European Journal of Operational Research 75 (1994) 567-581
North-Holland

Case Study

Developing and validating a decision support system for locating emergency medical vehicles in Louisville, Kentucky

John F. Repede *

Department of MIS / OM, College of Business Administration, University of North Carolina-Charlotte, Charlotte, NC 28223, USA

John J. Bernardo

Department of DSIS, College of Business and Economics, University of Kentucky, Lexington, KY 40506-0034, USA

Received May 1992; revised March 1993

Abstract: Decreasing the elapsed time to respond to an emergency is a goal of Emergency Medical Services (EMS). The size of the ambulance fleet and the location of ambulances within the service area are two factors that the EMS planners can control; these two factors directly affect the systems response time and contribute to the attainment of this goal. In this paper, a maximal expected coverage location model with time variation (TIMEXCLP) is developed and integrated into a decision support system (DSS) to aid EMS planners to allocate vehicles within their service area. In applying TIMEXCLP to the Louisville, Kentucky, EMS system, response time was decreased by 36%.

Keywords: Health services; Location; Mathematical programming; Simulation; Practice

Introduction

The goal of emergency medical services (EMS) is the reduction of mortality resulting from acute illness and trauma. Attainment of this goal relies upon the rapid response of an appropriately staffed and equipped ambulance, on-scene stabilization of the patient's medical condition, and transportation to a medical facility with appropriate definitive care capabilities. The size of the ambulance fleet and the location of ambulances within the service area are two factors within the control of EMS planners which directly affect the system's response time and contribute to the attainment of this goal. Numerous studies have demonstrated the relationship between decreases in response time and corresponding decreases in mortality. For example, Cretin and Willemain (1979), Eisenberg et al., (1979), Mayer (1979), D'Agostino and Pozen (1981), and Reines (1984), all found a direct relationship between response time and mortality. It is the direct relationship

Correspondence to: John F. Repede, Department of MIS/OM, College of Business Administration, University of North Carolina-Charlotte, Charlotte, NC 28223, USA.

SSDI 0377-2217(93)E0256-W

between mortality and response time that makes the ambulance location problem an important issue for EMS system planners. The model developed in this paper allowed the Louisville EMS planners to determine vehicle locations that result in a 36% reduction in response time com-

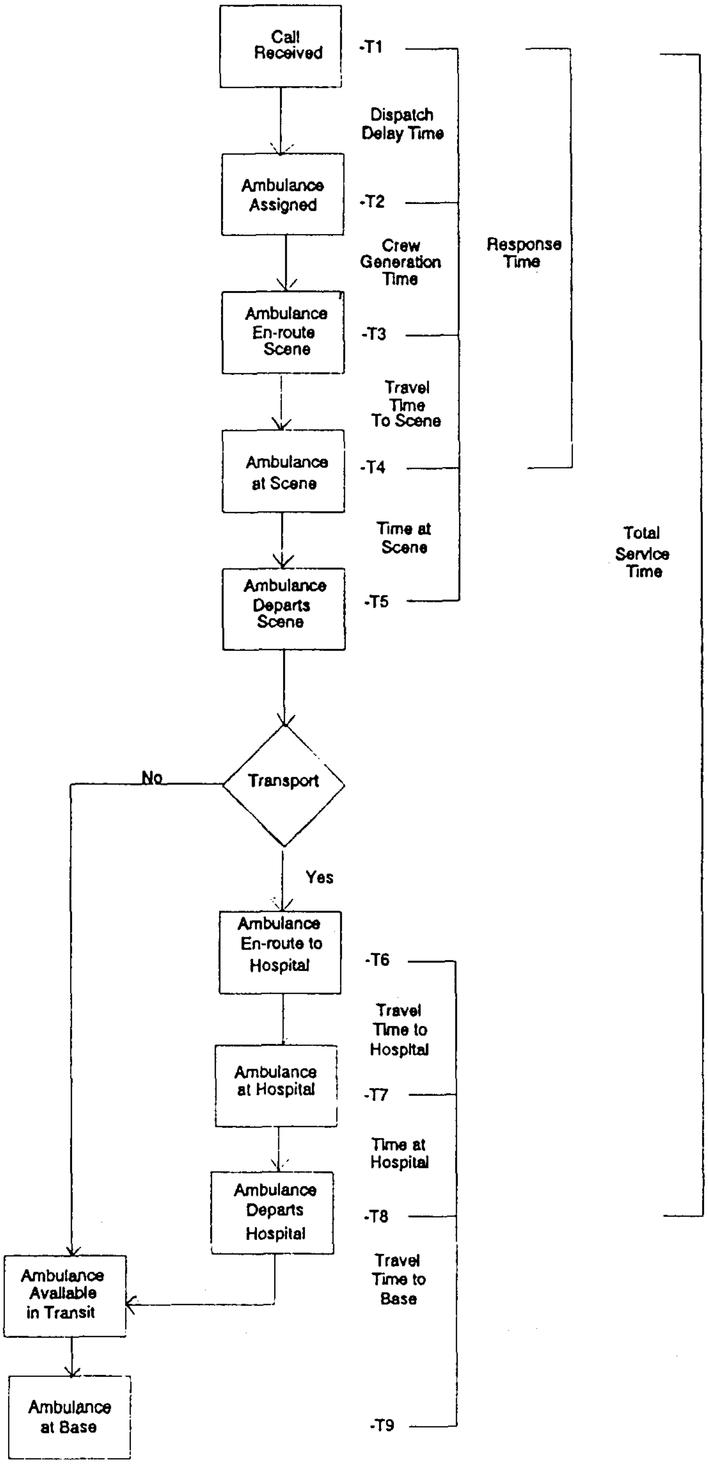

Figure 1. The EMS process

pared to the locations that result from present approaches.

In urban areas, the level of service sought by EMS planners is the de-facto national standard. The EMS system should respond to 95% of its daily demand for service within 10 minutes (Daskin, 1982). This service level is rarely attained. For example, it was found that the metropolitan Louisville area EMS actually responded to only 84% of its demand within ten minutes.

Present covering models, most notably those by Church and ReVelle (1974), and the improvement of their model by Daskin and Stern (1981), do not allow for time variation in the demands nor for the reallocation of vehicles over time. Because the Louisville EMS system faces a stochastic demand for its services and the parameters of the demand distribution change over time, the problem is non-linear and time variant. Applying present maximal covering approaches to a time varying situation resulted in a discrepancy between the model results and what actually occurs.

In this paper, a model that solves the maximal expected coverage location problem with time variation (TIMEXCLP) is developed and integrated with a simulation model into a decision support system (DSS) to aid EMS planners to allocate vehicles to demand nodes. The procedure gives EMS planners a tool that closely estimates service levels and that allows them to reallocate a changing active fleet size over time to respond to variability in demand. The DSS also allows the EMS planners to better utilize current capital equipment and to analyze alternative responses to changes in the demand pattern, the active number of vehicles at any given time, and other system characteristics. This is accomplished by giving the EMS planner a tool that can determine vehicle relocations for the new conditions that maximize the service level. By using a simulation model, Louisville can increase the accuracy of its estimate of expected coverage by 55% and can reduce its response time by 36% by using TIMEXCLP locations.

The work reported in this paper is part of a larger effort to improve the Louisville area EMS service. The purpose of this paper is to present the development and use of the TIMEXCLP model and the behavior of the model compared with existing approaches. In Section 1 of this paper, the decision problem facing EMS planners and the current literature that attempts to solve that problem is discussed. In Section 2, the results of the data analysis is presented. In Section 3, an overview of the DSS that was developed for the Louisville EMS system is presented. Section 4 discusses the development and testing of the simulation model. Section 5 contains the derivation of the TIMEXCLP model. In Section 6 the behavior of the TIMEXCLP model is investigated. Finally, the use of the approach to analyze alternative dispatching rules and sizes of districts is presented along with conclusions.

1. The EMS planning problem

The EMS planner faces decisions from both a strategic and a tactical viewpoint. Strategic decisions involve questions relating to the number of ambulances and their locations. Tactical decisions involve responses to situations that arise given a fixed number of vehicles. A decision support system was developed to aid a number of decisions faced by the Louisville EMS planners. Among these decisions were:

What are the best vehicle locations?
What adjustments must be made to locate an incremental vehicle?
How should maintenance be scheduled?
What are the best shift assignments?
Can better dispatching rules be used?

In this section, the EMS process along with existing work in the area will be discussed so that the complexity of the system can be appreciated.

The EMS process

A typical EMS service response is flow charted in Figure 1. At time T1, a communications center operator is initially made aware of a demand for service by way of a two-way radio or a telephone. An initial screening is made to determine: 1) whether or not an ambulance should be dispatched, and 2) the response code if a dispatch is indicated. The response code is 1 or 2 depending upon the perceived severity of the illness or trauma. The dispatcher then assesses the location and availability of the fleet and, based upon a predetermined dispatch assignment rule, assigns the call to one of the ambulance crews at time T2. The generally accepted dispatch assignment rule is to always send the closest available ambulance. However, this may be changed to alternative dispatch rules in the simulation model. The time elapsed from call receipt to ambulance assignment (T2 − T1) is referred to as the dispatch delay.

The time from when an ambulance is assigned (T2) until the vehicle is in route to the demand location (T3) is a delay which represents crew generation time. In systems which are staffed by full time personnel, the crew generation time is commonly included in the dispatch delay as T3 − T1.

The time interval between the crew's notification that they are traveling to the scene and notification of their arrival at the scene (T4) is the travel time to the scene. Although the system response time is more accurately represented by the time interval calculated as T4 − T1, it is common within EMS systems to view the response time exclusive of the dispatch delay. That is, response time is usually calculated as T4 − T2.

Although an initial screening for the appropriate use of an ambulance is usually performed by the dispatcher, this does not occur in all cases. At times, triage cannot be performed until an EMS crew is present at the scene. For this reason, the time at the scene is very low in a significant number of calls where it is determined by the crew that either their services are not needed or the patient refuses medical treatment and/or transportation. In these cases, the crew departs the scene at T5 and returns to their base location. During their return to base, the crew is available for another response if necessary.

In those cases where treatment has been rendered at the scene (for time length T5 − T4), the crew determines the destination hospital along with the travel code. After arriving at the hospital (T7), time is spent by the crew in transferring care of the patient to the hospital staff, completing reports, and cleaning and resupplying the ambulance until time T8. The crew then departs the hospital at time T8 and returns to their base location. The crew may not complete the return to the base location if they are assigned another call while traveling. For this reason, the total service time recorded for the previous call is computed as T8 − T1 rather than T9 − T1. Although crew utilization statistics include the travel time to base in computing service time (T9 − T2), the dispatch delay (T2 − T1) is excluded.

Work in the field

Current applications of management science models fall into three categories. Since calls arrive randomly, queuing approaches have been applied. Because the system is complex, many queuing models have not fully grasped the management problem. Consequently, simulation has been used to evaluate various location schemes. Since the basic problem is to cover demand, different forms of set covering and maximum covering models have also been formulated.

The most widely referenced queuing approaches for locating multiple facilities are the hypercube models of Larson (1974, 1975). Both models require the service time of the vehicles to be solely vehicle dependent. As we will see, the Louisville data suggests otherwise. Jarvis (1985) addressed the exponential service time assumption in the hypercube model by developing a procedure that allows general service time distributions which may depend on both the server and the customer type. However, as noted by Goldberg and Szidarovsky (1991), this method is computationally prohibitive when the Hypercube model is used and may be problematic when used with exponential correction factors in the Hypercube approximation model. Extensions of Larson's basic model by Jarvis (1985), Berman and Larson (1982), Benveniste (1985), Berman et al. (1987), Batta et al. (1989), and Goldberg et al. (1990a) are heuristic approximations for allocating multiple vehicles on a network. Because the assumptions of existing queuing models have not been found to exist in practice, many researchers have resorted to simulation.

One of the earliest applications of simulation to the ambulance location problem was conducted by the New York Scientific Center of IBM at the request of the Mayor of New York and reported by Savas (1969). Fitzsimmons (1973), Swoveland et al. (1973), and Berlin and Liebman (1974) used similar approaches. More recently, Liu and Lee (1988) used simulation to analyze a hospital EMS in Taipei based upon the earlier work of Uyeno and Seeberg (1984). The most recent application of simulation to the EMS loca-

Temporal Demand	Spatial Demand	
	Invariant	Variant
Invariant	Hakimi (1965) Torregas (1971) Volz (1971) Daskin (1981)	Church (1974) Siler (1977) Daskin (1983) Saydam (1985)
Variant		

Figure 2. A categorization of prescriptive EMS location covering models based upon demand assumptions

tion problem was by Goldberg et al. (1990b). There are two primary differences between the work reported here and that of existing simulation studies. First, the models developed in this application allow for more than one ambulance at a given location. Second, travel times are calculated as a function of call priority as was shown by the data analysis.

Figure 2 classifies the more relevant covering models based upon their demand assumptions. As can be seen, no models currently handle temporal variability in demand. The original set covering location model (SCLP) was developed by Torregas, Swain, ReVelle and Bergman (1971). This model minimizes the number of facilities required to cover all demand nodes within a specified distance. Application of the SCLP model to EMS systems has been limited because the model's formulation assumes that the volume of demand at every node is equal. The Maximal Covering Location Problem (MCLP) developed by Church and ReVelle (1974) was a significant extension to SCLP because it allowed for the specification of the portion of total demand that is located at each node. Because of this, MCLP was the first optimal location model to account for the spatial variation in the demand for service within the system. Charnes and Storbeck (1980) developed a multilevel, goal-oriented location covering model (MGLC) for siting vehicles in two-tier EMS systems by applying the MCLP approach within a goal programming framework. In 1982, Daskin (1982) developed the Maximal Expected Covering Location Problem (MEXCLP) which explicitly considered the possibility that the requested vehicle(s) are unavailable due to previous demands. Saydam and McKnew (1985) developed a separable programming approach to consistently solve MEXCLP.

Although MEXCLP is the most recent set covering approach to ambulance location, Batta (1989) and others found that due to the violation of the independence assumption, maximal covering models overestimate coverage. However, Bernardo and Repede (1988) found that MEXCLP also overestimates coverage because the model assumes that the relative proportions of demand at each node is constant over time. That is, overestimation results not only because the independence assumption is violated, but also because the demand for service varies over time. The TIMEXCLP model developed in this paper extends MEXCLP by incorporating temporal variation in the daily demand process in addition to spatial variation and multiple states of vehicle availability.

2. Data analysis

Regardless of the modeling approach chosen, the model must be based upon the system that it

Table 1
Results of the data analysis for the simulation input variables

Variable	Distribution	Groups	Required number of parameters
Demand inter-arrival time	Negative exponential	Weekday/District/Hour	1176
Dispatch delay	Gamma	Response code	2
Response Code	Binomial	District/Hour	168
Travel time to Scene	Gamma	Code/Path/Hour	2352
Time at scene	Gamma	Response code	2
Transport code	Empirical	Response code	6
Hospital destination	Empirical	District	21
Travel time to hospital	Gamma	Code/Path/Hour	2352
Time at hospital	Gamma	None	1

is to represent. To model the process of Figure 1, 47 499 calls received by the Louisville EMS were analyzed. Nine variables within the EMS service delivery process were identified as inputs to the simulation model. These were chosen because they represent random events within the delivery process which have an impact on the system's operating characteristics. The primary operating characteristics of concern within this study are the distribution of response times and the subse-

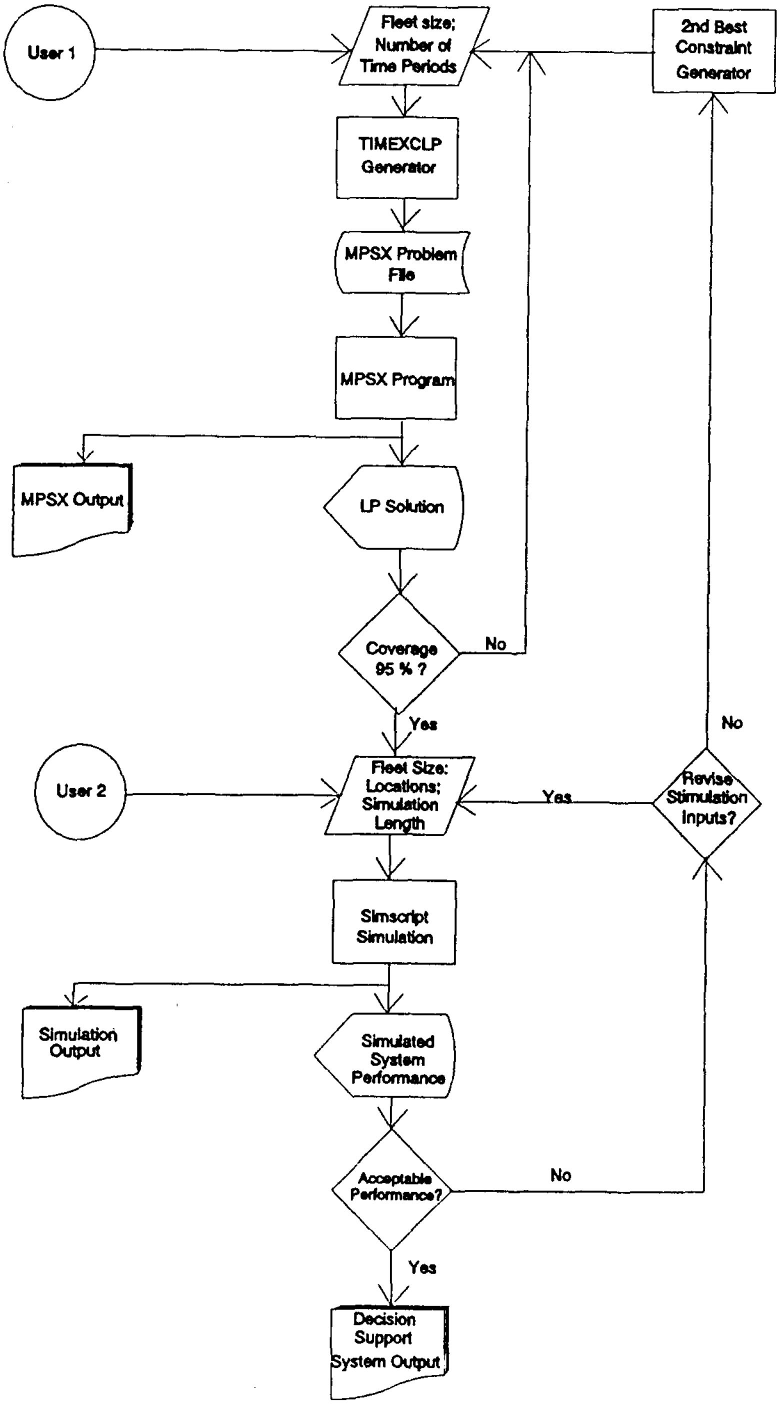

Figure 3. The EMS decision support system

quent level of coverage. The nine random processes are the demand process, the dispatch delay, the travel code to the scene, the travel time to the scene, the time at the scene, the hospital selection process, the travel code to the hospital, the travel time to the hospital, and the time at the hospital. For each of the variables, UNIFIT was used to determine the most appropriate distribution. Because this analysis determines the number of variables required in the simulation, a summary of the results of the statistical analysis is presented in Table 1. As can be seen, a total of 6 080 variables must be determined to account for the variability in the data and to accurately determine the EMS performance.

3. The decision support system

Figure 3 shows the logical flow of the decision support system. The DSS is composed of two models in the model base: the TIMEXCLP model and a simulation model. The TIMEXCLP model is a large scale expected covering model which is used to initially determine vehicle locations. Because such models have been found to overestimate coverage, the locations determined by TIMEXCLP are used as input for a simulation model that evaluates the locations for system characteristics. If the required system characteristics are not realized, the support system generates a new TIMEXCLP formulation which results in new locations that are again checked by the simulation. This procedure continues until the desired system characteristics are obtained. The simulation model can also stand alone to address other issues that are important to the EMS planner. This approach allows the EMS planner to analyze the EMS system characteristics under alternative scenarios.

The first step is to initialize the system with the appropriate information. Next, the TIMEXCLP model is repeatedly applied until the desired service level of 0.95 is obtained. Once this location scheme is determined, it is verified by using the TIMEXCLP vehicle assignments as the input to a simulation model. The simulation model yields coverage and response time data as output for each node and for the complete system. If the results are acceptable to the EMS planner, the vehicle assignments have been determined. If the results are not acceptable, for example side constraints on certain vehicle locations or the response time for a certain node, then the design procedure constructs a new TIMEXCLP formulation, and the procedure is repeated until the EMS planner is satisfied with the results.

The TIMEXCLP model itself is generated within the system to correspond to the system constraints supplied by the user. Because of the size of the model, this portion of the DSS was implemented on an IBM 3090 with three processors. The simulation model, which can also stand alone, was implemented on an IBM 3084. The effectiveness and efficiency of the DSS is related to the accuracy and usefulness of the simulation model and TIMEXCLP formulation. In the next section, the development of the simulation model is discussed along with a statistical evaluation of its behavior.

4. Development and validation of the simulation model

In this section, the EMS simulation model will be shown to be: 1) a valid abstraction of a generic EMS system, and 2) an accurate replication of the Louisville EMS system when operating characteristics of that system are used as model inputs. The simulation model was written in SIMSCRIPT II.5 and implemented on an IBM 3084.

Steady state

Before analysis can be made, a determination of the length of time required for the simulation

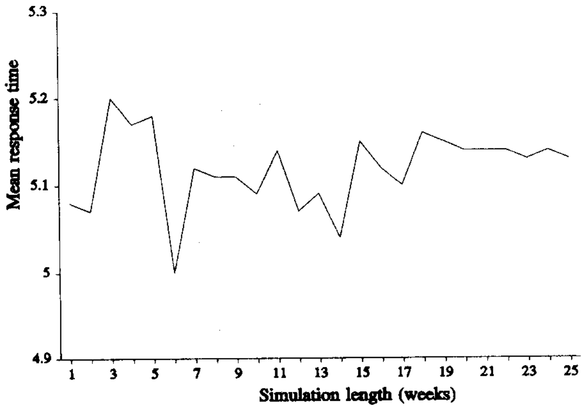

Figure 4. Mean response time by simulation length

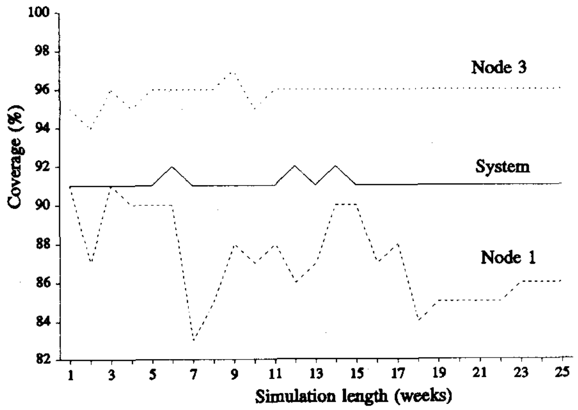

Figure 5. Coverage by simulation length

to reach steady-state behavior is required. Steady state behavior implies that the variance of the estimated performance measures of the EMS system is stationary. Calculated at simulated one week time intervals, the values of four different system performance measures are graphed in Figures 4 and 5. In Figure 4, system-wide mean response times computed for simulation lengths ranging from 1 to 24 weeks are plotted. A pattern of decay in the fluctuations of mean values is demonstrated. There is negligible variation in mean response times beyond 18 weeks.

In Figure 5, the mean coverage values for all nodes and the individual nodes 1 and 3 are plotted for simulation lengths of 1 to 25 weeks. Although node 1 coverage values had the greatest fluctuation, the variation of all nodes decayed after approximately 18 weeks. To be conservative, the first 24 weeks of each simulation were not used for accumulating statistics.

Variance of estimates

Two system performance measures were of particular importance in this study. These performance measures were the level of system-wide coverage and the mean response time. The time unit of analysis within this study was one day. Although, the simulation was run for 180 days, there is a limitation in drawing inferences from one extended run because of the lack of a confidence interval. The values of the performance measures used in the subsequent model validation represent the means of ten 180 day simulation runs where each of the ten runs utilized different random number streams for each of the random variables in the model.

Table 2
95% confidence intervals expressed as a percent of mean coverage by simulation length

Simulation length (days)	Coverage (C)	Standard deviation	Confidence interval (% of C)
30	0.9149	0.0056	1.09
90	0.9143	0.0054	0.98
180	0.9146	0.0019	0.33

Statistical tests

The approach used to validate the simulation model was to replicate the 180 day simulations ten times using the current Louisville data and locations as model inputs. In analyzing the resulting ten values of the percent of system coverage, the maximum difference between these values was less than 0.05%. In order to explore this further, thirty additional simulations were conducted.

Ten simulations of each of three lengths, 30, 90, and 180 days were subsequently run. The mean coverage for each of the three simulation lengths was 0.914. However, extending the simulation length from 30 to 180 days reduced the width of a 95% confidence interval by 70% as shown in Table 2. Table 3 shows similar results for the mean response time. The 95% confidence intervals are expressed as a percent of the simulation results. For example, we are 95% confident that the true system wide coverage is within one-third of one percent of the simulated coverage.

Accuracy of the estimates

Besides coverage and response time, the results of the ten simulations using current

Table 3
95% confidence intervals expressed as a percent of mean response time by simulation length

Simulation length (days)	Response time (RT)	Standard deviation	Confidence interval (% of RT)
30	5.48	0.185	3.37
90	5.44	0.123	2.26
180	5.46	0.057	1.04

Louisville ambulance locations were also compared to the actual total service time. Using appropriate statistical tests for the equality of means, none of the differences of the three performance measures were significant at the 0.05 level. Consequently, a close correspondence between the actual performance of the Louisville EMS system and the 180 day simulated performance was found.

5. Development of the TIMEXCLP model

In designing an EMS delivery system, the planner is faced with the task of selecting one set of ambulance locations from among a large set of feasible locations. The EMS planner seeks that set of locations which meets his goals with the minimum expenditure of resources. As previously stated, the goal most frequently sought by urban EMS planners is the national standard of 95% of the requests for service responded to within 10 minutes. The TIMEXCLP model provides the initial ambulance locations to attain that goal.

Objective function

Since the EMS planner desires to plan a location scheme that will give a desired level of system wide coverage, the objective function of the TIMEXCLP is expected system wide coverage. The objective function can be determined by utilizing the prior probabilities obtained from data analysis. Consider the EMS system as a network with K demand nodes. The expected demand at a given node, k, per unit time is expressed as $C_{t,k}$. Demand originating at node k during time period t is defined as $D_{t,k}$. The length of the time periods should be defined such that $D_{t,k}$ is constant during the period. Previous ambulance location optimization models have used one day as the unit of time within the model. However, analysis of the Louisville data has showed that a significant variation in the spatial distribution of demand can occur within the day. In order to capture this aspect of the demand pattern, the expected demand is partitioned into a two-dimensional matrix of T time periods and K demand nodes.

For a fleet size of S_t vehicles at time period t, the average system wide probability that during period t a randomly selected vehicle will be busy can be expressed as the ratio of the expected service time to total service time available. The system wide probability of an ambulance being busy is an estimate of any one ambulance being busy if the independence assumption is satisfied. If we define u to be the mean service time in minutes, and m to be the length of the time period t in minutes, we can represent this probability, P_t, as:

$$P_t = \sum_k (C_{t,k})(u)/S_t(m) \tag{1}$$

where:
u = Mean service time in minutes.
m = Minutes in period t.

$$S_t \geq \max_k \{\mathrm{INT}[(C_{t,k})(u)/m]\}, \tag{2}$$

where INT is the integer value of [...] rounded to the next highest value.

The constraint on fleet size, S_t, insures that the probability, P_t, is greater than zero and less than 1. This requirement also insures that the system's capacity exceeds its expected service demand. Consequently, the constraint also provides a lower bound for initiating the search for an optimal number of vehicles.

To determine the objective function for TIMEXCLP, Daskin's (1982) formulation is used. This formulation assumes that the independence assumption holds. Although Batta (1989) and others showed that this assumption does not hold, Birge and Pollack (1989) showed that the estimate is acceptable for planning models. A demand node is considered to be 'covered' by an ambulance if it can be reached from that ambulance's location within the required length of time. If one ambulance is placed such that it can 'cover' a given demand node, then that node is covered once. The expected coverage of that node is represented by the product of the probability of the ambulance being available during that time period and the portion of demand located at that node during that period of coverage. The complete TIMEXCLP formulation is:

$$\operatorname{Max} \sum_{t}^{T} \sum_{k}^{K} \sum_{j}^{S_t} (l - P_t)(P_t^{j-1})(D_{t,k})(Y_{t,j,k}) \tag{3}$$

subject to:

$$\sum_{j}^{S_t} Y_{t,j,k} = \sum_{i} (X_{t,i})(a_{t,i,k}) \quad \forall\, k \text{ and } t, \tag{4}$$

$$\sum_{i} X_{t,i} = S_t \quad \forall t, \tag{5}$$

where $X_{t,i}$ is integer and

$$a_{t,i,k} = \begin{cases} 1 & \text{if at time } t, \text{ location } i \text{ covers } j, \\ 0 & \text{otherwise.} \end{cases}$$

Discussion of the coverage matrix

In the development of the objective function, equation (3), $Y_{t,j,k}$ represents whether or not the j-th vehicle added to the fleet during period t covers node k. Since it is reasonable to expect that some nodes cannot cover other nodes, the value that $Y_{t,j,k}$ has depends upon where the j-th vehicle is located. If all nodes could cover each other, it would make no difference where any of the ambulances were located. The problem would simply be one of minimizing fleet size. The system's ability to provide inter-node coverage is represented by a coverage matrix. The coverage matrix is subsequently used to constrain the values of $Y_{t,j,k}$ to be zero when an infeasible coverage alternative is evaluated.

Within the network, let the number of feasible ambulance location nodes be represented by $i = 1, 2, \ldots, I$. The coverage matrix is a $T * I * K$ matrix whose elements $a_{t,i,k}$ take value 1 if at time t, node k is covered by location i, and value 0 otherwise. The coverage matrix is a deterministic representation of the ability of the network's vehicle nodes to cover demand nodes. Goldberg and Paz (1991) gave an example of the potential inaccuracy from using deterministic travel times and developed a model which incorporates stochastic travel times. However, their model requires a stationary call arrival rate and that all calls are answered by a vehicle originating from its base. Neither of these requirements hold for the Louisville EMS. In the Louisville EMS, 40% of the responses were answered by vehicles not located at their home base. However, response time is also a stochastic variable. Consequently, a method to determine the values in the coverage matrix is necessary. In the Louisville application, a node is considered covered if the median value of the response time distribution is less than or equal to 10 minutes. For consistency in evaluating model performance, this determination of the coverage matrix was used in all subsequent formulations of both the MEXCLP and TIMEXCLP models. Previous maximum covering approaches have used the mean response time to determine the values of the coverage matrix. However, since response time has a distribution that is significantly skewed, the use of the mean to estimate the coverage matrix resulted in an over estimation of response times by all models. Consequently, the median of the distribution was found to provide more accurate estimates.

Preservation of coverage constraint

The preservation of coverage constraint, equation (4), insures that the vehicles assigned do in fact cover the node as represented by $Y_{t,j,k}$. Let $X_{t,i}$ represent the number of ambulances located at node i during time period t. The number of ambulances that cover a given node, k, during period t can be expressed as the sum over j of $Y_{t,j,k}$. This number cannot be more nor less than the sum of the numbers of vehicles located at each of the nodes that cover k.

Conservation of fleet size

The number of ambulances located in each time period at each location node, $X_{t,i}$, cannot exceed the active fleet size, S_t. We can insure this constraint by summing over all vehicle locations at a given time period and equating this to the active fleet size. This is equation (5). The fleet size must also be an integer.

Extensions

A great variation exists between EMS systems in the configuration of the prehospital delivery and transportation subsystems. The service may be provided by municipalities, hospitals, or private companies. The system may use a combination of vehicles, or a tiered response (e.g., Pascarelli and Katz, 1978; Eaton et al., 1985). Figure 6 gives a categorization of the types of EMS services. Modern EMS services are no longer constrained by fixed locations, but rather, de-

Active Fleet Size	Ambulance Locations	
	Fixed	Variable
Fixed	Type I	Type III
Variable	Type II	Type IV

Figure 6. A classification of EMS delivery systems

pending upon the system, the vehicle locations can be changed. Although when $T = 1$, TIMEXCLP is identical to MEXCLP, extensions to TIMEXCLP allow for altering the active fleet size and its location that are not possible with MEXCLP.

The preceding formulation of the TIMEXCLP model represents a type IV EMS system. In a type IV EMS system, both the locations and the size of the active fleet can be varied because the personnel schedule allows alternative shift lengths and start and end times. In Louisville, temporary locations (such as shopping centers) can be utilized for vehicle placement to improve the ability of the system to respond to peak period demands. This is a type III EMS system. In a type III system, an additional constraint must be added to the above formulation to require a constant fleet size.

In a type II system, ambulances respond from permanent bases, such as hospitals or fire stations; however, the size of the active fleet can change over time. That is, additional units can be activated during periods of high demand. This system makes it possible for the EMS service to save on operational and personnel expenses when it is anticipated that demand for the ambulance service will be low. The type I system is represented by the traditional fire department operated service. In this type of facility location problem, the physical facilities housing the equipment and personnel cannot be relocated. In a type I EMS, both the number of ambulances and their locations remain constant.

Additional extensions are possible to accommodate vehicle relocation caused by breakdowns, personnel shortages, or demands for multiple vehicle responses to multi-casualty incidents. However, these extensions only require constraints to be added to the basic TIMEXCLP formulation in a manner similar to the above.

6. Statistical tests of TIMEXCLP behavior

Two questions were addressed to investigate the behavior of the TIMEXCLP model. First, does the implementation of TIMEXCLP vehicle locations result in improved coverage and response time performance when compared with MEXCLP vehicle locations? Second, does TIMEXCLP result in a smaller error than MEXCLP?

To answer the above questions, 35 scenarios were generated for the Louisville EMS system. These scenarios had seven fleet sizes ranging from 5 to 11 ambulances, and included 1, 2, 4, 8, and 24 time periods per day. TIMEXCLP was run for each scenario and the resulting optimal vehicle locations were determined. To estimate performance of the resulting vehicle locations in application, the locations were tested in the simulation model for the Louisville EMS system. Ten simulation runs were made for each scenario resulting in a total of 350 observations. In all situations, a performance comparison was made between the vehicle locations of TIMEXCLP with those of MEXCLP.

Table 4 shows the differences in the average error, system coverage, and response times for the results of MEXCLP and TIMEXCLP vehicle locations. As can be seen, there is an improvement in each of the three performance measures.

Coverage: The percent of calls that are responded to within ten minutes is referred to as coverage. The TIMEXCLP vehicle locations result in an increase of 0.7 in system wide coverage. This difference is significant at the $\alpha = 0.01$ level. For the Louisville application, nine ambulances are required by MEXCLP to reach 95% coverage; however, TIMEXCLP requires only eight ambulances to reach the 95% coverage level. Although the average difference in coverage between the two models is only 0.7, this difference

Table 4
Test for the difference between TIMEXCLP performance and MEXCLP

Performance measures	Mean difference	t	α
Error	1.60	2.39	0.03
Coverage	0.70	3.48	0.01
Response time	0.30	1.90	0.05

translates into an additional ambulance at a cost of \$250 000 per year for the Louisville EMS system.

Response time: The elapsed time between receiving an emergency call and the time that an ambulance is at the scene of the emergency is called response time. The response time of TIMEXCLP locations is 0.30 minutes lower than MEXCLP locations. Although the decrease is only seconds, it is significant at the $\alpha = 0.05$ level.

Model error: The difference between the coverage estimated by the model and the actual coverage is referred to as model error. As measured by the simulation model for all 35 scenarios, the average error of TIMEXCLP when compared with that of MEXCLP is 79% lower. The improvement of TIMEXCLP over MEXCLP was significant at the $\alpha = 0.03$ level.

7. Application

The DSS was initially used to perform three studies. The first use of the DSS was to investigate alternative dispatch rules. The EMS system was installing real time dispatching and control equipment which would make alternative and computationally difficult dispatching procedures realistic to employ. The EMS managers also wanted to approach the funding agency for monies to add an additional vehicle. Consequently, the second use of the DSS was to determine the increase in service levels that an additional vehicle would provide. The third initial use of the model was to provide insight into ambulance placement when special events took place.

Alternative dispatching rules

Since a real time dispatching and control equipment was being installed, the EMS manager desired an evaluation of an alternative dispatch policy. The EMS system management was concerned with a domino effect in backup coverage that was occurring during peak demand periods. This happens when vehicle b is called out of its primary response district to cover demand in another district, a, whose primary response vehicle is already engaged. Although vehicle b is unable to respond within the standard time it is dispatched because it is the closest available unit. While b is engaged in a's coverage zone, a demand occurs in b's district and must be answered by yet another vehicle, c, which is also unable to respond within the standard time. However, it is dispatched because it is the closest available unit. This results in two demands not being covered within the standard time. However, if vehicle c had initially been sent to cover the demand in a, then b would have been available to cover its demand and only one demand would not have been covered in the standard time. Management realized that this increase in coverage would result at the expense of overall mean system response time. But since the primary objective was to maximize coverage and not minimize response time, they were interested in exploring this concept further.

An alternative dispatch rule was proposed. If demand occurs in a district whose primary vehicle is unavailable to respond then the vehicle with the least likelihood of receiving a call in its primary district would be dispatched. If one or more units could respond with the standard time, the dispatched vehicle would come from this set.

The simulation component of the DSS was used to study the effects of the alternate dispatch policy. The results of the first phase of the analysis is shown in Table 5. It should be noted these results are from static, user specified locations for each fleet size and not TIMEXCLP output. As expected, the mean response time is greater for the alternative dispatch rule although there is a slight increase in coverage at higher utilization levels. The reported results are for the entire 24 hour day and not just for peak demand periods. It would be expected that the alternative dispatch rule would result in a larger improvement if it was applied to only peak demand periods. However, before additional analysis could be per-

Table 5
Preliminary analysis of the closest available vehicle and the alternative dispatch rules

Fleet size	Coverage		Mean response time	
	Closest	Alternative	Closest	Alternative
4	0.6247	0.6309	12.22	12.23
5	0.7812	0.7869	8.24	8.28
6	0.8705	0.8731	6.38	6.46
7	0.9147	0.9110	5.57	5.69
8	0.9307	0.9260	5.28	5.34

formed, management determined that a dispatch policy other than the closest available unit, even if it provided better overall performance, would not be politically nor legally acceptable. Consequently, additional analysis was not performed.

Affect of the additional vehicle on service levels

The second area of inquiry in the study was finding the best base location for an additional vehicle to achieve an expected coverage of at least 95%. Temporary locations were not being considered, and because of budgetary constraints on the construction or leasing of a new facility, existing location would have to be employed. Application of MEXCLP indicated that an expected coverage exceeding 95% could be achieved with the permanent placement of a second vehicle at location 6. The TIMEXCLP model was then formulated utilizing 24 hourly time periods. It was found that by using only ten vehicle relocations within a 24 hour period the desired coverage (95.3%) could be achieved with the existing fleet of seven active units. With relocations an additional unit is not required. However, an additional 2.5% increase in coverage could be obtained from an eighth vehicle being added to the fleet.

Vehicle locations for special activities

The model was also successful in determining vehicle assignments for special activities. Two types of special activities had to be considered. The first type of special activity is due to the requirements of the EMS itself. These activities included in-service education, crew evaluation, or vehicle inspection. These activities are characterized by the fact that other vehicles cannot cover the special activity other than the assigned vehicle(s). For the second type of activity, such as a sporting event, other vehicles can cover the special event node if demand warrants.

The first type of special activity is characterized by the fact that each crew and vehicle must be scheduled to participate in the special activity. To account for the special activity, a dummy node i is added to the model where X_{ti} represents the number of vehicles participating in activity i during period t. If only one vehicle is to participate at any one time, then $X_{ti} \leq 1$, for all t. Because the special activity is a dummy node, it has no demand; it does not cover any other nodes; and no other nodes can cover the special activity other than the assigned vehicle. Since the objective function is the maximization of coverage, vehicles will be assigned to this dummy node only when forced by the additional constraint

$$\sum_t X_{ti} = S.$$

The resulting application of the model identifies the scheduling of the vehicle(s) within the appropriate time which will minimize the reduction in expected coverage as well as suggest the optimal relocation of the remaining vehicles.

The second type of special event is characterized by an activity which is expected to have demand for emergency medical service but by its nature is atypical of normal demand patterns. Examples would include events such as a Fourth of July fireworks performance, outdoor festivals with large attendance, and sporting events such as the Kentucky Derby.

To analyze the second type of special application, an additional demand node is created. Demand estimates for each of the time periods are developed. The model is then re-solved with the additional node and time-dimensioned demand. It is not necessary for the additional demand node to be counted among candidate locations. That is, it's demand can be forced to be covered from other locations by adding a constraint that requires the number of vehicles located at the special node to be zero for all or some of the time periods. For example, Churchill Downs would become a candidate for location during the daytime hours but would be withdrawn as a location candidate during night time hours since the Derby's additional expected demand would not occur during those time periods.

Implementation

Both of the later recommendations were presented to the EMS management. The estimates of service changes that were possible by using the DSS resulted in a funding increase for the additional vehicle. The approach suggested for scheduling also proved to be fruitful. The EMS management accepted the model for scheduling EMS activities and for scheduling vehicles during special events.

A surprise result was a desire by the EMS management to obtain more detailed information. It was felt that if more service districts were added the data would be more reliable. Consequently, the number of districts were doubled with a corresponding decrease in the size of each district. At present, there is not enough history to calibrate the DSS and to make further evaluations.

8. Conclusions

In this paper, the TIMEXCLP model was developed and analyzed. The model is part of a decision support system to aid EMS planners to evaluate alternative ambulance locations. Two general statements can be made about the TIMEXCLP EMS vehicle location model.

1. The TIMEXCLP model results in a smaller model error when compared to the other set covering approaches. When compared to MEXCLP, TIMEXCLP resulted in an average decrease of model error of 79%.

2. The TIMEXCLP model yields ambulance locations that result in improved performance over existing approaches. In the Louisville application, this resulted in an annual savings of $250 000.

Regardless of how well a new approach compares to existing models, it is interesting to compare the approach in application. The TIMEXCLP model was developed by analyzing the Louisville, Kentucky, EMS system. Presently, the Louisville EMS system operates seven ambulances. With their present operations, the system responded to 84% of requests for an ambulance within ten minutes. The average response time was 7.45 minutes. By utilizing the TIMEXCLP vehicle locations, the system can obtain 95% coverage and a reduction of response time to 5.48 minutes. This corresponds to a 13% increase in coverage and a 36% decrease in response time with no increase in the number of ambulances nor operating personnel.

References

Batta, R., Dolan, J.M., and Krishnamurthy, N.N. (1989), "The maximal expected covering location problem: Revisited", *Transportion Science* 23, 277–287.

Benveniste, R. (1985) "Solving the combined zoning and location problem for several emergency units", *Journal of the Operational Research Society* 36, 433–450.

Berlin, G.R., and Liebman, J.C. (1974), "Mathematical analysis of emergency ambulance location", *Socioeconomic Planning Sciences* 8/6, 323–328.

Berman, O., and Larson, R. (1982), "The median problem with congestion", *Computers & Operations Research* 9, 119–126.

Berman, O., Larson, R. and Parkan, C. (1987), "The stochastic queue *p*-median problem", *Transportation Science* 21, 207–216.

Bernardo, J., and Repede, J. (1988), "Two-tier emergency medical service system: An analysis and extension of current location methodologies", Presented at the 1988 Annual Meeting of the Decision Sciences Institute, Las Vegas, NV.

Birge, J., and Pollack, S. (1989), "Using parallel iteration for approximate analysis of a multiple server queuing system", *Operations Research*, 37/5, 769–779.

Charnes, A., and Storbeck, J. (1980), "A goal programming model for the siting of multilevel EMS systems", *Socio-Economic Planning Science* 14/4, 155–161.

Church, R., and ReVelle, C. (1974), "The maximal covering location problem", *Papers of the Regional Science Association* 32, 101–118.

Cretin, S., and Willemain, T.R. (1979), "A model of prehospital death from ventricular fibrillation following myocardial infarction", *Health Services Research* 14/3, 221–234.

D'Agostino, R.B., and Pozen, M.W. (1981), "Distance in rural areas as a determinant of paramedic effectiveness", Presented at the 38th Annual National Meeting of the American Federation for Clinical Research, San Francisco, CA.

Daskin, M.S., and Stern, E.H. (1981), "A hierarchical objective set covering model for emergency medical service vehicle deployment", *Transportation Science* 15/2, 137–152.

Daskin, M.S., (1982), "Application of an expected covering model to emergency medical service design", *Decision Sciences* 13/3, 416–439.

Daskin, M.S. (1983), "A maximal expected set covering location model: Formulation, properties, and heuristic solution", *Transportation Science* 17, 48–69.

Eaton, D.J., Daskin, M.S., Simmons, D., Bulloch, B., and Jansma, G. (1985), "Determining emergency medical service vehicle deployment in Austin, Texas", *Interfaces* 15/1, 96–108.

Eisenberg, M., Berger, L., and Halstrom, A. (1979), "Paramedic programs and out-of-hospital cardiac arrest: Factors associated with successful prehospital resuscitation", *American Journal of Public Health* 69/1, 30–38.

Fitzsimmons, J. (1973), "A methodology for emergency ambulance deployment", *Management Science* 19/6, 627–636.

Goldberg, J., Dietrich, R., Chen, J.M., and Mitwasi, M.G. (1990a), "Validating and applying a model for locating emergency medical vehicles in Tucson, AZ", *European Journal of Operational Research* 49, 308–324.

Goldberg, J., Dietrich, R., Chen, J.M., and Mitwasi, M.G. (1990b), "A simulation model for evaluating a set of emergency vehicle base locations: Development, validation, and usage", *Socio-Economic Planning Sciences* 24/2, 125–141.

Goldberg, J., and Szidarovsky, F. (1991), "Methods for solving nonlinear equations used in evaluating emergency vehicle busy probabilities", *Operations Research* 39/6, 903–916.

Goldberg, J., and Paz, L. (1991), "Locating emergency vehicle bases when service time depends on call location", *Transportation science* 25/4, 264–280.

Hakimi, S.L. (1965), "Optimum distribution of switching centers in a communication network and some related graph theoretic problems", *Operations Research* 13/3, 462–474.

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

Larson, R.C. (1974), "A hypercube queuing model for facility location and redistricting in urban emergency services", *Computers & Operations Research* 1, 67–95.

Larson, R.C. (1975), "Approximating the performance of urban emergency service systems", *Operations Research* 23/5, 845–868.

Liu, M.S., and Lee, J.T. (1988), "A simulation of a hospital emergency call system using slam", *Simulation* 51/6, 216–221.

Mayer, J.D. (1979), "Emergency medical service delays, response time, and survival", *Medical Care* 17/8, 818–827.

Pascatelli, E.F., and Katz, I.B. (1978), "Planning and developing a prehospital mobile intensive care system in an urban setting", *American Journal of Public Health* 68/4, 389–393.

Reines, H.D. (1984), "The effect of transit time on mortality from major trauma", presented at the 44th Annual Session of the American Association For the Surgery of Trauma, New Orleans, LA.

Saydam, C., and McKnew, M. (1985), "A separable programming approach to expected coverage – an application to ambulance location", *Decision Sciences* 16/4, 381–398.

Savas, E. (1969), "Simulation and cost-effectiveness analysis of New York's emergency ambulance service", *Management Science* 15/12, 608–627.

Siler, K.F. (1977), "Level-load retrieval time: A new criterion for EMS facility sites", *Health Services Research* 12/4, 1055–1060.

Swoveland, C., Uyeno, D., Vertinsky, I., and Vickson, R. (1973), "Ambulance location: A probabilistic enumeration approach", *Management Science* 20/4, 686–698.

Torregas, C., Swain, R., ReVelle, C., and Bergman, L. (1971), "The location of emergency service facilities", *Operations Research* 19/6, 1363–1373.

Uyeno, D., and Seeberg, C. (1984), "A practical methodology for ambulance location", *Simulation* 43/2, 79–87.

Volz, R. (1971), "Optimal ambulance location in semi-rural areas", *Transportation Science* 5/2, 193–203.