MANAGEMENT SCIENCE
Vol. 15, No. 12, August, 1969
Printed in U.S.A.

SIMULATION AND COST-EFFECTIVENESS ANALYSIS OF NEW YORK'S EMERGENCY AMBULANCE SERVICE*

E. S. SAVAS†

Deputy City Administrator
Office of the Mayor, City of New York

Computer simulation was used to analyze the possible improvements in ambulance service that would result from proposed changes in the number and location of ambulances. The cost-effectiveness of several alternatives was examined. A particular alternative was shown to be of considerable value and it was concluded that low-cost improvements in service could indeed be achieved by redistributing ambulances in accordance with this alternative.

This marks the first time that the City of New York has utilized computer simulation as an aid to decision-making. In addition, this represents another step in the move to use computers more creatively in municipal management.

More generally speaking, the notion of applying the "space-age methods" of systems analysis, operations research, cost-effectiveness analysis, etc., to solve urban problems is a very popular one, frequently written about, discussed, and presented at conferences. The work reported here has translated this concept into practical results in a vital area of public service. This is a small but significant advance, of potential value to urban governments everywhere.

I. Introduction

Emergency ambulance service has been provided by the City of New York to residents and visitors since 1870. With the growth of the City, the service is now available to approximately ten million persons daily, 24 hours a day, every day of the year. In 1967, the City's ambulance service responded to more than half a million calls for emergency assistance, an increase of more than 43 percent in the past decade. This growing work load, together with an increasing concern about the adequacy and responsiveness of the system, led to a request by Mayor John V. Lindsay to Dr. T. W. Costello, Deputy Mayor-City Administrator, to analyze the service and to recommend and implement significant improvements.

Emergency ambulance service is provided by 109 ambulances, stationed at a total of 49 hospitals. Seventeen of these hospitals, with 60 of the ambulances, are municipal hospitals. The remaining 49 ambulances are supplied by, and stationed at, 32 voluntary hospitals, which provide emergency ambulance service under contract to the Department of Hospitals. Ambulance deployment is summarized in Table 1. Ambulances at a given hospital are generally assigned to cover a particular geographic district in the vicinity of that hospital.

The vast majority of requests for emergency ambulance service are made by the public via telephone calls to the Police Department Communications Division. Upon receipt of a call, the patrolman on duty determines the type of call and instructs the nearest hospital which provides emergency ambulance service to dispatch an ambulance.

* Received June 1968.

† The author wishes to thank the following members of his former staff at the New York Scientific Center of IBM for their crucial help in developing the simulation model: G. Gordon, K. Zelin, and Dr. P. C. Comba. This work would not have been possible without the assistance of Mrs. L. L. Burkholz and T. F. Loughlin of this Office, the initiative of M. L. Spiegel, First Deputy City Administrator and the cooperation of the Department of Hospitals under Commissioner J. Terenzio.

TABLE 1

Deployment of Ambulances

Type of Hospital	No. of Hospitals	Number of Ambulances on Duty		
		Midnight to 8 a.m.	8 a.m. to 4 p.m.	4 p.m. to Midnight
Municipal	17	43	56	60
Voluntary	32	45	49	49
Total	49	88	105	109

Ambulances are equipped with two-way radios, which enable them to communicate with the Communications Division. This communication link is utilized to redirect an ambulance from a low-priority sick call to a high-priority rush call, or to reassign an ambulance which is returning empty from a call, for example, when a patient does not require additional medical attention.

II. Scope of the Study

A. *Emergency Medical Care—An Overview*

Viewed in perspective, the emergency ambulance service fits within a more general framework of an overall emergency medical-care system. Such a system is composed of the following subsystems, with the first two comprising what is usually considered the ambulance system:

1. Communication
2. Transportation
3. Medical treatment

In addition, one can consider that preventive health care also enters into the total picture, for it clearly affects the requirements for and the nature of an emergency medical care system. For instance, improved preventive health measures and their ready availability to the community e.g., through neighborhood Health Care Centers, can be expected to reduce the demand for emergency care.

Communication Subsystem. This subsystem includes the means by which aid is summoned for a patient and the procedure for screening, assessing, and establishing priorities for such calls. It also encompasses the requirements and means for communicating among dispatchers, ambulances, and hospitals, and possibly even for contacting the Doctors' Emergency Service, Poison Control Center, etc.

Transportation Subsystem. This subsystem includes the means for conveying a patient to the medical facility, or for transporting medical facilities (doctor, first-aid attendant, oxygen, resuscitation equipment, stomach pump, antidotes, etc.) to a patient. Elements within this subsystem include such factors as the boundaries of ambulance service districts, the locations of ambulances and hospitals, and the number of ambulances. Other elements might involve the use of sirens and express lanes, the design and construction of ambulances, the location of first-aid stations, devices for carrying people down stairs, etc.

Medical Treatment Subsystem. This area encompasses the nature, speed, and adequacy of emergency medical treatment, in terms of the qualifications of personnel, their prompt availability, the organization, procedures, and equipment in an emergency room, the equipment carried on an ambulance, the possible utility of first-aid stations, etc. Improvements in the transportation subsystem could be vitiated, for example, if no doctor were available immediately after the patient is carried into the hospital.

B. *Systems Analysis of the Emergency Ambulance Service*

In light of the urgent need to improve the ambulance service itself, no effort was made initially to examine the prevention or treatment subsystems. For the purposes of this report, the communication subsystem can likewise be ignored. The major effort was focussed instead on particular elements of the transportation subsystem. Specifically, a quantitative analysis was made of the geographic distribution of emergency calls in the most severe problem area of the city, and the number and placement of ambulances needed to service these calls effectively. The merits of a proposed satellite ambulance station were examined in detail.

Systems analysis of a problem involves four classical steps:

(1) defining a specific objective which is to be achieved by the solution;
(2) formulating alternative ways of reaching that objective;
(3) establishing explicit criteria for evaluating the alternatives;
(4) selecting the best alternative, in terms of the criteria.

Objective. The concept of "improved ambulance service" can be described quantitatively by two related performance measures:

(1) response time—the period between receipt of a call at the ambulance station and arrival of an ambulance at the scene;

(2) round-trip time—the period between receipt of a call at the ambulance station and arrival of the assigned ambulance at the hospital with the patient.

Both of these related parameters are important from the public service point of view. Prompt arrival of an ambulance and trained attendant on the scene saves lives, reduces suffering, and produces confidence in the service on the part of the general citizenry. Round trip time is the vital parameter in those cases where the patient requires prompt professional medical treatment in the emergency room of a hospital.

The objective that was adopted was to decrease the response time in the Kings County Hospital district of Brooklyn. (It follows from the above definitions that a decrease in response time produces the same reduction in round-trip time.) However, no numerical target (e.g., a five-minute reduction in response time) could justifiably be set unless it were possible to relate time savings to the saving of lives. No such study has been reported and to tackle this problem was well outside the initial scope of the project. This remains as an important topic for future medical research.

Alternatives. Three alternatives that were considered initially were the following:

(a) redistribute the existing ambulances in the district by locating some of them at a satellite garage;

(b) increase the number of ambulances at Kings County Hospital:

(c) a combination of the above two alternatives.

Criteria. Both the cost and the effectiveness of the alternatives were considered. Costs include the capital and operating costs of additional ambulances and of a satellite garage. Effectiveness was measured in minutes of average response time and also in the percentage of calls whose response time exceeded a certain level.

C. *The Problem*

Figure 1 portrays a typical district which is served by a hospital, indicated by H in the figure. Under the present mode of operation, the ambulances serving that district are all stationed at the hospital. The dots on the map indicate the location and relative numbers of emergency calls from different points throughout the district. Calls are not uniformly and randomly distributed throughout the area. Due to varying population density and socio-economic characteristics, certain subsections of the district exhibit rather dense clustering of dots; i.e., there is a high demand for ambulance service from those areas.

A superficial look at Figure 1 suggests that a substantial improvement in ambulance service could be achieved by the relatively simple expedient of stationing ambulances at a satellite garage in the middle of one of the clusters, for example, at point S. Such a garage could consist of ordinary commercial garage space, or the garage of a police station or fire house. Proponents of this idea reasoned that an ambulance located at the satellite could pick up a patient in that vicinity and deliver him to the hospital in half the time that it would take an ambulance from the hospital to go and pick up that patient and return with him to the hospital; they envisioned a 50% reduction in round-trip time. However, a closer look shows that the situation is not so simple. In the first place, not all the time that elapses is travel time; various delays contribute to the total round-trip time (see Figure 2) and these would not be reduced by locating the ambulances elsewhere. Secondly, the ambulances will be called upon to service calls from anywhere in the district, not only those in the immediate vicinity of their

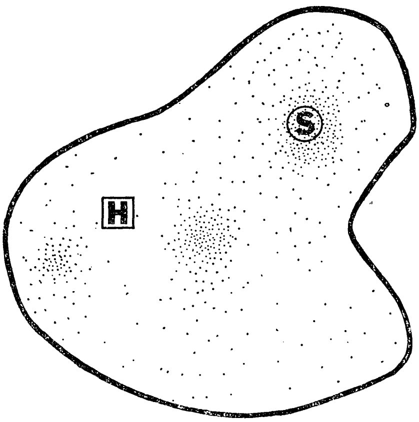

FIGURE 1. General representation of an ambulance district, showing location of hospital, H, and satellite garage, S, and showing points where calls were made.

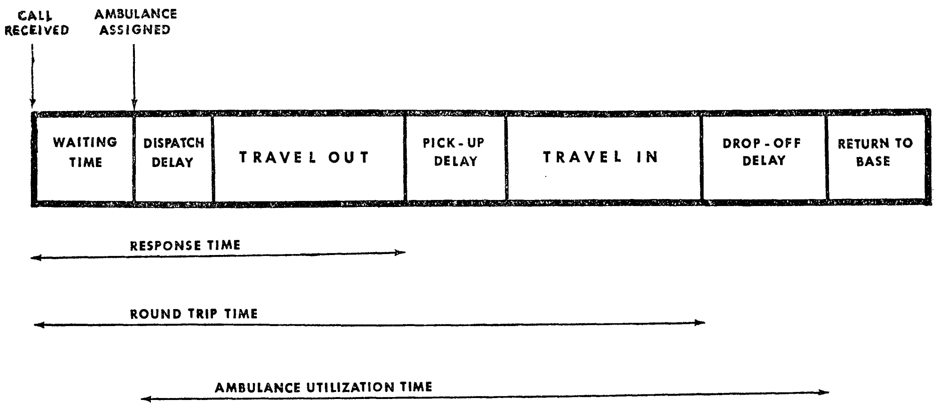

FIGURE 2. Sequence of events during a call, showing time relationships

satellite station, and it is difficult to forecast an improvement in handling those calls. Finally, the round-trip time is very sensitive to the frequency of calls; for example, infrequent calls from the area around the satellite can be assigned to waiting ambulances and in this case a substantial improvement would be realized. However, as the frequency rises, the ambulances would be spending more and more time shuttling back and forth between the hospital and the high-demand area around the satellite, calls would queue up to await an available ambulance, and in this case it would make no difference whether the busy ambulance were nominally stationed at the satellite, at the hospital, or at any point in between.

This qualitative analysis clearly shows that the picture is not so simple as appears at first glance, and that the level of service depends in a complex way on the following five major factors:

geographic distribution of calls throughout the district
frequency of calls
number of ambulances in the district
location of hospital
location of ambulance garage(s)

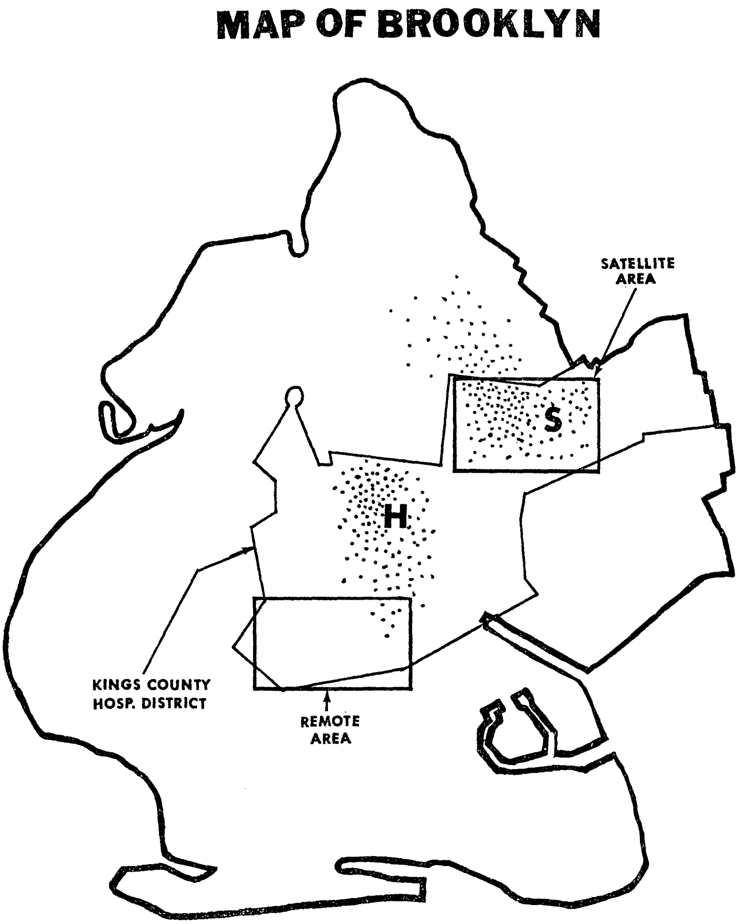

FIGURE 3. Map of Brooklyn showing hospital district and areas near hospital and satellite

Given the complexity of this system, no intuitive estimate can provide a sound guide. Nevertheless, the basic idea of a satellite station, that is, to put the ambulances where they are needed, is a sound one that warrants a detailed, quantitative analysis in order to provide valid estimates of the improvements to be expected.

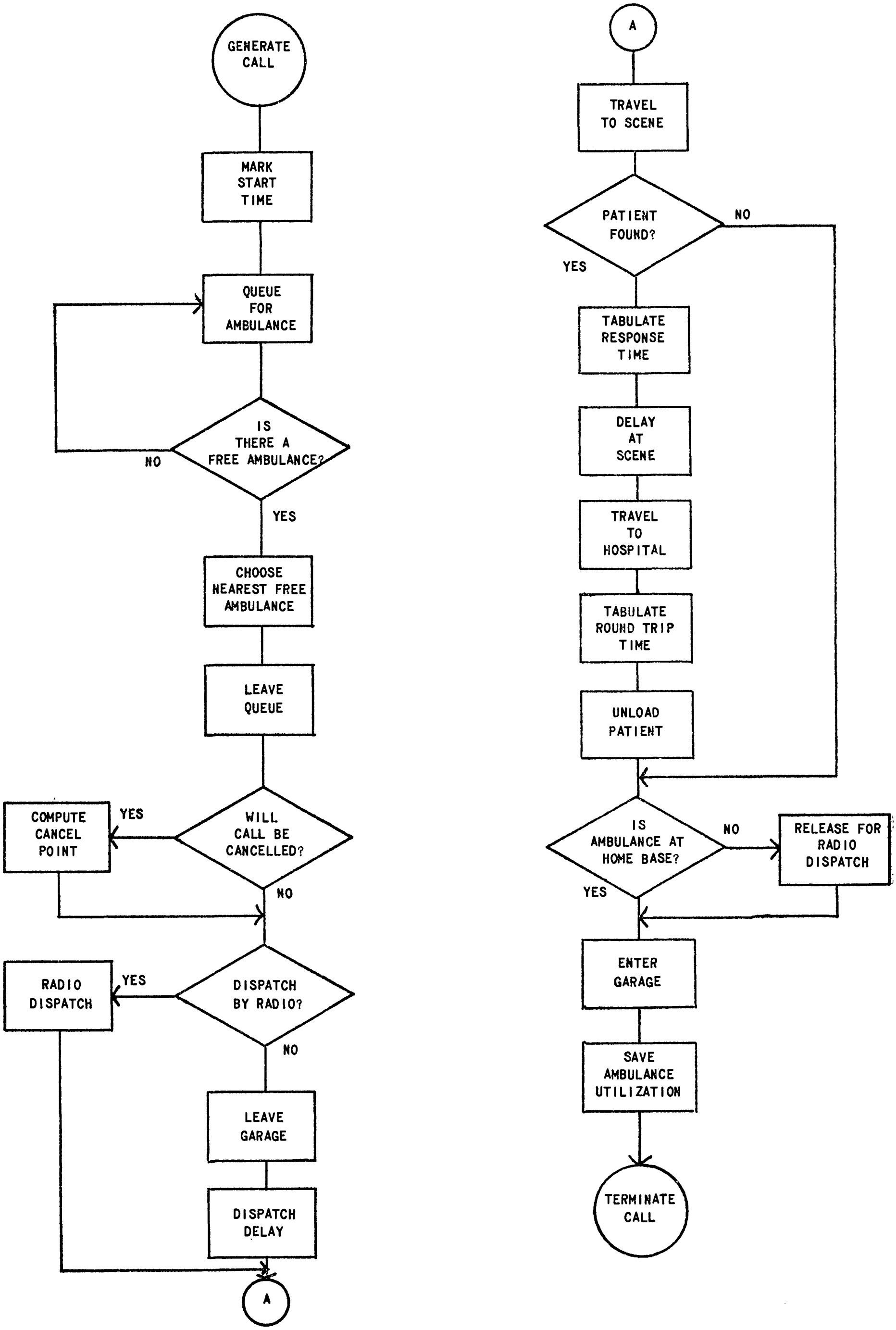

FIGURE 4. Generalized flow diagram of ambulance service model

D. *The Approach*

In light of the fact that the level of service is a complicated function of five variables, conventional computational approaches and simple mathematics will not suffice. Instead, the ideal analytical tool to use in this case is computer simulation. This is the method of choice where there are many interrelated factors, where the expected effects are complex, and where trial-and-error experimentation is costly or impractical. This definition describes the ambulance system perfectly.

The ambulance service system in the Kings County Hospital district was simulated on a digital computer using a mathematical model of the system. A map of the district, as it was in August, 1966, appears in Figure 3; the hospital is located at H and the proposed satellite is at S. About 175,000 calls were simulated, corresponding to almost four years of operation of that hospital's ambulance service. Attention was focussed on the peak-load period, the 4 p.m.-to-midnight shift. The inter-arrival time was set at 7.28 min., which characterises the peak-load period in an average month of 4570 calls. (This number of calls is 15% greater than the actual observed monthly load, to allow for the predicted future load.) General and technical details concerning the simulation appear elsewhere [1]. A general flow diagram of the model appears in Figure 4.

III. Results of the Simulation

A. *Effect of a Satellite*

The number of ambulances serving the Kings County Hospital district was retained at seven and the effect of a satellite station at the location indicated in Figure 3 was simulated. Figure 5 shows what happens to the average round-trip time and average response time in the district as the seven ambulances are redistributed between the hospital and that satellite in various proportions.

The first thing to notice is that the times decrease continuously as the ambulances are removed from the hospital, one by one, and placed at the satellite garage. In fact, if there are seven ambulances available to service the Kings County district, the optimum way to use them is to have all seven located at the satellite and none at the hospital. In other words, the satellite is at a better location for the hospital than is the hospital itself, at least in terms of ambulance service. This finding should not be interpreted as an argument for moving the hospital. A constructive conclusion is that redrawing of hospital district lines as well as redeployment of ambulances may be in order.

The second conclusion to be drawn from Figure 5 is a disappointing one: the average round-trip time is reduced a mere 5%, from 33 to 31.5 minutes, which is far less than the 50% improvement which seemed so obvious at first glance. (This reduction of 1.5 minutes applies to the average response time as well, and constitutes an 11% improvement over the existing time of 13.5 minutes.) This negative finding was not unexpected, in light of the discussion in §IIC above.

B. *Effect of Additional Ambulances*

In this case, the effect of placing additional ambulances at the hospital was studied. The results are evident in Fig. 6. Average response time drops by 0.3 minutes as the number of ambulances stationed at the hospital is increased from seven to ten, but thereafter virtually no improvement occurs no matter how many ambulances are added. Once one reaches the "elbow" of the curve, one is operating on a plateau and additional ambulances are wasted.

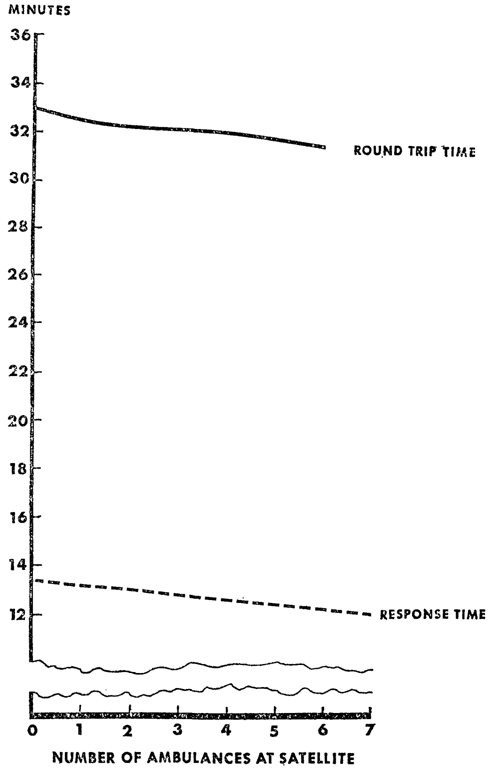

FIGURE 5. Effect of satellite on service, with a total of seven ambulances in the system

Waiting Time. The reason for this effect becomes clear upon inspecting Figure 7. The solid line shows the average waiting time as a function of the number of ambulances at the hospital. Waiting time is the period between receipt of a call at the ambulance station and assignment of an available ambulance to that call. (See Figure 2) Waiting time constitutes one identifiable segment of the response time. As more ambulances are added to the system, the waiting time drops essentially to zero and the response time therefore levels off (as in Figure 6) at a value which depends almost exclusively on the travel time. Travel time, in turn, is a fixed characteristic of a given district and depends upon its geometry (size and shape, and the location of its ambulances) and its traffic (routes and conditions).

Ambulance Utilization. The dashed line of Figure 7 shows how ambulance utilization declines as more ambulances are added. (Utilization is the fraction of time that an ambulance spends on a call; see Figure 2 for a graphic definition.) The increase in idle time (decreased utilization) is the price paid for reducing the average waiting time, that is, for assuring that an ambulance will be available for prompt assignment when a call comes in.

It should be noted that the minimum response time (at the "elbow" of Figure 6), which is achieved when the waiting time approaches zero (in Figure 7) corresponds to a utilization of 42%. This compares to the actual current utilization of about 60%.

This utilization factor is an important indicator of service and, because it is relatively easy to measure, it can be used to manage the ambulance system. For example, given the existing boundaries for Kings County Hospital (and no satellite), this analysis shows that if utilization is greater than 42%, improved service can be obtained by adding ambulances. On the other hand, if utilization is less than 42%, ambulances can safely be released from the district without fear that the level of service will be degraded. Furthermore, simple arithmetic suffices to calculate how many ambulances to add or remove in order to arrive at the 42% utilization figure.

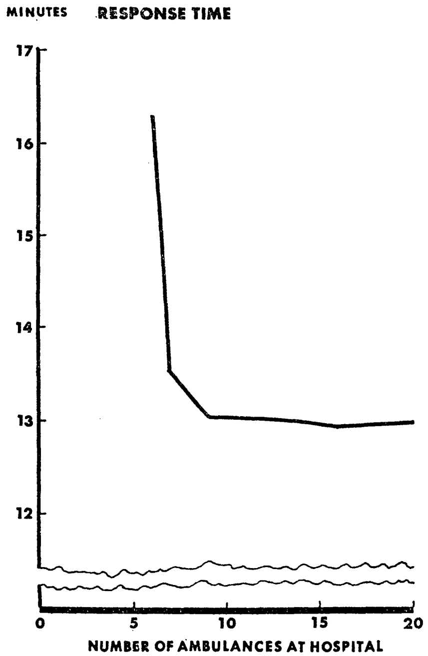

Figure 6. Effect of additional ambulances at the hospital on response time

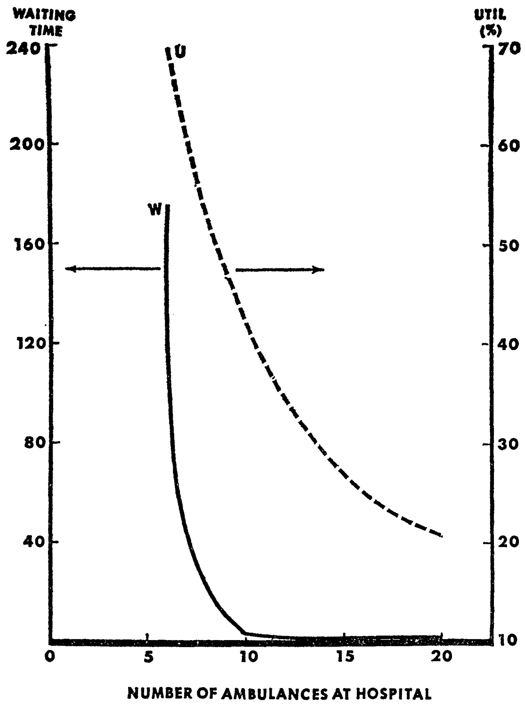

Figure 7. Effect of additional ambulances on waiting time and on ambulance utilization

Economy of Large Districts. Figure 8 displays the relationship between work load (in calls per month) and "ideal utilization" (the utilization corresponding to negligible waiting time, e.g., 42%). The significant observation here is that ideal utilization is not constant and independent of the load; as the load increases, the ideal utilization

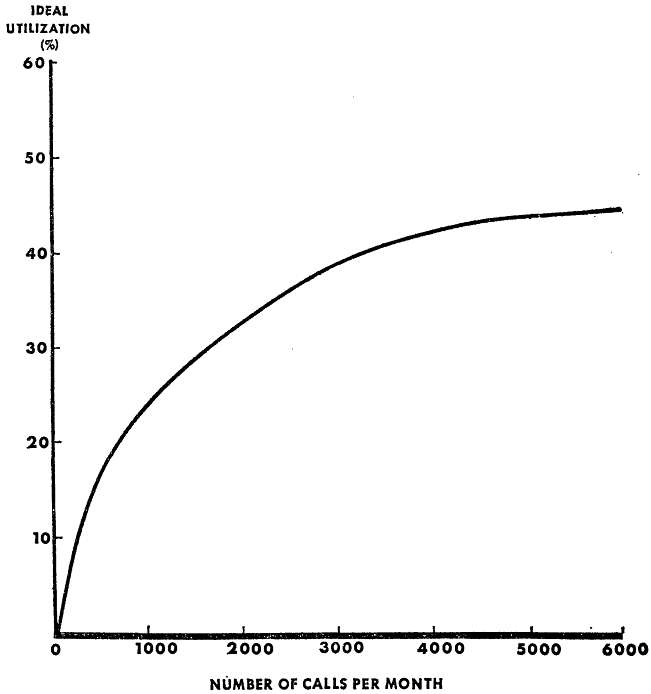

FIGURE 8. Relationship between ideal ambulance utilization and work load

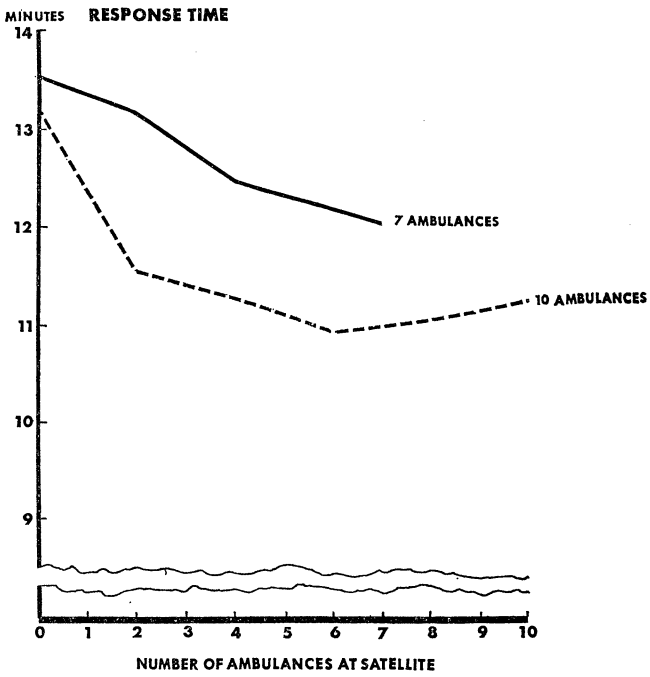

FIGURE 9. Effect of satellite on response time

rate also increases. In essence, this says that if the load were to be doubled, one would need less than twice as many ambulances in order to continue providing ideal service. This result has important policy ramifications. It means that a group of small districts, each with a small load and one or two ambulances, requires more ambulances to provide a given level of service than would be required if the districts were consolidated into a single large district with the ambulances pooled under a unified command in that district. The same effect is achieved by ignoring district lines and simply assigning the nearest available ambulance.

Figure 8 could be used to adjust the number of ambulances in the Kings County Hospital district as the work load fluctuates over time. With minor modification, the data could also be used to guide the staffing patterns of the three work shifts.

C. *Effect of Satellite with Additional Ambulances.*

The number of ambulances serving the district was increased to ten and their effect, with a satellite, was simulated. Figure 9 shows the results for various distributions of the ten ambulances between the hospital and the satellite. The corresponding curve for seven ambulances, taken from Figure 5, is also displayed here for comparative purposes.

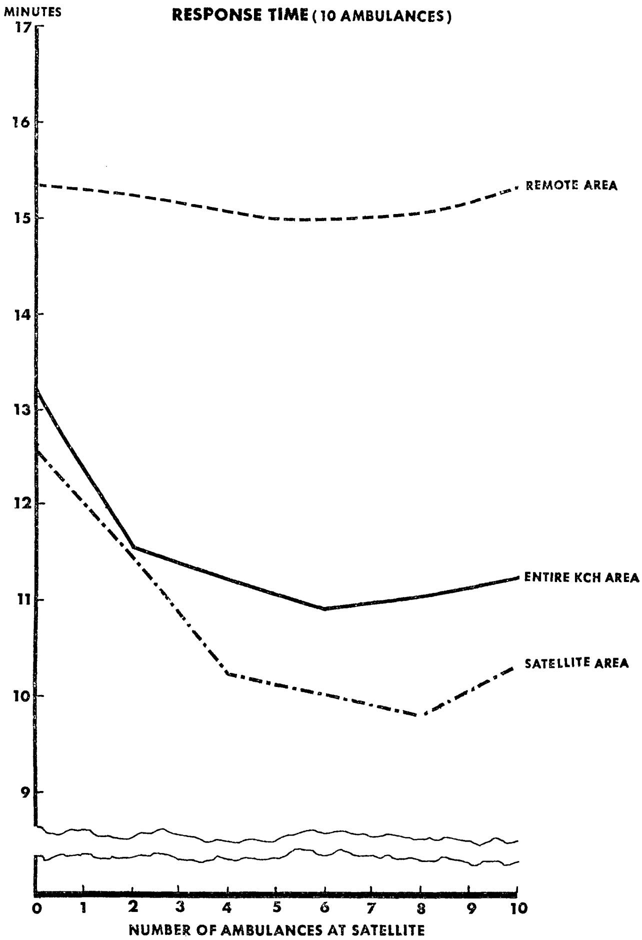

FIGURE 10. Effect of satellite on service to remote area and to area near satellite

The most important feature to observe is that the response time drops to a minimum of 10.9 minutes (with six ambulances at the satellite and four at the hospital), a reduction of 19% from the pre-existing 13.5 minute average. When more than six ambulances are at the satellite, the service gets worse, i.e., the area near the satellite becomes over-saturated with ambulances, just as too many ambulances at the hospital also wastes resources.

Service in Subareas. Up to this point, the discussion has centered on average response time for the entire district. The question can be raised whether certain subareas of the district will experience a *decline* in service, a decline which might be masked in the district average because of a more-than-compensating improvement in service in the subarea near the satellite. Accordingly, the remote subarea and the satellite subarea indicated by the two rectangular areas on the map of Figure 3 were examined. The results are shown in Figure 10. As would be expected, the satellite subarea has better service than the district average; with the six ambulances at the satellite station the satellite subarea has an average response time of 10.1 minutes, a 21% improvement over its pre-existing value of 12.8 minutes.

Even for the remote subarea, however, it was gratifying to note an improvement of 6% over the pre-existing situation, a drop from 16.1 to 15.1 minutes when the ambulances are divided 4:6 between the hospital and the satellite.

IV. Proposed System

On the basis of the simulation results it was concluded that the proposed satellite station, with additional ambulances, could be justified as an immediate way to realize substantial improvements. This satellite was in fact placed in operation on a pilot basis. However, it was felt that further improvements were possible.

In §IIIA it was brought out clearly that Kings County Hospital is not particularly well situated with respect to the district that its ambulances service. Undoubtedly this is true of many other districts in the city as well. This suggests that one source of improvement would be to redistrict the city, taking into consideration the locations of the hospitals and the distributions of calls in order to draw more rational district boundaries.

In considering this suggestion, however, a different and more fundamental recommendation emerged. Simply stated, *ambulances ought to be stationed where the patients are,* without regard to hospital locations, and *an ambulance ought to bring a patient to the nearest appropriate treatment center,* without regard to the ambulance's home station. In other words, the transportation and the treatment subsystems should be separated ("de-coupled"). The problem can therefore be stated as:

(a) where to locate ambulances so that they reach patients promptly; and

(b) where to deliver patients.

A. *Location of Ambulances*

Practically speaking, this recommendation to divorce the transportation service from the treatment service means that ambulances should be operated centrally, for example, by the Department of Hospitals. They should be distributed throughout the city in accordance with the observed demand, and they should be redistributed periodically as the geographic pattern of demand changes due to changes in the population. Hospitals are relatively permanent installations with no mobility and therefore it makes little sense for the transportation service to be attached so inflexibly to such facilities.

These statements should not be interpreted to mean that ambulances should not be stationed at hospitals; if the distribution of calls indicates that a particular hospital is well-situated to serve as an ambulance station, of course it should be used as such.

Furthermore, it is evident that to reduce response time ambulances should be completely dispersed; that is, rarely should there be two or more ambulances stationed at one location. The simulation showed clearly how the response time improves as existing ambulances from one station are apportioned properly among two stations to put them closer to high-demand areas. Further dissemination will result in further improvement, and the maximum decentralization, one ambulance per station, will probably produce the maximum improvement. This statement is in no way inconsistent with the statement in §IIIB concerning the economy of large districts. The crucial point here is the elimination of the entire concept of districts as far as the transportation subsystem is concerned. The nearest available ambulance will be assigned to a call, without regard to any real or hypothetical district boundaries. In fact, this approach is completely equivalent to making the entire city a single district, and having all the ambulances serve that district. This is in perfect harmony with the comments about the savings associated with large, multi-ambulance districts.

Satellite garages, provided they entail only short-term commitments in terms of leases and amortization of capital conversion costs, offer one relatively inexpensive way to provide a more rational dispersion and distribution of ambulances, as has been shown. However, a further extension of this concept is suggested. It is not at all clear that ambulances must be inside a garage while awaiting assignment. Just as there are taxi stands, bus stops and reserved parking places in front of hotels, consulates, hospitals, and post offices, one can conceive of on-the-street ambulance stations. Besides permitting optimal placement of ambulances, this would reduce costs, increase the visibility of the service, and probably reduce the dispatching delay. (On the other hand, additional unnecessary calls might be generated by virtue of the high visibility.) The problem of comfort facilities for the ambulance crew can be handled in the same way that it is handled for radio patrolmen, bus drivers, sanitation men, and taxi drivers. The problem of keeping warm in a standing ambulance in the winter should be surmountable; at worst it will be necessary to keep the motor running, although recognizing the cost in gasoline, noise, and air pollution.

Dispatching of ambulances in such a highly decentralized system must be performed centrally, as it is now, by the Communications Division of the Police Department, but without going through an intermediary dispatcher at a garage. This capability exists today. Each ambulance is already equipped with a two-way radio and is in communication with the police dispatcher. The forthcoming computer-based command-and-control system ("SPRINT") at the Communications Division will enable the ambulance dispatcher to provide even closer, minute-by-minute control over the status and activity of each ambulance in the city system. Furthermore, SPRINT can automate the process of selecting and assigning the nearest available ambulance to each call, despite the wide dispersion of ambulances among many individual locations.

In addition, by employing statistical estimates for the length of time that an ambulance is assigned to a call, the computer might even be able to advise the dispatcher whether to assign a call immediately to a relatively distant but available ambulance, or to wait a few minutes for another ambulance, currently on assignment, that is likely to become available at a location sufficiently close to the point of demand to warrant a brief delay in making the assignment.

Supervisory control, as distinguished from dispatching, would be exercised by the Department of Hospitals by matching information reports on each call from the police dispatcher, the ambulance crew, and the hospital emergency room. SPRINT makes such detailed reporting practical, and such reporting is strengthened by the accurate and prompt feedback on ambulance service that the public provides when it calls in to complain that no ambulance has yet arrived on the scene. This opportunity for feedback control, which is absent in the case of routine preventive police patrol and in the case of sanitation trucks, minimizes the need for on-site supervision. However, such supervision, if deemed necessary, could be performed by a borough commander driving around on inspection in a sedan, in the manner of a district sanitation superintendent.

This tentative recommendation for garage-free satellite operation requires more careful evaluation to determine its feasibility and to see if any necessary functions of a garage have been overlooked. For example, some garage space would still be needed for parking excess ambulances during low-load shifts. In any event, an ambulance from a street station will have to be driven back to a garage at shift change. Refueling and minor maintenance could be performed there. Locker facilities at that garage enable the attendants to change into uniforms and store personal belongings, just as they do now.

B. *Delivery of Patients*

The above recommendations, when implemented, will tend to minimize the response time—the time required for an ambulance to reach the scene. The next question is where to deliver the patient so as to minimize round-trip time. The ideal answer is to deliver him to the nearest appropriate treatment center. (The general phase "treatment center" is used here in order to leave open the possibility of providing some types of emergency medical care at neighborhood health clinics or first-aid centers. The term "hospital" will be employed for the sake of convenience, although the above option should be borne in mind.) The appropriateness of a hospital as a delivery point for ambulance patients depends on the following major factors:

(1) adequacy of its emergency room.

(2) if not a municipal hospital, its selectivity in terms of "interesting cases," a patient's economic resources, and other possible factors.

(3) capacity or bed availability.

One area of possible improvement lies in the first factor. If study shows that average round-trip time in an area could be reduced by bringing patients to a hospital which is not qualified at the present time solely because of inadequate emergency room facilities, the possibility, cost, and effectiveness of upgrading those facilities should be explored. (If such a hospital is too small to accept many emergency patients it should not be considered, as the *average* round-trip time will not show a marked improvement.)

The second factor, a non-municipal hospital's policy of selectivity, is outside the realm of practical systems analysis. Existing hospital district lines in some cases result from such selection criteria. Analysis can serve to identify those hospitals whose participation, or fuller participation, in the system would substantially improve the emergency ambulance service, thereby providing some direction for policy-making officials to negotiate and otherwise bargain with the private institution to secure its participation on mutually acceptable grounds.

It is the third factor, hospital capacity or bed availability, which presents the greatest problem. For the most part, the whole concept of a hospital's ambulance district is a crude attempt to match the hospital's capacity with the expected number of emergency cases in an area. Because it is such a rough approximation, patients are sometimes re-transferred elsewhere when there is no space, and, conversely, overcrowding occurs despite available beds at a nearby hospital. The ideal situation would be for the central dispatcher to have up-to-the-minute information on the actual number of beds available in each hospital in the system. In that case the following ideal sequence of events would occur:

(1) the dispatcher assigns the nearest available ambulance to a call;

(2) after picking up the patient, the ambulance driver informs the dispatcher whether hospitalization is required and if a specialist or particular equipment is needed immediately upon arrival at the hospital (to the extent that he is able to make such determination);

(3) if a hospital bed is required, the dispatcher determines the nearest hospital with an available bed;

(4) the dispatcher instructs the ambulance to proceed to that hospital;

(5) if the driver has requested special aid to be on hand, the dispatcher so advises the hospital.

The Department of Hospitals has already started developing a computer-based bed inventory system covering the municipal hospitals. Depending on the implementation timetable of various recommendations, the inventory could be made available initially to "its ambulances" and then, as decentralization of ambulances is carried out, the dispatcher at the Communications Division would receive this information, probably by telephone. When SPRINT is in operation, on-line input from the various municipal hospitals to the SPRINT computer can be considered. Extension of such an inventory system, where it does not yet exist, to non-municipal hospitals which provide emergency service would be encouraged.

C. *Increased Availability of Ambulances*

The simulation study showed that the combination of adding ambulances in a dis-district and shifting ambulances closer to the point of demand produces improved service. The discussion here centers on low-cost means for increasing the effective number of ambulances. The alternative of simply buying more ambulances is an obvious one that will be excluded from the discussion.

Improve Screening of Calls. About 15% of all calls turn out to be "unnecessary," according to a recent statistical study of New York's emergency ambulance service [2]. More diligent efforts by personnel at the Communications Division to question the caller before deciding to dispatch an ambulance is likely to reduce substantially the number of cases where ambulances are sent out on unnecessary calls. (This is the procedure followed in Baltimore, where only 8% of the calls turn out to be unnecessary.) By decreasing ambulance utilization in this way, service on true emergency calls will be improved.

Overtime Pay. Because drivers and attendants do not receive pay at overtime rates, crews are said to be reluctant to accept assignments a few minutes before their normal quitting time. If this is true, it would seem that the marginal cost of overtime labor is an inexpensive way to buy, in effect, more ambulances. In addition, by being able to offer overtime pay, an employee finishing one shift can be induced to work a second shift if his relief man fails to report to work. Such absenteeism effectively results in ambulances out of service.

Interchange of Crew Members. At present, ambulance attendants report to the nursing service in hospitals while the drivers are responsible to a garage foreman. This divided allegiance results in inflexible scheduling; for example, it is not possible to shift crews from one garage to another when there is a local shortage. Furthermore, if only a driver reports to work at one hospital, and only an attendant at another, pairing the two men to provide one ambulance, instead of keeping two ambulances idle, is administratively awkward. By divorcing the ambulance service from the hospital itself, the resulting centralized authority over crew members should simplify the handling of such problems.

The suggestion has also been made to provide the same training for both driver and attendant. This will permit complete interchangeability and, together with the change in policy on over-time pay, will result in fewer ambulances standing idle due to personnel absences.

Patient Acceptance Procedures. On certain classes of calls (e.g., psychiatric cases) the ambulance crew must wait at the hospital, after delivering the patient, for an inordinately long time before the ambulance is released and becomes available for reassignment. A change in the admission/acceptance procedure in such cases will increase the effective availability of ambulances.

V. Cost-Effectiveness Evaluation

A. *Effectiveness*

Average Response Time. Several simulation runs were conducted with ambulances stationed at various points in the district in order to compare the proposed system of dispersed ambulances to the other alternatives. The results are summarized in Table 2.

TABLE 2

Average Response Time

Alternatives	Number of Ambulances			
	7		10	
	Average Response Time	% Improvement	Average Response Time	% Improvement
a. All Ambulances at Hospital	11.9′	0		
b. Optimal Allocation of Ambulances be-between Hospital and One Satellite	10.2′	14%	9.3′	22%
c. Totally Dispersed Ambulances	9.7′	18%	8.4′	30%

TABLE 3

Fraction of Calls with Response Time Greater than Twenty Minutes

Alternatives	Number of Ambulances			
	7		10	
	Fraction	% Improvement	Fraction	% Improvement
a. All Ambulances at Hospital	.099	0		
b. Optimal Allocation of Ambulances between Hospital and One Satellite	.073	26%	.051	48%
c. Totally Dispersed Ambulances	.065	34%	.03	69%

It is clear that the dispersed system is superior to the other alternatives in terms of the improvement attainable; for example, ten dispersed ambulances will reduce the response time by 30% (from the base case) whereas the same ten ambulances distributed in optimal fashion (4:6) between hospital and satellite will produce only a

TABLE 4

Estimated Costs

	Purchase Price	Annual Cost	Total Annual Cost*
I. Vehicle (ambulance)	$5,700	$950	$950
Ambulance (6 yr. life)	4,900		
Equipment (6 yr. life)	800		
II. Vehicle (supervisory)		3,040	3,040
Sedan (2 yr. life)	2,000	1,000	
Equipment (5 yr. life)	200	40	
Maintenance & Supplies		2,000	
III. Vehicle Maintenance and Supplies		1,958	1,958
Maintenance and Repair Supplies		657	
Mechanics' Labor		505	
Gasoline and Oil		296	
Oxygen and Medical Supplies		500	
IV. Ambulance Crew		14,505	72,525
Motor-Vehicle Operator		8,175	
Salary		6,500	
Overhead (22%)		1,430	
Uniform (allowance)		65	
Food Allowance		180	
Attendant		6,330	
Salary		5,000	
Overhead (22%)		1,100	
Uniform (issued)		50	
Food Allowance		180	
V. Garage		13,600	13,600
Rent		12,000	
Heat		1,100	
Light		300	
Telephone		200	
VI. Garage Staffing		14,516	72,580
Foreman		9,395	
Salary and overhead		9,150	
Uniform and food allowance		245	
Clerk		5,121	
Salary and overhead		4,941	
Food Allowance		180	
VII. Cruising Supervisor		9,395	46,975

* For three shifts per day, seven days per week. Allowing for vacations, illnesses, etc., five crews are required to staff three shifts per day, seven days per week.

TABLE 5

Deployment of Ambulances

Alternative	Tour 1		Tour 2		Tour 3		Total No. Ambulances In System	Additional Ambulances	Total Tours*	Added Tours
	E	T	E	T	E	T				
Seven ambulances (original pattern)	6	0	7	2	7	0	9	—	20	—
Eight ambulances	6	0	7	2	8	0	9	0	21	1
Nine ambulances	6	0	7	2	9	0	9	0	22	2
Ten ambulances	6	0	7	2	10	0	10	1	23	3

E = Emergency service T = Transfer service

* Does not include transfer service

Note: A tour is defined here as an 8-hour work period for each of seven days.

TABLE 6

Incremental Costs of Alternatives

Alternative	Annual Cost	Monthly Cost
(A) 7 ambulances with a satellite	$86,180	$7,182
Garage	13,600	
Garage staffing	72,580	
(B) 10 ambulances with a satellite	161,613	13,468
Garage and garage staffing	86,180	
Additional ambulance	950	
Maintenance and supplies	1,958	
Ambulance crews (5)	72,525	
(C) 7 ambulances dispersed	16,700	1,400
cruising supervisor and vehicle	16,700*	
(D) 8 ambulances dispersed	40,800	3,400
cruising supervisor and vehicle	16,700*	
ambulance crews (1–2/3) (to staff one seven-day tour)	24,100	
(E) 9 ambulances dispersed	64,900	5,408
cruising supervisor and vehicle	16,700*	
ambulance crews (3–1/3) (to staff two seven-day tours)	48,200	
(F) 10 ambulances dispersed	92,133	7,700
Cruising supervisor and vehicle	16,700*	
Additional ambulance	950	
Maintenance and supplies	1,958	
Ambulance crews (5) (to staff 3 seven-day tours)	72,525	

* Because a supervisor can cover 20–30 ambulances, a district of 7–10 ambulances requires only one-third of his time; hence, only one-third of the cost is charged to this district.

Note: The cost of a dispersed system does not include credit for savings due to staff, space, and equipment reductions at the base garage.

22% improvement. (Due to an adjustment in the mathematical model, the absolute values of the response times are not identical to the values shown on the figures and discussed earlier for identical ambulance configurations. This improvement does not change the earlier conclusions nor do they alter significantly the percentage improvements.)

Reduction of Long Delays. The average response time in itself is insufficient to portray the effect of the alternatives on reducing the frequency of those unfortunate occurrences where a patient waits for an excessively long time before an ambulance appears. Because of the great desirability of reducing the fraction of calls which are subject to long delays, the effect of the alternatives on this factor was also examined. The findings are summarized in Table 3. Again, the dispersed pattern of operation is best by far: ten dispersed ambulances can be expected to reduce this fraction by 69% compared to a 48% reduction with a satellite.

Inasmuch as mathematical models never duplicate the real world exactly and because of statistical uncertainties in the findings, it is felt that although the absolute fractions of delayed calls, shown on Table 3, are not necessarily accurate, the relative improvements shown for the alternatives are indeed meaningful.

B. *Costs*

The simulation results presented in the preceeding section were devoted exclusively to portraying the effectiveness of the different alternatives. Now the reverse side of the coin, the costs, must be examined. Table 4 displays the capital and operating costs for the various resources required. Using Table 5, which indicates the staffing patterns, the costs shown on Table 4 can be combined to reflect the incremental costs in-

TABLE 7

Cost Effectiveness of Alternative Ways to Reduce Response Time

Alternative	Effectiveness: Minutes Saved	Cost: $/Month	Cost/ Effectiveness: $Per Minute	Cost Per Call
(A) 7 Ambulances with a Satellite	1.7	$7,182	$1.16	$1.96
(B) 10 Ambulances with a Satellite	2.6	13,468	1.42	3.68
(C) 7 Ambulances Dispersed	2.2	1,400	.17	.38
(D) 8 Ambulances Dispersed	(2.6)	3,400	.36	.93
(E) 9 Ambulances Dispersed	(3.0)	5,408	.49	1.48
(F) 10 Ambulances Dispersed	3.5	7,700	.60	2.10

TABLE 8

Cost Effectiveness of Alternative Ways to Reduce Excessive Delays

Alternative	Effectiveness: Percentage Points Reduced Below 20 Minutes	No. Calls Per Month Reduced Below 20 Minutes	Cost: $/Month	Cost/ Effectiveness: $ Per Call Reduced
(A) 7 Ambulances with a Satellite	2.6	95	$7,182	$75.50
(B) 10 Ambulances with a Satellite	4.8	176	13,468	76.50
(C) 7 Ambulances Dispersed	3.4	125	1,400	11.20
(D) 8 Ambulances Dispersed	(4.6)	168	3,400	20.20
(E) 9 Ambulances Dispersed	(5.8)	222	5,408	24.40
(F) 10 Ambulances Dispersed	6.9	252	7,700	30.60

Note: Figures in parenthesis are obtained by interpolation.

volved in going from the present configuration to each of the three alternative configurations; these are shown in Table 6.

It is assumed that:

(1) on-the-street stations, with zero cost, are used for the dispersed ambulance systems;

(2) equivalent levels of supervision are employed for all alternatives, thus permitting accurate and fair comparisons to be made;

(3) shift-to-shift staffing patterns are the same for each alternative, as shown in Table 5.

C. *Cost-Effectiveness*

The cost-effectiveness of alternative ways to reduce response time and to reduce excessive delays are shown in Tables 7 and 8. The dramatic superiority of the dispersed configurations is self-evident; eight dispersed ambulances (alternative D) are as effective as ten ambulances in a satellite system (alternative B) at about one-fourth the incremental cost per call. This is a significant conclusion and a compelling argument for a dispersed system. Furthermore, the relative ranking of the alternatives is clear and unambiguous even though the actual dollar figures in the cost/effectiveness columns of the tables may not be accurate enough for budgetary or accounting purposes.

These tables give the policy maker the opportunity to make an enlightened choice as to the degree of improvement he wishes to aim for, and the most efficient (least expensive) way to achieve that objective. Work is now underway to examine other areas of the city and to take the administrative steps necessary to translate these analytical findings into public policy.

References

1. Gordon, G., and Zelin, K., "A Simulation Study of Emergency Ambulance Service in New York City," Tech. Rept. No. 320-2935 (March 1968), IBM Corporation.
2. Dimendberg, D. C., "An Analysis of the Ambulance Service," Department of Hospitals, City of New York (June, 1967).