

**A Non-Linear Multi-Criteria Programming Approach for Determining
County Emergency Medical Service Ambulance Allocations**



Joanna R. Baker; Edward R. Clayton; Bernard W. Taylor, III

The Journal of the Operational Research Society, Vol. 40, No. 5. (May, 1989), pp.
423-432.

Stable URL:

<http://links.jstor.org/sici?sici=0160-5682%28198905%2940%3A5%3C423%3AANMPAF%3E2.0.CO%3B2-4>

The Journal of the Operational Research Society is currently published by Operational Research Society.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ors.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

A Non-linear Multi-criteria Programming Approach for Determining County Emergency Medical Service Ambulance Allocations

JOANNA R. BAKER, EDWARD R. CLAYTON and BERNARD W. TAYLOR, III

Virginia Polytechnic Institute and State University, Blacksburg, USA

In this paper an integer, non-linear mathematical programming model is developed to allocate emergency medical service (EMS) ambulances to sectors within a county in order to meet a government-mandated response-time criterion. However, in addition to the response-time criterion, the model also reflects criteria for budget and work-load, and, since ambulance response is best described within the context of a queueing system, several of the model system constraints are based on queueing formulations adapted to a mathematical programming format. The model is developed and demonstrated within the context of an example of a county encompassing rural, urban and mixed sectors which exhibit different demand and geographic characteristics. The example model is solved using an integer, non-linear goal-programming technique. The solution results provide ambulance allocations to sectors within the county, the probability of an ambulance exceeding a prespecified response time, and the utilization factor for ambulances per sector.

Key words: ambulance allocation, emergency medical service, goals, multiple criteria, non-linear programming, response time

INTRODUCTION

In 1974 the Emergency Medical Service (EMS) Act was passed by Congress, which mandated that 95% of all calls for emergency medical service must be responded to within 10 minutes in an urban area (30 minutes in a rural area).¹ Since the passage of this bill, numerous analytical and heuristic models have been developed that seek to allocate and/or locate emergency medical service ambulance units within a region in an efficient manner such that these government-mandated response times might be achieved. The majority of these models have been developed using one of several solution approaches, including simulation, queueing and set-covering. However, these models all have limitations, not the least of which is their ability to reflect multiple criteria in the allocation decision.

In this paper a multiple-criteria, mathematical programming model for allocating ambulances to sectors within a county is proposed. The model employs several criteria for allocation, including the EMS response-time criterion mentioned above and budgetary considerations. Because of these multiple-decision criteria, and the fact that unitary items (ambulances) are being allocated, an integer goal-programming model formulation is employed. However, although the ambulance-allocation problem can be conveniently described within a mathematical programming framework (i.e. allocating scarce units in an efficient manner given specific objectives and constraints), the actual process of ambulances responding to calls is best described as a queueing system. As such, this model includes queueing-system formulations defined as goal constraints within the mathematical programming framework, which, in turn, necessitates a non-linear solution approach. This model is demonstrated in the paper within the context of a case example. The model-solution results are presented and discussed, and several sensitivity analysis scenarios are analysed.

THE AMBULANCE-ALLOCATION PROBLEM AND EXISTING SOLUTION APPROACHES

There are two primary issues generally associated with the emergency medical service (EMS) planning: allocation of ambulances and the allocation of EMS units within a region. The type of allocation criterion employed depends, in part, on the type of region, i.e. urban, rural or a mixture. In an urban area, which is characterized by high demand for service, allocation is normally addressed using a work-load criterion; i.e. allocate enough units to minimize the probability that a

call must wait. Work-load approaches have been analysed using simulation^{2,3} and queueing models.⁴⁻⁷ In rural areas, where demand is typically low and the area serviced is large, response time is frequently the most important criterion of performance effectiveness for planners. When response time is used as a criterion, the efficient location of ambulances is of primary importance (and hence the modelling objective) rather than work-load.

The traditional approach to optimally locating ambulances within a geographic region in order to minimize response times is set-covering location modelling. Set-covering models for ambulance allocations were derived as a special case of the facility-location problem.⁸⁻¹⁰ Such models generally employ 0-1 integer, linear programming models to solve problems of location with a single criterion of minimizing average response time (see, for example, Daberkow,¹⁰ Daberkow and King,¹² Daskin,^{13,14} Daskin and Stern¹⁵ and Toregas and Revelle¹⁶). Set-covering is not generally directed at the differences in population, and therefore differences in demand levels, with a region. For this reason most allocation and location models have been applied to urban regions or regions with relatively homogenous demand. Studies which have attempted to examine the problem for non-homogeneous areas include Jarves *et al.*,¹⁷ who used a 'degenerative version' of the set-covering problem to locate ambulances in a semi-rural area. Groom,¹⁸ in an effort to meet ambulance activation and response-time criteria, combined coverage and queueing concepts in the problem of locating ambulances in West Glamorgan. A more recent approach is the maximal expected set-covering model and the non-linear maximal expected set-covering model developed by Saydam and McKnew.¹⁹ These more recent approaches use separable programming techniques and consider demand coverage in locating ambulance units in areas of mixed demographics. However, this modelling approach, like the other traditional set-covering approaches, employs a single criterion, average response time, to locate ambulances, and does not explicitly reflect the distribution of response times for different demographic localities.

An additional difficulty with location models in general is that they indicate the optimal location of EMS units but fail to reflect the fact that the location decision is frequently politically motivated as well as being a practical decision. A more realistic approach, as employed in this paper, is to determine the number of ambulances required to provide effective service within a limited geographic area, given that a governmental planner has already located the EMS base units.

A model which considers both work-load and average response-time criteria for allocation and location of EMS units is the hypercube or spatially distributed queueing model (see Larson,⁷ Bodily²⁰ and Larson and Odoni²¹). In this type of model, primary response areas (PRAs) are defined within a region *a priori*, and each PRA is subdivided into small areas (or atoms), which generate demand in proportion to population (using census tracts, for example). A predetermined number of EMS units are allocated to an atom within a PRA. An M/M/N (multiple channel) queueing model is employed to model the resulting system. Performance characteristics, such as unit work-loads, intersector despatches and average travel times, are obtained for each server and each PRA and/or atom. Once performance characteristics are obtained, changes to the original fixed allocations must be made manually. Most applications of the hypercube model have been for police patrols, fire-unit allocations and ambulance allocations within urban areas^{7,21} (although recently Burwell *et al.*²² have adapted the hypercube model for use by non-urban EMS planners).

The multi-criteria mathematical programming approach described in this paper is similar, in principle, to the hypercube modelling approach in that we subdivide a region (i.e. a county) into PRA-like areas that are referred to as sectors. The hypothetical county that will be employed as an example is predominantly urban, but it contains distinctly rural and mixed areas. Each sector is uniquely characterized as urban, rural or mixed on the basis of arrival rates of calls, area and travel speed of ambulances, and different types and mixes of EMS personnel are allowed in each sector. As previously noted, the model does not address the issue of location within a sector (as would a spatially distributed model). It assumes the EMS units will be relatively centrally located within a sector. Also, unlike the hypercube model, which considers average travel time for a predetermined number of units within an area, the mathematical programming approach presented in this paper determines the best allocation of EMS units and the distribution of response times for that allocation over all sectors. However, in addition, the model considers budget, work-load and response-time criteria in the determination of the most efficient allocation of ambulances.

MODEL FORMULATION

The modelling approach is presented within the context of a case example based on a large metropolitan county in the southeastern part of the United States. The county encompasses a geographic area of approximately 830 square miles. The county is divided into 10 sectors, and ambulances of the emergency medical service are allocated within these individual sectors. Sectors are determined primarily according to geographic characteristics, including natural boundaries, such as highways, rivers, etc., and distinguishable population areas, such as a township, a suburb, etc. Other factors that might affect the determination of a sector include political considerations and such items as census tracts. (For example, each sector in the example county encompasses several census tracts.) In general, these sectors have a historical basis, and there is a strong impetus on the part of the county government to maintain the sector identity across all forms of governmental operations.

The 10 sectors within our example county include two distinctly urban sectors and five clearly rural sectors, with the remaining three sectors constituting a mix of rural and urban areas. In general, urban sectors are characterized by high average demand for ambulances, shorter travel distances and total service times, higher costs per call and slower travel speeds. Alternatively, a rural sector is characterized by lower demand, larger geographic area, lower costs per call and higher travel speeds. However, despite the urban and rural mix, the urban components are pervasive and, as such, the county is viewed as 'urban' for planning purposes.

The function of the proposed model is to allocate a constrained number of ambulances to the ten sectors in the county subject to several internal and external constraints and objectives. The ambulances will be located at a predetermined facility, from which service calls will be initiated. (These facilities are usually centrally located in the geographic region, or located according to population density; however, this model assumes in either case that the ambulance facility location already exists.) A significant external objective is dictated by the Emergency Medical Service Systems Act,¹ which states that 95% of urban calls for medical service should be responded to within 10 minutes (or 30 minutes for a rural call). Internal objectives include system-cost considerations and ambulance-utilization factors. Because of multiple system objectives and resource constraints, multiple-criteria (goal) programming is employed as the modelling technique.

Decision variables

The example county employed in the model consists of 20 ambulances, which are assigned to 10 sectors encompassed by the county. Each of the sectors has a unique demographic composition, as follows:

Sectors 1 and 2: urban
Sectors 3, 4, 5: mixed
Sectors 6, 7, 8, 9 and 10: rural

The deployment of ambulances to sectors is represented by the following set of decision variables:

x_i = number of ambulances assigned to sector ' i ' ($i = 1, 2, \dots, 10$);
 x_i = integer ≥ 0 .

Also, goal constraint deviational variables are defined as

d^+ = overachievement of a goal;
 d^- = underachievement of a goal, where $d^+, d^- \geq 0$.

Model constraints

The general goal-programming model typically has two types of constraints: system constraints and goal constraints. The former represent a set of existing restrictions that define the feasible solution space. The latter are a set of functions representing the performance goals or objectives of the problem, some of which may be unable to be met. The general goal-programming technique is described in a number of sources, e.g. Charnes and Cooper,²³ Ignizio,²⁴ Lee²⁵ and Zeleny.²⁶

System constraints

The maximum number of available ambulances to be allocated to the 10 sectors is 20. This restriction is reflected in the following system constraint:

$$\sum_{i=1}^{10} x_i + d_1^- = 20. \tag{1}$$

The second system constraint in the model ensures that the ambulance-utilization factor in each sector does not exceed one:

$$\frac{\lambda_i}{x_i \mu_i} = d_j^- = 1 \quad \begin{matrix} i = 1, 2, \dots, 10 \text{ sectors,} \\ j = 2, 3, \dots, 11, \end{matrix} \tag{2}$$

where

λ_i = Poisson arrival rate of calls,
 μ_i = general service rate for calls.

The system of calls for ambulances is a queueing system with Poisson arrivals, negative exponential service times and N servers (M/M/N system). An M/M/N model approximates the service system under study but, as noted in Larson and Odoni²¹ (p. 303), no actual service system will ever conform exactly to a given model's assumptions.

The average arrival rate expressed as calls per hour is shown in the third column of Table 1 for each of the 10 sectors in the county. Data for the estimates were derived from ambulance-run data for counties in South Carolina defined as urban, rural or mixed. The demographic type of each sector is shown in column 2 of Table 1.

TABLE 1. Arrival rates for sectors

1. Sector i	2. Type	3. λ_i (calls/hour)
1	urban	4.000
2	urban	2.100
3	mixed	1.340
4	mixed	1.340
5	mixed	1.340
6	rural	0.610
7	rural	0.610
8	rural	0.610
9	rural	0.117
10	rural	0.117

Goal constraints

Budget. The budgetary goal constraint is based on fixed and variable cost components as follows:

$$\sum_{i=1}^{10} (FC_i \cdot x_i) + (\lambda_i \cdot 24 \cdot VC_i) + d_{12}^- - d_{12}^+ = \$ \text{ Budget}. \tag{3}$$

The fixed cost component is based on an hourly rate for a paramedic of \$15 and an hourly rate of \$10 for an emergency medical technician (EMT), which must be paid, regardless of whether ambulance runs are made or not. The average fixed cost per call for the urban, mixed and rural sectors is \$116, \$110 and \$90, respectively.

The variable cost component is based on the average cost per call, which has been derived from a survey of emergency medical service units in South Carolina.²⁷ The costs differ for sectors as a result of the number of different types of calls for the different sectors. The variable cost per call is dependent upon the type of respondent (a paramedic or EMT) and the type of call (emergent or non-emergent). If an emergency call is received (i.e. a stroke or heart attack), then a paramedic is licensed to use advanced life-support measures and drugs on the scene. An EMT is more limited and generally transports the patient. As a result, calls requiring a paramedic are more expensive. An EMT call is assigned a cost of \$100, and a paramedic call is assigned a cost of \$150. (If at least one of the two crewmen is a paramedic, then the call is assumed to cost \$150.) The 10 sectors

differ in the ratio of paramedics to EMTs responding. For the urban sectors, the ratio of paramedics to EMT calls is 9 to 1; for mixed sector, the ratio is reduced to an average call differential of 3 to 1; and for rural sectors, the ratio is 1 to 3.

The total variable cost component is computed by multiplying the hourly arrival rate for call, λ_i , by 24 hours and the cost per call, VC_i . The average variable costs per call for urban, mixed and rural sectors are \$145, \$135 and \$110, respectively.

The functional form used to model the total cost of the allocation provides only one possible alternative. The model can accommodate any non-linear cost-function approximation a decision-maker might find appropriate. Not included in the variable cost component is the effect of mileage cost. Although not shown here, average travel times, and therefore average travel distance for a sector, could easily be introduced into the total cost function.

Response Time. As noted previously, the Emergency Medical Services Act of 1974 states that 95% of urban calls for medical service should be responded to within 10 minutes. Since the county being examined has a significant urban population, this standard of 10 minutes (as opposed to the 30-minute standard for rural counties) is employed. Two goal-constraint formulations are developed in association with this requirement. The first goal constraint attempts to meet the EMS Act requirement for the individual sectors within the county, while the second goal formulation is for the county as a whole.

It is not specifically required that each sector in the county meet the response-time criterion designated by the EMS Act; i.e. it is a requirement for the county as a whole, but not for individual sectors. However, we are considering the 10-minute response-time criterion as a goal for each individual sector for political and public reasons. It is possible to achieve the requirement that 95% of the calls in the county as a whole be responded to within 10 minutes, while individual sectors do not meet this criterion. For those individuals living in a specific sector, the response time in that sector is quite as important (or more so) than the county response time, and their political representation will, in all likelihood, reflect that concern. As such, it is viewed as desirable to achieve both the overall county objective (in order to satisfy a governmental mandate) and the individual sector objectives (in order to satisfy a public mandate).

The goal constraints for response time for each sector are formulated as follows:

$$P(W_{q_i} > t_i) = [1 - P(W_{q_i} = 0)]e^{-x_i\mu_i(1-\rho_i)t_i} + d_k^- - d_k^+ \quad (4)$$

$$= 0.05 \quad \text{for } x_i > 1$$

$$P(W_{q_i} > t_i) = \rho_i e^{-\mu_i(1-\rho_i)t_i} + d_k^- - d_k^+ \quad (5)$$

$$= 0.05 \quad \text{for } x_i = 1,$$

where

$$t_i = 10 - T_i;$$

T_i = average travel time;

μ_i = service rate for sector i , where the service time ($1/\mu_i$) is computed by summing service time without travel time (column 5 in Table 2) and travel time;

W_{q_i} = the time a call must wait in the queue before an ambulance is despatched in sector i , i.e. the dispatch time;

$$\rho_i = \frac{\lambda_i}{x_i \mu_i} = \text{utilization factor for sector } i;$$

$$k = 13, 44, \dots, 22;$$

$$i = 1, 2, \dots, 10.$$

T_i is computed using the following formula:

$$T_i = \left(\frac{c_i}{V_{oi}}\right) \cdot (A_{oi})^{1/2}, \quad i = 1, 2, \dots, 10. \quad (6)$$

This formula for expected travel time is derived by Larson and Odoni²¹ (p. 136), and it enables travel time to be related to the geographic characteristics of a sector and the speed of the vehicle.

TABLE 2. Sector data for EMS units

1. Sector <i>i</i>	2. Type	3. <i>c_i</i>	4. Description	5. Service time w/o travel time (mins)	6. <i>Ao_i</i> (mi ²)	7. <i>V_{oi}</i> (m.p.h.)
1	Urb	0.50	Centre/R <	40.71	30	20
2	Urb	0.50	Centre/R <	40.71	25	20
3	Mix	0.38	Centre/Euc	52.59	75	30
4	Mix	0.38	Centre/Euc	52.59	75	30
5	Mix	0.50	Centre/R <	52.59	60	25
6	Rur	0.38	Centre/Euc	71.20	100	30
7	Rur	0.38	Centre/Euc	71.20	100	30
8	Rur	0.38	Centre/Euc	71.20	110	30
9	Rur	0.38	Centre/Euc	71.20	130	30
10	Rur	0.38	Centre/Euc	71.20	125	30

Note that expected travel time, T_i , is independent of the number of ambulances operating in a sector. It is assumed that bases are located centrally and are fixed. In general, ambulances respond to calls directly from the base facility (i.e. they do not patrol as would police cars, for example), although in rare instances a call will come in while an ambulance is on another call and the unit will go directly from one service location to another. Thus, T_i depends only on the geographic characteristics of a sector reflected by c_i , V_{oi} and Ao_i .

' c_i ' in the above formula is a proportionality constant which corrects for the metric used to determine the travel distance and the shape of the sector. The value of c_i is prescribed based on a description of the sector presented in column 4 of Table 2. 'Centre' implies that all EMS units are located in the centre of the sector, an assumption appropriate for most PRAs. 'R <' corrects for a right-angle travel metric, which is most appropriate for an urban sector or a township. 'Euc', or Euclidean measure, is appropriate for rural sectors, where the prevailing routes are secondary roads or highways. These area descriptions identify a constant value, c_i (shown in column 3 of Table 2), taken from Larson and Odoni²¹ (p. 135).

Ao_i is the area of the sector in square miles, and these values for each sector are indicated in column 6 of Table 2.

V_{oi} is the average velocity at which an ambulance travels. Urban areas have slower travel times owing to congestion, traffic lights, etc. The values for V_{oi} are shown in column 7 of Table 2.

The goal-constraint functions (4) and (5) representing a single goal criterion are based on the standard multiple-channel queueing model²⁸ (pp. 417-424). Response time encompasses two time-components—the time required to dispatch an ambulance and travel time. W_{qi} is the dispatch time, which can also be expressed as the time a call must wait before an ambulance is dispatched. By minimizing d_k^+ in the applicable equation, (4) or (5), these functions attempt to limit the probability that a call for an ambulance must wait longer than the difference between time and 10 minutes (before being dispatched) to 0.05. These queueing functions are non-linear.

Goal-constraint equations (4) and (5) attempt to achieve the response-time criterion of 10 minutes for each sector. The response-time goal formulation for the county which meets the EMS Act criterion is expressed as follows:

$$P(W_q > t) = [1 - P(W_q = 0)]e^{-\sum_{i=1}^{10} x_i \mu (1 - \rho)^t} + d_{23}^- - d_{23}^+ = 0.05, \tag{7}$$

where

$$t = 10 - T;$$

T = average travel time for the county;

μ = average service rate for the county;

W_q = the time a call must wait in the queue before an ambulance is dispatched in the county;

$$\rho = \frac{\sum_{i=1}^{10} \lambda_i}{\sum_{i=1}^{10} (x_i)(\mu)} = \text{average utilization rate for the county};$$

$$\sum_{i=1}^{10} \lambda_i = \text{average arrival rate for the county}.$$

Notice that goal equation (7) is basically the same multiple-channel queuing formulation as shown in equation (4) and taken from Hillier and Lieberman.¹⁸ Likewise, its explanation is similar. The primary differences are that ρ is for the county rather than for a sector, μ is the average service rate for the county, and the arrival rate is for the county. The average service rate, μ , was computed by summing the individual sector service rates for all ambulances, and dividing by the total number of ambulances in the county. In order to achieve the objective of responding to calls within 10 minutes with a probability of 0.95, d_{23}^+ is minimized.

Ambulance Utilization. As an internal county objective it is desired that a utilization factor for each sector of at least 0.40 be achieved. This goal formulation is expressed as

$$\frac{\lambda_i}{(x_i)(\mu_i)} + d_i^- - d_i^+ = 0.40, \quad l = 24, 25, \dots, 33, \quad (8)$$

where

$$\begin{aligned} \lambda_i &= \text{average arrival rate for each sector in the county;} \\ \mu_i &= \text{average service rate for each sector in the county.} \end{aligned}$$

This goal expresses the desire on the part of the county administrators that the sector ambulance services be utilized at least 40% of the time. This goal is achieved by minimizing the negative deviational values, d_i^- .

Objective function

The initial priority designation for the objective function will be for a standard case for the county case example which has been presented in this section. In this standard case, the top two priority goals, P_0 , will be for the two system constraints for available ambulances (1) and sector utilization (2). The first-priority regular goal, P_1 , is to achieve the response-time criteria set by the EMS Act for the county and expressed in equation (7). The second-priority goal is not to exceed the budget goal (3); the third priority is to achieve the individual sector response-time probability criteria (4) and (5); and the last priority is to achieve a utilization factor of at least 0.40, as expressed in equation (8):

$$\min \left\{ P_0 \left(d_1^+ + \sum_{j=2}^{11} d_j^+ \right), P_1 d_{23}^+, P_2 d_{12}^+, P_3 \sum_{k=13}^{22} d_k^+, P_4 \sum_{l=24}^{33} d_l^- \right\}. \quad (9)$$

The model solution for this priority designation will be given in the following section. In addition, an alternative priority designation and the corresponding solution will also be considered.

MODEL SOLUTION AND RESULTS

Since this formulation encompasses several functions that are non-linear, and also requires integer solution values, an integer, non-linear solution approach is employed which is referred to as INVOP. INVOP is a computerized algorithm that combines modified pattern search and gradient search techniques to locate a 'good' but not necessarily optimal solution.^{29,30} The program will run efficiently on any FORTRAN compiler. The user must develop a FORTRAN subroutine containing the model equations with deviations, and create a data input file that contains the problem characteristics (i.e. number of variables, priority level of goals, etc.).

The initial solution is shown in Table 3. Column 2 shows the manner in which the 20 available ambulances were allocated to the 10 sectors in the county. Column 3 indicates the probability that the response time will exceed 10 minutes (i.e. the probability that a call for service will have to wait longer than the difference between travel time and 10 minutes). It can be observed that five sectors were 0.05 or less, thus indicating that all other sector response times violated the criteria that there be a 0.95 probability that the response times be less than 10 minutes. The overall county probability that the response time is greater than 10 minutes is 0.09, thus indicating that the first-priority goal was not achieved (i.e. $d_{23}^+ = 0.04$). The actual budget required for this allocation scheme is \$46,804; thus the second goal was also not achieved (i.e. $d_{12}^+ = \$16,804$). The overall average travel time for the county is 7.85 minutes. The third goal was for the response-time

TABLE 3. Model solution results

1. Sector i	2. Ambulances x_i	3. $P(W_{qi} > t_i)$	4. T_i (mins)	5. $\lambda_i/x_i\mu_i$
1	4	0.157	8.22	0.836
2	2	0.291	7.50	0.844
3	2	0.050	6.58	0.661
4	2	0.050	6.58	0.661
5	2	0.370	9.30	0.691
6	2	0.026	7.60	0.401
7	2	0.026	7.60	0.401
8	2	0.037	7.97	0.402
9	1	0.086	8.67	0.156
10	1	0.086	8.50	0.156

probability in each sector, and it has already been noted in our discussion related to column 3 in Table 3 that this goal was not achieved. The final (priority 4) goal to achieve a 0.40 utilization rate for each sector in the county was not achieved. Column 5 in Table 3 indicates that a 0.40 utilization rate or greater was achieved in eight of the 10 sectors.

In order to increase confidence in the quality of the solution (since the INVOP program does not guarantee optimality), the model was run with a number of different starting-points. This started the search of the solution space from different directions at different locations. The heuristic program starts at one point, and moves until it finds a 'peak' which it cannot improve upon by moving in any direction. The danger is that there may be another peak some distance away that any stepwise move could not find. By starting at various points in the solution space, the possibility of not discovering these alternative peaks was reduced. Different solutions were occasionally obtained with different starting-points; however, none of these different solutions was better than the solution shown in Table 3.

The solution encompassed in Table 3 was obtained in approximately 3 seconds on an IBM 3090, so rapidly that computing time is not a deterrent to running the model interactively. Solutions to larger test problems encompassing over 100 variables and 10 goal constraints were obtained in less than 1 minute.

Alternative case 1

Since the initial example solution attained only minimal goal achievement and, in particular, did not achieve the EMS Act mandated response-time criterion for the county (the first-priority goal), the example parameters were altered in an attempt to determine a more satisfactory solution. (This type of model experimentation also replicates the process which planners would employ to locate a satisfactory allocation.) Using the initial solution as a guideline, the budget level was increased to \$49,000 and the number of available ambulances increased to 26; i.e. the right-hand-side value of equation (1) was changed to 26, and the r.h.s. value of equation (3) was changed to \$49,000. The solution results to this altered model are shown in Table 4.

The allocation in Table 4 results in the achievement of the first-priority goal for overall county response time ($p = 0.031$) and an overall budget allocation of \$48,484 ($d_{12}^- = \516). The third-priority goal for response time on the individual sectors was not achieved, as can be seen from column 3 in Table 4; however, only sectors 1 and 5 failed to meet the goal. The final goal for

TABLE 4. Model solution results: alternative case 1

1. Sector i	2. Ambulances x_i	3. $P(W_{q>} t_i)$	4. $\lambda_i/x_i\mu_i$
1	4	0.157	0.836
2	3	0.005	0.562
3	2	0.050	0.661
4	2	0.050	0.661
5	2	0.370	0.691
6	3	0.001	0.267
7	2	0.025	0.401
8	3	0.002	0.268
9	2	0.002	0.078
10	3	0.000	0.052

sector ambulance utilization was not totally achieved as sectors 6, 8, 9 and 10 had utilization rates below the goal level of 0.40.

This alternative case example solution is viewed as a more effective solution in that it substantially achieves the most important goals in the model; that is, both the response-time goal and the budgetary goal were met. Although additional cases will not be analysed, the model can be employed to experiment with alternative priority structures as well as differing parameter values.

SUMMARY

This paper described and demonstrated an integer, non-linear mathematical programming approach to allocating ambulances to sectors within a county, given the existence of multiple objectives and regions which exhibit mixed demand. The model is based on a multiple-channel queueing system. Although it is not possible to determine a guaranteed optimal solution, it was demonstrated that a satisfactory solution can be achieved. It was also demonstrated via a case example how the model can be employed in an experimental scenario to develop an effective allocation plan. Although not demonstrated, the model can also be used as a vehicle to experiment with alternative priority structures, as well as parameter levels. The model and INVOP solution technique are robust and flexible enough to consider a number of alternative system configurations.

The arrival and service rates for the model example are based on actual data for a county in South Carolina. The queueing formulas are standard, and the assumptions regarding the distributions (for arrival and service) are consistent with the previous research in this area and actual studies of ambulance allocation. Additional testing of the arrival and service rates did not indicate that the model was particularly sensitive to changes in these parameters; however, significant variations will change the sector allocations.

The size of the model (i.e. 10 sectors) is accurate for many counties in the southeast United States with the characteristics described in this paper (i.e. rural with some population density in cities). We attempted to model a generic type of county rather than an urban area. However, the model can be easily expanded to include more sectors, ambulances, etc. As already noted in our discussion of the INVOP solution approach, a model encompassing 104 variables and 10 goal constraints was solved in a test case.

A primary assumption in the development of the modelling approach was that the consideration of multiple criteria more closely reflects the actual decision-making environment for ambulance-allocation decisions than a single-criterion model. However, even if a single-criterion model is desired, the multiple-criteria model derivation is a more appropriate starting point since it will assist the decision-maker in determining what the best single-criterion approach is.

Goal constraints (4), (5) and (7) reflect equality of service to the community, while equation (8) reflects equality among servers (i.e. work-load). In the United States, equality of service is primarily addressed via the EMS Act, which requires that a minimal level of service be provided. Because discrete variables (i.e. ambulances) are the allocation measure, exact balancing of utilization is unrealistic. Rather, it was attempted to ensure, via goal constraints, a minimum utilization level.

The nature of the ambulance-allocation problem is such that it must be presumed that sectors exist *a priori*. These designs, in practice, are politically and geographically mandated. As such, it is easier to allocate within an existing design configuration than it is to redesign sectors subject to a particular allocation.

REFERENCES

1. Emergency Medical Service Systems Act. Federal Register, 29 March 1974, 39 (2), 11758–11766.
2. D. W. ASHLEY, D. G. GRAHAM and Q. B. WELCH (1980) A computer simulation study of emergency medical services in Kansas City, Missouri. National TIMS/ORSA Meeting, May 1980, Washington, DC
3. E. S. SAVAS (1969) Simulation and cost-effectiveness analysis of New York's emergency ambulance system. *Mgmt Sci.* **15**, 8608–8627.
4. J. A. FITZSIMMONS (1973) A methodology for emergency ambulance deployment. *Mgmt Sci.* **19**, 627–636.
5. J. A. FITZSIMMONS and R. S. SULLIVAN (1979) Establishing the level of service for public emergency ambulance systems. *Socio-Econ. Plan. Sci.* **13**, 235–239.

- 6 W. K. HALL (1971) The application of multi-function stochastic service systems in allocating ambulances to urban areas. *Opns Res.* **20**, 558–570.
- 7 R. C. LARSON (1974) A hypercube queueing model for facility location and redistricting in urban emergency services. *Comput. Opns Res.* **1**, 67–95.
- 8 L. J. SHUMAN, L. HARDWICK and G. A. HUBER (1973) Location of ambulatory care centers in a metropolitan area. *Hlth Ser. Res.* **8**, 121–138.
- 9 C. SWOVELAND, D. UYENO, I. VERTINSKY and R. VICKSON (1973) Ambulance location: probabilistic enumeration approach. *Mgmt Sci.* **20**, 686–698.
- 10 C. REVELLE, D. BIGMAN, D. SCHILLING, J. CONON and R. CHURCH (1977) Facility location: a review of context free and EMS models. *Hlth Ser. Res.* **12**, 129–146.
- 11 S. G. DABERKOW (1977) Location and cost of ambulances serving a rural area. *Hlth Ser. Res.* **12**, 299–311.
- 12 S. G. DABERKOW and G. A. KING (1977) Response time and the location of emergency medical facilities in rural areas: a case study. *Am. J. Agricultural Econ.* **59**, 467–477.
- 13 M. R. DASKIN (1982) Application of an expected covering model to emergency medical service system design. *Decis. Sci.* **13**, 416–439.
- 14 M. R. DASKIN (1983) A maximal expected covering location model: formulation, properties, and heuristic solution. *Transport. Sci.* **17**, 48–69.
- 15 M. R. DASKIN, and E. STERN (1981) A hierarchical objective set covering model for emergency medical service vehicle deployment. *Transport. Sci.* **15**, 137–152.
- 16 C. TOREGAS and C. REVELLE (1971) Location under time and distance constraints. *Papers of the Regional Science Association* **28**, 133.
- 17 J. P. JARVIS, K. A. STEVENSON and T. R. WILLEMAIN (1975) A simple procedure for the allocation of ambulances in a semi-rural area. Technical Report TR-13-75. Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts.
- 18 K. N. GROOM (1977) Planning emergency ambulance services. *Opl Res. Q.* **28**, 641–651.
- 19 C. SAYDAM and M. MCKNEW (1985) A separable programming approach to expected coverage: an application to ambulance location. *Decis. Sci.* **16**, 381–398.
- 20 S. E. BODILY (1978) Police sector design incorporating preferences of interest groups for equality and efficiency. *Mgmt Sci.* **24**, 1301–1313.
- 21 R. C. LARSON and A. R. ODoni (1981) *Urban Operations Research*. Prentice-Hall, Englewood Cliffs, NJ.
- 22 T. H. BURWELL, J. P. JARVIS and M. A. MCKNEW (1985) Ambulance location using a spatially distributed queueing model. In *Proceedings of the American Institute for Decision Sciences*, Las Vegas, November 1985 (B. HARTMAN and J. RINGUEST, Eds), pp. 705–707.
- 23 A. CHARNES and W. W. COOPER (1961) *Management Models and Industrial Applications of Linear Programming*, Vols I and II. Wiley, New York.
- 24 J. P. IGNIZIO (1976) *Goal Programming and Extensions*. Heath, Lexington, Mass.
- 25 S. M. LEE (1972) *Goal Programming for Decision Analysis*. Auerbach, Philadelphia.
- 26 M. ZELENY (1982) *Multiple Criteria Decision Making*. McGraw-Hill, New York.
- 27 J. R. BAKER, M. A. MCKNEW, T. R. GULLEDGE, JR. and J. RINGUEST (1984). An application of multiattribute utility theory to the planning of Emergency Medical Services. *Socio-Econ. Plan. Sci.* **18**, 273–280.
- 28 F. S. HILLIER and G. J. LIEBERMAN (1980) *Introduction to Operations Research*, 3rd Edition. Holden-Day, San Francisco.
- 29 E. R. CLAYTON, W. WEBER and B. W. TAYLOR (1982) A goal programming approach to the optimization of multi-response simulation models. *IIE Trans* **14**, 282–287.
- 30 B. W. TAYLOR, L. J. MOORE, E. R. CLAYTON, K. R. DAVIS and T. R. RAKES (1985) An integer non-linear goal programming model for the deployment of state highway patrol units. *Mgmt Sci.* **31**, 1335–1347.