RESPONSE TIME AND ITS SIGNIFICANCE IN MEDICAL EMERGENCIES*

JONATHAN D. MAYER

THE Emergency Medical Systems Service Act (EMSS Act) of 1973 was passed by Congress to promote systematic planning and evaluation of regional emergency systems. The Act represented an attempt to coordinate a number of components in prehospital emergency cases, such as transportation, personnel, and facilities. One of the provisions of the EMSS Act is that 95 percent of all ambulance response times must be within twenty minutes in rural areas and within ten minutes in urban areas. Compliance with response-time criteria is largely a geographical problem. In terms of any given population distribution, how may emergency units be most efficiently allocated? This allocation problem is such that its solution may be solved by location-allocation models and their various derivatives. Thus geographical attention to emergency medical systems has concentrated largely on the solution of the local problem.[1] The "objective function" has been structured in several ways: to minimize response time subject to a cost constraint; to minimize cost subject to a response time constraint; to ensure that a certain percentage of emergency calls are answered within a specified time period. Cost has been an increasingly important issue because it is theoretically possible, but obviously impractical, to locate one ambulance on every city block and hence to ensure a minimal response time. The two major variables that may be manipulated are the number of emergency vehicles and their locations.

An implicit assumption in the application of normative models to emergency medical problems is that patient outcome is partially determined by ambulance-response time. Previous research on the spatial aspects of emergency medical services leaned heavily on this assumption as justification for the use of location models.

> The efficient location of equipment and personnel is an imperative element in the design and operation of an EMS system. Poor siting could delay the provision of service, thus increasing the probability of mortality or disability in many circumstances.[2]

The assumption about efficient siting, as reflected in rapid ambulance-response times, has remained untested. If minimization of response time is not correlated with patient outcome, then the use of location models in emergency medical problems is

---

* I wish to thank the Seattle Fire Department and Lieutenant Ralph Maughan, in particular, for providing the data on which this paper is based. I also thank Dr. Alfred Hallstrom and Dr. Leonard Cobb for a discussion of ideas relevant to this paper.

[1] There is a massive literature about optimal location models. Of particular relevance are Geoffrey Berlin and Jon Liebman, Mathematical Analysis of Emergency Ambulance Location, *Socio-Economic Planning Sciences,* Vol. 8, 1974, pp. 323–328; Richard Church and Charles ReVelle, Theoretical and Computational Links between p-Median, Location Set-Covering, and the Maximal Covering Location Problem, *Geographical Analysis,* Vol. 8, 1976, pp. 406–415; S. Daberkow, Location and Cost of Ambulances Serving a Rural Area, *Health Services Research,* Vol. 12, 1977, pp. 299–311; and Charles ReVelle and others, Facility Location: A Review of Context-Free and EMS Models, *Health Services Research,* Vol. 12, 1974, pp. 6–21.

[2] Rex Honey and others, Spatial Analysis of Iowa's Emergency Medical Services, *Small-Area Analysis Series 3* (Iowa City: University of Iowa, Health Services Research Center, 1978), p. 1.

● DR. MAYER is an assistant professor of geography and an affiliate of the Center for Health Services Research at the University of Washington, Seattle, Washington 98195.

more an intellectual exercise than an important policy tool. On the other hand, if response time is an important factor in patient survival, then efficient ambulance location is crucial. It is obviously necessary to question and to test the importance of response time.

RESPONSE-TIME STUDIES

Response time is defined as the interval between the time at which an emergency vehicle is dispatched and its arrival time at the location of the emergency. This is one of several delays in emergency medical services. Others include the wait from the onset of symptoms or occurrence of the emergency until a decision is made to seek medical attention, the delay from the formulation of a decision until contact is made with the system, and the interval from the time that an ambulance dispatcher is contacted until an emergency unit begins its response.

In some areas the travel time between the scene of the emergency and a hospital constitutes a fifth delay. This occurs in areas where ambulances are staffed by untrained personnel or by emergency medical technicians (EMTs) who can administer first aid and perform other simple procedures, but who cannot provide definitive medical care. On the other hand, where paramedics who have received up to 1,500 hours of formal training staff the emergency vehicles, the delay between the scene of the emergency and the hospital is unimportant, because paramedics are legally and technically able to stabilize patients at the scene of the emergency.

Previous studies suggest that the first delay, or "patient decision time," is the lengthiest. In one study in Rochester, New York, decision time accounted for 65 percent of the prehospital delay.[3] Another study concluded that only 40 percent of heart patients sought medical assistance within two hours of the onset of their symptoms.[4] The psychological coping mechanism of denial and the ignorance of major symptoms were the main explanations of this delay.[5] The second delay—the one between the formulation of a decision and the action on the decision—is highly variable. When an accident occurs in a remote rural area, this delay may be quite lengthy; in urban areas it is normally a matter of seconds. The third delay, or "dispatch delay," is typically less than one minute. Finally response times and travel times to the hospital are also variable. As suggested by the criteria of the EMSS Act, travel times in rural areas may be as long as thirty minutes, owing to the distances that must be traversed, while response times in urban areas are infrequently greater than ten minutes.

Little research has been completed on the relationship between response time and patient outcome, and it has therefore been difficult to assess the potential impact of using location-allocation models on patient survival. In cases of cardiac arrest, when the heart ceases its basic functions, cardiopulmonary resuscitation (CPR) must be initiated in four to six minutes, or irreversible brain damage occurs. Resuscitation may be initiated either by technical personnel or by trained bystanders. This suggests

---

[3] A. Moss and others, Delay in Hospitalization during the Acute Coronary Period, *American Journal of Cardiology,* Vol. 24, 1969, pp. 659-665.

[4] S. Tjoe and M. Luria, Delays in Reaching the Cardiac Care Unit: An Analysis, *Chest,* Vol. 67, 1972, pp. 617-621.

[5] A. Moss and others, Denial of Chest Pain in 32 Patients with Acute Myocardial Infarction, *Journal of the American Medical Association,* Vol. 190, 1974, p. 977.

that response time is tremendously important in cardiac arrest cases. A linear relationship between the probability of successful resuscitation and long term survival, on the one hand, and the time from the onset of the arrest until the administration of definitive care, on the other, has been noted.[6] Unfortunately the research did not analyze the impact of response time per se on patient outcome. When all medical emergencies are considered, research suggests that response time is a poor predictor of survival. Many medical emergencies, however, are not potentially life-threatening, and it is desirable to include in the analysis only those emergencies that may result in death.[7]

The role of response time may be tested with data maintained by the Seattle Fire Department, which is also the operator of the paramedic program in Seattle. The emergency medical system of this city has received national and international recognition as a pioneering effort at developing strategies for paramedics to stabilize patients at the scene of the emergency. Known as the "Medic-I" program, the emergency system has been described in detail in technical journals and through popular media such as "Sixty Minutes."[8] The Medic-I program provides three types of responses for medical problems. The first level includes the thirty-five fire-engine companies in the city, all of which are staffed by EMTs. In emergencies such as unconsciousness, major trauma, and suspected heart problems, all of which are potentially life-threatening, these units will be dispatched to the scene when they are the closest units to the emergency, with the expectation that they will be the first to reach the scene. They are thus able to commence life-support measures, such as cardiopulmonary resuscitation, as soon as possible. The second response level consists of nine "aid units," also staffed by EMTs, that are spaced throughout the city and correspond to a more traditional notion of "ambulances." In nonlife-threatening emergencies, aid units provide a backup response. The third level of service consists of four paramedic units or "mobile intensive care units," located throughout the city. Paramedics are able to perform sophisticated, supportive measures in the field, such as cardiac defibrillation, administration of intravenous medication and fluids, and certain invasive surgical procedures like tracheotomies. The fire department responds to all reported medical emergencies in the city; private ambulance companies transport only patients who do not require close monitoring to the hospital, when they have been stabilized by the paramedic units. Thus in a life-threatening emergency, one or more EMT units are dispatched to provide early supportive measures prior to the arrival of a paramedic unit.

The Medic-I program operates within the response-time criteria stated in the Emergency Medical Service Systems Act. In 1976, for example, the mean-response times for fire engines, aid units, and paramedic units were respectively 2.6, 3.8, and 5.3 minutes. During the first half of 1978, the mean-response time for nonparamedic units was 5.1 minutes. In 1976 approximately 20,000 emergency calls were answered; half of them were considered to be potentially life-threatening and merited the dis-

---

[6] Mickie Eisenberg and others, Paramedic Programs and Out-of-Hospital Cardiac Arrest: I. Factors Associated with Successful Resuscitation, *American Journal of Public Health,* Vol. 69, 1979, pp. 30–38.

[7] Jonathan D. Mayer, Seattle's Paramedic Program: Geographical Redistribution, Response Times, and Mortality, *Social Science and Medicine,* Vol. 13D, 1979, pp. 45–52; and Jonathan D. Mayer, Emergency Medical Service Delays, Response Time, and Survival, *Medical Care,* in press.

[8] Leonard Cobb and others, A Rapid Response System for Out-of-Hospital Cardiac Emergencies, *Medical Clinics of North America,* Vol. 60, 1976, pp. 283–290.

patch of a paramedic vehicle. During the first half of 1978, approximately 10,500 medical calls were answered by the fire department, 5,700 of which merited a paramedic response.

RESPONSE TIME AND SURVIVAL

Because it is an explicit assumption of geographical allocation models that short response times are desirable, a reasonable expectation is that the probability of a patient surviving an emergency should be a decreasing function of response time. Longer response times should thus have fewer patient survivals associated with them than shorter response times, other things being equal. This may be tested by partitioning response time into intervals and computing the percentage of deaths corresponding to each response-time interval. Computerized data from the Seattle Fire Department on response times, patient status, tentative diagnosis information, and demographic statistics were analyzed with this test in mind.

The results of the analysis cover each of the three levels of medical response and all cases in which a paramedic unit was dispatched in 1977 (Table I). In addition, the percentage of prehospital deaths corresponding to the response time of the first arriving unit, regardless of its type, was computed. As will be seen subsequently, this statistic is particularly important in cardiac arrest cases. There is a noticeable relationship between response time and short-term patient survival. In particular, when the response time of the first arriving unit exceeds one minute, the probability of death doubles. Similarly the probability of death increases with longer paramedic-unit response times. The first minute appears to be crucial with respect to paramedic response time. Moreover, delays of nine minutes and longer raise the probability of death from .06 to .11. While the increased probability of death with response-time delays is not dramatic, it is nonetheless present. As was mentioned earlier, the effect of response time on fatality rates may be expected to be greatest for cardiac arrest cases. From an epidemiological viewpoint, it is desirable to sample only the cases that are most similar to one another.

CARDIAC ARREST AND RESPONSE TIME

Cardiac arrest is one of the most dramatic medical emergencies. It occurs when the heart is unable to pump blood to the rest of the body. The arrest is recognizable by lack of respiration and pulse, by rapid onset of cyanosis, and by absence of neurological response signs. Unconsciousness occurs within eight seconds of cardiac arrest. Cardiac arrests are of several forms, the most common being ventricular fibrillation (VF) in which the heart beats in a totally uncoordinated manner, thereby preventing circulation, and asystole, or total cardiac standstill, when there is no electrical activity in the heart, the usual result of uncorrected VF. Resuscitation is much more difficult in cases of asystole than in VF cases.

If circulation is not restored within four to six minutes after the onset of cardiac arrest, irreversible brain damage occurs even if the patient is successfully resuscitated. Resuscitation after long periods of time is highly unlikely because the probability of successful resuscitation diminishes rapidly. If artificial respiration is administered along with chest massage (cardiopulmonary resuscitation), enough circulation may be maintained so that definitive resuscitation may be accomplished by electrical

TABLE I—RESPONSE TIMES AND PERCENTAGE OF DEATHS, LIFE-THREATENING EMERGENCIES

RESPONSE TIME (minutes)	FIRST-IN		ENGINES		AID UNITS		MEDIC UNITS	
	N	% Deaths	N	% Deaths	N	% Deaths	N	% Deaths
0-.99	127	3.1	31	6.5	100	2.0	80	2.5
1-1.99	1184	5.9	510	7.5	694	4.9	284	4.6
2-2.99	2983	6.2	1135	7.5	1937	5.6	887	4.8
3-3.99	3133	5.3	931	7.5	2506	5.0	1768	5.4
4-4.99	1642	5.4	401	7.2	1672	5.0	1658	4.5
5-5.99	698	4.6	132	9.8	1039	4.9	1373	4.7
6-6.99	231	5.6	62[b]	11.3	640	5.3	1132	5.6
7-7.99	241[a]	4.4			483	6.4	898	5.7
8-8.99					333	6.3	636	6.0
9-9.99					247	10.5	470	8.3
10-10.99					277	11.0	427	8.4
11-11.99					121	8.3	211	6.2
12-12.99					92	4.3	168	8.9
13-13.99					70	5.7	107	6.5
14-14.99					29	17.2	56	8.9
15 or more					64	7.8	98	11.2

[a] No first-in times longer than 7 minutes
[b] No engine-response times longer than 6 minutes

TABLE II—RESPONSE TIME AND PERCENTAGE OF DEATHS, CARDIAC ARRESTS

RESPONSE TIME (minutes)	FIRST-IN UNIT		MEDIC	
	N	Percentage	N	Percentage
0-1.99	21	61.9	6	66.7
2-3.99	127	59.1	48	47.9
4-5.99	60	60.1	82	53.7
6-7.99	42	60.0	87	57.5
8-9.99	15	60.0	36	63.9
10-11.99	10	70.0	20	65.0
12-13.99	4	25.0	8	50.0
14 or more	40	60.0	32	93.8

countershock (defibrillation) and appropriate intravenous medication. For the patient to survive, resuscitation must be administered as soon as possible.

The variation of survival with response time should be quite pronounced in cardiac arrest cases, due to the necessity of administering aid in a maximum of six minutes. This problem was tested with several methodologies, with a statistical base of the 319 cardiac arrest cases in Seattle during the first half of 1978. Fire-engine and aid-unit responses are grouped together because neither response can provide definitive care, which must await the arrival of a paramedic unit. The data suggest that the death probability varies irregularly and erratically with the response time of nonparamedic units. (Table II). The pattern is understandable because nonparamedic EMTs can institute only basic, not advanced, life support. On the other hand, death probabilities increase regularly with increasing paramedic response time. With paramedic response times greater than fourteen minutes, nearly 94 percent of the patients are not resuscitated, in contrast to 47.9 percent when paramedic units arrive in the 2- to 3.99-minute interval. With the exception of the 12- to 13.99-minute interval, the increase in unsuccessful resuscitations is monotonic.

The hypothesis that paramedic response time exerts a significant influence on survival may be tested formally by simple cross tabulation. The null hypothesis that survival or death probabilities are equal for different time intervals is tested by deriving the chi-square statistic (Table III). The chi-square statistic indicates that the null hypothesis must be rejected. A comparison of the actual and the expected values suggests that the number of surviving patients, given response times of less than eight minutes, is greater than expected, while the number of surviving patients, when response times are greater than eight minutes, is much less than expected. The mean-response time for successfully resuscitated patients, furthermore, was 5.7 minutes, in contrast to 6.2 minutes for those who were not resuscitated successfully.

Although response time clearly affects outcome in cardiac arrest cases, the preceding analysis provides little indication of how significant response time is as a determinant of survival. The total time to definitive care clearly is an important predictor of survival, yet response time alone is not necessarily correlated with time to definitive care, because of factors such as bystander delays in recognizing the seriousness of sudden collapse, inaccessibility to a telephone, or absence of witnesses who may contact the emergency system. An important question is the degree to which response time is a predictor of survival in cardiac arrest cases.

Response Time, Demographic Variables, and Survival

Several demographic variables are traditional concerns in epidemiological studies and frequently prove to be of value in predicting survival. Of these, the most basic and available data include the patient's age, sex, and race. Because the immediate cause of every death is cardiac arrest and because mortality among the elderly is quite high due to advanced atherosclerosis and other degenerative diseases, age may be expected to vary inversely with probability of survival. From the Seattle data, this proposition may be tested directly. Sex and race are also of epidemiological concern. These three variables, along with response time, may be entered into a discriminant analysis to assess their power to discriminate between patients who are resuscitated from an arrest and those who are not resuscitated. The purposes of the discriminant analysis are twofold: to assess the degree to which the aforementioned variables can predict survival, and to ascertain the relative importance of each of the variables as a predictor of survival.

More specifically, the following variables are considered in the discriminant analysis. The medic (MEDICRES) and first-arriving-unit (INIRES) response times are interval-level variables that are measured in minutes, while age is measured in years. The race and sex are measured using dummy variables. If the patient is white, the variable WHITE assumes a value of 1; if the patient is black, the variable BLACK assumes a value of 1. If the patient is of another race, both racial dummy variables are set equal to 0. Finally, if the patient is male, the variable MALE is set equal to 1.

The results of the discriminant analysis include both the standardized and understandardized discriminant function coefficients (Table IV). The standard coefficients suggest that the medic-response time and the dummy variable BLACK contribute the greatest to the discriminant function, with the other variables contributing less to the function. The signs of these two coefficients confirm that longer response times are inversely related to survival and also indicate that a patient's being black adversely affects short-term survival chances. When the discriminant analysis is per-

TABLE III—CROSS TABULATION OF RESPONSE TIME AND SURVIVAL

	RESPONSE TIME					
	0–3.99 MINUTES		4–7.99 MINUTES		8 OR MORE MINUTES	
LIFE STATUS	Cases	Expected Value	Cases	Expected Value	Cases	Expected Value
Alive	27	21.67	75	67.81	26	38.52
Dead	54	32.33	94	101.19	70	57.48

TABLE IV—DISCRIMINANT FUNCTION COEFFICIENT

VARIABLE	STANDARDIZED COEFFICIENT	UNSTANDARDIZED COEFFICIENT
INIRES	−.24	−.14
MEDICRES	−.70	−.24
WHITE	.14	.39
BLACK	−.72	−2.5
MALE	−.14	−.30
AGE	.01	.01
CONSTANT		1.90
GROUP CENTROIDS		
Alive	.17	
Dead	−.14	

formed stepwise with Wilk's lambda as a criterion for entry into the analysis, significant improvement in discrimination comes only from the entry of the first two variables, BLACK and MEDICRES. Other stepwise additions to the analysis are significant only at the .5 level and greater.

Although the analysis indicates the relative contribution of each variable to the discriminant function, there is nothing indicative of the degree to which the function predicts patient outcome. Further analysis suggests that the discriminant function is a questionable predictor. For example, the canonical correlation is only .15 for the direct analysis. It should be noted, however, that 60.7 percent of the cases were correctly classified, with the corresponding chi-square significant at the .001 level.

When response times only are considered, the discriminating power of the derived functions is even less. When first-arriving and medic-unit response times are included, the canonical correlation is only .09, with the resulting chi-square test of Wilk's lambda significant at the .3 level. Only 53 percent of the cases, moreover, were classified correctly. It was therefore impossible to predict whether a patient was resuscitated successfully based only on the relevant response times. From a geographical viewpoint, the poor predictive power of the discriminant function is rather disconcerting, because it suggests that even when EMS response times and the patient's age, sex, and race are known, it is difficult to predict the outcome of the case.

POLICY IMPLICATIONS

A powerful and essential step in the evaluation of a theoretical construct is to question the assumptions of the construct. Much work in emergency medical services research has accepted the need to minimize response time, without testing the assumption or subjecting it to scrutiny. This study has provided a method and a case study by which the effects of response time may be tested.

The relationship between response time and patient survival is clearly a complicated relationship and appears to be less obvious than initially assumed. This study indicates that paramedic response time seems to influence short-term survival in life-threatening emergencies, particularly cardiac arrest cases. Even in arrest cases, response time, however, remains a poor predictor of successful resuscitation, when important demographic variables are included in the analysis. Why this is so is not totally clear. Total time to definitive care seems to be more important than response time, as does the particular pathology of a given case, because cardiac arrest has a set of diverse etiologies. If the public is unaware of the symptoms of cardiac arrest or of other emergencies, or if they are apathetic about taking definitive action to help others, quick response times can have little effect on patient outcome. This situation suggests that education of the public about the symptoms of severe cardiac problems and their initial treatment is more important than marginally productive improvements in ambulance response time.

> Substantial reductions in ambulance response time have a small but noticeable effect on the probability of pre-hospital death as long as patient delays are short . . . when patient delays reach the often reported median of 2 to 3 hours, the effects are negligible.[9]

Unfortunately most work by EMS researchers has not acknowledged the limited degree to which minimal response-time reductions may improve patient outcome, the final standard by which EMS systems must be judged. "Response time . . . does not measure the quality or quantity of service provided but only vehicle speed and service area geographic constraints."[10] What is needed is a more holistic approach to emergency services, where patient care in its entirety is maximized, rather than isolated components that may or may not have significance. The efficient allocation of emergency vehicles should clearly not be ignored. Certainly evidence exits to prove that rapid response times are better than slow response times. Prediction of survival, however, is much more complicated than calculating response times and then mapping the spatial distribution of response times. Epidemiological factors must be considered: for example, the distribution of diseases and patterns of demand, demographic factors such as the age and the ethnic distribution of the population, and behavioral patterns, including propensity to delay the search for medical assistance or to deny the existence of serious symptoms.

Location-allocation models may be a useful tool as one element in the design of a well-functioning emergency medical system. In a rural area, where travel distances and response times are long, any correlation between response time and survival may appear to be greater than in an urban area, but this question deserves further research. Response time, however, should not to be seen in isolation of the rest of the emergency system.

Any comprehensive analysis of response time must deal with a number of unanswered questions, all of which may be approached from a geographical framework, tempered by epidemiological knowledge and philosophical wisdom. How important is response time for survival of emergencies other than cardiac arrest? For which emergencies is response time in any way important? In nonlife-threatening emer-

[9] Shan Cretin, A Model of the Risk of Death from Myocardial Infarction (Cambridge: M.I.T. Operations Research Center, 1974).

[10] Geoffrey Gibson, Measures of Emergency Ambulance Effectiveness: Unmet Need and Inappropriate Use, *Journal of the American College of Emergency Physicians,* Vol. 6, 1977, pp. 389–392.

gencies, where there is no question of survival but of discomfort and pain, how may the significance of response time be assessed? There is a pressing moral and philosophical issue, because in the allocation of emergency medical units the fundamental question is not transportation efficiency, but life itself. The reallocation of units may save several lives, but it may also effectively prevent the survival of a patient in an area that becomes poorly served. And finally the question of cost cannot be ignored. Paramedic units are extremely costly. Addition of new paramedic units to an emergency system may save lives, but the cost of the system will also increase. Decision makers eventually must "draw the line" somewhere and decide when enough units are operating in the system. In so doing, they are comparing lives saved to dollars saved. Perhaps this should not be done, but the practice seems inevitable. Geographers who are involved in such studies must consider more deeply the validity of their assumptions underlying the models that are being used. These individuals must realize that the implementation of their efficient locational decisions are not only technical but also involve a moral dimension.