ReSTART: A Novel Framework for Resource-Based Triage in Mass-Casualty Events

Alex F. Mills, PhD; Nilay T. Argon, PhD; Serhan Ziya, PhD; Brian Hiestand, MD, MPH; James Winslow, MD

ABSTRACT

Objective: Current guidelines for mass-casualty triage do not explicitly use information about resource availability. Even though this limitation has been widely recognized, how it should be addressed remains largely unexplored. The authors present a novel framework developed using operations research methods to account for resource limitations when determining priorities for transportation of critically injured patients. To illustrate how this framework can be used, they also develop two specific example methods, named ReSTART and Simple-ReSTART, both of which extend the widely adopted triage protocol Simple Triage and Rapid Treatment (START) by using a simple calculation to determine priorities based on the relative scarcity of transportation resources. **Methods:** The framework is supported by three techniques from operations research: mathematical analysis, optimization, and discrete-event simulation. The authors' algorithms were developed using mathematical analysis and optimization and then extensively tested using 9,000 discrete-event simulations on three distributions of patient severity (representing low, random, and high acuity). For each incident, the expected number of survivors was calculated under START, ReSTART, and Simple-ReSTART. A web-based decision support tool was constructed to help providers make prioritization decisions in the aftermath of mass-casualty incidents based on ReSTART. **Results:** In simulations, ReSTART resulted in significantly lower mortality than START regardless of which severity distribution was used (paired *t* test, $p < .01$). Mean decrease in critical mortality, the percentage of immediate and delayed patients who die, was 8.5% for low-acuity distribution (range –2.2% to 21.1%), 9.3% for random distribution (range –0.2% to 21.2%), and 9.1% for high-acuity distribution (range –0.7% to 21.1%). Although the critical mortality improvement due to ReSTART was different for each of the three severity distributions, the variation was less than 1 percentage point, indicating that the ReSTART policy is relatively robust to different severity distributions. **Conclusions:** Taking resource limitations into account in mass-casualty situations, triage has the potential to increase the expected number of survivors. Further validation is required before field implementation; however, the framework proposed in here can serve as the foundation for future work in this area.

Keywords: *triage, mass-casualty event, prioritization*

Introduction

There is very little existing research to validate the reliability and effectiveness of triage systems and algorithms currently in use.[1] In a mass-casualty situation, triage is required to maximize the delivery of limited resources so as to benefit injured patients to the greatest extent possible. For example, within the START triage guidelines, patients are classified as expectant (not expected to survive injuries given severity or care available), immediate (survivable injuries but requires definitive medical treatment within 1 hour to survive), delayed (potentially serious but not expected to deteriorate over the next several hours), and minor.[2] In the START framework, all immediate patients in a mass-casualty incident should be transported before delayed patients, regardless of the resource availability. Previous authors have suggested, however, that there is potential to improve outcomes if prioritization decisions explicitly account for the resource limitations that may arise in various stages of the response effort.[1-5] Research to date has not addressed whether there is evidence to support this claim or how existing triage protocols such as START could be extended in a way that takes into account resource limitations. Despite the fact that one of the core concepts of mass-casualty triage is that triage systems should be resource dependent,[3] existing triage guidelines fail to describe how "the decision of whom to treat and/or transport first and how best to use the resources on hand," should be made.[2] Currently, there is no guideline or rule-of-thumb to help emergency responders make such decisions based on resource limitations. To fill this gap, we report the findings of a research project that uses operations research methodologies including mathematical analysis, optimization, and discrete-event simulation.

Such techniques, which have found many applications in many industries, have not been successfully applied to patient prioritization during emergencies.

To our knowledge, the only existing work that has developed a new method to make resource-based patient prioritization decisions is by Sacco et al.[5,6] So far, the method, which is named Sacco Triage Method (STM), has not found wide acceptance because complex calculations are required and because so many different triage classifications are introduced.[7] Importantly, however, the authors propose that first responders and triage personnel begin including resource availability in their transport decisions. Like Sacco et al., we will illustrate the benefit of making triage allocations based on transport and resource availability. However, unlike Sacco et al., our objective is not to propose a new triage protocol. Rather, the main objective of this report is to provide general guidance on how resource restrictions should be taken into account when determining priority levels for transporting patients in the aftermath of mass-casualty incidents. Our results can be used as an adjunct to commonly adopted triage classification system, such as START and SALT (see later). We hope that this report will introduce responders to the need that resource availability should be taken into account when making decisions on the order in which patients should be transported. Given the paucity of evidence to support the current triage systems and algorithms systems in use, we hope this study will help generate further research on patient prioritization. More broadly, this study demonstrates the value of applying operations research concepts that are commonly used in industry to provide decision support in emergency medicine.

To illustrate this concept of resource-based patient prioritization, we have developed a model that can inform and guide decision-making in the field or support efforts to develop more sophisticated decision support systems. The illustrative model is based on rigorous mathematical and computational analysis of a mass-casualty patient triage and transport problem, which consisted of the following steps: we first developed a mathematical representation, which captured the fundamental features of the problem. Then, using various mathematical techniques, we identified the solution that maximizes the expected number of survivors. Following that, we carried out an extensive computer simulation study, which considered realistic mass-casualty conditions, and compared the expected number of survivors under the current practice versus those using our model. Complete details of this study that serves as the scientific foundation for the proposals here have already been published.[8] This current report is a product of our translational research effort that aims to provide practical insights for medical providers based on the earlier derivation study.

The authors are of the opinion that, for a number of reasons that are fundamentally related to the complexity of the problem, infancy of research in this area, and lack of data, developing very specific prioritization methods that would be universally valid is not a realistic goal at this stage. Nevertheless, to demonstrate the potential benefits of the results of this report, we have developed a specific prescription that describes what action to take in the field when prioritizing patients. We name this new method Resource-based START (ReSTART). As the name suggests, ReSTART is a patient prioritization protocol that can be added to the START classification criteria (or similar color-coded classification system), allowing providers to work within a familiar framework. As in START, patients are classified as expectant, immediate, delayed, or minor, and patients classified as minor and expectant receive the lowest priority. However, ReSTART differs from the standard implementation of START in that whether priority transportation of the immediate or delayed may change based on the resource availability. ReSTART is based on the solution of a mathematical optimization problem, but it does not require performing a complex calculation in real time following a mass-casualty incident. The solution, and thus the priorities, can be obtained quickly by using a simple predetermined formula.[8] Simulation studies point to significant potential benefits of adopting ReSTART or similar prioritization methods over the standard practice.

Methods

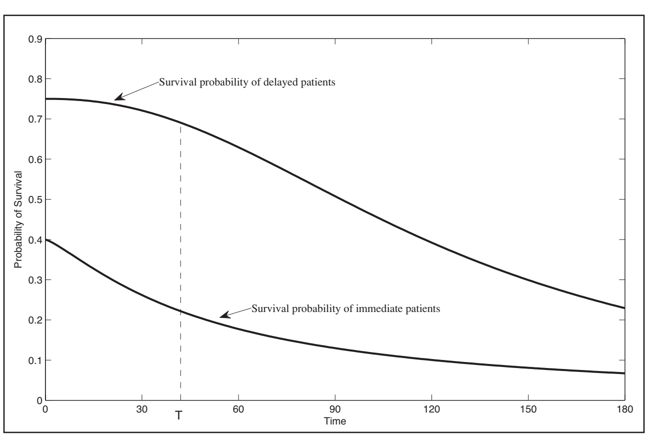
Modeling Assumptions

Determining priorities in the aftermath of mass-casualty incidents mainly concerns patients who are classified as immediate or delayed. There is no question that patients classified as minor or expectant have lower priority than the others as long as there is no reason to believe that they might have been misclassified in the first place. The key difference between the immediate and the delayed patients is that while patients from both classes have serious injuries and need to be attended to urgently, the delayed patients should be able to wait a little longer than the immediate patients. In other words, it is assumed that the chance of surviving (which we refer to as *survival probability*) of the immediate patients starts declining rapidly very soon after the incident, whereas the rapid decline in survival probability of the delayed patients starts some time later. While field management may prolong survival of delayed patients, it is expected that immediate patients will die quickly without definitive care, which is assumed to require transport from the mass-casualty scene. Figure 1 is an illustration of this anticipated difference between the survival probabilities for the two classes. This structural difference between the two curves is the underlying assumption that justifies the different treatment of immediate and delayed patients.

Main Findings From the Mathematical and Computational Analysis

Figure 1 communicates the higher degree of urgency for immediate patients. However, our mathematical and computational analysis, details of which can be found in a previously published article,[8] revealed that, depending on resource limitations, giving priority to immediate patients may in certain cases lead to poor outcomes. For a relatively small-scale incident in which resources are not too restrictive, giving priority to immediate patients would be reasonable because delayed patients will not experience a very significant decline in their survival probability while they wait for the immediate patients to be treated or transported. However, if the resources are constrained (because of having either too many patients or too few ambulances), by the time the disposition of immediate patients is over, it might be too late for at least some of the delayed patients to have a realistic chance of surviving. If the goal of emergency response is to maximize the total number of survivors, this result would indicate a poor use of resources because, compared with the immediate patients who received priority, these delayed patients (who ended up having low survival probabilities due to the delay in response) in fact had a higher chance of survival to begin with and therefore were likely to benefit more from the use of limited resources. This observation suggests that

Figure 1 *Structure of survival probability curves for immediate and delayed patients. The curves indicate the probability that a patient classified as immediate/delayed will survive if he/she is transported/treated at the corresponding time. The figure is used for the purpose of illustrating the shape of the curves; probabilities chosen are arbitrary. T represents the time point when the rate of deterioration of probability of survival in the delayed patient group begins to accelerate and the slope of the survival curve in the delayed patients is now worse than the slope of the survival curve in the immediate patient.*



which classification receives priority should depend on the relative availability of the resources in comparison with the size of the mass-casualty event. In particular, in severely constrained environments, giving priority to the delayed as opposed to the immediate patients would maximize the total number of patients who survive. Increased specificity can be gained by examining whether the resource restriction is caused mostly by a large number of immediate patients, a large number of delayed patients, or both. Adding this distinction to our model leads to Figure 2, which describes the solution for our mathematical representation, and illustrates the relationship among the number of patients, their composition, resource availability, and the priority policy that has a better chance of maximizing the expected number of survivors. The reader can refer to the technical version of this report for more details on our mathematical analysis and its results.[8]

Figure 2 *A model for taking resource limitations into account when determining priorities.*

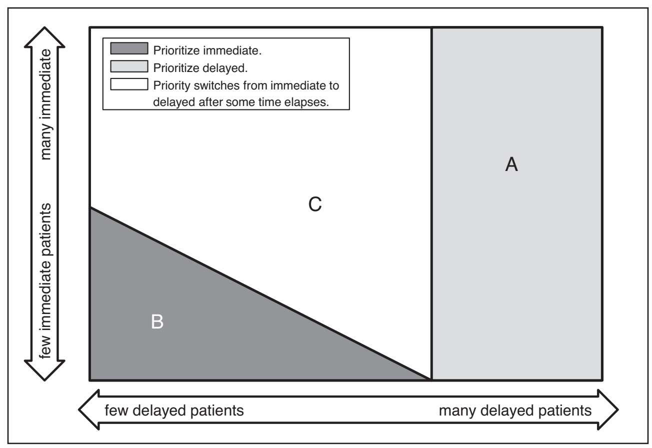


Figure 2 establishes how the prioritization strategy will depend on the number of patients in each triage class as prescribed by our mathematical findings. In particular, there are three different types of scenarios, each corresponding to one of the three regions in the figure:

Scenario 1. When there are many delayed patients and/or few ambulances (region A), delayed patients should get priority. In this case, the number of immediate patients is irrelevant because there are so many delayed patients relative to the available ambulances that there is no reason to move resources away from the delayed patients even for a short period of time. Transportation resources are severely restricted even when only delayed patients are considered. Without prioritizing delayed patients from the start, there will be no feasible way of avoiding the survival deflection point where delayed patients begin dying at a faster rate. Given that, in this scenario, there are a large number of delayed patients to begin with, this results in the decreased probability of survival affecting a proportionately large cohort of patients.

Scenario 2. When there are few immediate and few delayed patients (region B) and/or relatively ample supply of ambulances, immediate patients should get priority. This is because transportation resources are not severely restricted compared with the number of patients, which makes it possible to attend to the immediate patients first without significantly compromising the lives of the delayed patients. Thus, the current practice, which prioritizes immediate patients, works well when transportation resources are *not* severely limited.

Scenario 3. When there are not too many delayed patients but a significant number of immediate patients with respect to the number of available ambulances (region C), immediate patients should have priority initially, but priority should switch to the delayed patients at some point in time during the response effort. In this case, because there are not too many delayed patients, there is no need to start with the delayed patients. There is some time that can be devoted to the immediate patients immediately after the incident. However, because there are relatively a large number of immediate patients, it is necessary to switch to the delayed patients at some point even if there are still immediate patients waiting, in order to improve their chance of surviving. Continuing to dedicate all resources to immediate patients will face diminishing returns, as their place on the survival curve has continued to descend. At the same time, the deflection point for the survival curve for the delayed patients is approaching, and resources should be redirected to the delayed patients as in scenario 1.

The model depicted in Figure 2 provides a broad description as to how priorities should be determined but is kept intentionally nonspecific due to the many complexities of the patient prioritization problem in practice, some of which are difficult or impossible to quantify mathematically. Thus, Figure 2 could be useful in providing a model for guiding decisions in the field, but it cannot replace triage personnel, who will be able to assess the overall situation much better, interpret the model's suggestions, and override them as he or she sees fit. However, it is possible to extend the analysis and make the model more specific by providing precise definitions for "many" delayed patients and "many" immediate patients relative to the availability of transport, and when exactly to switch priorities from the immediate to the delayed (if at all). It is clear that more research is needed to define these in a way that is scientifically credible. However, as a proof of concept for how Figure 2 can be helpful in practice, we derived simple formulas using our mathematical analysis that can be used to determine in which of the three regions a given mass-casualty situation lies, which in turn implies the prioritization strategy to be used. These formulas lead to two different methods, which we name ReSTART and Simple-ReSTART.

Description of ReSTART and Simple-ReSTART

For the purposes of the illustrative examples in this report, we assume all the patients will be transported to a single hospital (or if there are multiple hospitals, the travel time to each hospital is approximately the same). Both methods make use of the following information to determine the prioritization policy:

V = total number of transport resources (e.g., ambulances) available for transporting patients.

R = expected round-trip travel time (between the incident location and the hospital).

I = number of patients classified as immediate.

D = number of patients classified as delayed.

T = the estimated point in time when the survival rate of the delayed patients begins to decrease faster than the survival rate of the immediate patients.

Note that when determining V, a transport resource is a unit that can transport one patient. If each ambulance can transport two patients, then V would be two times the number of ambulances because each ambulance would represent two resources. Except for T, estimating the above information does not require any extensive effort on the part of providers. Estimation of T is more difficult, and it is a very important topic for future research. Nevertheless, this is a task that should ideally be carried out in advance, not on the field, so that T would be defined prior to the occurrence of a mass-casualty incident. In an actual disaster setting, the triage personnel could also use their clinical gestalt based on the type of injuries seen to estimate the time at which the delayed patients' survival rate would begin to markedly deteriorate as compared to the immediate patients while waiting for transport. Given the variety of injuries and types of disasters this may be the only realistic way to estimate T. To get a better understanding of the time T, see Figure 1. As indicated on the figure, T is the time point at which the rate of decline for the delayed patients surpasses the rate of decline for the immediate patients.

The only calculation needed to determine the priority policy is

$$S = T - \frac{DR}{2V}.$$

(An online tool that can perform the above calculation can be found at http://www.restarttriage.com/. A mobile web site that can perform the same calculation can be found at http://bit.ly/ZR7667 or by scanning the QR code found in Figure 3.) Even though we are making

Figure 3 *QR code for accessing the mobile app for the ReSTART calculator.*

this calculator available online, this is not meant to signify that calculations must be done to at least use the concepts illustrated here.

Then, the ReSTART triage and prioritization policy can be described as follows:

Step 1: Classify patients according to START criteria, with management decisions being made per START protocols until the number of casualties is known.

Step 2: Calculate the parameter S.

Step 3: Determine the priorities among the immediate and delayed patients as follows:

(I) If $S \leq 0$, transport all the delayed patients first, and then any immediate patients that still survive.

(II) If $S \geq IR/V$, transport all immediate patients first and then all remaining surviving delayed patients.

(III) If $0 < S < IR/V$, transport immediate patients for S minutes or until there are no more immediate patients. Then, start transporting delayed patients and continue until there are no more delayed patients. Finally, continue with the transportation of any remaining immediate patients.

Note that steps 2 and 3 can be repeated as often as necessary as additional patients are found or as triage classifications are corrected. Although the largest benefit is obtained when steps 2 and 3 are completed as soon as possible after the incident, in most scenarios ReSTART will begin by prioritizing the immediate patients just as START does; thus, the standard implementation of START can be used until it is possible to obtain the information needed to complete steps 2 and 3. In this way, one can see that ReSTART is not a replacement for START, but rather a decision support tool that can be used *in addition to* START to help improve prioritization once sufficient information about the patient distribution and resource availability is available.

In this description, S can be seen as a measure of how restricted transportation resources are in comparison with the size of the mass-casualty incident and composition of the patient population and IR/V is a rough approximation for the expected total time needed to transport all immediate patients if all available ambulances are allocated to them. Negative values of S (case I) indicate that transportation resources are severely limited in comparison with the number of delayed patients and thus ReSTART prioritizes those patients (region A in Figure 2). This is because when resources are severely limited, it is not possible to transport all the patients within a time window that will allow a reasonably high survival probability for all the patients; therefore, it is best to use the resources first for the delayed patients, who have a higher chance of benefiting. In effect, in these situations of severe resource constraint, the immediate patients are now essentially reclassified as expectant by default (although they still may be prioritized prior to minor injury patients, who have almost no risk of mortality). If S is positive and large (case II), this means that resources are relatively abundant and one should use START as currently practiced by giving priority first to immediate patients and then to the delayed patients (region B in Figure 2). When S is positive but small (case III), the resource limitation is at a medium level. This means that there is some amount of time during which it is beneficial to give priority to the immediate patients, but at some point, priority should switch to the delayed patients. Specifically, this switch should occur at time S, which is some time before the deterioration rate of the delayed patients exceeds that of the immediate patients.

ReSTART, as described earlier, makes sense intuitively: observe that S tends to be larger (and thus the resources tend to be less limited) when there are few delayed patients, when travel time to the hospital is short, and when the number of ambulances is large. Precisely why this particular calculation of S should be expected to give good results is supported by mathematical analysis we have described previously.[8] Because S can be calculated quickly using only a few pieces of information, it can be updated as frequently as needed: for example, when patients are re-triaged and change classification from immediate to delayed or vice versa, or when the number of available transportation resources changes. With this recalculation, providers can better assess how conditions are changing at the scene—that is, whether resources are becoming more restricted (decreasing S) or less restricted (increasing S). In short, calculation of S allows providers to get a "snapshot" of resource availability by combining several pieces of information into a single parameter that measures the extent to which resources are restricted.

It is important to emphasize the fact that ReSTART does not change the way patients are classified but calls for using patients' triage class information more intelligently by considering resource restrictions. Calculating the resource limitation (S) is a simple method to use since it requires a single computation to determine what specific prioritization scheme to employ. The only possible complicating factor could be that if the resource limitations (S) are at a medium level, providers must switch priorities between the immediate and the delayed at some time during the response effort. It may be the case that a policy that maintains the same priority all the time would be easier to implement. To give providers this option, we propose Simple-ReSTART, which is determined from ReSTART by dividing region C, which corresponds to medium level of resource scarcity (region C in Figure 2) into two, designating one part as belonging to region A, where resources are severely limited, and the other as belonging to region B, where resources are not restrictive. See Figure 4 for a visual depiction of the model behind Simple-ReSTART. Specifically, the steps for applying Simple-ReSTART are as follows:

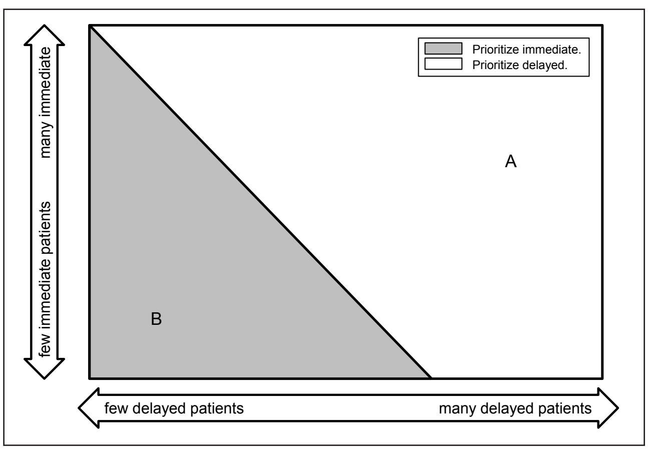
Step 1: Classify patients according to the START criteria.

Step 2: Determine the priorities among the immediate and delayed patients as follows:

(I) If $I + D > 2VT/R$, transport all the delayed patients first and then the immediate.

(II) If $I + D \leq 2VT/R$, transport all the immediate patients first and then the delayed.

Figure 4 *A simpler model for taking resource limitations into account when determining priorities.*



Like ReSTART, Simple-ReSTART also is an intuitively reasonable policy: when there are few patients or more ambulances, or when travel times are shorter, resources are less likely to be restricted and the inequality in case II is more likely to be true, in which case we use the standard practice of prioritizing the immediate patients. Otherwise (in case I), prioritizing the immediate patients would increase the response time for the delayed patients so much that they could not benefit from their initially high chances of survival. In that case, priority is given to the delayed patients. Simple-ReSTART is simpler than ReSTART in two respects. First, it always prescribes a fixed prioritization scheme, eliminating any potential confusion on the field. Either immediate patients have priority over delayed patients all the time or delayed patients have priority over the immediate patients all the time. Second, the formulas that determine which one of the two prioritization scheme to be used involves the total number of critical patients (I+D), not necessarily the respective numbers in each class, making the method more robust with respect to classification errors across the two classes, which can lead to incorrect estimates for I and D, but not their sum $I + D$.

Demonstration of Prioritization Policies via an Example

We now use an example to compare three prioritization methods: START, ReSTART, and Simple-ReSTART. We deliberately chose an example so that all three methods will prescribe a different set of actions. In this hypothetical example, police and emergency medical services providers have responded to a mass shooting at a crowded theater where people are watching a play. The theater is located 15 minutes from the only trauma center in the area. Ten ambulances are able to stage adjacent to the theater, but it takes ten minutes to secure the area and gain access to the theater. Two mutual aid ambulances are also dispatched at the time of the incident, but it will take about 30 minutes for them to arrive. About 200 additional people are able to walk away and are directed out of the theater. Among those who remain, EMS providers find 20 dead or expectant patients, 30 immediate, 30 delayed, and 20 with minor injuries.

We will assume that it has been estimated that T, the time at which the survival rates for delayed patients start declining faster than those for the immediate patients is 67 minutes from the time of the shooting based on the survival probabilities given in Figure 5. Note that these survival probability functions given in the figure are for penetrating injuries, and were obtained by averaging the survival probabilities corresponding to Sacco et al.'s RPM scores 9–12 for delayed and 4–8 for immediate.[6] Based on the other information given above, we have $I = 30$ immediate patients, $D = 30$ delayed patients, $V = 12$ ambulances, and $R = 30$ minutes. Then, the ReSTART parameter S equals 29.5, indicating a medium level of resource scarcity and according to the ReSTART policy we should prioritize immediate patients only for the first 29.5 minutes after triage starts. The Simple-ReSTART policy, on the other hand, recommends always prioritizing delayed patients. For this example, Table 1 shows how the value of S, the priority suggested by ReSTART, and the priority suggested by Simple-ReSTART change based on different numbers of ambulances (i.e., V). As shown in the table, as the number of vehicles becomes more restricted (i.e., V becomes smaller), both ReSTART and Simple-ReSTART give higher priority to delayed patients, while START continues prioritizing immediate patients.

Figure 5 (A) *Survival probability functions used in the simulation; height of curve indicates the probability that a patient will survive if taken to the hospital at the indicated time.* **(B)** *Severity distributions used in the comprehensive simulation study; height of bar indicates the relative likelihood that a patient in that distribution has injury characteristics classified as Expectant* (**E**), *Immediate* (**I**), *Delayed* (**D**), *or Minor* (**M**).

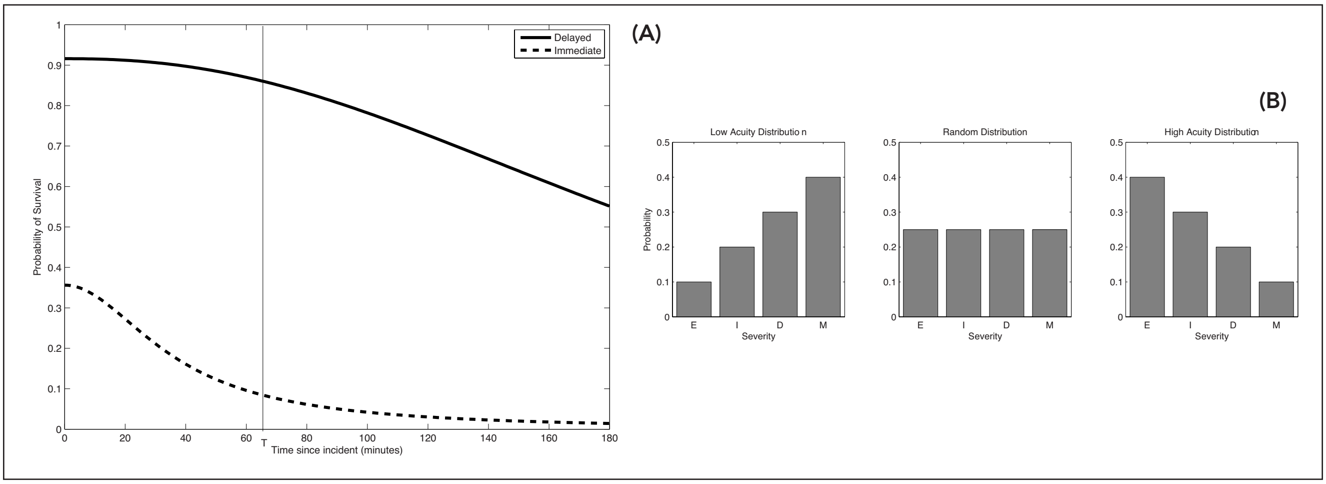


Table 1 *Comparison of START, ReSTART, and Simple-ReSTART for the Example Under Different Levels of Resource Scarcity*

Entries in the table indicate which triage class is prioritized or at what time priority switches from immediate to delayed.

V	S	START	ReSTART	Simple-ReSTART
3	–83.0	Immediate	Delayed	Delayed
6	–8.0	Immediate	Delayed	Delayed
9	17.0	Immediate	Switch @ 17 min	Delayed
12	29.5	Immediate	Switch @ 29.5 min	Delayed
15	37.0	Immediate	Switch @ 37 min	Immediate
18	42.0	Immediate	Switch @ 42 min	Immediate
21	45.6	Immediate	Immediate	Immediate
24	48.3	Immediate	Immediate	Immediate

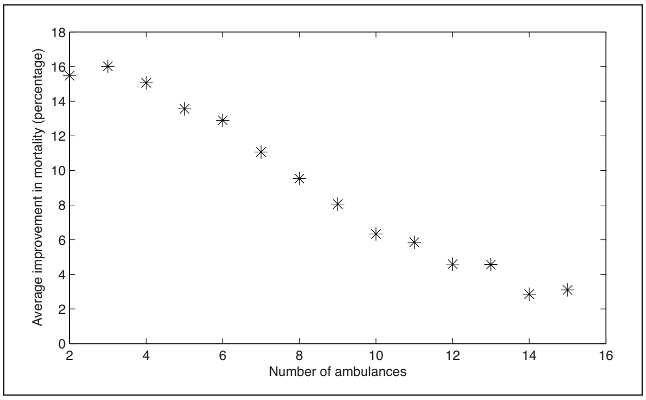
Comprehensive Simulation Study

For the simulation study, we constructed 3,000 scenarios using a random number generator. Each scenario could differ in the total number of patients (chosen randomly between 25 and 125), the number of ambulances (chosen randomly from 2 to 15), and the average one-way trip time (chosen randomly from 10 to 45 minutes). While these choices obviously do not encompass every possible scenario, they represent a wide range of resource scarcity. From each scenario, we created three different incidents by varying the distribution of the casualties. The distributions used are given in Figure 5: in the low-acuity distribution, casualties were more likely to be less severe; in the random distribution, casualties had equal likelihood of any severity; and in the high-acuity distribution, casualties were more likely to be more severe. A total of 9,000 incidents (3,000 scenarios multiplied by three severity distributions) were used in the simulation study. For the per-trip travel times for each ambulance, we used a lognormal distribution, which has previously been used to model ambulance travel times.[9] A Poisson process was used to model the initial arrivals of the ambulances to the scene.[10] Using the same randomly generated travel times for each of the three models, we simulated the hospital arrival times of the patients under three policies: START, ReSTART, and Simple-ReSTART. When the patient arrived at the hospital, the survival probability function was checked and it was determined whether that patient died or survived. The simulation code was written in the MATLAB programming language. The code counted the total number of survivors for each simulated scenario and then reported the critical mortality rate (i.e., the fraction of immediate and delayed patients who did not survive).

Results

When comparing START vs. Re-START, the mean decrease in critical mortality, the percentage of immediate and delayed patients who die, was 8.5% for high-acuity distribution (95% confidence interval [CI] 8.3% to 8.8%, overall range –2.2% to 21.1%), 9.3% for uniform distribution (95% CI 9.0% to 9.6%, overall range –0.2% to 21.2%), and 9.1% for low-acuity distribution (95% CI 8.9% to 9.4%, overall range –0.7% to 21.1%). ReSTART provided significantly lower mortality than START regardless of which severity distribution was used (paired t test, $p < .01$). Although the critical mortality improvement due to ReSTART was different for each of the three severity distributions, the nominal difference was less than 1 percentage point, indicating that the ReSTART policy is relatively robust to different severity distributions. However, improvement in critical mortality clearly depended on the resource availability. As the number of ambulances increased, the improvement from using ReSTART declined from a peak of 16.0% mean improvement in scenarios with three ambulances to a minimum 2.9% mean improvement in scenarios with 14 ambulances (Figure 6).

Figure 6 *Mean improvement in critical mortality using ReSTART versus START for scenarios with randomly distributed severities, by number of ambulances available.*



Similar results were obtained for Simple-ReSTART, which also provided significantly lower critical mortality than START (paired *t* test, $p < .01$). Mean decrease in critical mortality was 8.1% for high-acuity distribution (95% CI 7.8% to 8.4%, overall range –5.8% to 21.2%), 8.6% for uniform distribution (95% CI 8.2% to 8.9%, overall range –6.3% to 21.2%), and 8.1% for low-acuity distribution (95% CI 7.8% to 8.4%, overall range –6.2% to 21.3%). Note that in addition to the slightly lower average improvement, the main disadvantage of using Simple-ReSTART as opposed to ReSTART is that in the small number of scenarios where ReSTART is outperformed by START, Simple-ReSTART results in an increase in critical mortality compared with ReSTART. A closer examination of these outliers revealed that these scenarios tended to be those that fall somewhere close to the line that divides the two regions in Figure 4. Otherwise, the results from Simple-ReSTART are similar to those from ReSTART; in particular, the effect of resource limitations on the percentage improvement in critical mortality is structurally similar.

While we do not provide details in this paper for brevity, it is important to note that our sensitivity analysis revealed that the good performances of ReSTART and Simple-ReSTART are fairly robust. This is because we found that the expected number of survivors is statistically larger under ReSTART or Simple-ReSTART than it is under START even when the "true" survival probability curves are somewhat different from what ReSTART and Simple-ReSTART calculations assumed to be. The reader can find more details on this study in the previously published paper.[8]

Discussion

The need to incorporate information on resource availabilities in determining patient priorities for treatment or transportation in the aftermath of mass-casualty incidents has been recognized.[2] For example, a recent research effort to standardize triage protocols resulted in SALT, a guideline that contains four parts: Sort, Assess, Lifesaving Intervention, and Treatment/Transport. While there was largely consensus on the first three parts (one important agreement being on the need for triage classes to include *immediate, delayed, minor,* and *expectant*), the problem of prioritizing patients for treatment or transportation in the fourth step was largely left as an open question, which the authors acknowledged required more evaluation.[1] To date, very few specific ideas have been put forward to address this issue. This is understandable, as this is a complex problem at multiple levels. The prioritization decision involves so many variables that it is very difficult for a human being to factor in all the details when deciding how to prioritize patients. A human mind is simply not well equipped for such a task. On the other hand, mass-casualty events are not well-structured events. Each event has its unique characteristics and emergency response teams typically have to deal with unexpected developments that derail attempts for organization. This makes it very difficult if not impossible to model the problem at a level of detail that is sufficient to rely purely on mathematical real-time solution methods. Therefore, the best chance to improve the current practice is with a method that uses a mathematical approach in a way that also recognizes the need to involve human decision makers due to operational realities. This is the basic principle behind the development of the prioritization model and ReSTART.

The developers of the Sacco Triage Method made an important contribution by introducing mathematical modeling and optimization as a tool to improve patient prioritization decisions.[4] We incorporate some ideas from the Sacco Triage Method's formulation, namely the decline in survival probability with the passage of time. However, the Sacco Triage Method has a number of critical limitations, some of which have already been discussed in the literature. We believe that the fundamental problem with the method is its overreliance on the solution to a complex mathematical program coupled to a very granular patient risk stratification scheme. The developers make the reasonable argument that START's use of only two different classes for critical patients (immediate and delayed) does not provide much discriminatory power to differentiate patients with respect to their survival probabilities. To overcome this limitation, the Sacco method calls for categorizing patients according to their RPM (respiratory rate, pulse rate, and motor response) scores, which can take one of 13 possible values implying potentially 13 different triage classes. While the question of whether RPM scoring is superior to START classification is beyond the scope of this article, this more detailed classification comes with a cost: it is difficult to manage the whole response effort logistically when there are many patients belonging to one of 13 different classes,[7] and it is more difficult to incorporate human input (e.g., to decide when to override and/or how to modify the prioritization policy suggested by the Sacco Triage Method in response to changes in the field).

We take a different approach: Recognizing that it may be difficult to reliably estimate all of the parameters in a large model, we model patient prioritization in a way that works within existing triage frameworks, requires relatively simple information, provides a strategy that serves as a broad guideline for patient prioritization, and encourages providers to incorporate clinical judgment. The model proposed in this paper was developed by making parsimonious use of mathematical analysis. Because of the widespread adoption and acceptance of color-coded triage systems using a small number of triage classes, our goal from the beginning was to investigate how systems such as START or SALT can be extended so that priority decisions are made with resource limitations in mind. Having a small number of triage classes to deal with is not helpful purely for operational efficiency. The simplicity also makes it possible for triage officers or emergency responders to have an intuitive understanding of why the prioritization policy makes sense. It is also important to point out that for the proposed model and the methods to work classification based on START is not a must. Any triage protocol meeting the uniform criteria described by Lerner et al.[2] will use the immediate and delayed classes. As long as the survival probability functions for these classes follow the intuitive structure shown in Figure 1, our model and policies point to a reasonable way of factoring in resource limitations.

Implementation Support

The model depicted in Figure 2 provides broad guidance for making prioritization decisions. It can serve as a template for developing more specific protocols to handle future mass-casualty events, or it can help decision makers who mostly act based on intuition make more informed decisions in real time in the field. The insights provided by this model should be taken as the main contribution of this work. When making decisions in the field based on these insights, providers will have to answer two difficult questions: *Are the resources restricted enough to deviate from the standard practice of prioritizing immediate patients? If yes, when exactly should we deviate?* While extensive research is needed to answer these questions confidently, we describe two specific methods, called ReSTART and Simple-ReSTART, mainly to illustrate one way of using the broad guidelines depicted by Figure 2 in practice. Both methods are supported by our mathematical analysis, and simulation studies show that they are superior to the current standard practice, but they should serve as a starting point for more research on this topic and not as a definite word on how prioritization decisions should be made.

To make the ReSTART policies accessible to a wide audience, we developed a web-based decision support tool, available at http://www.restarttriage.com/. The tool allows users to set input parameters such as survival probabilities, travel time, and number of ambulances on sliding scales. The model output is displayed in a user-friendly interface that demonstrates the effects of changing these parameters without requiring the use of any mathematical expressions, by displaying a dynamic, interactive chart similar to Figure 2. The chart displays the policy suggested by ReSTART as a function of the number of delayed patients (on the horizontal axis) and the number of immediate patients (on the vertical axis). The user can hover over any point in the chart to see the exact policy. The tool also provides a proof of concept for the use of ReSTART in a practical setting. If medical providers wish to adopt ReSTART, or a similar policy, in the field they can use the mobile version of the web site, which provides a quick way to calculate S and provides a visualization of the policy that has been adapted for display on small screens. A mobile app is currently in development so that on line connectivity will not be required for real time use.

Because the calculations involved in determining the policy are very simple, the web-based implementation does not require any delay for "solving" the problem, so it can offer instant feedback by showing how the transportation prioritization should change if, for example, the classification of one of the patients changes due to retriage or if additional resources arrive.

Study Limitations

The method described above is meant to introduce personnel making triage decisions to the fact that scarcity of transport resources should be taken into account when triaging patients and that in resource constrained environments this approach might suggest the need to transport delayed patients before immediate patients in order to maximize survival. In a perfect world, we would have accurate parameters for T and know the shapes of the survival curves. As with Sacco we have had to estimate these numbers. More research is needed to better define these parameters.

Arguably the biggest challenge in using a scientific approach to develop better emergency response methods is that there are no available data that can be used in estimating some of the crucial model parameters. While this challenge may be less of an issue when exploring the problem at a highly abstract level, as we did when we developed the model given in Figure 2, the practical usefulness of the specific methods we propose, namely ReSTART and Simple-ReSTART, depends on the quality of the estimates. In fact, one of the best features of ReSTART and Simple-ReSTART is that they require estimation of very few parameters, and except for the parameter T (the time point at which deterioration rate of delayed patients exceeds that of immediate patients), they are all easy to determine. To the best of our knowledge, there are simply no data available to estimate parameter T, and thus an important next step is to conduct research that specifically aims to estimate this parameter, or more broadly, to estimate the survival probability functions depicted in Figure 1, which might in turn be used to determine T. Estimation of the whole survival probability function would be more challenging but nevertheless much more useful because these functions can be used in simulation studies that compare different prioritization policies. The approach the developers of STM used to get around this problem was using the Delphi method, that is, asking a panel of experts to give their best estimates on the survival probability of a patient given the patient's RPM score and how this score changed with time. A similar study can be carried out to estimate survival probabilities given a patient's START class. These estimates can also differ depending on the type of the event or the nature of the patents' injuries. While the scientific credibility of this approach is uncertain, it might serve as a good starting point and the estimates can be improved as more data are gathered in future mass-casualty events.

Conclusions

This study puts forth a novel framework for taking resource limitations into account when prioritizing patients in the aftermath of mass-casualty incidents and gives precise descriptions for two new methods that are based on this framework. Uncertainty exists as to defining key parameters such as expected survival times, and this is a key area for future research.

Disclaimers

This research was supported by National Science Foundation grants CMMI-0927607, CMMI-0927668, CMMI-1234212, and CMMI-1234260.

Disclosures

The authors have no relevant financial conflicts of interest in regard to the submitted work.

References

1. Jenkins JL, et al. Mass-casualty triage: time for an evidence-based approach. *Prehosp Disaster Med*. 2008;23:3–8.
2. Lerner EB, Schwartz RB, Coule PL, et al. Mass casualty triage: an evaluation of the data and development of a proposed national guideline. *Disaster Med Public Health Prep*. 2008;2(Suppl 1):S25–S34.
3. Lerner EB, et al. Mass casualty triage: an evaluation of the science and refinement of a national guideline. *Disaster Med Public Health Prep*. 2011;5:129–137.
4. Frykberg ER. Triage: principles and practice. *Scand J Surg*. 2005;94:272–278.
5. 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.
6. Sacco WJ, Navin DM, Waddell RK, et al. A new resource-constrained triage method applied to victims of penetrating injury. *J Trauma*. 2007;63:316–325.
7. Cone DC, MacMillan DS. Mass-casualty triage systems: a hint of science. *Acad Emerg Med*. 2005;12:739–741.
8. Mills AF, Argon NT, Ziya S. Resource-based patient prioritization in mass-casualty incidents. *Manuf Serv Oper Manage*. 2013;15:1–17.
9. Ingolfsson A, Budge S, Erkut E. Optimal ambulance location with random delays and travel times. *Health Care Manage Sci*. 2008;11:262–274.
10. Green L. Queuing analysis in healthcare. In Hall RW, editor. *Patient flow: reducing delay in healthcare delivery*. New York, NY: Springer; 2006:281–307.

Dr. Mills is affiliated with the Department of Operations and Decision Technologies, Kelley School of Business, Indiana University. Bloomington, Indiana; this research was partially conducted while he was a student at University of North Carolina at Chapel Hill, North Carolina.

Dr. Argon is affiliated with the Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, North Carolina.

Dr. Ziya is affiliated with the Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, North Carolina.

Dr. Hiestand is affiliated with the Department of Emergency Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina. E-mail: bhiestan@wfubmc.edu.

Dr. Winslow is affiliated with the Department of Emergency Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina.