Evaluating Emergency Medical Service Performance Measures

Laura A. McLay*
Department of Statistical Sciences & Operations Research
P.O. Box 843083
1001 W. Main Street
Richmond, Virginia 23284
Virginia Commonwealth University
lamclay@vcu.edu

Maria E. Mayorga
Department of Industrial Engineering
Freeman Hall Box 340920 Clemson University
Clemson, South Carolina 29634
mayorga@clemson.edu

February 19, 2009

Abstract

The ultimate goal of emergency medical service systems is to save lives. However, most emergency medical service systems have performance measures for responding to 911 calls in a fixed timeframe (i.e., a response time threshold), rather than measures related to patient outcomes. These response time thresholds are used because they are easy to obtain and to understand. This paper proposes a methodology for evaluating the performance of response time thresholds in terms of resulting patient survival rates. A model which locates ambulances to optimize patient survival rates is used for base comparison. Results are illustrated using real-world data collected from Hanover County, Virginia. The results indicate that locating ambulances to maximize seven and eight minute response time thresholds simultaneously maximize patient survival. Nine and ten minute response time thresholds result in more equitable patient outcomes, with improved patient survival rates in rural regions.

Keywords: Discrete optimization, emergency medical service, patient outcomes, performance measures

*corresponding author: TEL (804) 828-5842 , FAX:(804) 828-8785

1 Introduction

The ultimate goal of emergency medical service (EMS) is to save patient lives. Analyzing EMS performance measures is an important problem, since performance measures are used to determine how EMS resources are used, and hence, they ultimately determine patient survival. Nearly all EMS systems have performance measures based on the response time interval, the time from when an ambulance is dispatched until it arrives at the scene, rather than explicitly measuring patient survival. System performance is most frequently measured as the number (or fraction) of calls that can be reached in a fixed timeframe, denoted as the *response time threshold* (RTT). RTTs are used to measure system performance, since they are easy to evaluate, easy to explain, and unambiguous. The most common EMS performance measure is to respond to 90% of Priority 1 (life-threatening) calls in less than nine minutes (i.e., a nine minute RTT) [11]. According to this performance measure, a call is considered to be covered if the response time is eight minutes and 59 seconds, whereas a response time of one second longer (i.e., nine minutes) is not considered to be covered. However, according to the medical literature, no difference in patient outcomes would be expected between these two scenarios.

There are many practical reasons to evaluate EMS systems using RTTs as opposed to patient survival. Estimating patient survival is difficult, since patient survival is defined as survival to hospital discharge, it is assessed at the hospital, and a patient may be discharged several days after delivery to the emergency room. Furthermore, patient survival information is not readily available, due to medical privacy regulations. Even when an EMS department obtains patient survival information, it can be difficult to retroactively determine the EMS response that results in these outcomes. RTTs, on the other hand, are easy to obtain and evaluate. For the foreseeable future, it appears that RTTs will be used to evaluate EMS performance. Therefore, it is important to understand how resources allocated according to standard measures, such as RTTs, affect patient outcomes.

Erkut et al. [10] illustrate the importance of explicitly linking ambulance location to patient survivability, by providing the following example. Two demand locations A and B are located eighteen minutes apart, with three potential stations located at A, B, and X, a station exactly eight minutes and 59 seconds minutes away from both A and B. The demand at A is 10, and the demand at B is 1. Exactly one ambulance is available. In order to maximize the number of calls responded to in less than nine minutes, the ambulance is located at X, which covers the entire demand of 11. If the probability of survival is given by $\exp(-t_R)$, where t_R denotes the response time (measured in minutes), then the expected number of survivors when an ambulance is located as X is $11 \times \exp(-9) = 0.001358$, whereas the expected number of survivors when the ambulance is located at A is 10, which differ by a factor of 700 [10]. This indicates that locating ambulance according to RTTs may result in arbitrarily bad patient survival rates and motivates modeling efforts that explicitly focus on patient outcomes. However, 911 call locations are more evenly dispersed in EMS systems than in this example, and hence, the disparity in patient survival is likely to be smaller. This paper explores such relationships between EMS response time intervals and patient survival.

For decades, EMS systems have been measured according to how they respond to and care for cardiac arrest patients [11, 7]. Emergency medical 911 calls are typically classified as Priority 1, 2, 3, where Priority 1 calls are life threatening emergencies (cardiac arrest calls are a subgroup of Priority 1 calls), Priority 2 calls are emergencies that may be life-threatening, and Priority 3 calls do not appear to be life-threatening emergencies. Responding to emergency cardiac arrest 911 calls is a major focus of EMS systems. EMS systems are often evaluated by how effectively they respond to and perform care for cardiac arrest calls because: 1) effective treatment is available (i.e., CPR, early defibrillation); 2) treatment is highly time dependent; 3) each element in the community emergency response (early access, early CPR, early defibrillation, and early advanced life support) must be strong for optimal survival to occur; and 4) if the EMS system can respond effectively to cardiac arrest and achieve a good survival rate, the same elements will usually yield similar outcomes for other life or death conditions such as major traumatic injury, heart attack, or stroke. This has been well-documented by medical research [23]. In other emergency medical 911 calls, patient survivability is not as strongly correlated with EMS response and patient care provided at the scene [8, 7]. The EMS response time interval is highly correlated with cardiac arrest patient survival, whereas the link between the EMS response time interval and patient survival from other types of emergency medical conditions is unclear [11]. However, the survival to hospital discharge rate after out-of-hospital cardiac arrest is low, with approximately five to seven percent of cardiac arrest patients surviving [22, 21].

This paper proposes a discrete optimization model for evaluating a class of EMS performance measures (namely, RTTs) in the context of locating ambulances at a fixed set of stations in order to assess which RTT performance measures are robust with respect to their corresponding patient survival rates. The RTTs measure the expected number of calls that can be reached in a fixed timeframe, where the length of the fixed timeframe varies from four to ten minutes. When optimizing using a RTT, the corresponding patient survival rates are retrospectively assessed to compare RTTs in terms of patient outcomes. The proposed discrete optimization model evaluates random response times that depend on the distances between stations and call locations. In addition, the issue of equity is addressed for all cases in order to compare the number of patients expected to survive in rural and urban areas, since patient survival rates in rural areas are observed to be significantly lower than in urban areas.

The results indicate that RTTs can act as proxies for patient survivability performance measures. For an example with real-world data, we find that seven and eight minute RTTs always locate ambulances such that patient survival is also optimal. Similarly nine and ten minute RTTs result in more equitable patient outcomes. These results are counterintuitive, since medical experts often recommend an EMS response time intervals of less than four minutes. Although a shorter response time interval improves patient survival, systematically utilizing resources to cover the most calls in less than four minutes tends to use resources in such a way that does not save the most patient lives overall.

This paper is organized as follows. A literature review on models used for EMS systems is performed in Section 2. EMS performance measures and a patient survival model are introduced in Section 3. The proposed discrete optimization model is formulated in Section 4. The results are applied to a scenario using real-world data collected from Hanover County, Virginia in Section 5. Concluding remarks, including policy implications as well as directions for future research are presented in Section 6.

2 Literature review

Although numerous models have been analyzed for locating ambulances at a set of stations, nearly all focus on meeting RTT performance measures or other proxies for patient survivability, rather than explicitly modeling patient survival [2, 12, 13, 24]. Recently, McLay [20] and Erkut et al. [10] examine how to locate medical units when focusing on patient survival. McLay [20] considers the impact of the first responder by extending the Maximum Expected Coverage Location Problem [6] for a nine minute RTT. In particular, McLay [20] investigates how to locate both ALS and BLS medical units such that response time is determined by the first responder, rather then the ALS medical unit, as traditionally done in EMS systems. This paper also examines the effect of dispatching policies for ALS and BLS medical units to Priority 1, 2, 3 calls to illustrate how to use and locate medical units.

Erkut et al. [10] apply patient survivability models reported in the medical domain to ambulance location models using discrete optimization models (linear and nonlinear) and Monte Carlo simulation. The probability of patient survival is a function of response time, the time to initiate CPR, pretravel delays, and the time to defibrillation. They demonstrate the benefit of using patient survival location models using data extracted from Edmonton, Canada. They explore how patient survival affects the location of ambulance by using survivability models as the objective function in set covering and maximal coverage models that consider one type of medical unit (ALS ambulances).

Many facility location models for EMS systems have incorporated variations or extensions to a basic facility location model, the Maximal Covering Location Problem (MCLP), where the objective is to maximize the population that can be served in a specified amount time or distance

[5]. In MCLP, there are a set of demand nodes (i.e., a geographic area where calls for service arise) and a set of location nodes (i.e., potential locations to stage ambulances) with a single time period with deterministic travel and service times, and no adjustments are made for busy vehicles. Daskin [6] develops an extension to MCLP that considers each vehicle being busy with some probability and maximizes the expected number of calls covered in a given amount of time. Batta et al. [1] introduce a model that lifts the assumptions made by the Daskin [6] model by embedding the Hypercube model [17, 18] in an optimization heuristic. The Hypercube model analyzes vehicle location and response district design for urban environments, and it assumes the underlying dynamics are that of a multi-server queuing system with indistinguishable servers, and has been extended several times [17, 14, 15, 4, 3, 19].

Much of the previous work in this area has determined how to locate or use ambulances to maximize a fixed RTT. In contrast to previous work in the area, this paper challenges the assumption that RTTs are effective proxies for patient survival, and it explores the relationship between RTTs and patient outcomes.

3 Performance measures in EMS systems

Medical research has shed light on some of the factors that save patient lives. The response time interval has been identified as an important predictor for short-term patient survival. Other factors that have been identified by the medical literature that affect short-term patient survival include the age of the patient, whether a bystander is present at the time of collapse, and whether bystanders attempted CPR and defibrillation before the first responder arrives [16, 23, 9].

There are several medical models that relate EMS response time intervals to patient survival probabilities in cardiac arrest patients, see for example Larsen et al. [16], Valenzuela et al. [25], Waaelwijn et al. [26], and De Maio et al. [9]. We select one such model for use in this work, which results in a patient survival probability that is a function of response time[1]. In particular, the patient survival model used in this paper is based on a study by Larsen et al. [16], which performs multiple linear regression for data from King County, WA. The times are measured relative to the time of collapse and are measured in minutes. This model is interpreted to assume that EMS personnel do not witness a cardiac arrest. Let S denote the patient survivability, which is assumed to represent a probability (i.e., $0 \leq S \leq 1$). Then,

$$S(t_{CPR}, t_{AED}, t_{ALS}) = \max\{0.67 - 0.023t_{CPR} - 0.011t_{AED} - 0.021t_{ALS},\ 0\}, \tag{1}$$

where t_{CPR} denotes the time from collapse until CPR is performed on the patient, t_{AED} denotes the time of defibrillation, and t_{ALS} denotes the time when advanced life support treatment as provided in Larsen et al. [16]. We cannot apply this model directly since we are interested in

[1] See Erkut et al. [10] for a discussion on choice of model.

the probability of patient survival as a function of the response time t_R. This relationship can be defined under the following assumptions that are confirmed to be reasonable by real-world data.

- It takes exactly one minute for a call to EMS to be made and an ambulance to be dispatched after the patient collapses.
- CPR is performed and an AED is used by a paramedic or EMT immediately upon arrival, and CPR is not performed earlier by a bystander, resulting in $t_{CPR} = t_{AED} = t_R$.
- Advanced life support is provided one minute after arrival, resulting in $t_{ALS} = t_R + 1$.

These assumptions simplify the survival probability (1) to

$$S(t_R) = \max\{0.594 - 0.055t_R,\ 0\}. \tag{2}$$

Note that patient survival is significantly less than one, even for response times of zero, and at nine minutes, patient survival is less than ten-percent.

4 The performance measure location problem

In this section, a discrete optimization model is proposed for evaluating RTT performance measures. It is assumed that the performance measures (either RTTs or patient survival) are functions of response time, which is itself a function of distance between stations and call locations, and hence, the performance measure is ultimately a function of distance. Our approach evaluates EMS performance measures in the decision context of locating ambulances at a given set of stations.

In the proposed model, the RTTs and patient survival performance measures are evaluated only over Priority 1 calls, since these calls are life-threatening. The closest available ambulance is dispatched to a Priority 1 call. If two ambulances are equidistant from the call, one is randomly selected. Ambulances are dispatched to Priority 2 and 3 calls in such a way that balances the workload among the ambulances and results in ambulances being busy approximately the same fraction of time. Therefore, how ambulances respond to all calls determines the availability of the ambulances, and hence, all calls are implicitly taken into account. The following list summarizes the parameters used:

n = total number of nodes (i.e., locations where calls for service originate),

h_i = demand generated at node i, $i = 1, 2, ..., n$ (i.e., number of Priority 1 calls),

r_i = proportion of demand generated at node i that is classified as rural (the remaining calls are considered urban), $i = 1, 2, ..., n$,

m = total number of stations,

N = number of ambulances available,

N_U = upper bound on the number of ambulances that can be placed at a facility location,

$D+1$ = number of *distance groups* that quantify the distance between node i and station j, with distance group 1 representing the shortest distance and distance group $D+1$ representing the longest distance, (for example, group 1 might be 0-1 miles, group 2 might be 1-2 miles, etc.)

$Z_{i,d} \subseteq \{1, 2, ..., m\}$ = subset of stations whose distance from node i is strictly in distance group d, (i.e., $Z_{i,d}$ are mutually exclusive for each i), $i = 1, 2, ..., n$, $d = 1, 2, ..., D + 1$,

$F_d(T)$ = cumulative probability that an ambulance responds in T minutes or less given that it is located at a station in distance group d relative to the call, $d = 1, 2, ..., D + 1$, with $F_1(t) \geq F_2(t) \geq ... \geq F_{D+1}(t)$ for all t,

S_d = cumulative probability that a cardiac arrest patient survives, given that the responding ambulance is located at a station in distance group d relative to the call, $d = 1, 2, ..., D+1$, with $S_1 \geq S_2 \geq ... \geq S_{D+1}$,

τ = average service time per call,

λ = total arrival rate of calls for service per unit time (i.e., Priority 1, 2, 3 calls),

p = proportion of time any ambulance is busy.

All parameters are interpreted as deterministic values except for $F_d(T)$, S_d and p, which are interpreted as probabilities. We use the Hypercube model to determine the dependence between ambulance availabilities [17, 18]. This application of the Hypercube model makes the following assumptions.

1. Calls for service arrive according to a Poisson process with rate λ, independent of server status.
2. Exactly one ambulance is assigned to every call.
3. The expected service times of an ambulance for a call for service has mean τ.
4. Calls are serviced by a random available ambulance.
5. Any call for service that arrives while all ambulances are busy is entered at the end of a queue and are serviced in first come first serve manner.
6. The busy probability of all ambulances are the same.
7. The system is operating in steady state.

In steady state, the proportion of time that all medical units are busy under these assumptions, is given by $p = \lambda\tau/N$. Considering a general M/M/N queueing system operating in steady state, let P_0 denote the probability that all servers are available,

$$P_0 = \frac{(N)^N p^N}{N!(1-p)} + \sum_{j=0}^{N-1} \frac{(N)^j p^j}{j!}. \tag{3}$$

The correction factors $Q(N, p, j)$ quantify the "correction" to the probability of obtaining j busy servers followed by an available server when assuming that servers operate independently,

$$Q(N, p, j) = \sum_{k=j}^{N-1} \frac{(N-j-1)!(N-k)(N)^k p^{k-j} P_0}{(k-j)!(N)!(1-p)}, \tag{4}$$

$j = 1, 2, ..., N-1$, with $Q(N, p, 0) = 1$ [18].

Regardless of the objective, the solution to each problem provides a set of ambulance locations. Let X denote the set of all feasible ambulance locations, with $x \in X$ denoting a single set of ambulance locations. For each $x \in X$, let x_j denote the number of ambulances located at station j, with $x_j \leq N_U$, $j = 1, 2, ..., m$, and $\sum_{j=1}^{m} x_j = N$.

Two types of performance measures are considered (1) the marginal increase in the expected number of Priority 1 patients who survive and (2) the marginal increase in the expected number of Priority 1 calls that are responded to within T minutes, $T = 4, 5, ..., 10$ (i.e., the T minute RTTs). We see that the former explicitly considers patient survival outcomes, while the latter uses response time as a proxy for patient survival. Both performance measures are evaluated relative to a base case. The base case scenario captures value of the performance measure when there are no ambulances in distance groups $1, 2, ..., D$ relative to any nodes. In other words, it is assumed that "distant" first responders (such as fire fighters, police officers, or ambulances from outside of the geographic area) are used to respond to and treat patients, these first responders are located far from the node or outside of the geographic area of consideration, and hence, are located in distance group $D + 1$. These distant responders have a small but positive probability of saving a patient's life. Ambulances located closer (i.e., in distance groups $d = 1, 2, ..., D$) are assumed to increase patient survival relative to this base case. The objective function is interpreted in this way, since many EMS systems leverage many types of resources to treat patients, and patients receive some kind of treatment even when no ambulances are available.

First, we describe the objective used for evaluating the optimal marginal increase in a performance measure, which is

$$\max_{x \in X} \sum_{i=1}^{n} h_i Y_i(x), \tag{5}$$

where $Y_i(x)$ denotes the marginal increase in the fraction of Priority 1 calls at node i that meet the performance measure (either a RTT or patient survival) given the set of ambulance locations x, and hence, $0 \leq Y_i(x) \leq 1$. Let $J_{i,d}$ denote the number of ambulances located in distance group d relative to node i given the candidate set of ambulance locations x, with

$$J_{i,d} = \sum_{j \in Z_{i,d}} x_j, \;\; i = 1, 2, ..., n, \;\; d = 1, 2, ..., D + 1. \tag{6}$$

Then, $Y_i(x)$ is given by

$$Y_i(x) = \sum_{d=1}^{D} \sum_{j=1}^{J_{i,d}} M_d \; Q\left(N, p, j + \sum_{k=1}^{d-1} J_{i,k}\right) (1 - p) p^{j-1+\sum_{k=1}^{d-1} J_{i,k}}, \;\; i = 1, 2, ..., n, \tag{7}$$

where M_d denotes the value of the marginal increase in the performance measure given distance group d (defined below), multiplied by the probability that an ambulance is available from distance group d. Note that in the expression for $Y_i(x)$, a randomly selected ambulance in distance group 1 is first dispatched to a Priority 1 call, and ambulances in sequentially further distance groups are randomly dispatched if closer ambulances are busy.

Note that (7) can be used to evaluate a range of EMS performance measures, depending on how the values of M_d are defined. The two types of performance measures evaluated in this paper which result in two types of definitions for M_d are: (1) maximizing the marginal increase in the expected number of Priority 1 patients that survive and (2) maximizing the marginal increase in the expected number of Priority 1 calls that meet a T minute RTT, with $T = 4, 5, ..., 10$. Note that M_d can be interpreted more generally as a utility function of response times to capture a broad range of performance measures.

The marginal increase in the expected number of Priority 1 patients who survive is computed using the survival rate expression using (2). Let S_d denote the probability that a patient survives, given that an ambulance in distance group d relative to the call location is dispatched to the call, $d = 1, 2, ..., D + 1$. Then the value of the patient survival performance measure for each distance group is given by

$$M_d = P(S_d) - P(S_{D+1}), \tag{8}$$

with

$$P(S_d) = \int_{t_R \geq 0} S(t_R) dF_d(t_R),\ d = 1, 2, ..., D + 1. \tag{9}$$

The marginal increase in the expected number of Priority 1 calls that meet a T minute RTT yields

$$M_d = F_d(T) - F_{D+1}(T),\ T = 4, 5, ..., 10. \tag{10}$$

Furthermore, the issue of *equity* is addressed as follows. Models for ambulance locations tend to locate ambulances in areas of high population density, leading to longer response times and low patient survival rates in rural areas. The expected number of patients that survive in rural areas given the set of ambulance locations x is given by

$$\sum_{i=1}^{n} h_i r_i Y_i(x). \tag{11}$$

5 Illustrative example

Hanover County, Virginia is used as a test bed for the proposed research models. Hanover County is a rapidly growing, semi-suburban, semi-rural county near Richmond, Virginia with a population of approximately 100,000 and an area of 474 square miles. It is described as being 70% rural with areas of suburbs. At present, Hanover County uses a nine minute RTT to evaluate their EMS system. One year of computer aided dispatch (CAD) data was provided to assist in the computational portion of this paper. The CAD data contains 9711 calls for service, and it includes information regarding the location, response time, and service times for all calls.

In Hanover County, the location of each call is reported as a grid location, not a physical address. Likewise, each station is located in a grid location. The grids are defined as two mile by two mile squares. There are a total of 175 grid locations in Hanover County. Each station and each call (i.e., node) is assumed to be located at the middle of a grid location, and distances are computed as the Euclidean distance between the centers of two grid locations. The four distance groups are defined as follows for each node.

1. Stations strictly less than two miles from nodes (i.e., the station is the same grid location as the node),
2. Stations at least two and less than four miles from nodes (i.e., the station is adjacent to the node),
3. Stations at least four and less than six miles from nodes,
4. Stations more than six miles from nodes.

It is assumed that stations in distance groups 1, 2, 3 can be used to evaluate the expected marginal increase in the performance measure, whereas distance group 4 is used for the base case. Hanover County contains sixteen fire and rescue stations where ambulances can be located. Since two of the stations (1 and 16) refer to the same physical address, $m = 15$ stations are used. Figure 1 shows a map of Hanover County and its stations.

The results are illustrated for the 12pm − 6pm time period, Monday through Friday, with $\lambda = 1.2$ calls/hour. Data analysis suggests that these times operate in steady state, with the call frequency approximately constant. This time period also represents the peak hours of service over the week. Hanover County currently uses approximately six ambulances during these times. For the analysis, between three and fifteen ambulances are located. For this time period, there are 3087 Priority 1 calls, with 720 Priority 1 calls in rural areas and 2367 Priority 1 calls in urban areas.

Table 1 summarizes the performance measure input parameters based on the Computer Aided Dispatch (CAD) data analysis. Note that all response times are measured in minutes. The cumulative probability that an ambulance responds in less than T minutes is determined by fitting a lognormal curve to the Hanover County data set. It is estimated that 26.1% of calls can be responded to in less than nine minutes in the base case ($d = 4$) and that 82.4% of calls can be responded to in less than nine minutes when the responding ambulance is less than two miles from the call location ($d = 1$). All survival probabilities are computed using (9). Table 1 indicates that the in base case, 6.3% of Priority 1 patients survive. Since there are 3087 Priority 1 calls, an average of 194 patients survive in the base case. The four distance groups can be used to aggregate the 175 grid locations into 91 nodes, summarized in Table 2. Table 2 contains the demand associated with each node and the stations located in distance groups 1, 2, 3.

Table 3 summarizes the marginal increase in the number of Priority 1 patients that survive and the marginal increase in the number of Priority 1 patients who are responded to within nine minutes when locating ambulances according to the optimal patient survival solution. The marginal increase in the expected number of patients that survive and the expected number of

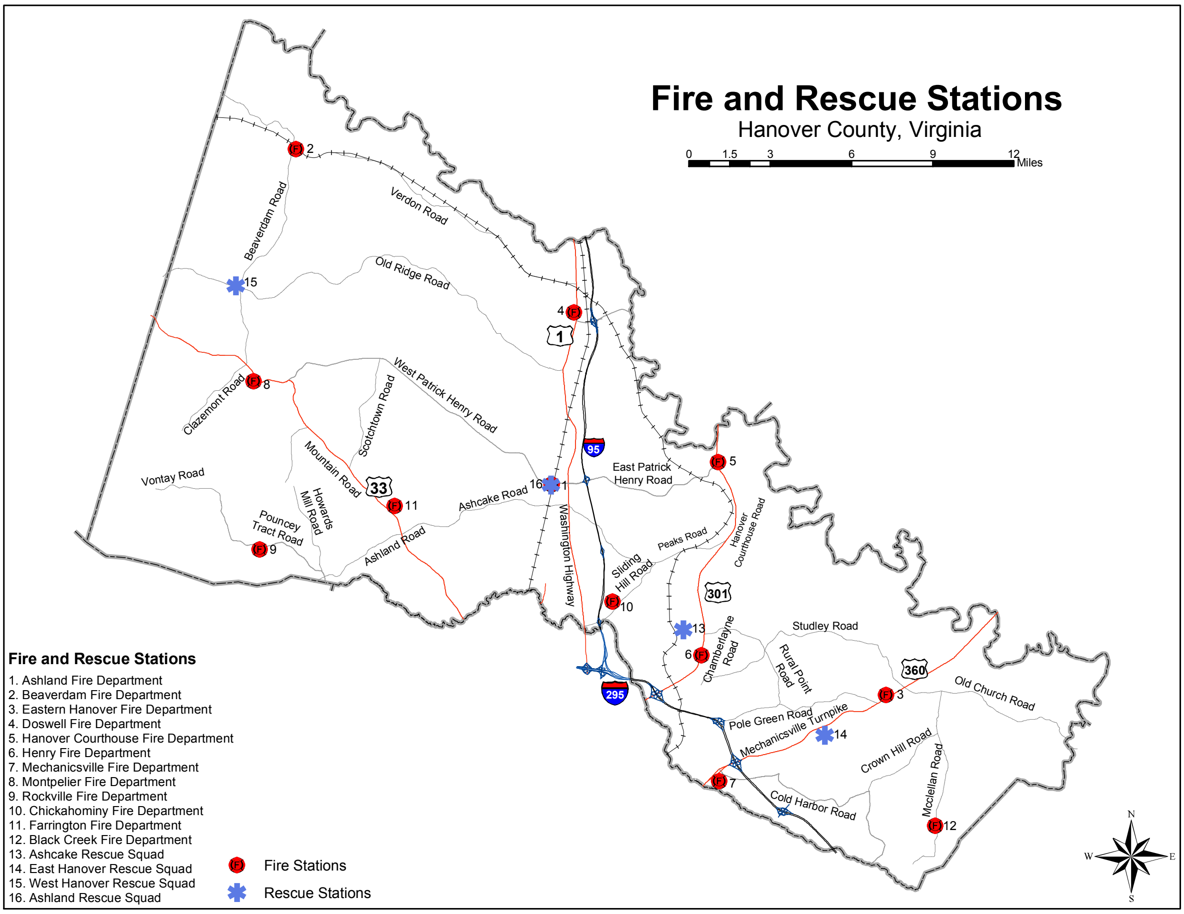

Figure 1: Fire and rescue stations in Hanover County, Virginia

calls responded to in less than nine minutes are reported for the rural and urban areas to illustrate the disparities between the rural and urban responses. All numbers are rounded to the nearest integer. The overall number of patients that survive is equal to the sum of the number of patients that survive in rural and urban areas. Table 4 reports the corresponding ambulance locations for the optimal patient survival solutions. In Table 4, each row lists the number of ambulances located at each of the fifteen stations given N.

Let S_N^* denote the marginal increase in the expected number of patients that survive for the optimal patient survival solutions with N ambulances, and let S_N^T denote the marginal increase in the number of patients that survive for the optimal T minute RTT solutions when N ambulances are located, $T = 4, 5, ..., 10$. Note that $S_N^T \leq S_N^*$ since maximizing the RTT produces a patient survival rate no greater than the optimal patient survival rate. Table 5 reports the difference in the expected number of patients that survive for the optimal patient survival solutions and for the optimal T minute RTT solutions, $T = 4, 5, ..., 10$. The negative values in Table 5 correspond

Table 1: Performance measure input parameters for demand groups

	Value			
Parameter	$d = 1$	$d = 2$	$d = 3$	$d = 4$
$F_d(4)$	0.185	0.079	0.025	0.015
$F_d(5)$	0.348	0.182	0.070	0.040
$F_d(6)$	0.508	0.309	0.142	0.079
$F_d(7)$	0.643	0.440	0.231	0.132
$F_d(8)$	0.747	0.559	0.329	0.261
$F_d(9)$	0.824	0.661	0.427	0.261
$F_d(10)$	0.879	0.742	0.519	0.330
$P(S_d)$	0.242	0.174	0.103	0.063

to the expected number of lives lost when locating ambulances according to a RTT as compared to patient survival, given the same number of ambulances. Differences in the expected number of patients that survive are due to different sets of ambulance locations. Note that only the seven and eight minute RTTs found the optimal patient survival solutions for all number of ambulances $N = 3, 4, ..., 15$. This suggests that locating ambulances according to seven and eight minute RTTs is an effective proxy for maximizing patient survival. Shorter RTTs (four to six minute) and longer RTTs (nine to ten minutes) did not always result in a set of locations that optimize patient survival.

The results reported in Table 5 are counterintuitive since several reports suggest that improving responses within four to six minutes is necessary for saving lives [23, 11, 7]. Although there are relatively few differences in the expected number of lives saved between the different performance measures, which suggests that having more ambulances is more important for saving lives than where ambulances are located, and the results in Table 5 also indicate that locating ambulances to optimize patient survival can save lives at no additional cost.

To compare how optimizing over different RTTs affects the set of ambulance locations, Table 6 reports the set of ambulance locations for the $N = 7$ scenarios for the optimal RTT solutions as compared to the optimal patient survival solutions. The urban areas of Hanover County are primarily in the South-East, ranging from Station 7 to Station 10 and near Station 1. Relative to the optimal patient survival solutions, shorter RTTs (four and five minutes) move an ambulance from Station 8 to Station 1, which locates all of the ambulances in urban areas and removes the only ambulance in the western half of the county. Longer RTTs (nine and ten minutes) move an ambulance from Station 7 to Station 14, which removes an ambulance from the area of highest call volume to an area at the outskirts of the urban area.

To further examine the ambulance locations, the marginal increase in the expected number of patients that survive in rural areas are compared to the urban areas. Recall that Hanover County is partitioned into urban and rural areas. Table 7 reports the difference in the expected number of patients that survive in the rural and urban areas when optimizing the T minute RTT as compared to optimizing patient survival, $T = 4, 5, ..., 10$. In Table 7, a negative value indicates

Table 2: Input parameters: nodes, demand, and stations

Node i	h_i	r_i	$Z_{i,1}$	$Z_{i,2}$	$Z_{i,3}$	Node i	h_i	r_i	$Z_{i,1}$	$Z_{i,2}$	$Z_{i,3}$
1	273	1	1			47	11	0		12	3
2	19	0	2		15	48	14	0		12	3 14
3	10	0	3	14	12	49	16	0		12	
4	93	1	4			50	22	0		13	5 6 10
5	11	0	5			51	15	0		13	5 6
6	103	1	6	13	7 10	52	13	0		14	3 6 13
7	652	1	7		6 14	53	9	0		15	2 8
8	13	0	8		15	54	41	1		10 13	1 6
9	11	0	9			55	13	0		2 15	
10	109	1	10		6 13	56	7	0		3 12	14
11	19	0	11			57	15	0		3 12 14	
12	13	0	12		3 14	58	10	0		3 14	7
13	40	1	13	6	10	59	27	1		3 14	7 12
14	35	0	14	3	7 12	60	11	0		3 14	
15	14	0	15		2 8	61	86	1		6 10 13	7
16	19	0.7368		1	4	62	123	1		6 10 13	
17	61	1		1	10	63	17	0		6 13	14
18	15	0		1	11	64	35	1		6 13	7 14
19	18	1		1	10 11	65	165	1		6 7	10 13
20	150	1		1	10 13	66	275	1		6 7	13 14
21	22	0		1	4 11	67	122	1		7 14	3 6
22	222	1		1		68	166	1		7 14	3 6
23	22	0		2	15	69	39	0		8 15	
24	5	0		2		70	7	0			2
25	10	0		3	14	71	10	0			3
26	5	0		3	12 14	72	29	0			4
27	16	0.4375		4	1	73	2	0			8
28	22	0.5		4		74	20	0			9
29	35	0		5	13	75	21	0.762			11
30	3	0		5	10 13	76	9	0			12
31	17	0		5	10 13	77	18	0			1 5 10 13
32	18	1		6	7 10 13	78	18	1			1 10 11
33	134	1		7	6	79	25	0			1 11
34	64	1		7	3 14	80	4	0			1 4
35	8	0		8	11	81	2	0			1 5
36	73	0		8	15	82	3	0			2 8 15
37	24	0		8		83	22	0			3 6 13 14
38	20	0		9	11	84	15	0			3 7 12 14
39	23	0		9		85	8	0			3 12
40	37	1		10	1	86	21	0			3 14
41	88	1		10	1 6 13	87	1	0			5 6 13
42	5	1		10		88	9	0			8 9 11
43	27	0.4074		11	1	89	9	0			8 11
44	13	0		11	9	90	14	0			8 15
45	19	0		11	8 9	91	14	0			8 9
46	15	0.5333		11							

Table 3: Marginal increase in the number of Priority 1 patients that survive and the nine minute RTT for optimal patient survival solutions

	Optimal Marginal Increase in Number of Patients that Survive			Marginal Increase in 9 Minute RTT		
N	Overall	Rural	Suburban	Overall	Rural	Suburban
3	216	8	208	741	29	712
4	295	10	285	1009	39	970
5	353	12	341	1199	46	1154
6	390	14	376	1323	53	1270
7	418	33	385	1422	124	1298
8	443	45	398	1507	166	1341
9	463	47	415	1571	173	1399
10	476	56	420	1617	203	1415
11	488	61	427	1654	219	1435
12	497	68	429	1685	246	1439
13	505	76	430	1714	270	1443
14	513	82	431	1740	294	1446
15	520	88	432	1762	313	1449

Table 4: Number of ambulances located at stations for optimal patient survival solutions

	Number of ambulances located at stations														
N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
3	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0
4	1	0	0	0	0	1	1	0	0	1	0	0	0	0	0
5	1	0	0	0	0	1	2	0	0	1	0	0	0	0	0
6	2	0	0	0	0	1	2	0	0	1	0	0	0	0	0
7	1	0	0	1	0	1	2	1	0	1	0	0	0	0	0
8	1	0	0	1	0	1	2	1	0	1	0	0	0	1	0
9	2	0	0	1	0	1	2	1	0	1	0	0	0	1	0
10	2	0	0	1	0	1	2	1	0	1	1	0	0	1	0
11	2	0	0	1	0	1	2	1	0	1	1	0	0	1	0
12	2	1	0	1	0	1	2	1	0	1	1	0	1	1	0
13	2	1	0	1	0	1	2	1	0	1	1	1	1	1	0
14	2	1	0	1	0	1	2	1	1	1	1	1	1	1	0
15	2	1	0	1	1	1	2	1	1	1	1	1	1	1	0

lives lost whereas a positive value indicates lives saved (both are in terms of the difference in expected number of lives). Table 7 indicates that the nine and ten minute RTTs save rural patient lives at the expense of urban patient lives whereas four to six minute RTTs save urban patient lives at the expense of rural patient lives (with the exception of the four minute RTT with $N = 6$). Note that most lives saved are urban patient lives, and these patient survival patterns reduce the disparity in patient outcomes between urban and rural areas. This suggests that although the nine and ten minute RTTs do not always maximize patient survival, they may be more equitable in terms of patient survival rates between urban and rural areas.

6 Conclusions

Optimally locating ambulances to improve patient survivability is a challenging problem. This paper proposes a model to evaluate a class of EMS performance measures—response time thresholds—

Table 5: Difference in expected number of patients that survive for optimal RTT solutions as compared to optimal patient survival solutions

	$S_N^T - S_N^*$ Overall						
N	$T=4$	$T=5$	$T=6$	$T=7$	$T=8$	$T=9$	$T=10$
3	0	0	0	0	0	0	0
4	-4.28	0	0	0	0	0	0
5	0	0	0	0	0	0	0
6	-0.99	0	0	0	0	-0.87	-0.87
7	-1.49	-1.49	-0.81	0	0	-2.10	-2.10
8	-3.42	0	0	0	0	0	-2.33
9	0	0	0	0	0	0	0
10	-1.39	-1.39	0	0	0	0	0
11	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0
14	-0.84	0	0	0	0	0	0
15	0	0	0	0	0	0	0

Table 6: Number of ambulances located at stations for $N = 7$ cases

	Number of ambulances located at stations														
Objective Function	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Patient Survival	1	0	0	1	0	1	2	1	0	1	0	0	0	0	0
4 minute RTT	2	0	0	1	0	1	2	0	0	1	0	0	0	0	0
5 minute RTT	2	0	0	1	0	1	2	0	0	1	0	0	0	0	0
6 minute RTT	1	0	0	1	0	1	2	0	0	1	0	0	0	1	0
7 minute RTT	1	0	0	1	0	1	2	1	0	1	0	0	0	0	0
8 minute RTT	1	0	0	1	0	1	2	1	0	1	0	0	0	0	0
9 minute RTT	1	0	0	1	0	1	1	1	0	1	0	0	0	1	0
10 minute RTT	1	0	0	1	0	1	1	1	0	1	0	0	0	1	0

Table 7: Difference in expected number of patients that survive in rural and urban areas for optimal RTT solutions as compared to optimal patient survival solutions

	$S^T - S^*$ Rural							$S^T - S^*$ Urban						
N	$T=4$	$T=5$	$T=6$	$T=7$	$T=8$	$T=9$	$T=10$	$T=4$	$T=5$	$T=6$	$T=7$	$T=8$	$T=9$	$T=10$
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	-2.24	0	0	0	0	0	0	-2.03	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1.38	0	0	0	0	15.48	15.48	-2.37	0	0	0	0	-16.35	-16.35
7	-16.55	-16.55	-6.52	0	0	10.84	10.84	15.06	15.06	5.72	0	0	-12.94	-12.94
8	-17.22	0	0	0	0	0	0.76	13.79	0	0	0	0	0	-3.09
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	-4.20	-4.20	0	0	0	0	0	2.81	2.81	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	-0.84	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0

in terms of their resulting patient survival rates. EMS performance measures are assessed in the decision context of locating ambulances at a given set of stations.

The results are obtained using data from Hanover County, a rural/urban county in Virginia. The results, therefore, should not be interpreted to provide a general policy for all types of EMS systems, since the results depend on travel times and call locations that may not be characteristic of other geographic resgions. However, the proposed model is transferable and can be used to evaluate EMS performance measures for other EMS systems.

The results suggest that locating ambulances to maximize RTTs can serve as effective proxies for patient survival. In particular, for Hanover county results indicate that seven and eight minute response time thresholds simultaneously maximize patient survival. Interestingly, these results are counterintuitive, since experts have suggested that response times should be four to six minutes in order to save patient lives. However, while medical evidence clearly suggests that shorter response times increase patient survival—calls are spread out over a geographical area such that it is only possible to respond to a small fraction of calls within four minutes because of ambulance travel times. Thus, EMS systems must balance the ability to respond to a small number of calls in a very short period of time (e.g., four minutes) with a high probability of patient survival with being able to respond to a greater number of calls in a longer period of time (e.g., eight minutes) with a lower probability of patient survival.

The results also indicate that longer RTTs (9 and 10 minutes in the case of Hanover county) result in more equitable patterns of patient survival. In these solutions, patient lives were saved in the rural areas at the expense of losing patient lives in the urban areas, which reduced the disparity between patient survival rates in urban and rural areas. Since a 9 minute RTT is the most common performance measure used by EMS systems in the United States, this suggests that many EMS systems implicitly consider patient equity or "fairness."

Analyzing EMS performance measures is an important problem, since performance measures are used to determine how EMS resources are used. RTTs are one of the most commonly used performance measure for EMS systems nationwide. This measure is easily understood, and much research has been to study optimal decisions in EMS for given RTT measures. This paper shows that by choosing the right RTTs, decision makers can simultaneously optimize patient survival rates.

We point out that this paper examines EMS performance measures in the context of locating ambulances at stations, which is just one resource allocation decision. Combining this approach to take into account the effect of dispatching policies and scheduling, for example, is important for optimally using EMS resources for saving patient lives.

Acknowledgements

It is a pleasure to acknowledge Battalion Chief Henri Moore, Jr., Chief Fred C. Crosby, II, Mr. Lawrence Roakes, and Mr. Edward Buchanan of the Hanover County Fire and EMS Department in Hanover County, Virginia, for the knowledge, experience, and data they provided to support this research effort. Suggestions for the medical component of this research from Dr. Joseph P. Ornato, M.D., are gratefully acknowledged.

References

[1] R. Batta, J. M. Dolan, and N. N. Krishnamurthy. The maximal expected covering location problem: Revisited. *Transportation Science*, 23(4):277 – 287, 1989.

[2] L. Brotcorne, G. Laporte, and F. Semet. Ambulance location and relocation models. *European Journal of Operations Research*, 147:451 – 463, 2003.

[3] K. R. Chelst and Z. Barlach. Multiple unit dispatches in emergency services: Models to estimate system performance. *Management Science*, 27(12):1390 – 1409, 1981.

[4] K. R. Chelst and J. P. Jarvis. Estimating the probability distribution of travel times for urban emergency service systems. *Operations Research*, 27(1):199 – 204, 1979.

[5] R. Church and C. ReVelle. The maximal covering location problem. *Papers of the Regional Science Association*, 32:101 – 108, 1974.

[6] M. Daskin. A maximum expected covering location model: Formulation, properties and heuristic solution. *Transportation Science*, 17(1):48 – 70, 1983.

[7] R. Davis. Atlanta becomes a template for improving cardiac-arrest survival rates. *USA Today*, 23 August 2007.

[8] R. Davis. Only strong leaders can overhaul EMS. *USA Today*, 20 May 2005.

[9] V. J. De Maio, I. G. Stiell, G. A. Welss, and D. W. Spaite. Optimal defibrillation response intervals for maximum out-of-hospital cardiac arrest survival rates. *Annals of Emergency Medicine*, 42:242 – 250, 2003.

[10] E. Erkut, A. Ingolfsson, and G. Erdogan. Ambulance location for maximum survival. *Naval Research Logistics*, 55(1):42 – 55, 2008.

[11] J. Fitch. Response times: Myths, measurement and management. *Journal of Emergency Medical Services*, 9:46 – 56, 30.

[12] J. B. Goldberg. Operations research models for the deployment of emergency service vehicles. *EMS Management Journal*, 1(1):20 – 39, 2004.

[13] L. V. Green and P. J. Kolesar. Improving emergency responsiveness with management science. *Management Science*, 50:1001 – 1014, 2004.

[14] J. P. Jarvis. Optimization in stochastic systems with distinguishable servers. Technical report tr-19-75, MIT, Cambridge, MA, 1975.

[15] J. P. Jarvis. Approximating the equilibrium behavior of multi-server loss systems. *Management Science*, 31(2):235 – 239, 1985.

[16] M. P. Larsen, M. S. Eisenberg, R. O. Cummins, and A. Hallstrom. Predicting survival from out-of-hospital cardiac arrest-a graphic model. *Annals of Emergency Medicine*, 22:1652 – 1658, 1993.

[17] R. C. Larson. A hypercube queuing model for facility location and redistricting in urban emergency services. *Computers & Operations Research*, (67 – 95), 1.

[18] R. C. Larson. Approximating the performance of urban emergency service systems. *Operations Research*, 5(845 – 868), 23.

[19] R. C. Larson and M. A. McKnew. Police patrol-initiated activities within a systems queuing model. *Management Science*, 7(759 – 774), 28.

[20] L. A. McLay. A maximum expected covering location model with two types of servers. *IIE Transactions*, to appear, 2009.

[21] V. N. Mosesso, Jr., M. M. Newman, J. P. Ornato, and P. M. Paris. Law enforcement agency defibrillation (LEA-D): proceedings of the national center for early defibrillation police aed issues forum. *Resuscitation*, 54:15 – 26, 2002.

[22] G. Nichol, A. S. Detsky, I. G. Stiell, K. Or, G. Wells, and A. Laupacis. Effectiveness of emergency medical services for victims of out-of-hospital cardiac arrest: a metaanalysis. *Annals of Emergency Medicine*, 27:700 – 710, 2002.

[23] I. G. Stiell, G. A. Wells, and B. J. Field. Improved out-of-hospital cardiac arrest survival through the inexpensive optimization of an existing defibrillation program: Opals study phase ii. ontario prehospital advanced life support. *The New England Journal of Medicine*, 351(7):647 – 656, 1999.

[24] A. J. Swersey. *Operations Research and the Public Sector*, volume 6 of *Handbooks in Operations Research and Management Science*, chapter The Deployment of Police, Fire, and Emergency Medical Units, pages 151 – 200. ElseVier, North-Holland, Amsterdam, 1994.

[25] T. D. Valenzuela, D. J. Roe, S. Cretin, D. W. Spaite, and M. P. Larsen. Estimating effectiveness of cardiac arrest intervention-a logistic regression survival model. *Circulation*, 96:3308 – 3313, 1997.

[26] R. A. Waaelwijn, R. de Vos, J. G. P. Tijssen, and R. W. Koster. Survival models for out-of-hospital cardiopulmonary resuscitation from the perspectives of the bystander, the first responder, and the paramedic. *Resuscitation*, 51:113 – 122, 2001.