-corrected bootstrap confidence intervals 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

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

A non-proprietary index to the Social Security Death Master File was searched to determine if any known victims of the train collision died in the subsequent thirty 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 2). With the exception of one on-scene fatality, these patients were transported to thirteen separate 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 six hours of

arrival. Of these patients, 22 were triaged as red, 68 were triaged as yellow, and 58 were triaged as green. Using the a priori outcomes criteria as the determinant, two patients were truly red, 26 patients were truly yellow, and 120 patients were truly green. (Table 2) This represents three patients who were under-triaged by one level (i.e., had a yellow outcome but were triaged as green), 79 patients who were over-triaged by one or two levels, and 66 patients whose outcomes matched their triage levels.

The overall accuracy of START was 44.6%. The Obuchowski statistic is 0.81 (95% CI 0.71-0.89) 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 3. 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 3.

The median elapsed time from the moment of the train collision to patient arrival at receiving hospitals was 1.29 hours (95% CI 1.17-1.67) for patients triaged red, 2.35 hours (95% CI 2.25-2.5) for patients triaged yellow, and 2.33 hours (95% CI 2.33-2.33) for those triaged green. (The confidence interval 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 three 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 collision in the subsequent thirty days which 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 prehospital and hospital care records, such as documentation of patients “walking on scene” for six persons triaged as yellow and one 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 seven individuals previously mentioned. These types of errors probably contributed to an over-triage bias, decreasing the apparent specificity of the system. Also, there are a small number of lost records. Investigators could not review 14 charts because they were missing or contained no data. This is consistent with the NTSB’s finding that 20 records could not be located for patients who were identified as having been transported; the NTSB’s presumption is that these patients were uninjured.¹⁴ The lack of children in the study population did not affect the analysis, as START is not intended for use in triaging

children. Due to 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 collision 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 over-triage which may be wasteful of potentially limited resources.

The three patients triaged as green that met yellow outcomes criteria were each admitted for over 24 hours, but did not require any resource-intensive intervention. One patient suffered an anterior chip fracture of a vertebra, one had a central cord syndrome (noted to be walking in the emergency department), and the last was diagnosed with concussion and a rib fracture. Of the two patients meeting red criteria, one died upon

arrival at the receiving hospital; the second, who had serious injuries and a prolonged hospital course, died approximately six 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 with labored and shallow respirations (who was noted to be “anxious” in the ED), 1 hit his or her head and had a history of brain surgery, but was only complaining of back pain without neurological deficits, and 1 had no indication why he or she was triaged as red.

The 20 red-triaged, non-red outcomes patients did not have any time-dependent immediate interventions, although this is by necessity a subjective account given the non-strict definition of “time-dependent”. One did go to the OR within six hours of the collision, 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 regarding the potential for under-triage by START emerged. The “walking filter” which defines the green triage level appears to have functioned well in identifying a less-severely injured group of victims, with a specificity of approximately 90%. Although there were only a small number of critically injured victims, each received an appropriate red triage designation, resulting in a sensitivity of 100%. Therefore, the risk for under-triage, represented by the sensitivity of the red category and the specificity of the green category, is acceptable as defined by an outcome of at least 90% for each parameter.

Over-triage was a frequent occurrence, noted in 79 of our 148 patients. Although on-scene information for individual patients is sparse, common themes noted on review

of these charts including patients walking on scene assigned to non-green triage levels, conscious patients with inability to recall the collision being assigned to the red triage level (presumably for mental status reasons), and patients being placed on backboards being considered non-ambulatory; this last category includes 18 yellow- and 4 red-triaged patients who met green outcomes criteria, and 3 yellow- and 7 red-triaged patients who did not meet green outcomes criteria.

The use of START by paramedics was moderately effective in prioritizing patients for transport to receiving hospitals. The elapsed time for victims triaged as red was significantly less than for the yellow and green groups. Investigators found no difference in times between the latter two victim categories. The fact that essentially all green patients were transported together simultaneously in two large buses instead of individually or in small groups may have contributed to this finding.

Investigators' determination that the Obuchowski statistic equaled 0.81 means that of any randomly selected pair of subjects, there is an 81% chance that the subject meeting the higher triage outcome criteria also received the higher field triage score. Although this statistic has less clinical value in comparison to sensitivity and specificity, it serves as a useful overall standard which will allow meaningful comparison to other triage systems when studied in a similar fashion.

Additional studies have examined the ability of paramedics and other emergency response personnel to learn START and apply it accurately in the setting of "tabletop" or other drill scenarios, and found that training improves the ability to perform START in an artificial setting.¹⁵ The accuracy of START category assignments during a drill with no previous refresher training has also been examined and shown to be moderate

(approximately 78% accurate).¹⁶ However, these studies do not address the application of START at an actual mass casualty incident.

In fact, clear evidence supporting the efficacy of any system of mass casualty triage employed by any nation after a disaster is limited. An investigation published by Aylwin, et al in 2006 attempted to examine the British system of triage after the London subway bombings.¹⁷ They measured the accuracy of triage decisions using an ISS score of 16 or greater to define the red category. However, no objective criteria for the yellow patients were identified, and no final outcome evaluation of patients triaged as green was described. Investigators also combined the red and yellow patients into one category for purposes of prioritizing field transport to area hospitals. This is problematic, as the two groups represent different severities of injury. Lastly, victim triage and outcome data were reported for only one hospital, so detailed information on victims sent to other hospitals is missing.

The method of determining outcomes for this START triage study deserves discussion. Investigators selected the previously validated modified Baxt criteria to establish victim acuity. A National Transportation Safety Board review of the train collision used a different set of criteria for description of injury outcomes (49 CFR 830.2), and reported 265 victims, of which two died, 22 were seriously injured, 119 had minor injuries, and 122 were not injured.¹⁴ As these criteria have not been validated in the context of appropriate utilization of limited resources, they were not selected.

Another method of determining outcomes utilizes an Injury Severity Score (ISS) of 16 or greater to differentiate red from yellow victims.¹⁷ This method was considered but rejected. It is not clear from a resource utilization perspective how significant the

difference is between a victim with an ISS of 15 (yellow) compared to one with a score of 16 (red). Does one require an ICU while the other does not? This is akin to the problems observed in the NASCIS 2 trial, where improvements in a neurologic score were observed in spinal cord injury patients after treatment with high dose methylprednisolone. Although scores improved, a clear improvement in clinically relevant outcomes was difficult to discern (e.g., ability to feed oneself, return of bladder/bowel function, ability to transfer). Those involved with the initial management of disaster victims need information on potential resource requirements; accordingly, the modified Baxt criteria were selected, as they provide more relevant clinical information on potential patient acuity.

Although the modified Baxt criteria were felt to represent the best existing outcomes standard for evaluating mass casualty triage algorithms, it is likely that a standard more specifically suited to this end could be developed. This would provide a more accurate method for evaluating triage algorithms, ultimately leading to improved care of disaster victims. Such a criterion standard would need to consider resource utilization and availability, victim condition, and outcomes to be of significant utility.

By sheer coincidence, the collision occurred approximately one hour prior to the scheduled commencement of a county-wide mass casualty drill, located within two miles of the collision site. Paramedics were prepared to triage mass casualties as part of the drill. Many private ambulance and other personnel (e.g., ham radio operators, American Red Cross volunteers) had already gathered at the drill site, and were immediately redirected to the collision scene. It is therefore improbable that there was any significant delay in emergency response which may have affected the triage outcomes.

It is critical that mass casualty triage algorithms produce accurate and reliable outcomes. Otherwise, the resultant mistriage 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 utilize data from actual disasters, as 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.

References

1. Guha-Sapir D, Hargitt D, Hoyois P. *Thirty Years of Natural Disasters 1974-2003: The Numbers*. 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*. Feb 15 1996;334(7):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: Chemical and Biological Defense Command*, Department of Defense; 1998.
5. Frykberg ER. Medical management of disasters and mass casualties from terrorist bombings: how can we cope? *J Trauma*. Aug 2002;53(2):201-212.
6. Hirshberg A. Multiple casualty incidents: lessons from the front line. *Ann Surgery*. Mar 2004;239(3):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 Medicine and Public Health Preparedness*. Sep 2007;1(1 Suppl):S14-24.
8. Garner A, Lee A, Harrison K, Schultz CH. Comparative analysis of multiple-casualty incident triage algorithms. *Ann Emerg Med*. Nov 2001;38(5):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*. Dec 1990;19(12):1396-1400.

10. Obuchowski NA. Estimating and comparing diagnostic tests' accuracy when the gold standard is not binary. *Academic radiology*. Sep 2005;12(9):1198-1204.
11. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine*. May 15 2000;19(9):1141-1164.
12. Sacco WJ, Navin DM, Fiedler KE, Waddell RK 2nd, Long WB, Buckman RF Jr. Precise formulation and evidence-based application of resource-constrained triage. *Academic Emergency Medicine*. Aug 2005;12(8):759-770.
13. Social Security Death Index Interactive Search. *Rootsweb.com* (Accessed April 6, 2008 at <http://ssdi.rootsweb.ancestry.com>.)
14. Engleman EG, Rosenker MV, Goglia JJ, Carmody CJ, Healing RF. 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, Arcona S. A two-hour intervention using START improves prehospital triage of mass casualty incidents. *Prehosp Emerg Care*. Apr-Jun 2001;5(2):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*. Sep-Oct 2006;27(5):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*. Dec 23 2006;368(9554):2219-2225.

Figure Legends

Figure 1. START algorithm.

Figure 2. Flow diagram.

Figure 3. Time analysis.

To be inserted below the figure:

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.

Table 1. Modified Baxt Criteria

- Chest decompression (needle or tube thoracostomy)
- Intravenous fluids for a systolic blood pressure <90 , or absence of radial pulse
- Blood transfusion
- Assisted ventilation or airway procedure
- Invasive central nervous system monitoring with brain imaging or other evidence of elevated intracranial pressure
- Non-orthopedic operation (except pelvic stabilization) with positive findings

Table 2. Frequencies of triage and outcome levels.

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

Table 3. Descriptive statistics by triage level.

	Sensitivity (%, 95% CI)	Specificity (%, 95% CI)	Positive predictive value (%, 95% CI)	Negative predictive value (%, 95% CI)	Likelihood ratio – positive (LR+, 95% CI)	Likelihood ratio – negative (LR-, 95% CI)
Red	100 (15.8-100)	77.3 (67.1-85.5)	9.1 (1.1-29.2)	100 (94.7-100)	4.4 (3.0-6.5)	0*
Yellow	39.1 (19.7-61.5)	11.9 (5.3-22.2)	13.2 (6.2-23.6)	36.4 (17.2- 59.3)	0.44 (0.26-0.75)	5.1 (2.5-10.6)
Green	45.8 (36.7-55.2)	89.3 (71.8-97.7)	94.8 (85.6- 98.9)	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 confidence interval for a value of zero.

Figure 1. START algorithm.

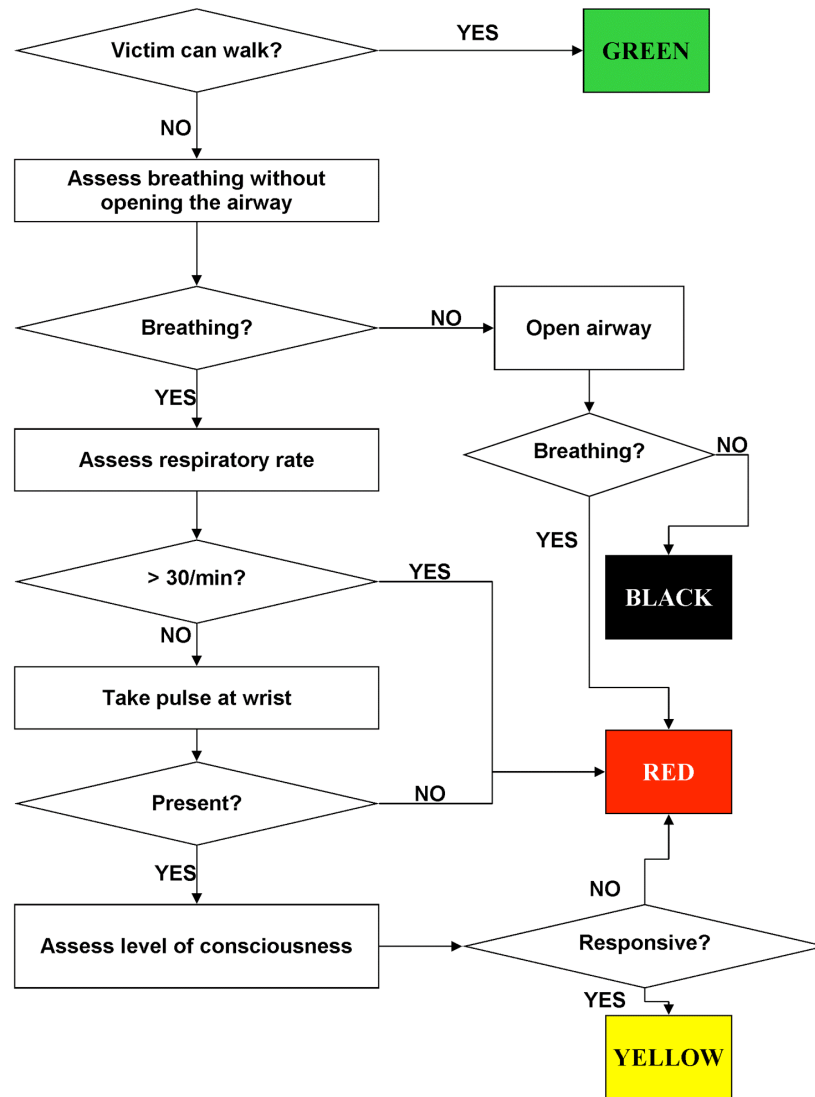


Figure 2. Flow diagram.

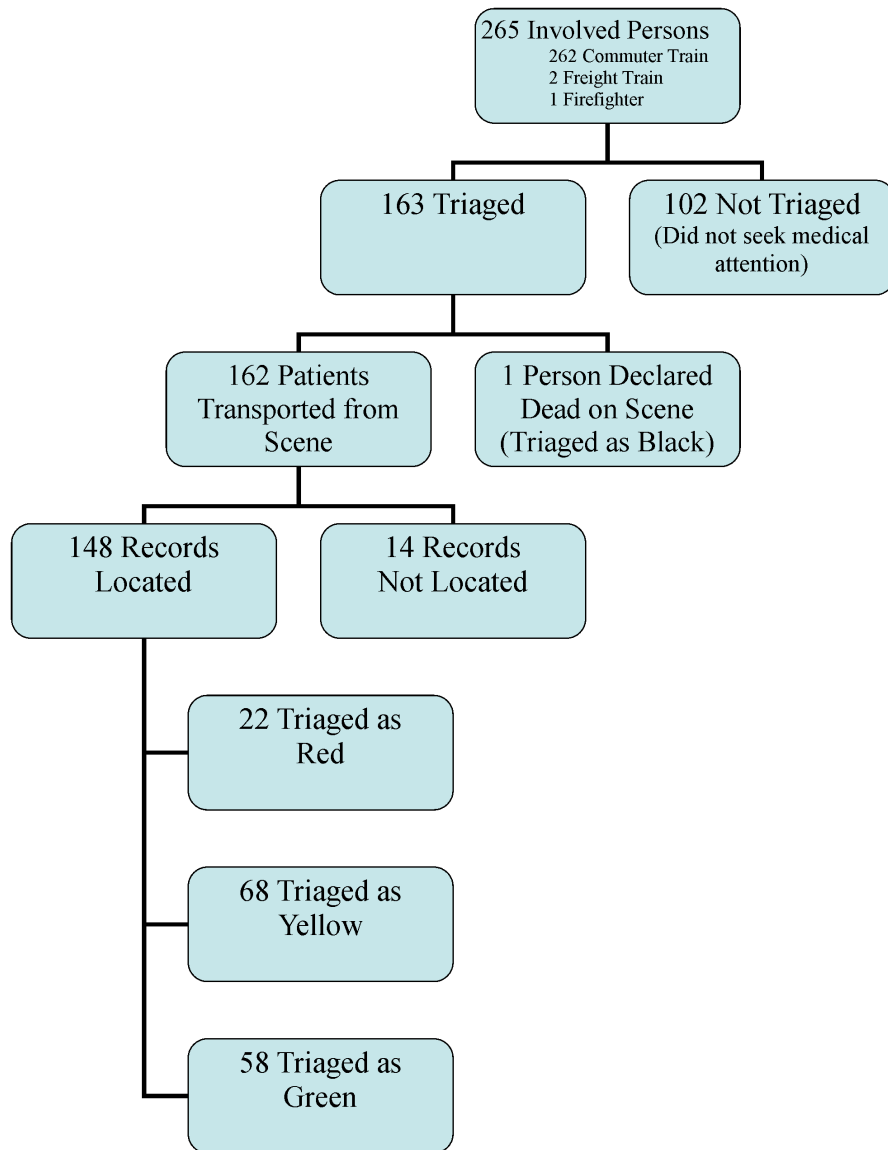


Figure 3. Time analysis.

