MULTI-LEVEL AMBULANCE SYSTEM DESIGN

by

RONALD VICTOR KITT
B. Comm., University of Alberta, 1976

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE (BUSINESS ADMINISTRATION)

in

THE FACULTY OF GRADUATE STUDIES
THE FACULTY OF COMMERCE AND BUSINESS ADMINISTRATION
(MANAGEMENT SCIENCE DIVISION)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
May, 1979

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Commerce and Business Administration

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date April 27, 1979

ABSTRACT

The objective of this research was to study the operation of the ambulance service in the Greater Vancouver Regional District. Attention has been focused on operational policies which affect the system's ability to respond to calls, especially to emergency calls.

The stated objective was accomplished by first reviewing the current demand for emergency ambulance services and the present operations. Next deterministic models were investigated and used to give "initial" locations of depots. Finally a computer simulation model was developed and used to conduct experiments examining alternative ambulance systems.

This research revealed that

1. computer simulation is an effective tool for analysing ambulance systems, and

2. there is a need for more emergency ambulance services, including more paramedics, in the Greater Vancouver Regional District.

TABLE OF CONTENTS

LIST OF TABLES

List of Tables (continued) Page

LIST OF FIGURES

List of Figures (continued) Page

ACKNOWLEDGEMENTS

The writer would like to express his sincere thanks to Dr. D.H. Uyeno for his encouragement, advice and guidance throughout the months of thesis research and preparation. My sincere thanks also to Dr. I. Vertinsky for his constructive suggestions.

The writing of a thesis requires the assistance of numerous individuals. In particular, the writer would like to thank Mr. Carson Smith, Head of Operations of the Emergency Health Services Commission for his cooperation and assistance. He provided much of the voluminous amount of data and assisted in the development of experiments.

CHAPTER 1

1.1 INTRODUCTION

Emergency ambulance services are an important component of the general health care system. Each and every day ambulance personnel provide life-giving support to people in need. Victims of serious motor vehicle accidents, of cardiac arrests, of burns, of industrial accidents and the like all depend upon ambulance services for prompt medical attention. A prompt ambulance service results in lives saved.

Two critical factors in saving lives are time and speed. These elements are a function of the availability of ambulance crews and their location. The main focus of this research is on the availability of ambulance crews and their location in Vancouver.

During the past five years ambulance services in the Greater Vancouver Regional District (G.V.R.D.) have changed dramatically. The changes include:

- The takeover of the responsibility for the provision of emergency ambulance service by the Provincial Government in January of 1975,
- a significant increase in the demand for ambulance services,

- the expansion of the service area, and
- the introduction of paramedic units.

These new circumstances have brought about a need for the examination of the operational aspects of the ambulance service.

After the takeover by the Province, the charges for ambulance service were significantly reduced to $5.00 per call and recently increased to $15.00 per call. This change in structure explains, in part, the increase in demand from 61,700 calls in 1974 to 78,000 (estimated) calls in 1977. With an average of over 200 calls per day ambulance availability is low. New service areas were also added, thus requiring that a larger geographical area be serviced.

Another change within the ambulance service is that paramedic units are now being utilized. However, their function lies mainly in attending to cardiac arrest victims.

1.1.1 SCOPE AND OBJECTIONS OF RESEARCH

The primary objective of this research is to study the operation of the ambulance system. We shall focus on operational policies which affect the system's ability to respond to calls, especially to emergency calls.

Analytical models are used in conjunction with simulation experiments to examine alternative ambulance systems.

The following sub-objectives were chosen:

- to determine the effects on ambulance effectiveness of changes in the number of ambulances and their location, and

- to determine how to effectively utilize the services of paramedic teams.

1.1.2 RESEARCH OUTLINED

Section 1.2 of this chapter reviews the recent literature on ambulance system models. Included is a brief discussion on the relevance of these models to the situation in the G.V.R.D.

Chapter 2 looks at the existing ambulance services in the G.V.R.D. and discusses the data gathered for the study. Topics discussed are:

- network structure of the G.V.R.D. and travel time data used,

- current number and current locations of ambulances,

- current demand for ambulance services,

- temporal and spatial distribution of calls,

- distribution of call destinations,

- current "response time" levels, and

- ambulance utilization.

In Chapter 3 the simulation model used for the study is described. Topics include:

- the need for a simulation,
- the description of the simulation model,
- assumptions made, and
- validation of the model.

The fourth chapter presents the ambulance location models, in particular the following subjects are examined:

- the essential features of the location problem,
- two analytical methods used for the study, and
- the evaluation of the methods used.

Chapter 5 comprises experiments with the analytical models and the simulation. The statistical output from the simulation is discussed.

In the last chapter, Chapter 6, the experimental results are presented along with implications for policies. The chapter presents recommendations with respect to:

- optimal locations,
- the use of paramedic teams,
- variations in response times by municipality and type of call;
- alternative dispatch rules,
- sensitivity of response times to changes in demand, and

1.1.3 FUNCTION OF THE AMBULANCE SYSTEM

Ambulance systems have five principle functions: rescue, life support, preliminary emergency care, transport to emergency health care facilities, and treatment at emergency facilities. (Functions of the paramedic ambulance service will be discussed later.) Although responsibility for these functions lies mainly with the ambulance service, police and fire departments compliment the ambulance service.

With an effective ambulance system, many lives are saved. This can be illustrated by a simple numerical calculation. Currently there are approximately 78,000 calls per year for ambulance services in the G.V.R.D. About 25% of all calls are emergency calls. Of these, roughly 2% involve life sustaining circumstances. Therefore some 390 people per year depend on the ambulance service to save their life.

1.1.4 THE ROLE OF PARAMEDIC UNITS

Paramedic teams are a group of specialized personnel trained to provide emergency service and first aid to victims who are in near-death circumstances. Such cases include:

- suspected heart attack,
- sudden collapse,
- major trauma,
- drowning,

- obstetric problems, and
- unconsciousness.

(Cardiac arrest cases furnish the majority of calls the paramedic units attend to.)

In these circumstances victims may require shock treatment, oxygen, drugs or minor surgery in order to stay alive. Ordinary ambulance personnel do not have the necessary qualifications or carry the special equipment (e.g. defibrillation machine) to properly treat the victim.

When a life-threatening situation occurs and a paramedic unit is dispatched, on arrival the paramedics evaluate the patient's conditions and in appropriate instances communicate with a physician. For cardiac arrest cases, the paramedics may perform cardiac pulmonary resuscitation. Early therapy of circulatory arrests has proven to result in many saved lives. Heart attack victims may also be placed under electrocardiographic monitoring. Drugs are administered if necessary. For other life-threatening circumstances minor surgery may have to be performed or oxygen may have to be given to the patient.

Special extraction equipment may also be carried by paramedics to get victims out of "hard to get places" such as from automobile wrecks.

Members of the paramedic team can also provide valuable assistance in the emergency ward.

Summarizing the functions of a paramedic team we have:

- rapid response,
- resuscitation from cardiac arrest,
- early therapy of acute trauma and other life-threatening emergencies, and
- transport of the patient to the hospital.

1.2 REVIEW OF AMBULANCE SYSTEM MODELS

1.2.1 REVIEW OF RECENT LITERATURE

Ambulance systems' models in the literature can be classified into two categories. There are those that examine the queuing aspect of the problem and those that focus upon the location problem.

With respect to the queuing aspect of the ambulance system both analytical queuing models and simulation models have been developed and used. Fitzsimmons (2) developed both a queuing model and a simulation model to estimate activity levels. The queuing model predicted response time was distributed for an actual operating system and the mean response time was iteratively improved by the reallocation of the ambulances on the basis of a pattern search method. Global optimality could not be proven.

Savas (24) studied New York City's emergency ambulance service. In the referenced system many ambulances were being dispatched from the hospitals. By using simulation he showed

that a substantial improvement in mean response time could be attained by dispersing ambulance depots away from hospitals and throughout the service area.

Chaiken and Larson (4) examined a number of urban emergency service systems and their operational problems. They suggest policies for: locating units, allocating numbers, designing response areas, and reallocating units. Their policies are basically derived from queuing models.

A study by Swoveland et al (29) uses simulation and a probabilistic branch and bound procedure to allocate ambulances. Output from the simulation provided information on system characteristics under various ambulance assignments. The output was used to construct the objective function for the optimal location problem. Simulation was in turn used to verify solutions from the branch and bound procedure.

Kolesar (16) has developed square root laws for determining the number and location of emergency unit depots. His conclusion rests on the assumptions of low demand rate, and on arrival rates and service times being Poisson and exponentially distributed, respectively. The model can be used to give quick and rough approximations of response times.

Literature on location analysis is abundant, especially in the private sector. (I.E., Weber (35), Cooper(6), Kuhn

and Kuenne (18), etc.)[1] The warehouse problem is a classic example. Procedures to handle this problem have been applied to public facility location problems. Public facility location problems differ in that their objective function is to maximize the benefit to society as opposed to minimization of costs.

Recently, most interest work in public sector facility location analysis has been done by Revelle (19, 20, 21), Church (5) and Toregas (32, 33, 34). They have studied the use of the p-median, set covering, and maximal covering procedures in locating ambulance depots. Although these models ignore the dynamic nature of the problem they do give an optimal solution to the static problem. Using the maximal covering model one can find the trade-off between the level of service and the number of ambulances used.

For the ambulance problem the p-median model can be formulated so as to minimize average response time. Minimization of average response time is one of the major goals of the ambulance system in the G.V.R.D. The maximal covering model finds the maximum number of people serviced within a specified maximum response time by the ambulance system. By experimentation one can find the trade-off between the level of service and the number of ambulances needed.

---

[1]For an excellent bibliography on plant location see "Plant Location Family" Working Paper No. WP-12-77, Krarup, Jekob, Pruzan, and Mark, University of Calgary.

The use of simulation (stochastic) models in conjunction with linear programming (deterministic) models would appear to be a good approach in solving ambulance system problems. Berlin and Leibman (2) have previously attempted such an approach where they used the set covering model to locate depots and simulation to determine utilization levels. A p-median model incorporates more information than a set covering model and, therefore, should give better initial solutions.

1.2.2 APPLICABILITY TO THE G.V.R.D.

Most of the queuing models ignore the problem of congestion (i.e. a call arrives while the ambulance station closest to it is empty). They work best, therefore, when demand rates are low. Many models utilize theoretical distributions which may not correspond to observations.

The combination of mathematical programming models to provide "good" initial solutions, and the employment of a simulation to test for optimality and provide data necessary for local searches, is called for in systems characterized by congestion problems. The G.V.R.D. provides an excellent example where such a combination of techniques is needed.

CHAPTER 2

2.1 EXISTING AMBULANCE SERVICE

At present, the emergency ambulance system services the total G.V.R.D. area. This district consists of the lower mainland extending east to Aldergrove. The municipalities serviced by the system include: Vancouver, North Vancouver, West Vancouver, Burnaby, New Westminster, Richmond, Delta, Surrey, Langley, White Rock, Coquitlam, Port Coquitlam, Port Moody, Pitt Meadows and Maple Ridge. The total population of the area is about 1,172,000. During 1977 about 78,000 calls were serviced by the ambulance system.

2.1.1 THE NETWORK

The physical layout of the region can be described from the ambulance system's point of view by a connected network of locations and travel times between nodes in the network. In this study we have adopted the regionalization scheme of the G.V.R.D. Board, dividing the area into 169 subregions. Population centroids of each subregion mark the nodes on the ambulance network. A matrix giving travel times between all

pairs of nodes in the system (the travel time matrix) was obtained from the G.V.R.D. Board. These times represented average travel times during the "morning rush hours". To examine the appropriateness of the information contained in the matrix we have compared it to travel times obtained from ambulance travel records (24-hour "a day records"). Travel times within each municipality were very similar. However, significant differences were found between the two sources of data for travel times between municipalities. Using the G.V.R.D. Board data (slower travel between municipalities) for planning ambulance services amounts to taking a more conservative service posture. Clearly ambulance allocations for periods of congested traffic will make the service more equitable among municipalities.

2.1.2 RESOURCES

In the beginning of 1978 the ambulance service operated thirty-one cars, two of which were equipped as "paramedic units". Recently a third paramedic car has been added to the fleet. In addition to the fleet of the provincial ambulance service there are several other ambulance services which have locations at U.B.C., Port Coquitlam and Port Moody. There are also volunteer services located in West Vancouver. These peripheral services account for less than 2% of the total

calls serviced in the lower mainland.

2.1.3 COORDINATION AND MANAGEMENT

The takeover by the provincial government in 1974 led to the adoption of a central dispatch system for the entire district. (Previously services were "Balkanized".) The service operates 18 cars, 24 hours a day, nine additional cars are operated from 7:30 a.m. to 6:30 p.m., and five cars from 12:00 a.m. to 11:00 p.m. Three of the cars operating 24 hours a day are paramedic units. The cars, which provide a continuous service, have two shifts from 8:00 a.m. to 6:00 p.m. and from 6:00 p.m. to 8:00 a.m. The other vehicles operate during a one 11-hour shift (see table 2.1).

2.2 DEMAND FOR AMBULANCE SERVICES

2.2.1 DATA COLLECTION

The data used to analyze patterns of current demands and production components was obtained from dispatching forms which record the movements of each ambulance while in service. The periods covered are from August 15 to August 24, 1977, and from September 1 to September 20, 1977. The specific information obtained for each call includes:

- origin (location of patient),

TABLE 2.1

CURRENT LOCATION OF AMBULANCES IN GVRD AND SHIFTS OPERATED

Station	Location	# of Cars	Shifts 0800-1800 1800-0800	0730-1830	1200-2300
G1	Vancouver	4	1 car	1 car	1 car
G2	Vancouver	1	1		
G3	Vancouver	1	1		
G4	Vancouver	1	1		
G5	Vancouver	1	1		
G6	Burnaby	3*	1*	1	1
G7	New Westminster	3*	1*	1	1
G7B	New Westminster	1		1	
G8	Vancouver	3	1	1	1
G9	Surrey	2	1	1	
G10	Richmond	3	1	1	1
G11	Ladner	1	1		
G12	Langley	1	1		
G14	White Rock	1	1		
G15	West Vancouver	1		1**	
G16	North Vancouver	2	1	1	
G17	Haney	1	1		
G18	Vancouver	1	1		
G19	Coquitlam	1	1		

* one paramedic car located among these

** operates from 0700-1800

NOTE: This table does not include the non-provincially operated ambulances.

- destination of travel (location to which patient is to be delivered),
- sex and age of the patient,
- priority rating of the call (i.e. transfer, non-urgent, emergency or paramedic call),
- time and date the call was received,
- time the ambulance left in response to the call,
- time the ambulance arrived at the scene,
- time the ambulance left the scene,
- time the ambulance arrived at its destination,
- time the ambulance left for its station,
- time the ambulance arrived at its station,
- time of cancellation of the call (if applicable).

2.2.2 TABLE OF DEFINITIONS

ANU (Ambulance Not Used): The situation in which the ambulance responds to the scene but is not used for transport (first aid may or may not be administered).

Ambulance Utilization: The percent of time an ambulance is actually servicing a call.

Cancelled Call: A call for an ambulance which is cancelled before the ambulance reaches the scene.

High Priority Call (Code 3 or 4): Calls in which the siren is used en route to the scene.

Loading Time: The amount of time spent at the scene (time spent applying first aid, moving the patient into the ambulance, etc.).

Non-Primary Response: The situation in which a call is serviced by some ambulance other than the one whose depot is closest to the scene of the call (this may occur when the latter is busy and therefore unavailable to respond).

Response Time: The elapsed time between arrival of the call by the dispatcher and the arrival of the ambulance at the scene.

Service Time: The elapsed time between arrival of the call by the dispatcher and the time the ambulance clears the destination.

Start-Up Time: The amount of time between the receipt of the call by the dispatcher and the time the ambulance departs for the scene.

Transfer Call: A non-urgent (often scheduled) call in which a patient is transported between two points (e.g. from hospital to home) and does not receive first aid treatment.

Unloading Time: The elapsed time spent at the point to which the patient is transported (generally, time spent admitting patients into emergency wards).

2.2.3 CURRENT DEMAND

A. TEMPORAL VARIATIONS

i) Seasonal Variations:

Analysis of monthly demand levels for 1976 and 1977 indicates that the demand in the month of December is the largest, probably due to the high stress and activity that the holiday season induces. Otherwise, no pattern in month to month variations was discovered.

ii) Hourly Variation of Demands During the Day:

Figure 2.1 provides average call rates by time of day for the total

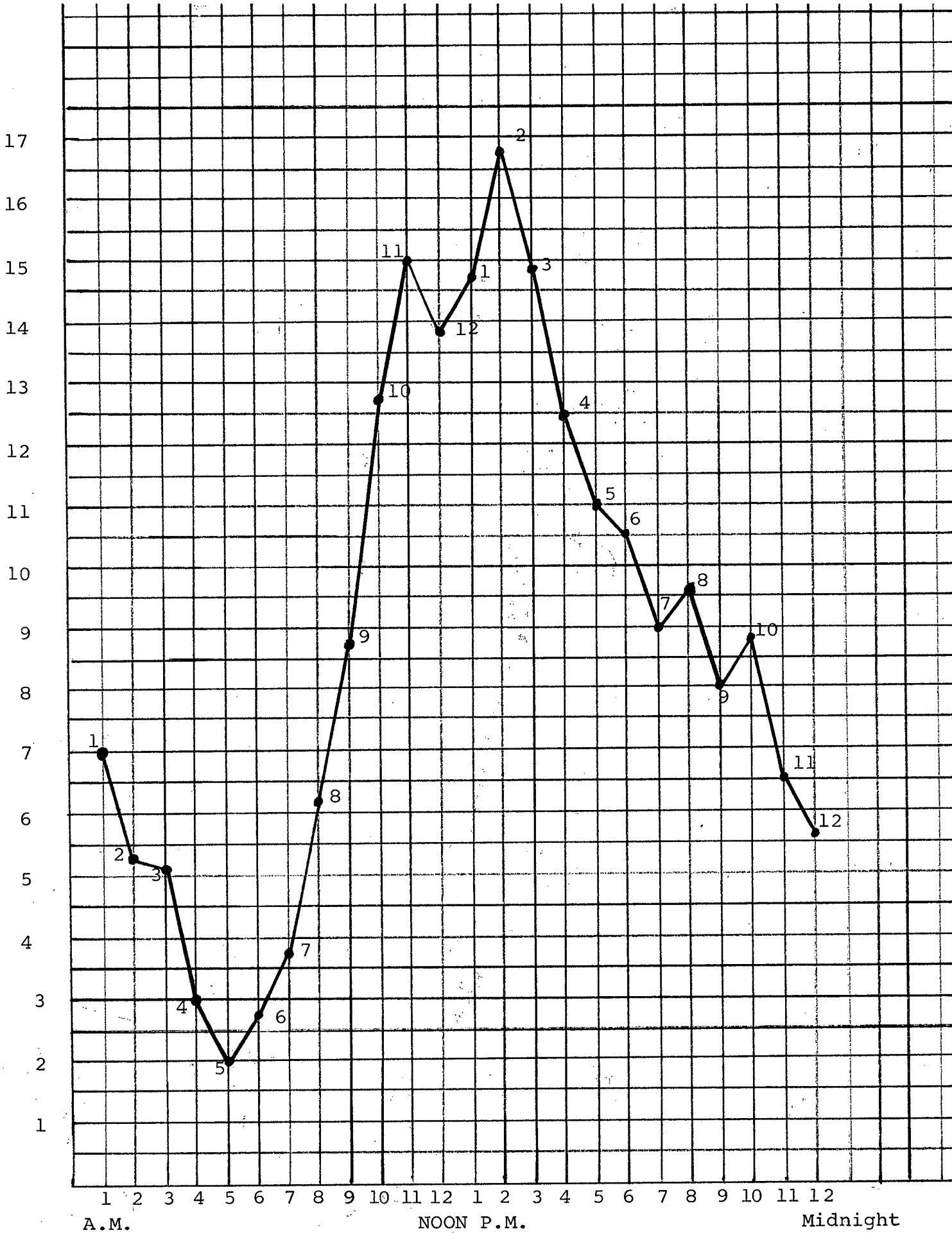


FIGURE 2.1

Average Call Rate (calls/hour) by Time of Day

region. The lowest demand for the day is approximately at 5:00 a.m. From 5:00 a.m. until 11:00 a.m. there is a sharp and steady rise in the number of calls. Demand for transfer calls seem to create a local peak around 11:00 a.m. Following a small decline, demand level increases to its global peak at approximately 2:00 p.m. From that time on, demands generally decline.

B. DEMANDS FOR PARAMEDICS

At the present time there are three paramedic cars, with the third car being added only recently. The two original paramedic cars are stationed in Burnaby and New Westminster. These two cars were initially introduced to serve those calls requiring special procedures for which they were equipped. However, the heavy general demand for ambulance services required their deployment for all types of demands.

A sample of their service records obtained for February 1978 revealed that the cars responded to an average of 19 calls a day. These calls consisted of 9% transfer calls, 40% non-urgent calls and 51% emergency calls. Two to three calls a day (all emergency calls) required the use of the specialized equipment of the paramedic units and the specialized training of their attendants.

2.2.4 DEMAND GROWTH PATTERN

The variation in hourly call rates for each municipality are provided in table 2.2. The day period 8:00 a.m. to 6:00 p.m. is the highest demand period in all municipalities. Comparing current day rates with those reported four years ago, one notes that demands in Vancouver increased less rapidly (about 50%) than demands in New Westminster, Richmond, Delta, Surrey,

Municipality	Average # calls/hour 12M-8AM	8AM-6PM	6PM-12M	Average Number of calls/day
Vancouver	2.55	6.25	4.87	112.17
North Vancouver	0.23	0.54	0.37	9.46
West Vancouver	-	0.23	0.18	4.00
Burnaby	0.42	1.02	0.80	18.33
New Westminster	0.22	0.78	0.49	12.42
Richmond	0.16	0.70	0.43	10.92
Delta	-	0.24	0.12	3.62
Surrey	0.30	0.58	0.60	12.75
Langley	0.15	0.25	0.19	4.86
Coquitlam	0.11	0.23	0.17	4.25
Pitt Meadows	-	-	-	-
Maple Ridge	-	0.20	0.12	3.33
White Rock	-	0.22	0.09	3.21
Port Coquitlam	-	-	-	0.96
Port Moody	-	-	-	-

For missing values, sample less than 10 calls.

TABLE 2.2 Daily Ambulance Demand Rate by Time of Day and Municipality

Coquitlam and White Rock (60% - 100% increases). These relative changes represent recent shifts in the spatial distribution of population in the lower mainland.

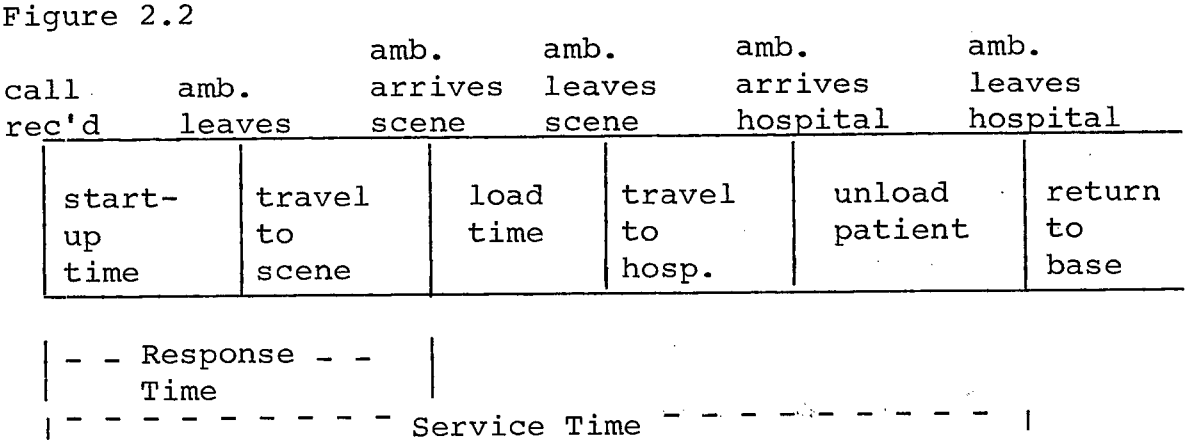
2.3 PRODUCTION AND EVALUATION

2.3.1 PRODUCTION COMPONENTS

Emergency ambulance service (EMS) is one element of the overall emergency medical care system. The focus of this research, is on the transportation function of an EMS system and, in particular, the response time of ambulances used solely for ground transportation of patients. Other functions of the EMS such as treatment and rescue add to the quality of service provided but are not examined here.

Response time, a key attribute of service quality, is calculated as the time the call is received by the dispatcher to the time an ambulance arrives at the scene. (See figure 2.2.)

Figure 2.2



2.3.2 ANALYSIS OF COMPONENTS

The average response time per time period per municipality is shown in tables 2.3 and 2.4. For most municipalities, the average response time is greatest during the middle of the day and is smallest during the morning hours of the day. These figures are highly correlated with the number of calls during the same period. Congestion of demand is clearly the dominant problem of the system as it operates now.

In Vancouver, Burnaby and New Westminster, the response times, compared to four years ago, have all significantly increased. Longer response times in Richmond, Delta, and Langley reflect the low population density of the area (hence, resulting in a longer travel time within subregions).

Also given in tables 2.3 and 2.4 is the average service time by municipality. There is little variation in these times. Richmond, Delta, Langley, Pitt Meadows and Port Coquitlam all have high service times due to the long travelling distances. Service times in Burnaby and New Westminster include paramedic calls which are usually longer than normal calls. Vancouver's long service time can be explained, in part, by the high number of transfer calls which take longer to service (about an average of 70 minutes, see table 2.5).

Figures 2A, 2B, 2C, 2D, 2E, and 2F give frequency and

Municipality	Average Response Time* (in minutes) Time of Day 12M-8AM	8AM-6PM	6PM-12M	Overall (24 Hours)	Average Service Time
Vancouver	8.8	14.7	9	12.0	47.2
North Vancouver	8.9	11.2	8.5	10.0	44.2
West Vancouver	-	11.8	9.9	11.3	50.6
Burnaby	11.0	16.4	13	14.5	51
New Westminster	11.2	18.6	18.8	17.7	55.5
Richmond	12.3	15.6	26.6	17.7	57
Delta	-	17.7	13.5	15.8	60.4
Surrey	12.1	12.2	10.4	11.7	46.8
Langley	15.5	24.9	15.9	20.8	67.5
Conquitlam	10.4	11.3	10.2	10.8	48.2
Pitt Meadows	-	-	-	-	-
Maple Ridge	-	13	10.2	11.6	57.1
White Rock	-	10	14.8	10.9	43.4
Port Coquitlam	-	-	-	21.4	65.8
Port Moody	-	-	-	-	-

TABLE 2.3 Average Response Time and Service Time by Municipality
(For missing values, sample less than 10 calls)

* does not include transfer calls having a zero response time.

Municipality	Average Response Time* (in minutes)				
	Time of Day			Overall	Average
	12M-8AM	8AM-6PM	6PM-12M	(24 Hours)	Service Time
Vancouver	8.8	14.7	9.0	12.1	47.2
North Vancouver and West Vancouver	9.8	11.4	8.9	10.5	46.2
Burnaby and New Westminster	11.0	17.4	15.3	15.8	52.8
Richmond and Delta	11.3	16.2	23.4	17.1	58.0
Surrey, Langley and White Rock	12.6	14.8	12.0	13.6	50.9
Coquitlam, Port Moody, Port Coquitlam, Pitt Meadows and Maple Ridge	10.8	13.2	12.4	12.6	54.2

TABLE 2.4 Average Response Time and Service Time by Municipality Groups

* does not include transfer calls having a zero response time.

cumulative distributions for response times for the whole G.V.R.D. and for the area of Vancouver, Burnaby, and New Westminster. (The frequency distributions show the percentage of calls taking a certain number of minutes. The cumulative distributions display the percentage of calls in which the response time did not exceed a specific value.) The cumulative distributions indicate that:

- 50% of transfer calls are responded to within 20 minutes
- 84% of all non-urgent calls are responded to within 20 minutes
- 99% of all emergency and paramedic calls are responded to within 20 minutes and further,
- 85% of all calls are responded to within 10 minutes.

In figure 2E one can observe the impact of response times to transfer calls on the total distribution of response times.

Frequency and cumulative distributions for service times are displayed in figures 3A, 3B, 3C, and 3D. Service times for all calls (except transfers) is very similar to emergency calls.

Let us now examine average operation times by type of call and time of day. These statistics are noted in tables 2.5 to 2.8. Table 2.5 contains data derived from observations for the total G.V.R.D. Average start-up times are lowest in

the morning period and highest during the middle part of the day when the large majority of calls occur. The high average start-up time for transfer calls is due to ambulance service servicing more urgent calls first (transfer calls are low priority). Response times for all categories is highest during the middle part of the day when the call load is heaviest. Average service times transfer calls have the highest average service times.

Table 2.6 has statistics for Vancouver, Burnaby, and New Westminster. The average response times for normal and emergency calls are lower than for the total G.V.R.D. and indicate that the outlying regions have a larger response time due to longer travel distances.

In table 2.7 the paramedic calls are excluded from the data. Note that the average response time for emergency calls is 5.96 minutes (5 minutes and 58 seconds).

Table 2.8 gives figures for average operation times for weekdays only. Average response times are higher than the average for all days. This was expected as weekdays furnish the periods of heaviest demand for calls.

Variations for average time spent at the scene do not differ significantly between tables 2.5 to 2.8. When paramedic calls are excluded, average time at the scene goes down for emergency calls. This indicates that the paramedic procedures take longer to perform than the normal call procedures.

	Time of Day			Priority			
	12AM-8AM	8AM-6PM	6PM-12M	Transfer	Normal	Emergency	All Priorities
% of calls	17.8%	55.6%	26.7%	24.4%	51.6%	24.0%	100%
Average start-up times	2.09	7.6	4.4	19	4.9	1.1	5.7
Average response times	10	15.1	11.4	26.5	13.2	6.8	13.2
Average time at scene	10.4	10.2	10	13.2	9.8	9.8	10.2
Average service time	46.2	51.7	48	70.9	47.8	43.7	49.7

TABLE 2.5 Average Operations Times (in Minutes) by Type and Time of Call*
(All Days)

* Startup and Response times do not include transfer calls which had a zero startup or response time.

	Time of Day			Priority			
	12AM-8AM	8AM-6PM	6PM-12M	Transfer	Normal	Emergency	All Priorities
% of calls	18.1%	54.7%	27.1%	24.0%	53.7%	22.3%	100%
Average start-up times	2.2	8.2	4.1	20.1	4.9	1.2	6
Average response time	9.4	15.5	10.6	27.1	12.7	6.2	13
Average time at scene	9.9	10.2	9.7	12.7	9.7	9.3	10
Average service time	44.2	51.7	45.7	69.6	46.8	41.9	48.7

TABLE 2.6 Average Operations Times (in Minutes) by Type and Time of Call*
(Vancouver, Burnaby and New Westminster Only)

* Startup and Response times do not include transfer calls which had a zero startup or response time.

	Time of Day			Priority			
	12AM-8AM	8AM-6PM	6PM-12M	Transfer	Normal	Emergency	All Priorities
% of calls	17.89%	55.19%	26.9%	24.8%	55.4%	19.8%	100%
Average start-up times	2.2	8.45	4.3	20.1	4.89	1.1	6.2
Average response times	9.4	15.7	10.75	27.1	12.7	5.96	13.2
Average time at scene	9.65	10.2	9.49	12.7	9.7	8.8	9.9
Average service time	43.2	51.77	45.4	69.5	46.7	40.4	48.5

TABLE 2.7 Average Operations Times (in Minutes) by Type and Time of Call*
(Vancouver, Burnaby and New Westminster Only. No paramedic calls)

* Startup and Response times do not include transfer calls which had a zero startup or response time.

	Time of Day			Priority			
	12AM-8AM	8AM-6PM	6PM-12M	Transfer	Normal	Emergency	All Priorities
% of calls	15.7%	58.9%	25.4%	28.7%	48.5%	22.8%	100%
Average start-up times	2.1	8.8	4.1	19.9	5.5	1.1	6.6
Average response times	10.3	16.5	11.2	27.6	14	6.8	14.2
Average time at scene	10	10.2	10.2	13.2	9.7	9.5	10.2
Average service time	46	53.1	48.5	72.6	48.5	43.4	50.9

TABLE 2.8 Average Operations Times (in Minutes) by Type and Time of Day*
(Only Weekdays)

* Startup and Response times do not include transfer calls which had a zero startup or response time.

The spatial distributions of call destinations is given in table 2.9 (i.e., where the ambulances go after picking up their patients). From left to right the hospital abbreviations stand for: Vancouver General, St. Paul's, Burnaby General, St. Mary's, Lion's Gate, Royal Columbian, Shaughnessy, Richmond General, Langley Memorial, Maple Ridge, Peace Arch, and other (including Surrey hospitals).). Hospitals whose emergency rooms are staffed to provide virtually any kind of emergency treatment necessary are: Vancouver General, St. Paul's, Lion's Gate and Royal Columbian. Table 2.10 is the complement of table 2.9, giving the distribution of call origins by hospital.

HOSPITALS

Municipality	VGH	SHY	SPH	BGH	SMH	LGH	RCH	SMY	RGH	LMH	MRH	PADH	Othe
Vancouver	46.46	8.36	26.42	2.77	0.20	1.24	0.20	0.64	0.69	0.15	0.1	0.1	12.6
North Vancouver	6.95	0.53	2.67	0.0	0.0	71.66	0.0	0.53	0.0	0.53	0.0	0.0	17.1
West Vancouver	0.0	0.0	2.35	1.18	0.0	83.53	0.0	0.0	0.0	0.0	0.0	0.0	12.9
Burnaby	10.83	2.56	1.14	39.89	0.85	0.0	4.84	25.93	0.0	0.0	0.28	0.0	12.8
New Westminster	6.35	1.19	0.79	3.97	4.37	2.20	18.25	42.86	0.0	1.59	1.59	0.0	19.0
Richmond	25.11	5.73	5.29	0.88	0.44	0.0	0.44	2.20	47.14	0.0	0.44	0.0	10.1
Delta	11.27	2.82	0.0	1.41	23.94	0.0	0.0	8.45	43.66	0.0	0.0	0.0	8.4
Surrey	2.73	1.17	0.39	0.0	56.25	1.02	0.78	16.80	0.0	3.52	0.39	9.77	8.2
Langley	6.12	2.04	3.06	0.0	2.04	0.0	0.0	19.39	0.0	63.27	0.0	1.02	2.0
Coquitlam	1.25	2.50	0.0	0.0	1.25	0.0	6.25	77.50	0.0	0.0	1.25	0.0	10.0
Pitt Meadows	0.0	0.0	0.0	0.0	0.0	0.0	0.0	33.33	0.0	0.0	66.67	0.0	0.0
Maple Ridge	3.03	1.52	1.52	0.0	0.0	0.0	0.0	7.58	0.0	0.0	78.79	0.0	7.5
White Rock	4.55	0.0	1.52	0.0	3.03	0.0	0.0	4.55	0.0	0.0	0.0	75.76	10.6
Port Coquitlam	5.56	0.0	0.0	0.0	0.0	0.0	22.22	66.67	0.0	5.56	0.0	0.0	0.0
Port Moody	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0

TABLE 2.9 Distribution of Ambulance Destinations by Municipality
(Rows sum to 100%)

unicipality	Hospitals											
	VGH	SHY	SPH	BGH	SMH	LGH	RCH	SMY	RGH	LMH	MRH	PADH
ancouver	86.07	82.44	94.51	26.67	2.16	10.46	5.06	3.50	9.21	3.75	3.13	2.56
orth Vancouver	1.19	0.49	0.89	0.0	0.0	56.07	0.0	0.27	0.0	1.25	0.0	0.0
est Vancouver	0.0	0.0	0.35	0.48	0.0	29.71	0.0	0.0	0.0	0.0	0.0	0.0
urnaby	3.48	4.39	0.71	66.67	1.62	1.26	21.52	24.46	0.0	0.0	1.56	0.0
ew Westminster	1.47	1.46	0.35	4.76	5.95	0.0	58.23	29.03	0.0	5.00	6.25	0.0
ichmond	5.23	6.34	2.12	0.95	0.54	2.09	1.27	1.34	70.39	0.0	1.56	0.0
elta	0.73	0.98	0.0	0.48	9.19	0.0	0.0	1.61	20.39	0.0	0.0	0.0
urrey	0.64	1.46	0.18	0.0	77.84	0.0	2.53	11.56	0.0	11.25	1.56	32.05
angley	0.55	0.98	0.53	0.0	1.08	0.42	0.0	5.11	0.0	77.50	0.0	1.28
oquitlam	0.09	0.98	0.0	0.0	0.54	0.0	6.33	16.67	0.0	0.0	1.56	0.0
itt Meadows	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.27	0.0	0.0	3.13	0.0
aple Ridge	0.18	0.49	0.18	0.0	0.0	0.0	0.0	1.34	0.0	0.0	81.25	0.0
hite Rock	0.28	0.0	0.18	0.0	1.08	0.0	0.0	0.81	0.0	0.0	0.0	65.10
ort Coquitlam	0.09	0.0	0.0	0.0	0.0	0.0	5.06	3.23	0.0	1.25	0.0	0.0
ort Moody	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.81	0.0	0.0	0.0	0.0
verage # calls/day	45.46	8.54	23.54	8.75	7.71	9.96	3.29	15.50	6.33	3.33	2.67	3.25

TABLE 2.10 Distribution of Call Origin by Destination (i.e., by Hospital)
(Columns sum to 100%)

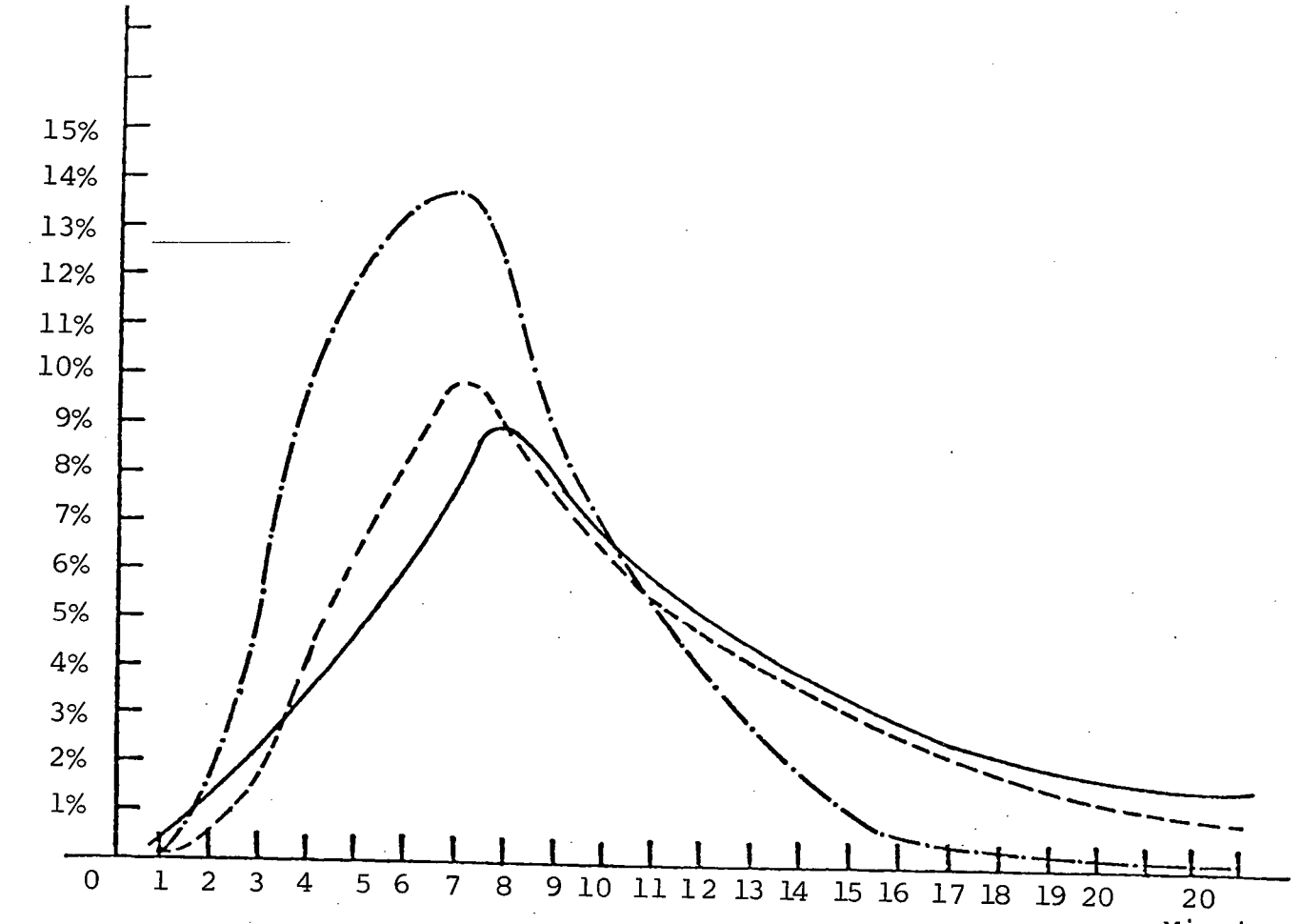


FIGURE 2A Frequency Distribution

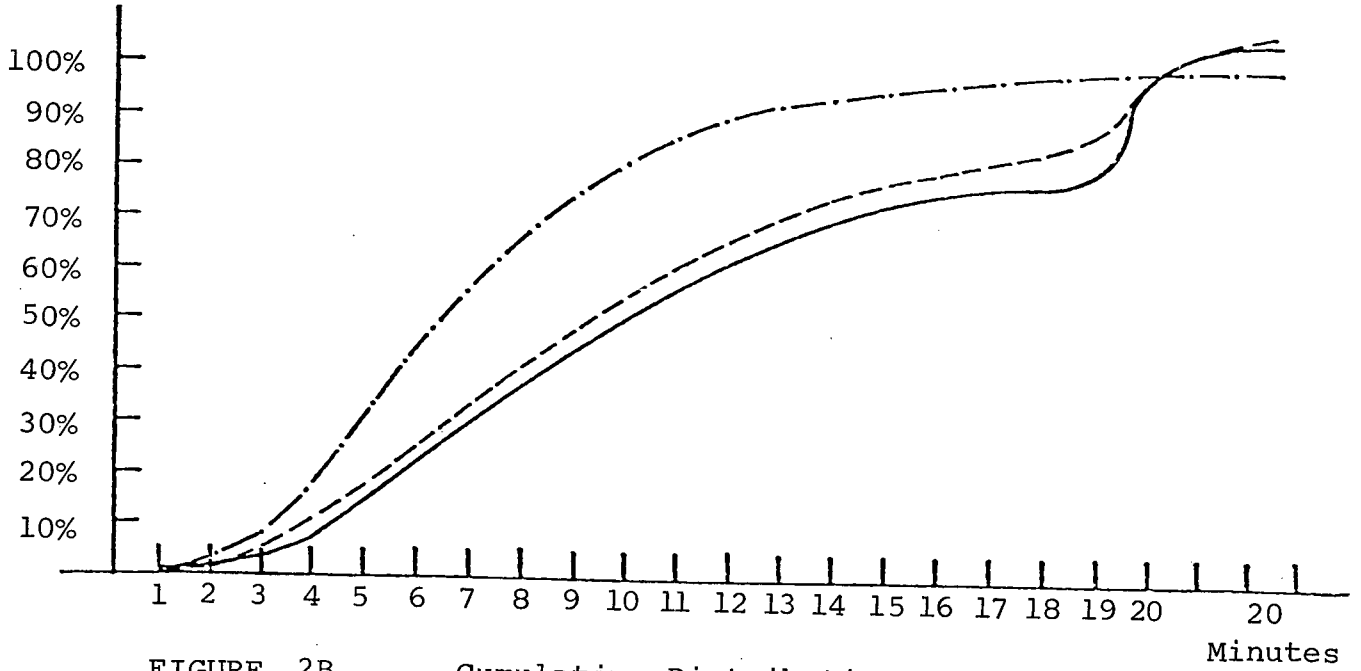
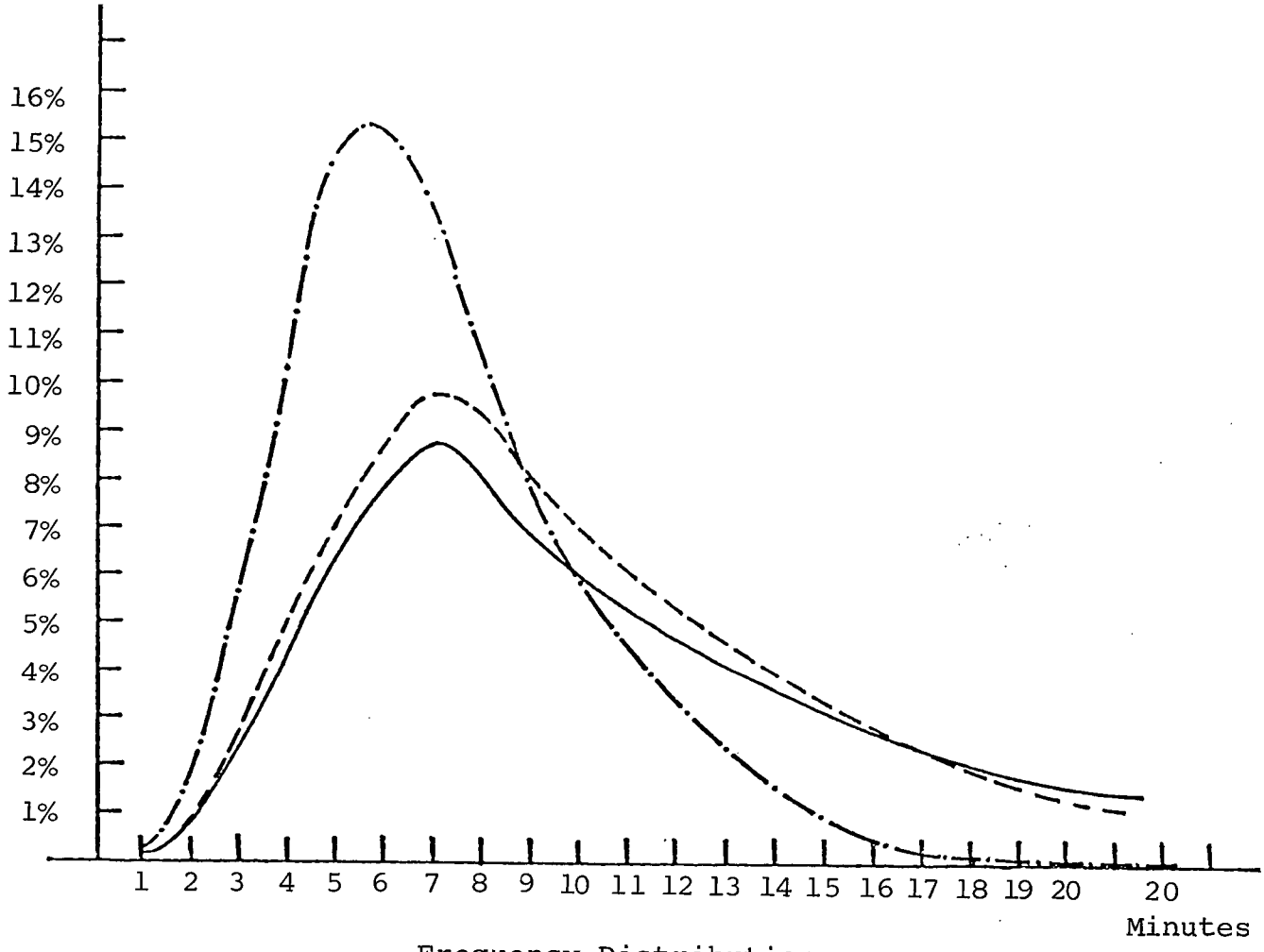


FIGURE 2B Cumulative Distribution

Distribution of Response times
(All Municipalities)

——— all calls.
------ all, except transfer calls.
-.-.-. emergency calls only.



Frequency Distribution

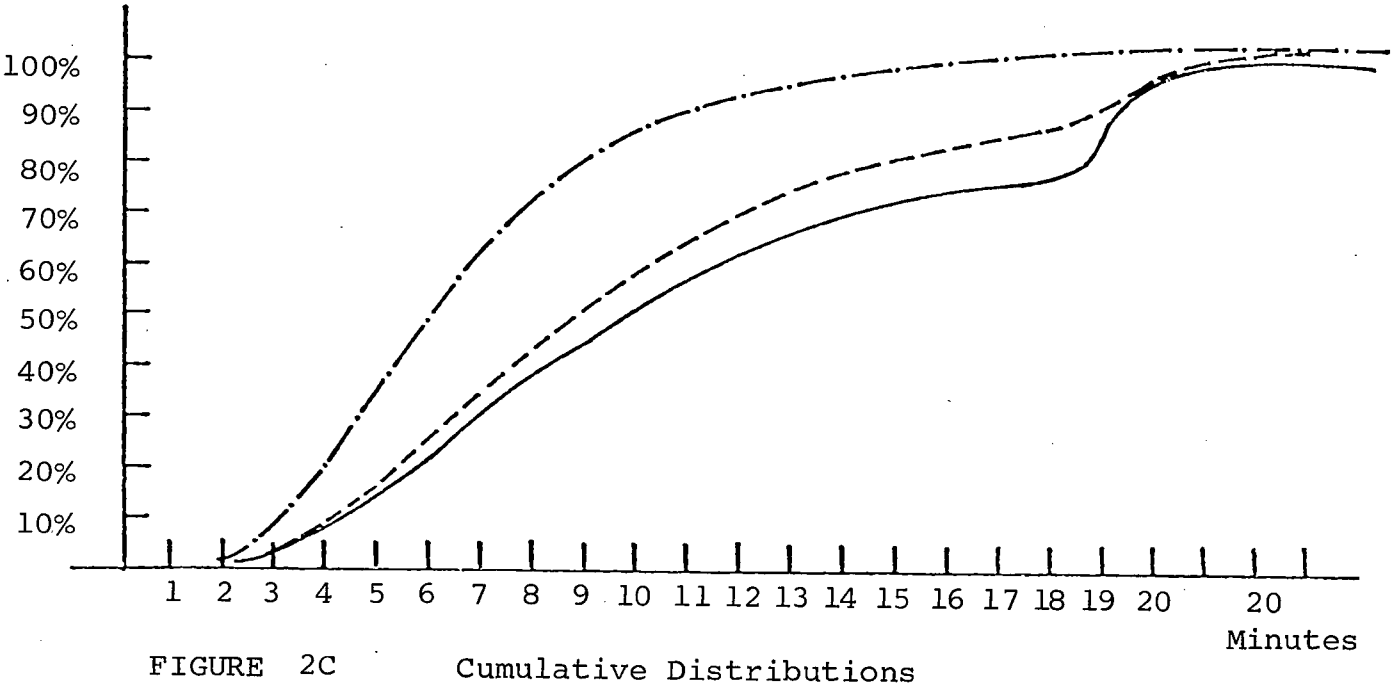
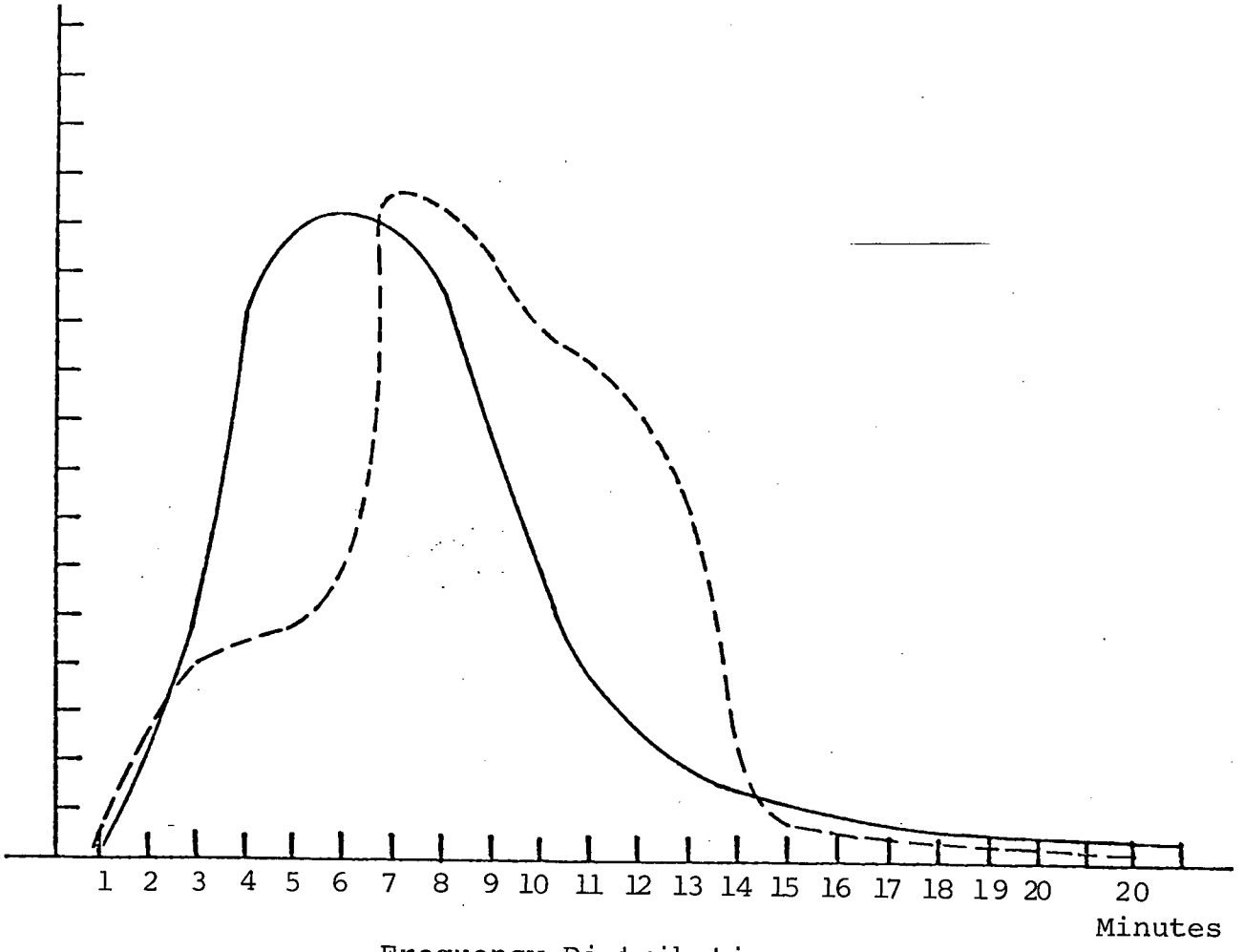


FIGURE 2C Cumulative Distributions

Distributions of Response times (Vancouver, Burnaby and New Westminster only)

——— all calls.
------ all calls, except transfer.
-.-.-. emergency calls only.



Frequency Distribution

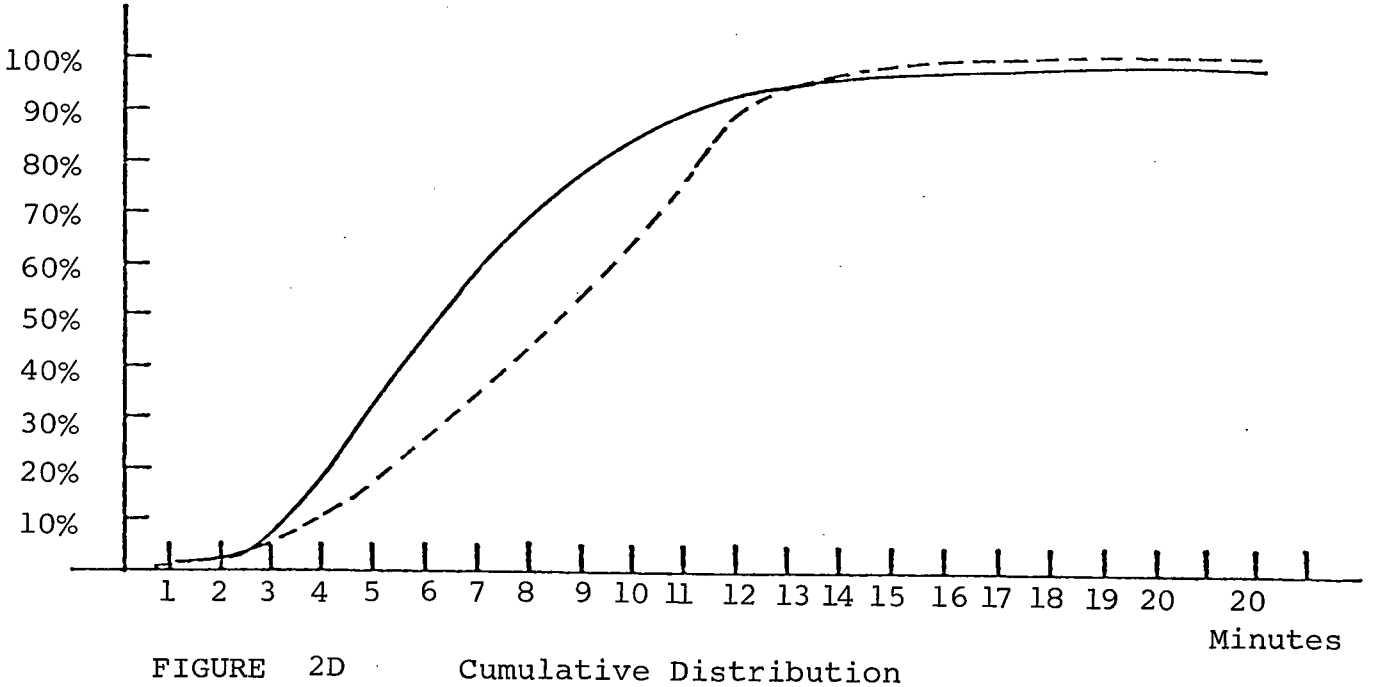


FIGURE 2D Cumulative Distribution

Distribution of Response times
(All Municipalities)

——— emergency calls only

------ paramedic calls only

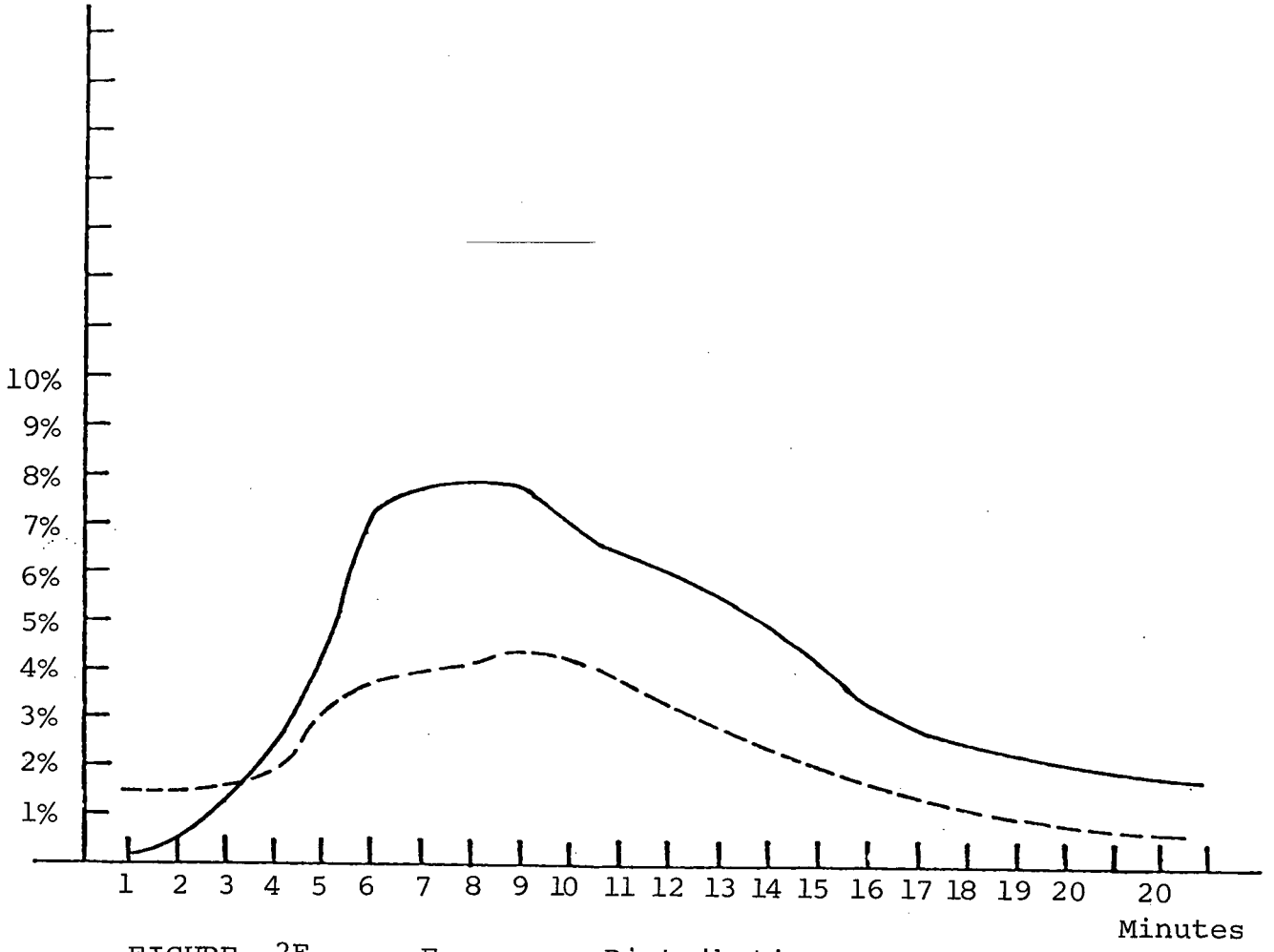


FIGURE 2E Frequency Distribution

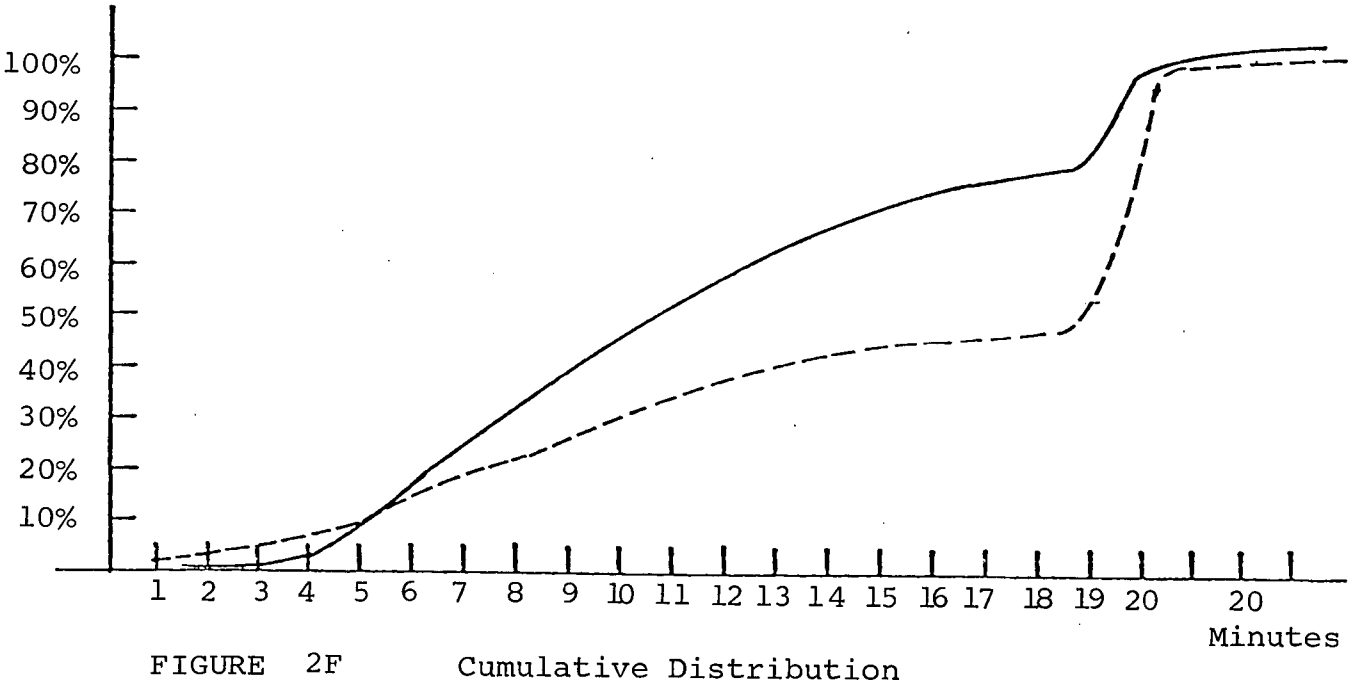


FIGURE 2F Cumulative Distribution

Distribution of Response times (All Municipalities)

------ transfer calls only

——— ordinary calls only.

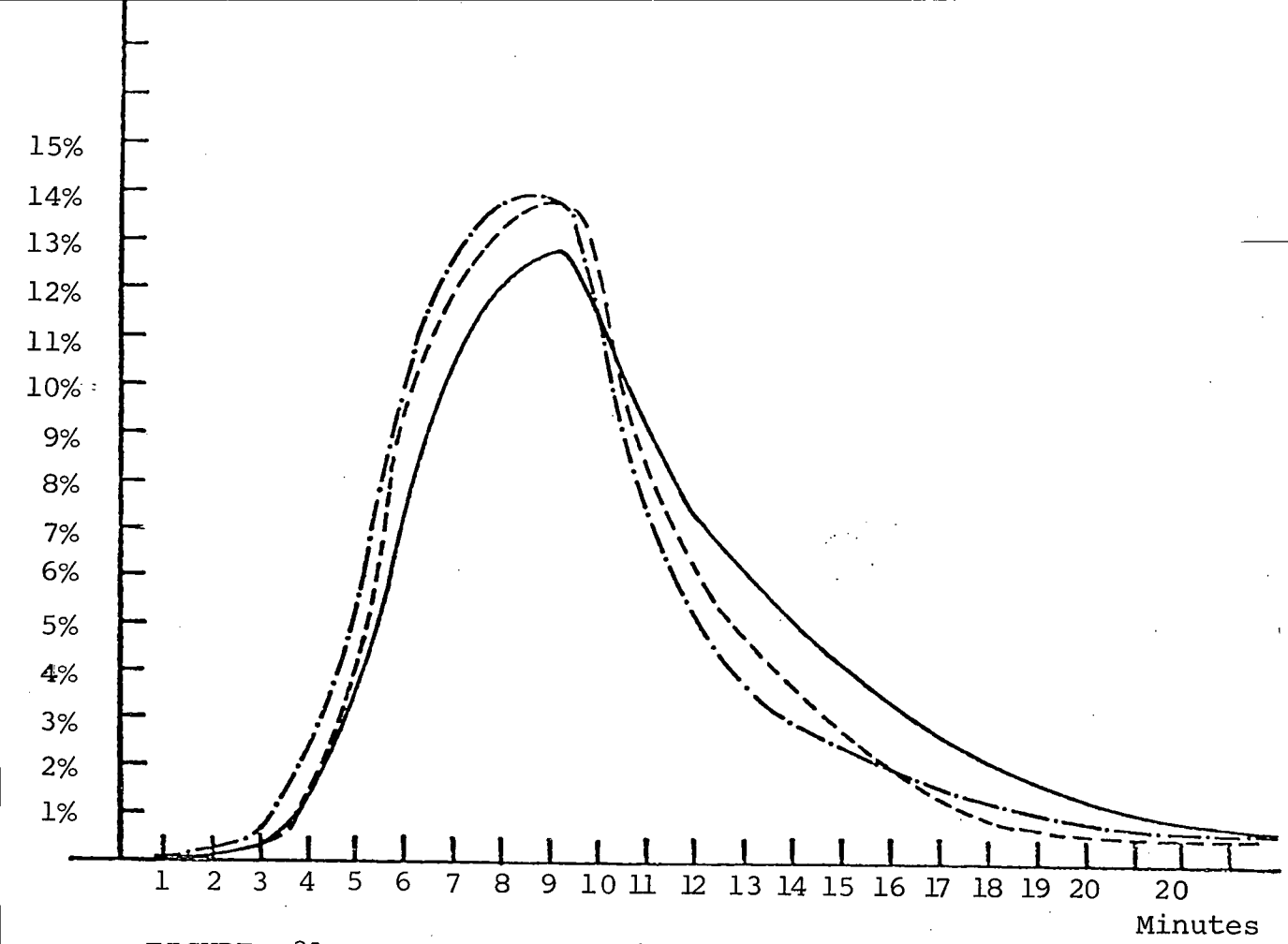


FIGURE 3A Frequency Distribution

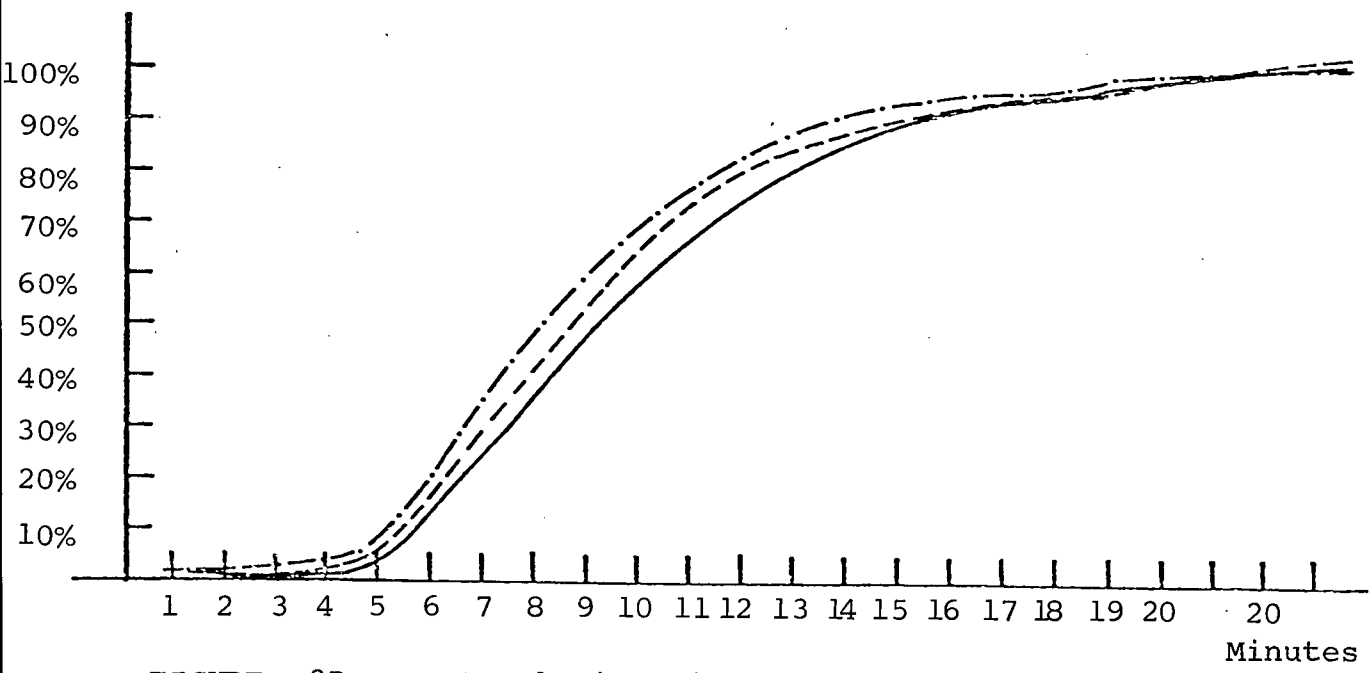


FIGURE 3B Cumulative Distribution

Distribution of Service times
(All Municipalities)

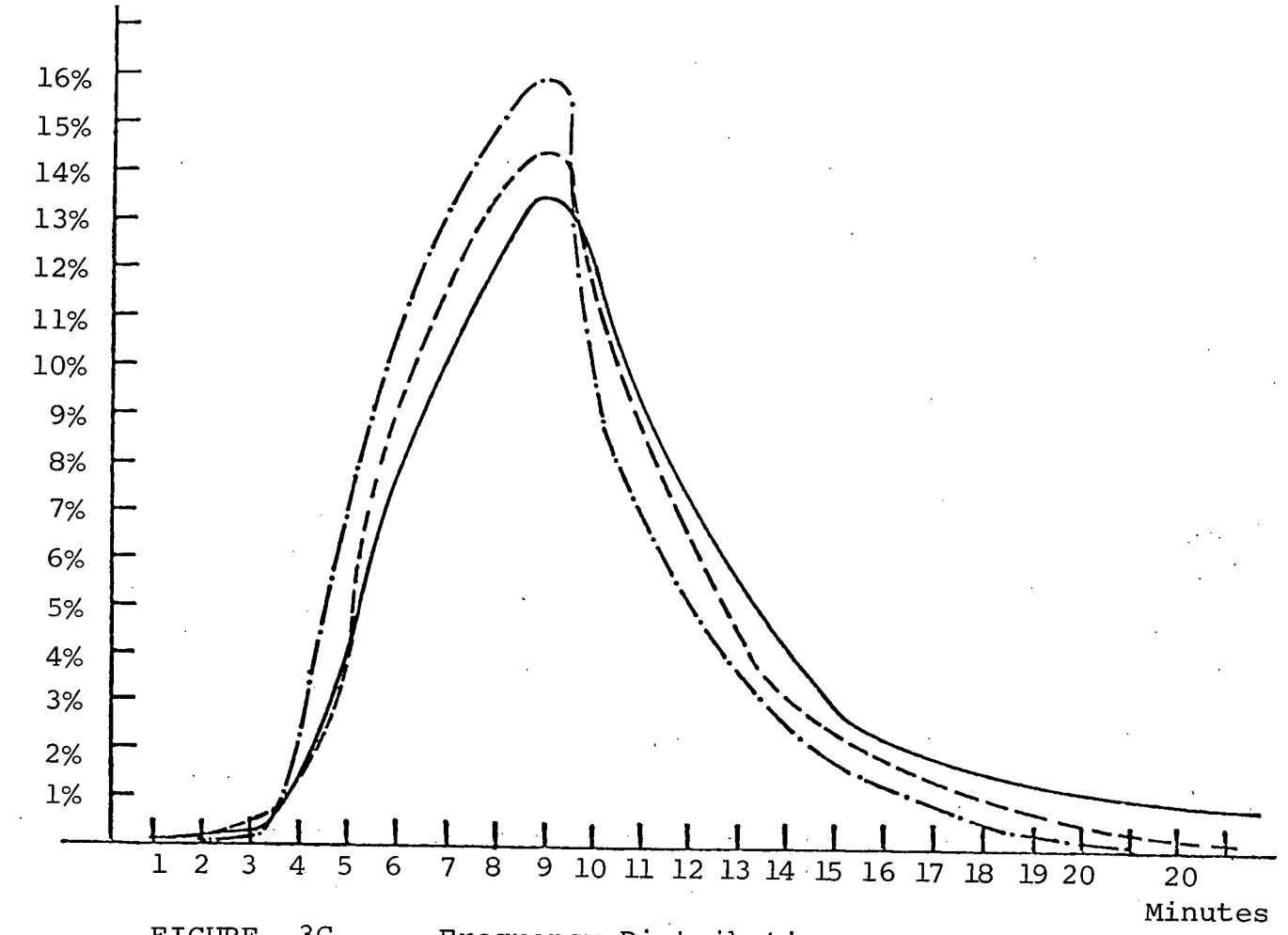


FIGURE 3C Frequency Distribution

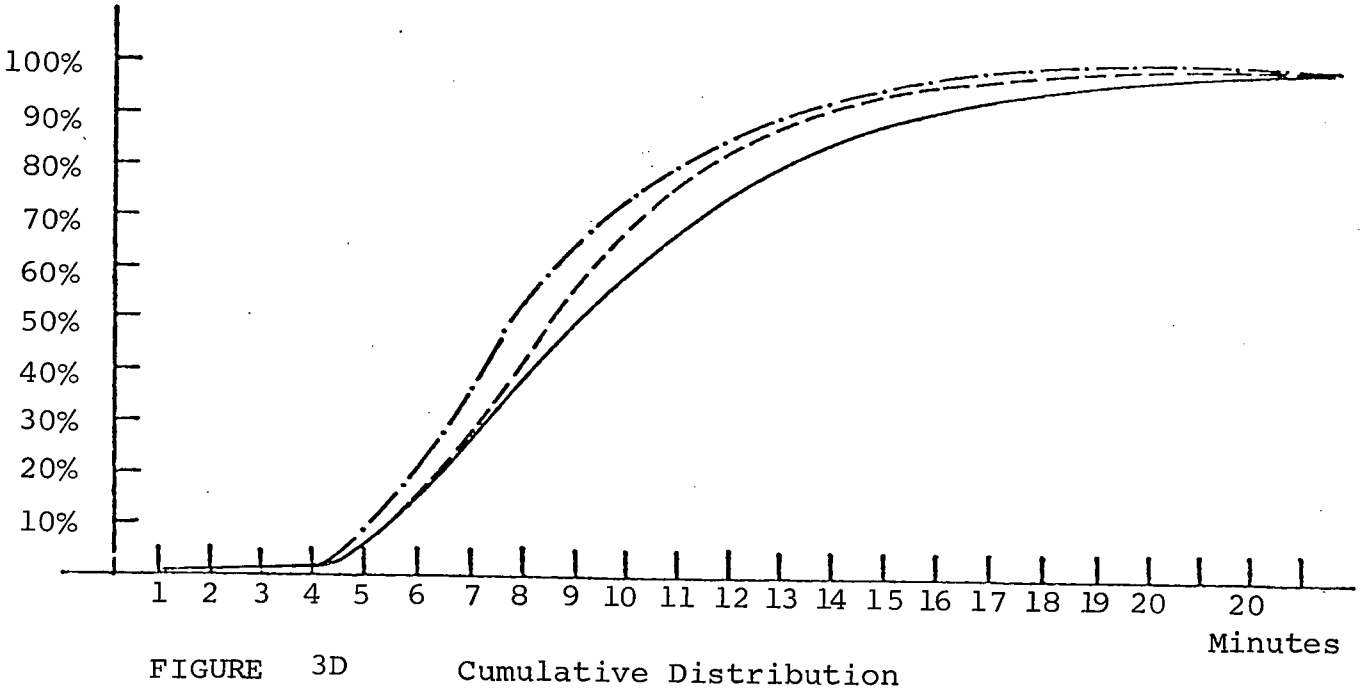


FIGURE 3D Cumulative Distribution

Distribution of Service times (Vancouver, Burnaby and New Westminster only)

——— all calls
------ all calls, except transfer
-.-.-. emergency calls only

CHAPTER 3

THE COMPUTER SIMULATION MODEL

3.1 USE OF SIMULATION IN THE ANALYSIS OF EMERGENCY AMBULANCE SYSTEMS

One important evaluation criterion of an ambulance system is its response time. There are six major factors which affect response time of ambulances:

- the frequency of calls,
- the geographical distribution of calls,
- the number of ambulances available for service,
- the location of hospitals,
- the location of ambulance depots, and,
- the general policies which govern ambulance services.

The ambulance system is a complex stochastic system with many possible management options. Management of an ambulance system involves an array of interdependent decisions concerning resource inputs to the system and their deployment. As the performance of an ambulance system is multi-faceted it is difficult to capture the complexity of the system through

simple analytical models. Computer simulation offers an alternative method. Through computer simulation it is possible to obtain predictions of alternative demand contingencies and policy responses. The simulation offers a "laboratory world" in which alternative policies can be tested instantaneously and changes introduced without the high costs of trial and error in the real world.

3.1.1 THE STOCHASTIC ELEMENTS OF AN AMBULANCE SYSTEM

Secondary responses (i.e., ambulances responding to demands that are regularly served by other ambulances whose depots are closer) resulting from congestion in the system inhibit the use of analytic models. Furthermore, many of the stochastic variables of the system are distributed without correspondance to any known theoretical distribution. Therefore, in the ambulance simulation, empirical distributions are used directly.

The simulation uses as inputs, the empirical distributions of the following variables:

- inter-arrival times of calls (by type and location, and
- loading and unloading times.

3.2 GENERAL DESCRIPTION OF THE COMPUTER SIMULATION MODEL

A GPSS (General Purpose Simulation System) computer

simulation model was built to analyze the behavior of the ambulance system in the G.V.R.D. GPSS was chosen as the simulation language since its elements are homomorphic to the ambulance system. Two basic elements of GPSS are: movement (units of traffic) and service stations (facilities). In the ambulance system model the unit of traffic is the patient and the service stations are the ambulances. A patient or call represents a transaction. A complete transaction is a movement of the ambulance from the location where a demand originates to the destination, and back again to the depot.

Two other basic elements of GPSS are queues and a time clock. Queues in the ambulance model are represented by the following: response times, service times, and time spent waiting for an ambulance to become free. The time clock automatically sequences all events. In an ambulance system the sequence of events for each transaction is similar.

3.2.1 GENERAL OUTLINE

The simulation model consists of two major parts. Part one handles the non-paramedic calls, while part two handles the paramedic calls. Both non-paramedic and paramedic cars can be used to service calls of either type.[1]

---

[1] This policy is under control of the user and the model can be altered so designated cars answer only one type of call.

Let us first consider the operation of part one of the model. A non-paramedic call enters the model through the job-tape, and is routed to part one where a check is made to see if any ambulances are available for service within a radius of 30 minutes travel from the scene. The closest non-paramedic ambulance is designated to service the call. If no ordinary ambulances are available a search is made to see how many paramedic ambulances are available. If no ambulances can service the call the transaction is queued by priority (emergency, normal, and transfer) and by call order.

For the ambulance designated to service the call, the travel time to the scene is determined. If the ambulance was at the depot at the time the call was received, a start-up time is added to the travel time. For emergency calls the travel time is multiplied by a constant (<1) re speed-up to indicate that it will be using the siren while en route to the call. While travelling to the scene the call may be cancelled. Cancellation may also occur once the ambulance arrives at the scene. In either case, the call is terminated and the ambulance checks for other calls to be serviced. If the call is cancelled after the ambulance arrives at the scene its response time is calculated.

Once the ambulance arrives at the scene the patient is loaded and transported to a specific destination (e.g., the hospital) where the patient is unloaded. Response time is

calculated when the car arrives at the scene and service time is calculated after the patient is unloaded. After unloading the patient the ambulance checks for other waiting calls. Waiting paramedic calls are serviced first, then emergency calls, then normal calls and finally transfer calls. If there are no calls waiting to be serviced, the ambulance travels back to the depot and waits for the next call.

When a paramedic call is processed through the model it is routed to part two of the model. Since both[1] paramedic and non-paramedic cars may answer these calls, part two is split into a stream for paramedic cars and a stream for non-paramedic cars. The following decision rules are used:

1. Contingency: there is an ambulance of each type available response -

   (A) If an ordinary ambulance is closer it is dispatched. After it arrives a decision is made whether to

      I. cancel the ordinary ambulance and wait, if necessary, for the paramedic ambulance; paramedic ambulance carries patient, or

      II. cancel the paramedic ambulance; ordinary ambulance carries patient.

   (B) If the paramedic ambulance is closer, it is dispatched and it carries the patient.

2. Contingency: ordinary but no paramedic ambulance

---

[1]The model builder can control the policy regulating which cars can answer paramedic cars via a test block.

is available

Response: dispatch ordinary ambulance and queue for a paramedic one

(A) If an ordinary ambulance arrives before a paramedic ambulance is sent - cancel the call for the paramedic ambulance and let the ordinary one carry the patient.

(B) If a paramedic ambulance is sent before an ordinary arrives and

I. paramedic ambulance arrives first - cancel the ordinary ambulance, or

II. ordinary ambulance arrives first

1) cancel the ordinary ambulance and wait, if necessary, for paramedic ambulance, paramedic ambulance carries patient, or

2) cancel the paramedic ambulance and let ordinary carry patient.

3. Contingency: paramedic ambulance but no ordinary ambulance is available

Response: dispatch the paramedic ambulance.

4. Contingency: no ambulance of any type is available

Response: queue for both types

(A) If paramedic ambulance available first, only dispatch that ambulance

(B) If ordinary ambulance is available first, do the same as in the situation described in 2.

Let us examine the case where there is an ambulance of each type available and the ordinary ambulance is closer.

A paramedic call enters the model. It is routed to the

paramedic section. Two transactions are created. One transaction (call it transaction A) enters the ordinary car stream and the other transaction (call it transaction P) enters the paramedic car stream. First transaction A enters the queues and finds an ordinary ambulance that is available within thirty minutes. The facility number and the time to the scene are saved and then the transaction stops. Now transaction P in the paramedic car stream is executed. It notes the ordinary facility, enters the queues, and finds a free paramedic ambulance. Then response time is determined and is compared to the response times for the two ambulances. In this case the ordinary ambulance is closer. Transaction P then creates a duplicate (say transaction Q). Transaction P is then loaded on a user chain. Transaction Q advances to the scene. While the paramedic is travelling, transaction A checks to see if a paramedic was available and subsequently moves to the section where both cars are available. Then it compares the response times and finds that the ordinary ambulance is closer. The facility (ambulance) is captured (made unavailable for other calls) and also advances to the scene. After the ordinary ambulance arrives it leaves the queues
and waits for another minute to decide whether to cancel or not. If the paramedic ambulance has not arrived, the ordinary ambulance cancels the paramedic en route in 75% of the time. Transaction P is removed from the user chain and sent to look

for another call.

Transaction A proceeds to load the patient, carry and unload him in the hospital. Then it looks for another call. In the meantime, transaction Q on the paramedic stream arrives at the scene and searches the user chain for transaction P. Since it does not find it, it concludes that the call for a paramedic ambulance is cancelled and the transaction terminates.

Twenty-five per cent of the time the ordinary ambulance randomly cancels itself. This reflects the actual proportion of times this occurs in real-life situations. Transaction A is then put on a user storage chain. When the paramedic car arrives at the scene it searches the user chain for transaction P. It finds it and transaction Q terminates. Transaction P departs the queues, tabulates response time and checks to see if the ordinary ambulance has decided to cancel itself. When this is the case, transaction P takes transaction A off the user chain and subsequently services the call. After unloading at the hospital, it searches for another call. Transaction A counts the cancellation as an ambulance not used and the ordinary car searches for another call.

In the above illustration it is shown that there are two cars simultaneously pursuing one call and thus there is a need for two streams.

The paramedic section of the model is also represented in figure 3.1. The possible states of an ordinary ambulance

when a car arrives are: unavailable (N/A) and available. These are the headings for columns one and two. Rows one and two give the state for a paramedic ambulance.

Column three represents an arrival of an ordinary ambulance at the scene.

- If no paramedic ambulance has been dispatched, the paramedic ambulance is cancelled and the ordinary ambulance carries the patient.

- If a paramedic ambulance has been dispatched, the ordinary ambulance makes a decision whether to cancel the paramedic or not.

Row two (columns two and three) states that if neither ambulance is available and the paramedic is available first, then the ordinary is cancelled.

Row three, column two, represents the situation where an ordinary ambulance was dispatched and a paramedic was dispatched later but arrived at the scene before the ordinary ambulance, the ordinary ambulance is cancelled.

FOR A PARAMEDIC CALL

ORDINARY AMBULANCE

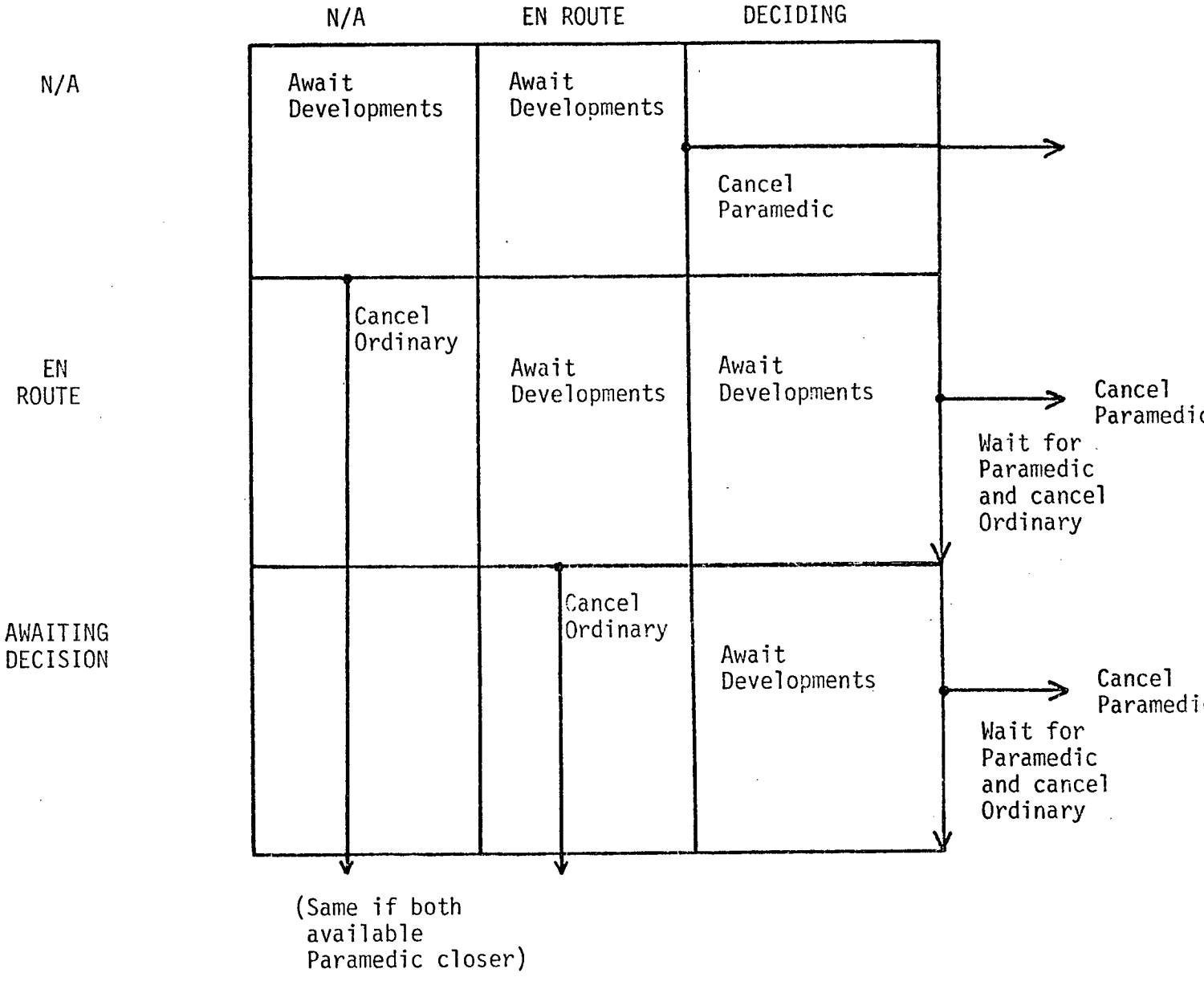


FIGURE 3.1

3.2.2 INPUTS

The inputs to the model consisted of information concerning 4800 actual calls which took place between August 16, 1977 and August 23, 1977 and between September 3, 1977 and September 19, 1977. These calls were entered on a file for the use of the simulation model. For each call or transaction the following information was provided: scene of the call, municipality of the call, destination (hospital), type or priority of the call (1 - transfer, 2 - normal, 3 - emergency and 4 - paramedic). The inter-arrival time, the time it took to load and unload the patient, and in the case where the call was cancelled the time from dispatch to cancellation was recorded. For missing loading or unloading data, a value was estimated from the empirical distribution.

The model also used travel time data. This data was used to determine the time needed to travel from the place of dispatch to the node representing the scene and from there to the hospital.

The model also contained general information on all the ambulances. This information included the location of depots, the number of ambulances at each depot, the type of ambulance located at each depot and the periods when each ambulance is available.

3.2.3 OUTPUTS

The simulation recorded response times, service times, number of calls answered by each ambulance, the average utilization per ambulance, and statistics on length of queues.

Outputs of the simulation were: the overall average response times per type of call, a distribution of response times per type of call, the average response time per type of call per municipality and the average response time for all calls per subregion. For service time only the overall averages and their associated frequency distribution were recorded. For each subregion and each municipality, statistics on the maximum and average contents of the queue were recorded along with the total entries. Other statistics gathered included the total number of calls, by type, that had to stand in the queue and the average time spent waiting in these queues.

3.2.4 MODEL VALIDATION

The conventional measure of validity is the degree of correspondence between the simulated system and the referent system. Higher correspondence can be achieved by increasing the complexity and detail of the model. The appropriate level of correspondence is decided by the value of the added information. This pragmatic criterion of validation states that refinements should be stopped when no net benefits can be

obtained by refining the model to improve its correspondence to reality.

In the ambulance simulation several measures were taken to ensure validity. First, the model and its desired output were discussed with management (in particular the desired degree of spatial resolution). Second, when possible, empirical distributions were used, and therefore, loss of information derived from data inputs was decreased. Finally, test of correspondences between simulated and actual time series were conducted. This was done for each component of the model and the model as a whole.

In the computer simulation model start-up times were adjusted for each type of call to calibrate the model. Start-up time is the time from when a call is dispatched to leave for the scene and the actual time it did leave. Data gathered from the ambulance service records contained information on the time a call was received, the time the call was dispatched and the arrival time at the scene. With this information one could only estimate response times which included start-up times. The start-up time could not be isolated and was, therefore, used as a residual to calibrate the model.

The travel times used in the simulation were obtained from the G.V.R.D. These were compared against actual data records from the ambulance service. It was found that travel times for short distances were very similar. (See Appendix.)

A test was also conducted to compare travel times by day and night for both urban and rural municipalities. It was found that the travel times for rural municipalities during the day needed to be adjusted downward by 10%, travel times for urban municipalities during the evening needed to be adjusted downward by 10% and travel times for rural municipalities needed to be adjusted downward by 25%.

The final set of variables for which correspondences were tested were:

- actual and simulated average response times,
- actual and simulated emergency call distributions,
- actual and simulated response times by type of call by municipality, and
- actual and simulated number of calls answered by each facility.

The comparison of the actual and simulated average response times by type of call and municipality are presented in table 3.1.

The comparison of the average response time for each type of call shows that the simulated times differ from the actual by 12 seconds for emergency calls, by 19 seconds for ordinary calls, by 12 seconds for transfer calls, and by less than 10 seconds for paramedic calls. Average response times by type of call for Vancouver only, which produced by far the greatest block of calls in any municipality, were almost

COMPARISON OF RESPONSE TIMES BY PRIORITY AND MUNICIPALITY

(all times are in minutes)

	TRANSFER CALL		ORDINARY CALL		EMERGENCY CALL	
	Actual	Simulated	Actual	Simulated	Actual	Simulated
Average	19.46	19.25 21.02	12.69	12.43 13.00	6.62	6.21
Municipality						
Vancouver*	19.91	19.68	11.35	10.64	5.49	5.47
North Vancouver	10.94	18.23	9.54	11.51	6.55	6.15
West Vancouver	32.53	20.61	10.94	15.93	8.17	10.28
Burnaby	22.51	18.86	15.4	12.02	7.72	6.58
New Westminster	20.97	16.87	17.12	11.63	6.53	6.28
Richmond	22.55	19.19	17.2	13.26	7.1	7.68
Delta	19.13	19.61	18.39	13.91	10.19	8.85
Surrey	10.54	17.34	13.02	12.02	7.3	7.27
Langley	16.3	23.79	23.78	17.85	11.24	11.29
Coquitlam	13.15	19.64	12.38	12.16	6.57	6.72
Pitt Meadows	*	-	23.0	20.20	*	-
Maple Ridge	14.0	22.40	13.5	18.94	6.42	7.17
White Rock	22.86	21.16	11.44	9.16	7.67	7.70
Port Coquitlam	*	-	21.44	15.71	*	-
Port Moody	*	-	15.5	14.62	5.0	9.78
Number of Calls	1015		2541		996	
Vancouver* proportion of calls	47%		62%		54%	

	Actual	Simulated
Paramedic	8 min.	6.10 min.
Service time	49.7 min.	49.80 min.

*Less than 10 calls

TABLE 3.1

identical for emergency calls, differed by 43 seconds in ordinary calls and differed by 13 seconds in transfer calls.

In general, simulated response times for emergency calls for each municipality were very similar to actual response times. Since emergency calls have the highest priority and thus are of the highest concern to management, these results are very encouraging. Comparing response times by municipality for ordinary calls, one finds that the differences range from 43 seconds up to 5.93 minutes in Langley (a rural district). For transfer calls (lowest priority calls) one finds differences of up to 12 minutes in West Vancouver. But, on the average, the differences are small, since there were only 227 paramedic calls a comparison of response time by municipality would not be meaningful. The overall average service time difference between actual and simulated data was less than three minutes.

The comparison of the distribution of simulated emergency response times with actual times is displayed in figure 3.2. Due to a ½ minute start-up time and a minimum inter-node travel time of 1.2 minutes, the simulation started to record response times from the 2 minute level and upwards by 1 minute intervals. The resulting discrepancy, however, narrows down for higher response times.

With a difference of less than 5% for the range of 5 minutes to 20 minutes, it was concluded that the two distributions are similar.

COMPARISON OF CUMULATIVE DISTRIBUTION
OF ACTUAL EMERGENCY RESPONSE TIME VERSUS SIMULATED
EMERGENCY RESPONSE TIME

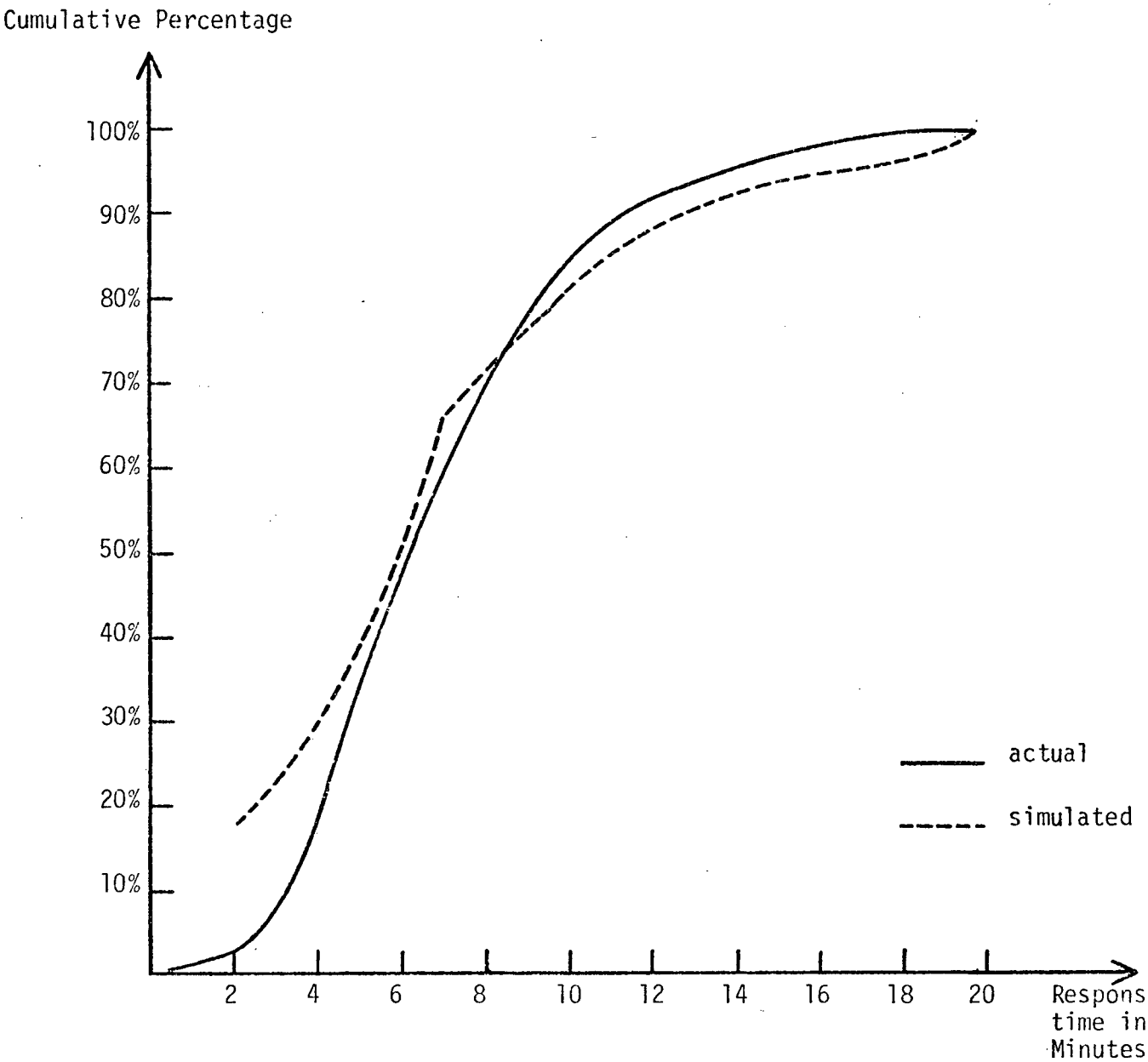


FIGURE 3.2

Another test for validation was the comparisons of simulated and actual calls per day per station. These are shown in table 3.2. The average total number of calls per day answered in the simulation is almost the same as in the actual data.

In a comparison of actual and simulated data conducted for each station, two stations with significant differences were discovered. One is the Burnaby station. Burnaby and New Westminster have the only paramedic cars. Often when a paramedic call occurs, both an ordinary and a paramedic car will travel to the scene, and therefore, the call is counted as two calls by the simulator. This, in part, explains the discrepancy. Station G8 (located at the south end of the second narrows bridge) also shows a large discrepancy. This is due to the fact that the ambulances from this station are being called to handle some of the numerous Vancouver calls, thus being routed away from home. Also these cars are crossing into North Vancouver where the travel times are longer. In the present system these cars may rarely go into North Vancouver, but this constraint was not inserted in the simulation. For all other stations the average numbers of simulated calls per day were very similar to the numbers of actual calls.

A final test was to evaluate the impact of the length of the run. It was found that the standard deviations lowered with more calls and time series started to stabilize after

COMPARISON OF CALLS/DAY BY STATION: VALIDATION

	Simulation Model Calls per day	August Data Calls per day	September Data Calls per day
G1 - Vancouver	31.85	31.8	34.5
G2 - Vancouver	18.19	20.5	20.8
G3 - Vancouver	8.47	10.4	9.5
G4 - Vancouver	13.17	12.1	12.43
G5 - Vancouver	12.77	12.7	12.47
G6 - Burnaby	19.98	15.2	13.33
G7 - New Westminster	20.07	18.0	20.47
G8 - Vancouver	22.98	29.3	28.40
G9 - Surrey	11.11	12.2	12.17
G10 - Richmond	13.22	10.7	11.70
G11 - Delta	2.91	4.1	2.73
G12 - Langley	4.93	5.1	5.50
G14 - White Rock	4.93	4.7	4.33
G15 - West Vancouver	2.78	4.1	4.17
G16 - North Vancouver	11.20	9.4	10.40
G17 - Haney	3.54	4.8	4.17
G18 - Vancouver	10.53	10.2	9.0
G19 - Coquitlam	7.75	5.6	6.43
Calls per day	220.38	220.9	222.5

TABLE 3.2

3500 calls.

Sensitivity tests were also conducted. The following parameters were examined: start-up times, speed-up factors for emergency calls, number of calls per day and travel times. The response variation of start-up times were linear, as expected, the model was sensitive to changes in the speed-up factor for emergency calls and changes in the travel times. The model was very sensitive to changes in the number of calls per day indicating that the real system operates near capacity.

CHAPTER 4

THE AMBULANCE LOCATION PROBLEM

This chapter considers the problem of locating ambulances so as to minimize mean response time.

4.1 ESSENTIAL FEATURES OF THE AMBULANCE LOCATION PROBLEM

The deterministic location problem assumes the following:

- the closest ambulance is always dispatched,
- an ambulance is always available at the station when a call is received,
- all calls have the same priority, and
- the rate of calls is constant throughout the day.

In contrast, in the stochastic formulation it is recognized that:

- ambulances are not always available at their station when a call is received,
- the probability that an ambulance is at its station depends upon the incidence rate,
- the response time for a call depends upon the number of ambulances which are busy when a call is received, and,
- since during peak periods the hospital becomes a departure point for ambulances

responding to calls, the location of hospitals affect mean response time.

The stochastic formulation indicates that the assignment of ambulances is a dynamic problem. The deterministic-static formulation can serve only as a crude approximation to the stochastic system. This approximation is valid only for systems with little or no congestion problems. In the next section of this chapter two deterministic approximations to the problem are considered.

4.1.1 ANALYTIC-DETERMINISTIC APPROXIMATIONS

Two models are considered:

1. Maximal covering location model, and,
2. P-Median Model.

The two models are crude approximations to the stochastic assignment problem and are used to identify "good" initial solutions.

4.2 THE MAXIMAL COVERING LOCATION PROBLEM

In solving public facility location problems two surrogate measures of value are: total weighted distance and the distance of the most distant user from a service facility. Total weighted distance is used in the p-median model which will be discussed later. The maximal covering location problem (MCLP) uses maximal service distance as the surrogate measure

for the value of a given configuration of locations. The model allocates p facilities to positions on the network such that the maximum number of people will find service within a stated distance or response-time standard. If all of the population is serviced within the response-time standard then the maximum distance which an ambulance would have to travel to reach the patient would reflect the worst possible performance of the system.

With the MCLP model a trade off curve may be developed showing various levels of coverage within desired distances for a range of facilities. For example, ten depots may service 80% of the population within a specified distance while 15 depots may service 100% of the population within a specified distance.

Another aspect of the model is that of 100% population coverage within a maximum time. If only 90% of the population is within five minutes of service but 2% are 30 minutes away then the quality of service to these 2% is much lower than to the majority. A constraint may be added in the model to ensure that all citizens get service within a specified maximum distance or response-time.

Advantages of the MCLP model are:

- it can be solved by linear programming
- it does not entail an excessive burden of computations (180 constraints in this study)
- population and maximal service distance are both used

- it indicates the maximum extent of demand coverage to be expected from a number of facilities less than the minimum number needed to cover all demand.

Disadvantages of the MCLP model are:

- it assumes one ambulance per depot
- it assumes the ambulance is always available
- it does not take into account congestion (in areas of high demand the solution may show that one ambulance can cover the entire region)
- one ambulance can cover several high demand areas if these areas are in close proximity of each other (such as the case in downtown Vancouver).
- fractional solutions may be encountered.

Next is a mathematical representation of the model followed by a discussion of the use of the model and an evaluation of the model.

4.3 MATHEMATICAL FORMULATION OF THE MCLP

DEFINE:

I = set of nodes

P = number of facilities (depots)

a_i = frequency of demand of node i (population at node i or # of calls per day at node i)

X_i = 1 if demand node i is covered by a facility
0 otherwise

Y_j = 1 if a facility is allocated to site j
0 otherwise

$N_i = \{j \mid d_{ij} \leq S \quad , \quad j \in J\}$

J = a set of potential facility sites

S = a specified maximum response distance (or time)

d_{ij} = smallest distance (or time) between node i and node j

Z_i = number or percentage of population served or "covered" within the desired service distance

Problem 1

1. Maximize $Z_i = \sum_{i \in I} a_i \; X_i$

subject to

2. $\sum_{i \in N_i} Y_j \geq X_i \quad \forall_i \in I$

3. $\sum Y_j = P$

4. $X_i, Y_j \in (0,1) \; , \forall_j \in J \; , \forall_i \in I$

Interpretation

1. - Maximize number of people served or "covered" within the desired service distance.

2. - Allow X_i to be covered only when a facility is within the desired service distance.

3. - A specified number of depots.

A third constraint may be added namely:

5. $\sum_{j \in M_i} Y_j \geq 1 \quad , \forall_i \in I$

where $M_i = \{j \mid d_{ij} \leq T > S\}$

Which will ensure that no one is farther than a response distance T to his closest facility. Adding this last constraint provides some degree of equity to the population not served within service distance S. Problem 1 has an equivalent formulation derived by substituting $1 - X_i = \bar{X}_i$.

Problem 2

$$\text{Minimize } Z_2 = \sum_{i \in I} a_i \bar{X}_i$$

$$\text{subject to } \sum_{j \in N_i} Y_j + \bar{X}_i \geq 1 \quad \forall_i \in I$$

$$\sum_{j \in J} Y_j = P$$

$$\bar{X}_i, \forall_j \in Y_j \in (0,1), \forall_i \in I, J$$

$$\bar{X}_i = \begin{cases} 1 \text{ if demand node } i \text{ is not covered by a facility within S distance} \\ 0 \text{ otherwise} \end{cases}$$

This problem seeks to minimize the population left "uncovered" if p facilities are to be located on the network. Problem 2 is the dual of Problem 1. Either formulation may be programmed. For the purpose of this study Problem 2 was programmed.

4.4 USE OF THE MCLP MODEL

The MCLP model was adopted for this study to provide:

- initial "good" depot locations which could be starting points for the simulation, and

- a trade off curve of the percentage of population covered versus a range of number of facilities.

Input data for the dual of the MCLP was the demand at each of the 169 nodes and a travel time matrix. In each particular subregion the demand was equal to the number of calls in that subregion divided by the total number of calls for the whole of the G.V.R.D. Thus the demand was a percentage of the total calls. The travel time matrix contained in the shortest time to travel between subregions and the average internode travel time was roughly two minutes.

At the time of the study there were 19 depots in Vancouver which are indicated in figure 4.1. Also shown are the high density regions. It appeared reasonable to initially solve the MCLP model for 19 depots and at a maximum response time of two minutes. The solution was interesting.

Practically all of the depots were in the high density region and virtually none in the outlying regions. Further, only 57% of the population was covered. Thus it was decided to increase the maximum response distance. The results were similar with only 64% of the density covered. Further increases in maximum response time were added and two depots were added. See figures 4.2, 4.3, 4.4, 4.5, 4.6 and table 4.1. As the maximum response time increased one depot covered a few very densely populated regions. This revealed some serious drawbacks in using the MCLP model.

The analysis revealed that:

1. At a very small maximum response time each depot was only covering one node. Thus the MCLP model allocated the depots to the 19 nodes with the highest demand many of which were very close together.

2. As the maximum response time increased one depot now could cover a few or several subregions of high demand. Outlying depots were allocated to nodes of highest demand in those regions.

The deficiencies of the MCLP model are therefore the following:

1. At a low maximum response time outlying regions are not considered.

2. In outlying regions no consideration is being made of nodes surrounding the high population node, i.e. travel time to these other nodes - subregions were larger and one depot covered one subregion.

3. As maximum response time increased, one depot could cover several high demand regions. In places, one depot covered 10% of demand whereas other places one depot only covered ½% of demand.

4. No consideration of time of day of calls or type of calls. (Depots were also allocated at hospitals since they have many transfer calls.)

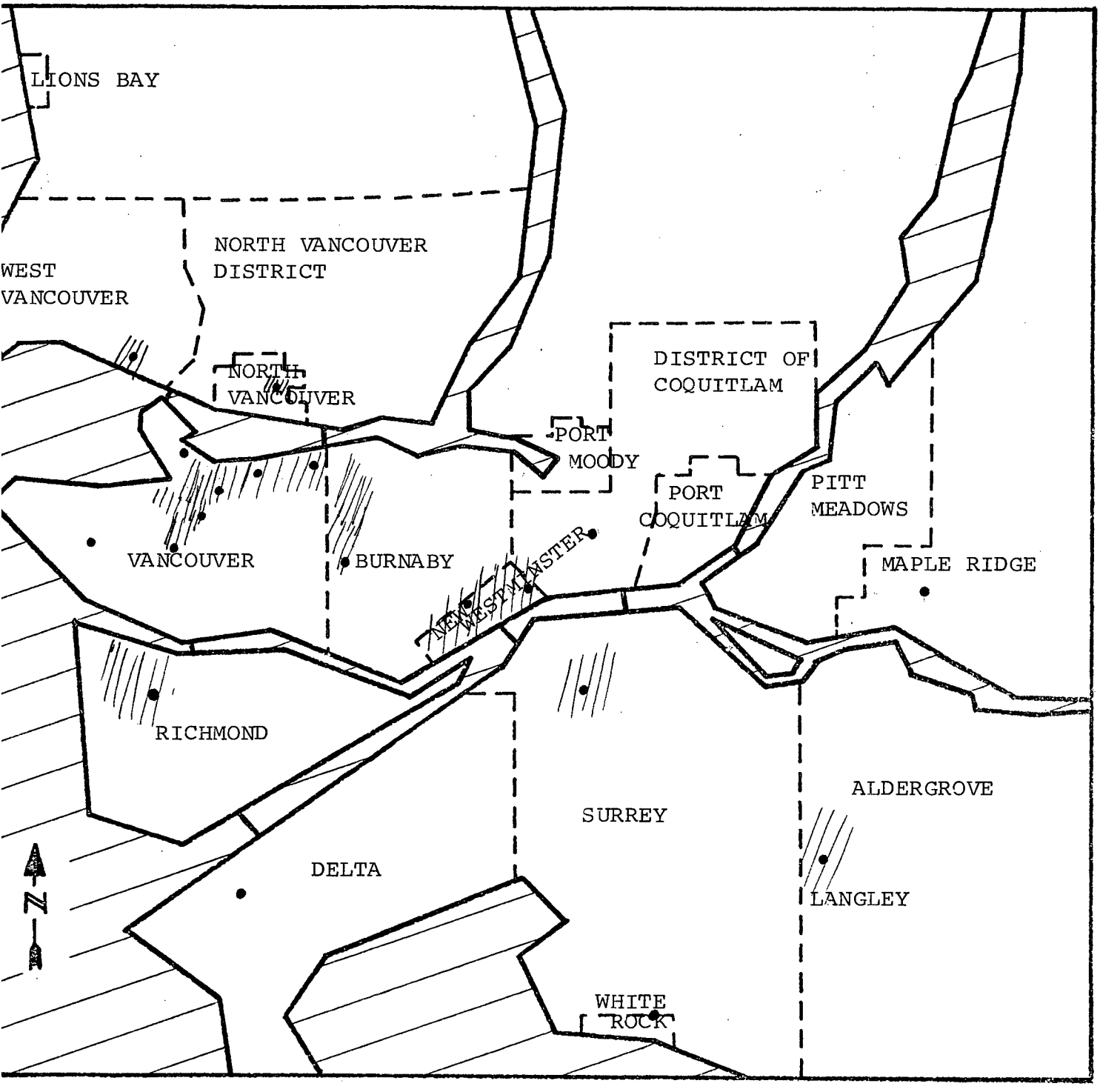
These results were discouraging since one would need a very high number of depots and a possible rearrangement of subregions to get a better fit. The p-median model was tried and compared to the MCLP model.

TABLE 4.1

MAXIMAL COVERING LINEAR PROGRAMMING SOLUTIONS

Number of Facilities	Maximum Response Time	% Demand Covered	% Demand Not Covered
19	2 (minutes)	57%	43%
19	4	64%	36%
19	5	75%	25%
19	6	86%	14%
19	7	92%	8%
19	8	96%	4%
19	10	99%	1%
19	15	100%	0
21	4	66%	34%
21	5	78%	22%
21	6	88%	12%
21	7	94%	6%

PRESENT AMBULANCE DEPOTS

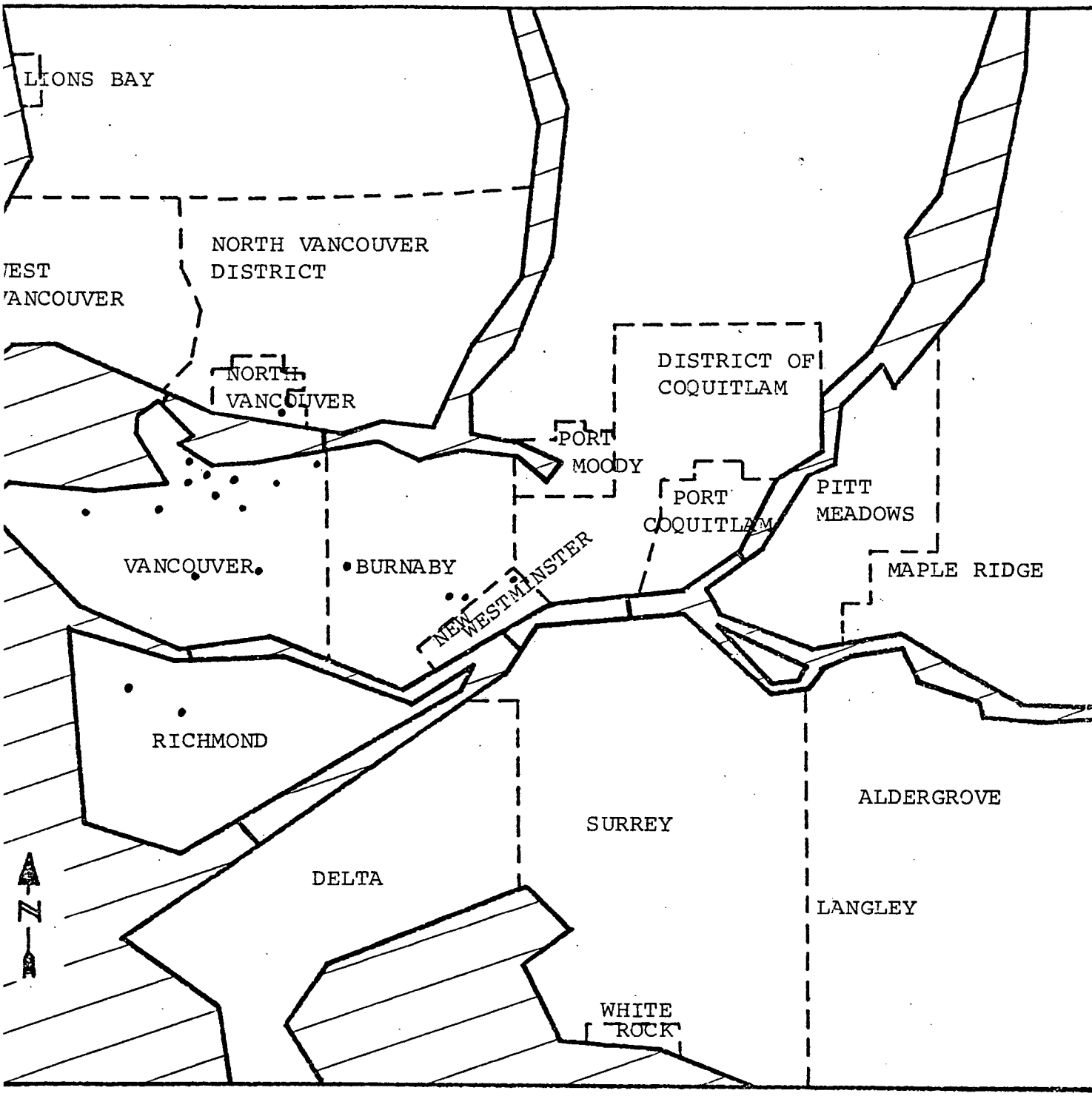


• Present depot locations

High demand zones

FIGURE 4.1

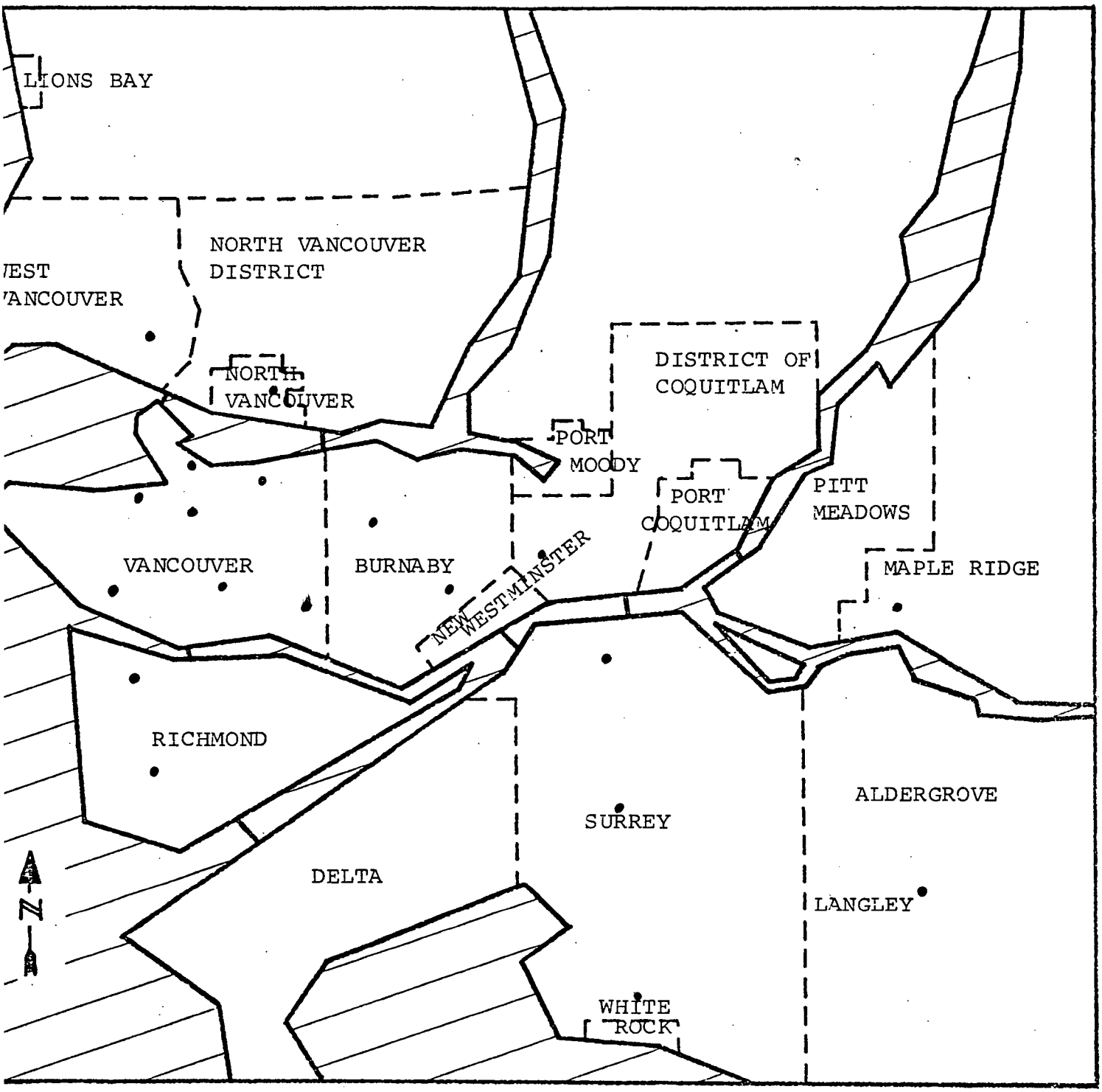
MAXIMAL COVERING SOLUTION FOR 19 DEPOTS
AND A FOUR MINUTE MAXIMUM RESPONSE TIME



64% of Demand Covered

FIGURE 4.2

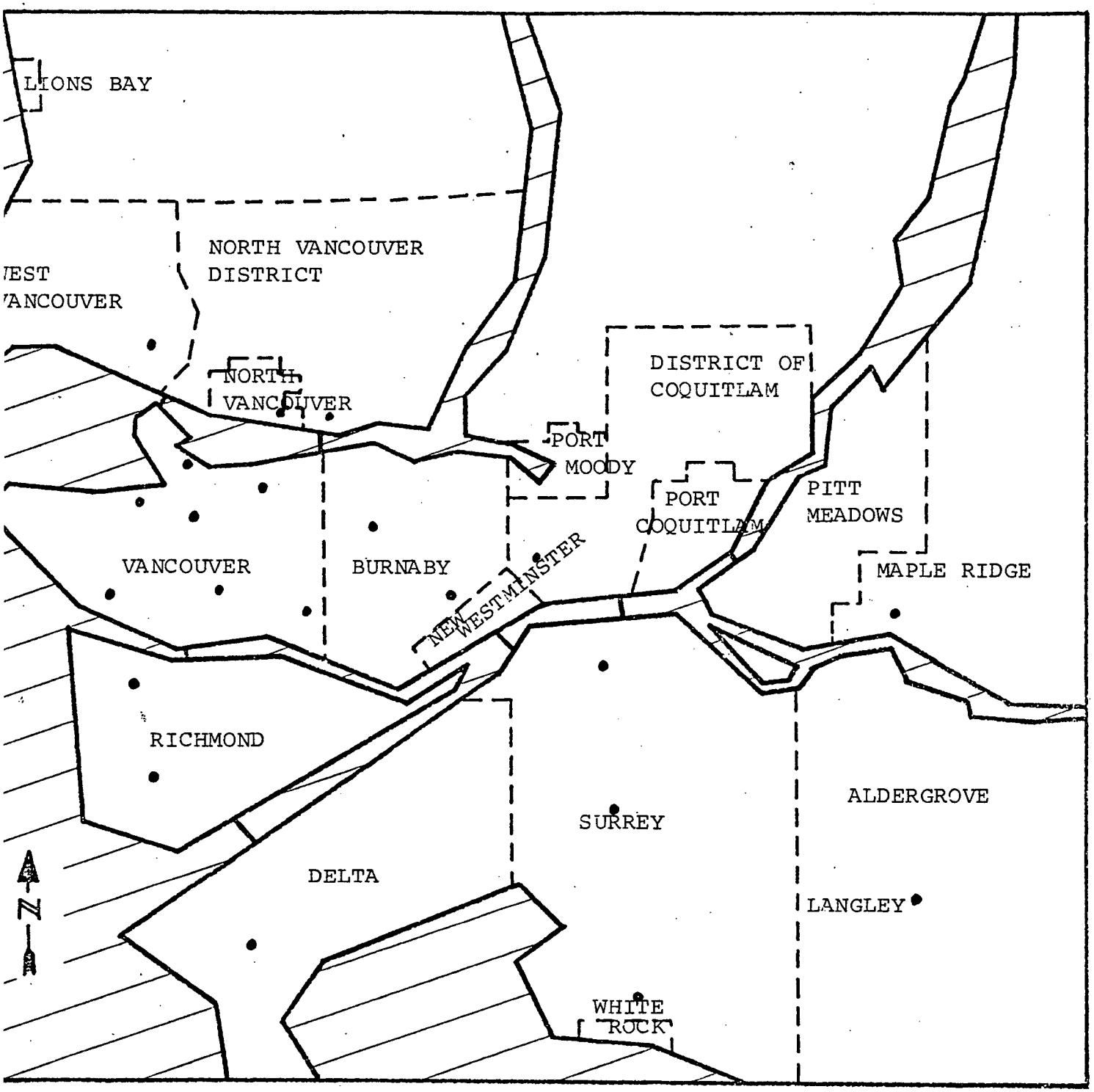
MAXIMAL COVERING SOLUTION FOR 19 DEPOTS
AND A SIX MINUTE MAXIMAL RESPONSE TIME



86% of Demand Covered

FIGURE 4.3

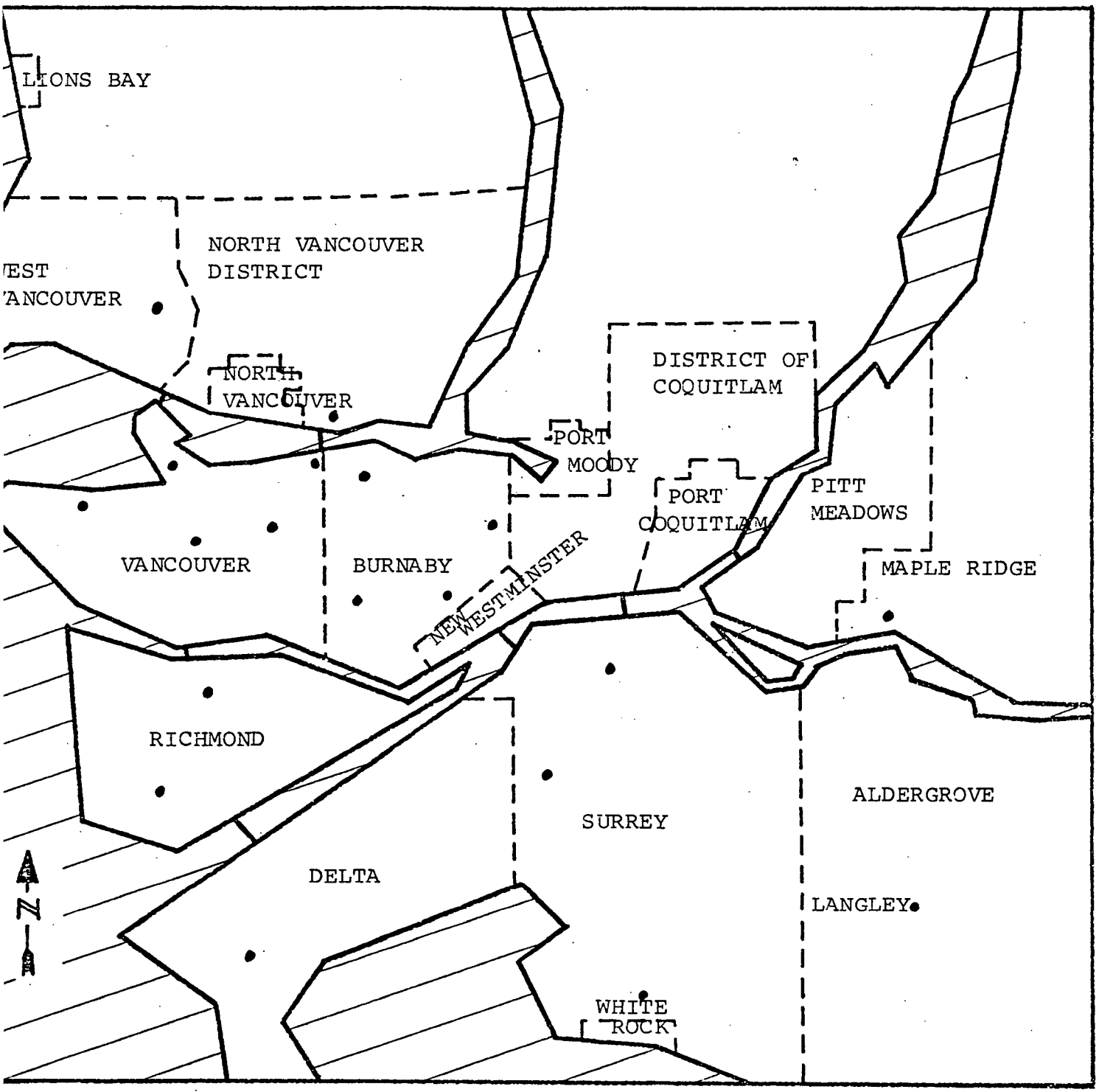
MAXIMAL COVERING SOLUTION FOR 21 DEPOTS
AND A SIX MINUTE MAXIMAL RESPONSE TIME



88% of Demand Covered

FIGURE 4.4

MAXIMAL COVERING SOLUTION FOR 19 DEPOTS
AND A SEVEN MINUTE MAXIMUM RESPONSE TIME



S = 7 minutes 12.78
P = 19 depots

FIGURE 4.5

MAXIMAL COVERING SOLUTIONS
PERCENTAGE OF POPULATION COVERED

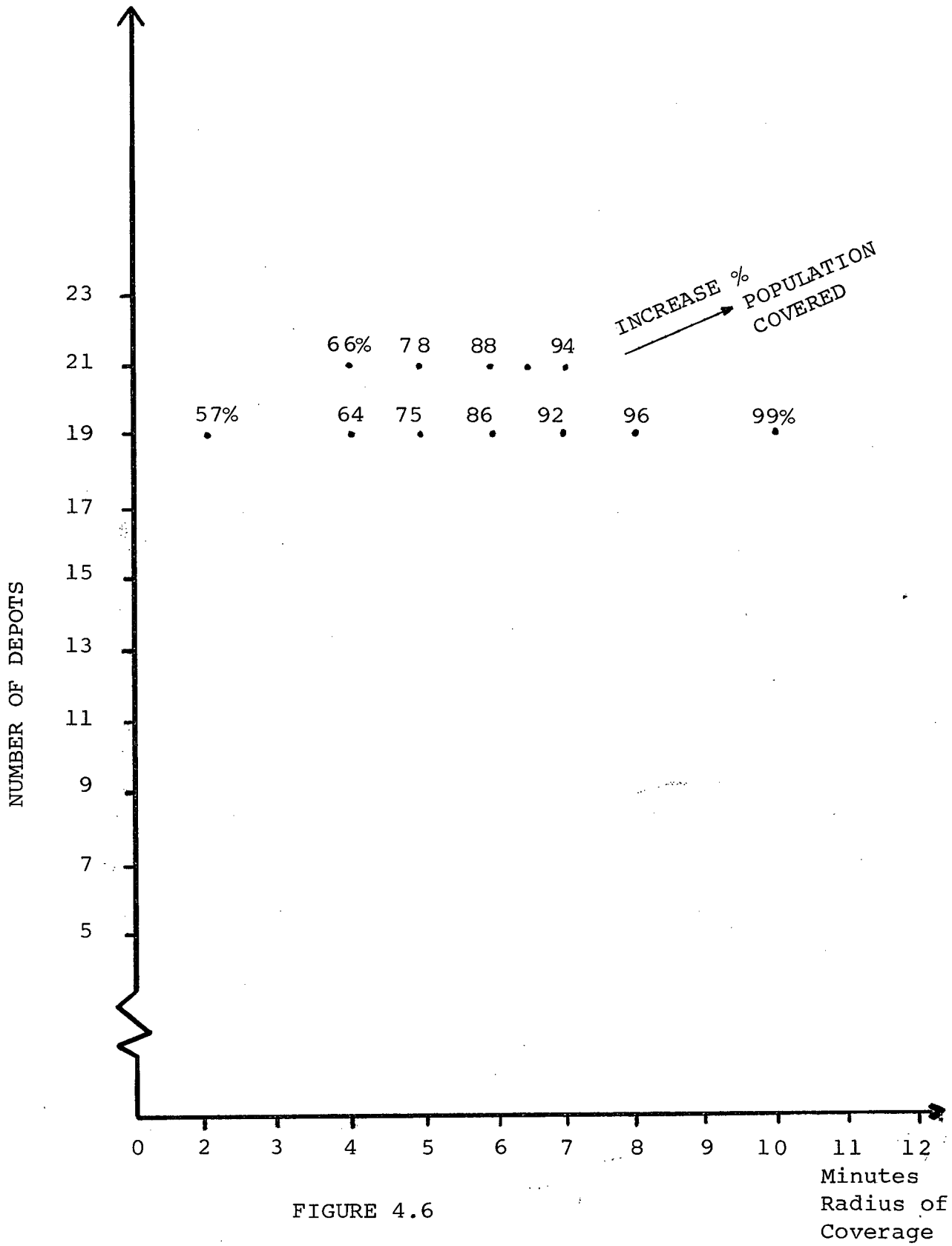


FIGURE 4.6

4.5 THE P-MEDIAN LOCATION MODEL

This network model is designed for determining facility locations and has a structure similar to the integer formulation of the p-median problem of Revelle and Swain (1970). The optimization problem is to minimize total response time.

4.6 MATHEMATICAL FORMULATION OF THE P-MEDIAN MODEL

DEFINE:

I = set of nodes

P = number of facilities (depots)

a_i = frequency of demand at node i - (population at node i or # of calls per day at i)

d_{ij} = smallest distance (or time) between node i and node j

x_{ij} = $\begin{cases} 1 \text{ when a facility } j \text{ serves node } i \\ 0 \text{ otherwise} \end{cases}$

J = a set of potential facility sites

S = a specified maximum response distance (or time)

For each node i, the service set N_i is defined as

$$N_i = \{ j/d_{ij} \leq S,\ j \in J \}$$

$$Y_j = \begin{cases} 1 \text{ when a depot is placed at node } j \\ 0 \text{ otherwise} \end{cases}$$

Problem 3

1. Minimize $Z_1 = \sum_{i \in I} \sum_{j \in J} a_i d_{ij} x_{ij}$

s.t.

2. $\sum_{j \in N_i} x_{ij} = 1, \quad \forall i \in I$

3. $\sum_{j \in J} y_j = p$

4. $y_j - x_{ij} \geq 0 \quad , j \in N_i, \; i \in I$

5. $x_{ij} \in (0,1) \quad , \forall_j \in J, \; \forall_i \in I$

$y_j \in (0,1) \quad , \forall_j \in J$

Interpretation

The problem is to locate p facilities on a network of demands so as to minimize total distance (response time) from depots to demand nodes. The first constraint demands that each community be assigned to one and only one depot for service; that is, the assignment cannot be partial. A second constraint specifies and fixes the number of depots (facility sites) available. The third constraint ensures that communities are served by locations designated as depots and that depots satisfy some constraint of maximum response time.[1] With a maximum response time constraint the solution guarantees that no points of demand will be far away from the nearest facility.

---

[1]The definition of a feasible service set N_i reduces the size of the problem from an otherwise larger problem with N^2 variables and $N^2 + 1$ constraints. The amount of reduction depends upon the sizes of N_j's.

One deficiency in the above formulation is that it does not consider congestion. An alternative formulation is to take into account the problem of demand congestion. This formulation uses a surrogate index to the impact of queues. The measure chosen in the alternative formulation implemented in this study (problem 4) was the distance or response time from the nth closest demand node. Thus the surrogate index acts as a subsidy for each facility node.

<u>Problem 4</u>

1. Minimize $$Z_2 = \frac{\sum_{i \in I} \sum_{j \in J} a_i d_{ij} x_{ij}}{\sum a_i} - \beta \sum d'_{ij} y_j$$

subject to

2. $$\sum_{j \in N_i} x_{ij} = 1 \quad , \forall_i \in I$$

3. $$\sum_{j \in J} y_j = P$$

4. $$y_j - x_{ij} \geq 0 \quad , j \in N_i, \ i \in I$$

5. $$x_{ij} \in (0,1) \quad , \forall_i \in I, \forall_j \in J$$

$$y_i \in (0,1) \ , \quad \forall_j \in J$$

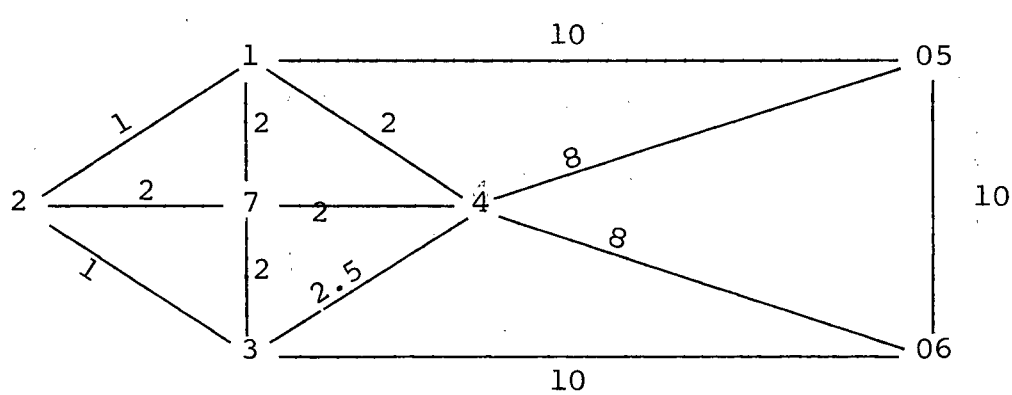
where d_{ij} is the distance (or time) to the nth closest node. This may be determined by

$K \cdot \frac{n}{p}$ where K is a constant, n is the # of demand nodes and p is the # of facility nodes.

is a constant which adjusts for geographical and terrain

factors.

Problem 4 gives the facility locations after adjusting for congestion. The adjustment is subtracted from the first term (total distance) to give the model the effect of moving depots from congested areas towards less congested regions and remote areas of sparse population. Often in high demand regions an ambulance is busy and is unable to respond to calls immediately. Servicing of such a call depends on how far away the second closest ambulance is (assuming he is free). Without an adjustment term, facilities will tend to be located in the center of demand regions. Outside regions are considered "secondary". With the subsidy term, depots move towards outside regions. This is illustrated by the following example. Let the network be defined as follows:



Thus, the shortest distance matrix would be:

	1	2	3	4	5	6	7
1		1	4	2	10	10	2
2	1		1	4	11	11	2
3	4	1		2.5	10.5	10	2
4	2	4	2.5		8	8	2
5	10	11	10.5	8		10	10
6	10	11	10	8	10		10
7	2	2	2	2	10	10	

And let the demand at each node be:

dm =	1	2	3	4	5	6	7	$\sum$ dm
	100	175	100	50	50	50	200	725

We will let d'_{ij} be determined by using the time to the second closest node.

We are trying to solve the following objective function for problem (4).

Minimize $\sum dm \cdot ds \cdot X - (\sum dm)\ \beta \sum ds\ X$

where $\sum dm = 725$

ds = distance between nodes

X = location of depots - find 2

$\beta = 0.10$

Four feasible solutions are as follows:

depots	2, 7	(subsidy)	4, 7	(subsidy)	5, 7	(subsidy)	2, 4	(subsidy)
2-1	100	20	200	20	200	100	100	20
2		17.5	350	52.5	350	192.5		52.5
3	100	20	200	25	200	100	100	25
7-4	100	15		10	100	40		75
5	500	55	400	50		50	400	55
6	500	55	400	50	500	50	400	55
7		40		40		200	400	40
no subsidy answer	1300	min.	1550		1350		1400	1st closest
subsidy	-222.5		-247.5		-732.5		-322.5	2nd closest
total	1077.5		1302.5		627.5	min.	1077.5	

By use of a subsidy term, one of the depots moves to an outlying region.

Using the same network let us assume that calls occur at node 2 first and then at node 5. The solution of problem 3 places depots at node 2 and node 7. For problem 4 the solution would be at nodes 5 and 7. In problem 3, if an ambulance at node 2 services a call at node 2, a second call at node 5 would require the ambulance at node 7 to travel a fair distance. However, with depots at nodes 5 and 7 a call at node 2 would be serviced by the ambulance at 7 and a second call at node 5 would be serviced immediately. Certainly most calls occur in nodes 1, 2, 3 and 7 and this is where ambulances are needed.[1]

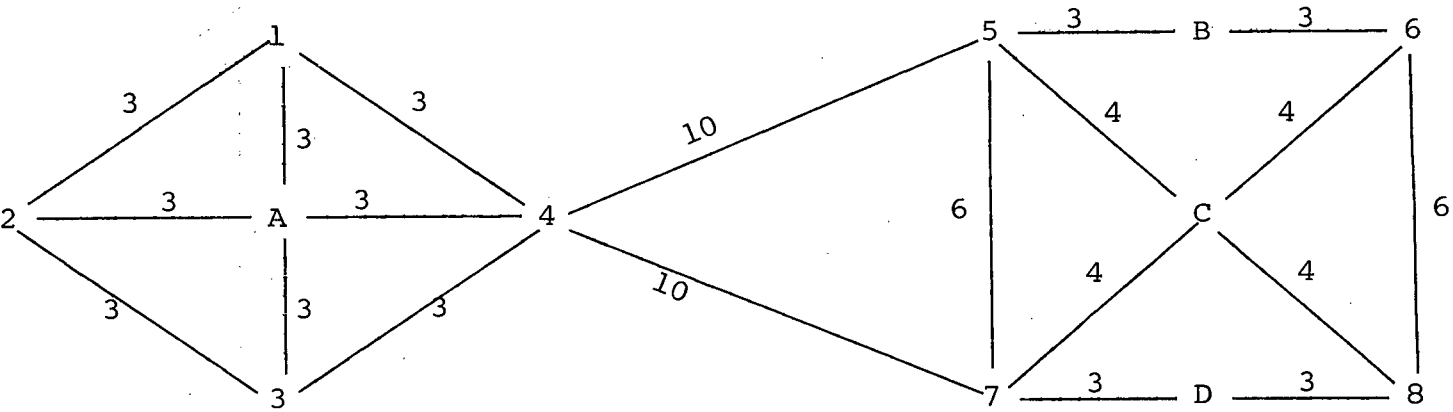
The p-median model has some advantages over the maximal covering model. This can be well illustrated in the following

---

[1]The choosing of β and the "2nd" closest distance versus say the "3rd" or "4th" has a definite bearing on the solution.

simple examples.

Consider the following network:



Demand	1	100	5	10	A - no demand
	2	100	6	10	B - no demand
	3	100	7	10	C - no demand
	4	100	8	10	D - no demand

Maximal covering solution for three ambulances with maximum response of 5

Solution depots at A, B, D

Objective function 100% of population covered.

P-median solution for three ambulances

Solution depots at 2, 4, C

Objective function = 4.75 or average response time = 4.75 minutes.

The major congestion problem is at 1, 2, 3, 4 and this is where ambulances are needed. If the maximal response time is lowered to 3, the maximal covering will still place only one ambulance in the congested region. If the maximal response time is decreased any lower than the maximal covering solution will be at three of the first four depots and ignore the outlying region. It does appear that the p-median model, which determines the response time itself, uses time and distance more effect-

ively.

When the p-median algorithm was solved (without subsidy) for 19 depots, depots were located in all outlying regions and not congested in downtown Vancouver as solutions from the maximal covering model.

The disadvantages of formulating the ambulance problem as a p-median are:

- It involves a very large number of constraints. The LP problem may have to be solved in two parts or by making initial guesses at depot points. If N is the number of demand nodes and P is the number of facility nodes, then the possible lower bound on the number of constraints P(N-P) = 20(170-20) - 3000.

- The formulation requires experiments for calibrating and for the determination of the appropriate closest node.

- The service distance (or time) constraint may eliminate all feasible solutions.

- The formulation may result in fractional solutions.

- There are many possible optimal solutions, and,

- The formulation assumes one ambulance per depot.

4.7 USE OF THE P-MEDIAN ALGORITHM

The LP formulation of the p-median problem requires some N^2+1 constraints. In the G.V.R.D. N is equal to 169, thus the number of constraints is over 25,000. Due to the sparseness of

the matrix, the cost of solving this LP is prohibitive. (Solving the problem for N equal to 40 costs $50.00.) An alternative to solving the problem is to include every third or fourth node in the model instead of all 169 nodes. But since the nodes already represent large regions the approximation is almost unrealistic. An efficient algorithm for solving the p-median problem is available at U.B.C. R.A. Whitaker in research undertaken while in the geography department at U.B.C. formulated a programme which computes solutions to the p-median problem subject to some side constraints. Heuristic procedures were utilized to estimate the medians. The basic procedure is as follows in p-median estimation problems.

1. Pick an initial set of nodes at which to locate P service bases.

2. Assign all other nodes to be serviced by the nearest service base.

3. Alter service base locations in such a way as to check for possible improvements.

Both steps 1 and 3 require Heuristic procedures. Singer (1968)[1] has created a method to obtain initial set of nodes (Step 1) by which the P service bases are spaced the maximum distance apart. This spread is particularly useful in the ambulance setting since it ensures that the initial feasible solution for ambulance locations covers less dense rural areas.

---

[1]Singer, S. multi-centers and multi-medians of a graph with an application to optimal warehouse location, mimeographed paper presented at the annual meeting of Tims and Orsa, San Francisco, June, 1968.

For step 3 Singer employed cyclic perturbation of one node service base at a time. In his adaptation of Singer's algorithm, Whitaker allows this as an option since several of his examples did not require perturbation to obtain optimality.

Whitaker allowed constraints to be placed on the amount of service provided at each base. This was handy for the ambulance location problem since the number of ambulances was not particularly large and thus service was rather "lumpy". With the constraints, it was possible to solve for an initial set of base nodes through the application of the transportation method of linear programming. Whitaker used the Ford-Fulkerson algorithm.

Whitaker also mentioned a maximum distance constraint in his dissertation. His suggested method was simply based on restriction of the nodes for which demand was allocated to any base to the neighbourhood of the base. This was not implemented in the published version of the program.

A method to employ the maximum distance constraint is through the use of the travel time matrix. By multiplying travel times for each node outside a given radius by a δ factor, ($\delta > 1$) one is implementing a distance penalty. Thus travel times to and within rural areas are increased. Since the p-median algorithm is minimizing, the depots are spread out. By increasing travel times the effect is that the rural areas are subsidized and thus is the same as adding

the surrogate in the LP formulation. Thus by adding the maximum distance constraint it is also taking into account congestion.

Attempts were made to run the Whitaker p-median algorithm (WPMA) with the service constraint in place over all possible base locations. Results were uniformly poor, running from no solution after computer time had been spent under fairly moderate constraints to an absurd solution in which five ambulances were in the west point grey area.

It was decided instead to selectively apply constraints to specific nodes in the otherwise unconstrained problem, expecting that the workload (number of nodes serviced) at these specific nodes would be below the maximum specified and the workload overall to be spread more equitably among the depots. However, there were great differences in the workload per station, with a range of 1 to 31 nodes serviced by a specific station.

The next step was to force some nodes into the solution. Inclusion of different sets of nodes in the solution resulted in radical differences of response time and in number of nodes serviced by particular stations. Unfortunately, there was no way of placing upper bounds on these "forced in" nodes.

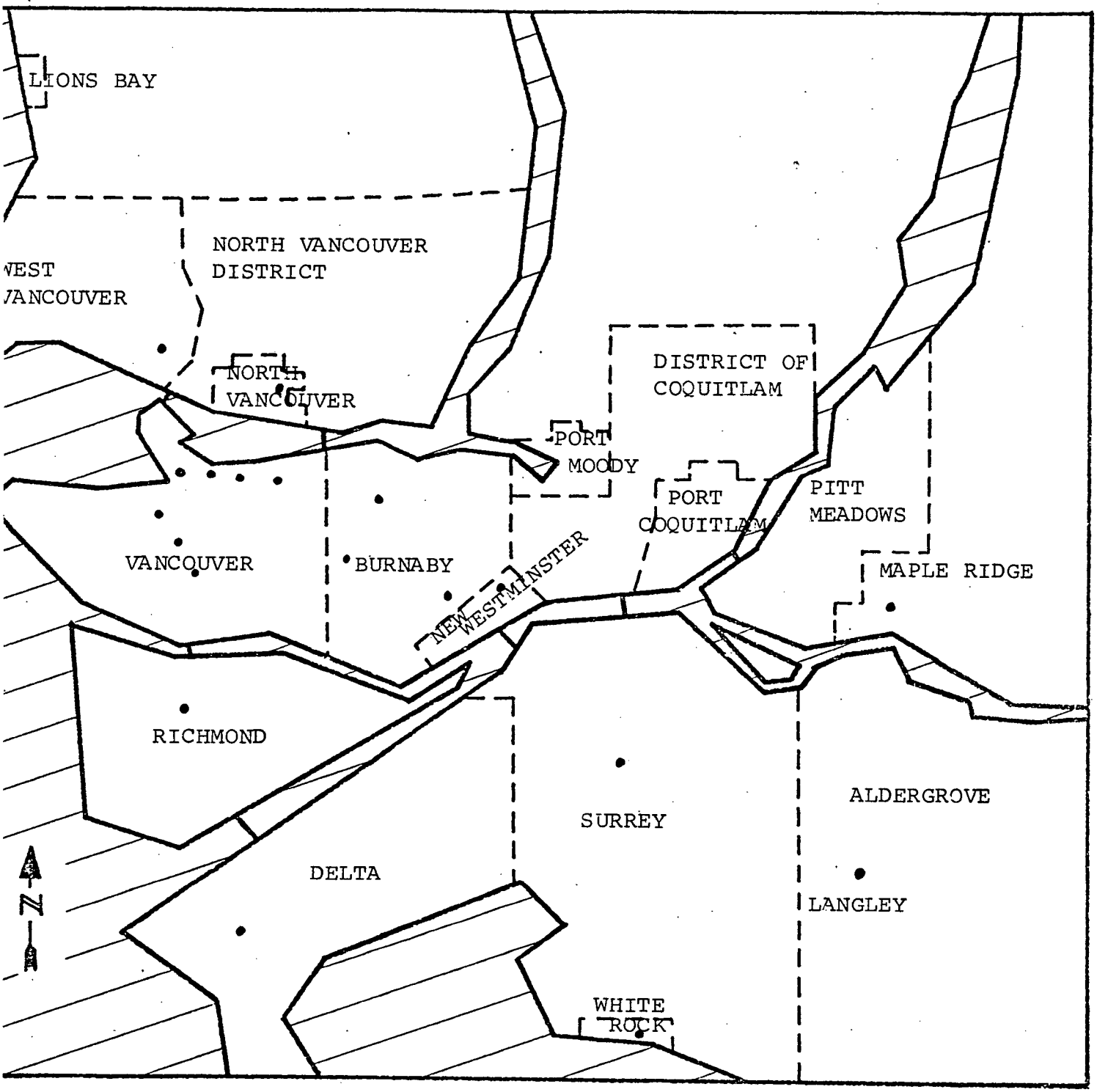
Further options exercised were the extended search, with weight shift and perturbation. The extended search option calls for searching throughout the whole region locating medians

instead of searching only in the immediate area surrounding a specific node. Perturbation tests call for checking surrounding nodes to see if there would be any improvement in response time.

Although the service constraints, applied overall or to individual nodes could not be used successfully, the remaining unconstraint problem, with perturbation, produced "good" results as initial solutions. This became especially evident as the number of facilities (depots) in the solution increased. The congested areas were well "covered" and more depots were placed in low density areas. See figures 4.7 and 4.8.

The solutions from the p-median were inputted into the computer simulation. Now the congestion aspect on the p-median solution was observed. With these results the p-median solution was altered to account for congestion and inputted back into the simulation. The p-median solutions were used in the experiments which are discussed in Chapter 6.

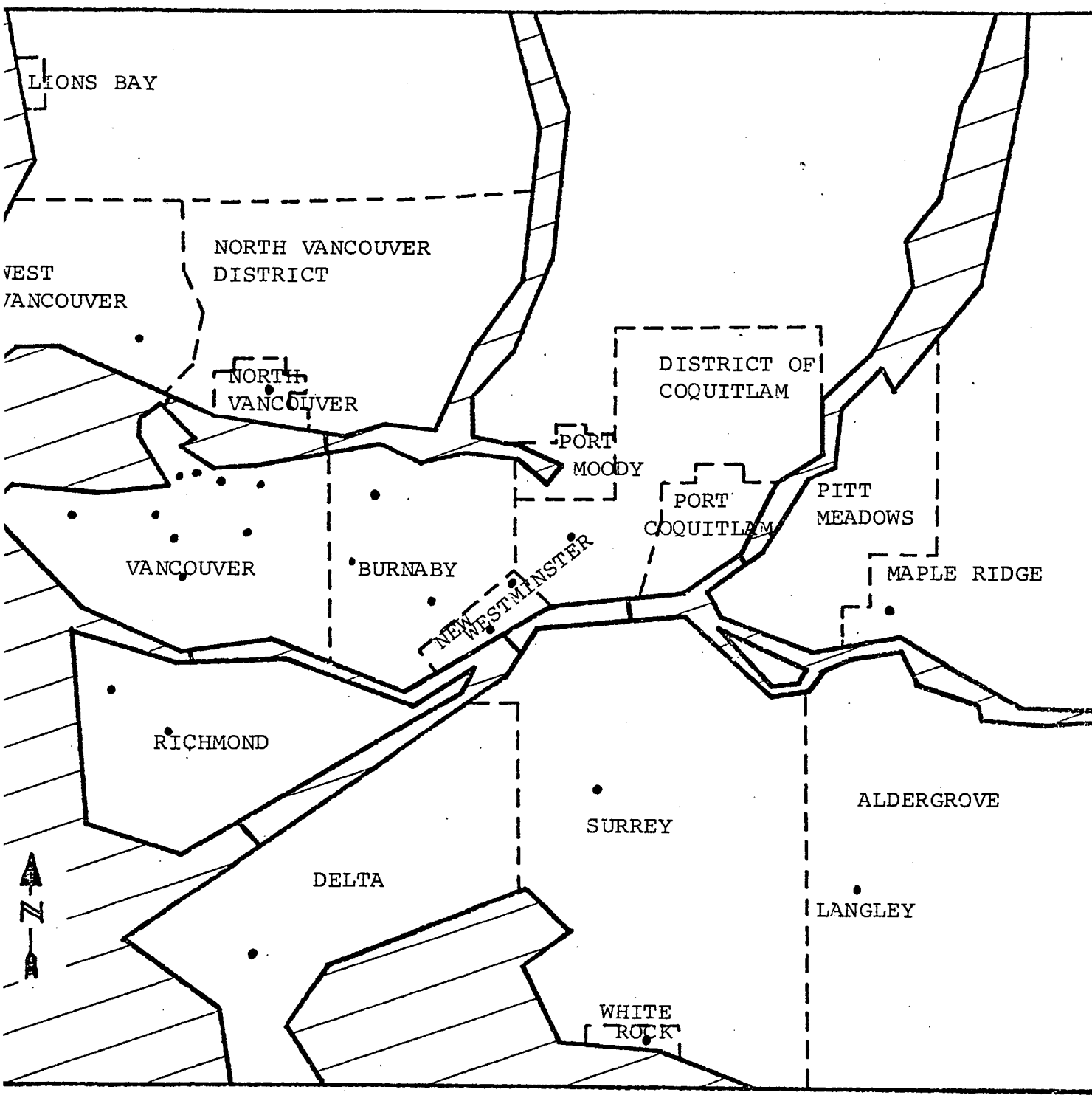
P-MEDIAN SOLUTION FOR 19 DEPOTS



Average response time 4.07 minutes

FIGURE 4.7

P-MEDIAN SOLUTION FOR 24 DEPOTS



Average Response time 3.58 minutes

FIGURE 4.8

CHAPTER 5

DESIGN OF EXPERIMENTS

5.1 INTRODUCTION

To conduct experiments a number of decisions must be made:

- determination of starting conditions of the model,
- determination of length of the run,
- determination of parameters so that the effect of one variable can be studied without compounding its effect with simultaneous changes in other system variables,
- determination of parameters to expose different system responses, and,

In addition one must ensure that the model represents the real world system and has just sufficient detail to examine the important features of the system.

In Chapter 3 the length of the runs was discussed. With over 4500 transactions run, and the model beginning at midnight when there were few calls statistics could be gathered from the start of the run without having any significant effect.

Sensitivity analysis was also discussed in Chapter 3. The ambulance model was designed to be an abstraction of the real system, with the emphasis being on the evaluation of response time and utilization. The model is sufficiently detailed so as to highlight changes in behavior caused by experimentation with system parameters.

5.2 EXPERIMENTAL DESIGN

There are five experiments reported here:

1. Determining the effects of moving the West Vancouver ambulance westward.

2. Determining the effects of removing the three night crews.

3. Moving one Burnaby paramedic ambulance to the Vancouver General Hospital (V.G.H.).

4. Replacing one V.G.H. ordinary ambulance with a paramedic ambulance under two conditions.

5. Determining the number of depots and ambulances required to have an emergency response time of five minutes for at least 90% of the calls.

Simulation was used in all five experiments. The p-median and maximal covering models give answers on where to locate depots optimally whereas the first four experiments are not concerned with optimality.

In experiment (1) a simulation run was made for each move of the ambulance from its present location to a westward

node. For experiment (2) two different sets of night crews were removed and a separate simulation run was made for each. Experiment (3) required only one run. Replacing an ordinary ambulance with a paramedic in experiment (4) was done initially with the paramedic unit only on emergency and paramedic calls. A test was made with the paramedic unit helping out on ordinary (priority 2) calls when not otherwise busy.

For the third experiment the following procedure was employed:

1. The p-median algorithm was run, with perturbation and extended search, to determine the optimal locations for a large number of ambulances.
2. The number of depots is the same as the number of night crews. Thus only the night shift transactions were run on simulation for a range of ambulances until 90% of the emergency calls were answered within five minutes. This established the number of depots.

3. For each depot ambulances were added (0-3) depending upon the demand surrounding the depot location. The simulation model was now run with transactions for the full day. Ambulances were added until 90% of the emergency calls were answered within five minutes.

The results obtained from these experiments cannot be considered exact. Response times and the placement of depots and ambulances during various parts of the day may vary slightly since a computer simulation cannot take into account all issues. Some of these are:

- Small adjustments in individual locations are often necessary to secure adequate

adequate accommodation for the depot to account for local traffic conditions, etc.

- Considerations other than response time may influence the location of ambulance depots. For example, for reasons of training and liaison, it may be desirable to locate a certain number of ambulance depots adjacent to emergency rooms.

- Some or all of the depot locations selected for the evening or early morning periods may be restricted to the depot locations selected for the daytime period.

- The objective of minimizing average response time may be disputed. For example, it may be felt that a small decrease in response time in an outlying region may be worth more than a relatively larger increase for the region as a whole. This is clearly a political decision.

5.3 SIMULATION STATISTICS

For each call the simulation recorded the type of call it was, the subregion and municipality of occurrence, the time in queue, the response time for the ambulance, and the service time. The number of calls each ambulance answered and type of call answered was also recorded.

The output statistics consisted of the following:

- mean response time per type of call,

- a distribution of response times per type of call,

- the average response time per type of call per municipality,

- the average response time for all calls

per subregion,

- the service time only - the overall averages and their associated frequency distribution,

- maximum, average and total calls for each subregion and municipality, and

- total number of calls, by type, that had to wait in queue and the average time spent waiting in these queues.

CHAPTER 6

EXPERIMENT RESULTS AND THEIR IMPLICATIONS

6.1 INTRODUCTION

In this chapter the results of five experiments are presented. All of the experiments use simulation. With simulation different aspects of the ambulance service can be investigated as well as the total system. It should be noted here that in the experiments the dispatch policy was to send the closest available ambulance. Each ambulance was restricted to a 30 minute radius. Thus if all ambulances were busy within a 30 minute radius of a call that call is queued.

The experiments highlight several problems (e.g. - congestion) and permit evaluation of alternative strategies and policies. The chapter concludes with identification of areas which merit research in the future.

6.2 THE WEST VANCOUVER AMBULANCE EXPERIMENT

The objective of the West Vancouver ambulance experiment was to determine the effects on the system of moving the West

Vancouver depot westward. All of factors in the simulation were kept constant.

Currently the West Vancouver ambulance is located at node 43 (see figure 6.1). It services the nodes in the West Vancouver region and helps out on calls in North Vancouver when that ambulance is busy. Also, if required, it may travel into Vancouver to service a call. The hospital is located at node 51 (see figure 6.1) where the North Vancouver depot is located. The experiment was called upon since calls west of node 43, especially at node 39, required the West Vancouver ambulance to travel from node 43 to node 39 and then to node 51. Thus patients at node 39 wait quite awhile before reaching the hospital. The results are given in table 6.1.

The effect of moving the ambulance from node 43 to node 41 decreases the average emergency response time in West Vancouver. However, movements to other nodes have the opposite effect of increasing the average emergency response time for West Vancouver. The explanation requires understanding how the West and North Vancouver ambulances interact and also the unique road network in West Vancouver.

First as the ambulance is moved westward the ambulance in North Vancouver answers some calls that previously would have been answered by the unit at node 43. The North Vancouver car needs to travel further westward resulting in both increased response time and total service time. Its availability is thus

lessened and other calls are now queued. A few hours may be lost before the North Vancouver ambulance is back to its normal period of availability. Overall, the average emergency response time for North Vancouver increases.

Now with the North Vancouver unit covering more territory, West Vancouver's ambulance has some pressure relieved. Its availability increases. It is also more centrally located in West Vancouver and consequently, the West Vancouver average emergency response time decreases. Horseshoe Bay, node 39, also has its response time lowered with the depot located at node 41. This result was fully expected and desired.

The average response time for ordinary and transfer calls in West Vancouver enlarged. Some of these calls occur in nodes 40 and 42 which are located on the high road in West Vancouver. The unique road network in West Vancouver has a lower road through nodes 45, 43 and 41. The upper road goes through nodes 44, 42, 40, 39 and 38. These two roads run somewhat parallel to each other and getting from the upper road to the lower or vice versa is very difficult. There are very few direct roads connecting the major arteries. With the ambulance at node 41, it has more difficulty getting to the upper nodes than it did with the depot at node 43. Another partial explanation for the augmented response time for ordinary and transfer calls is that the West Vancouver ambulance now has to travel further to answer calls in North Vancouver if that ambulance is busy.

Also, any transfers from the hospital to a West Vancouver location require the West Vancouver ambulance to spend more time on the road. Spending more time servicing calls leads to less availability and an increase in response time. These reasons apply to the deterioration in service level when the ambulance is moved to other West Vancouver subregions. The road network is a major factor in increasing response times. At nodes 40, 38 and 39 all response times and West Vancouver's service time appreciate, with the exception of the response time at Horseshoe Bay.

Locating an ambulance at nodes 38 or 39 significantly increases response and service times due to: the difficulty of getting to the lower road, the long response time required in answering calls in North Vancouver, and the increased service time as a result of the ambulance having to travel a long distance back to its depot. Nodes 43, 44, 47 and 48 have the majority of calls in West Vancouver. A depot at either node 38 or 39 obviously will increase the response time for all calls. The travel time from the hospital to the depot is increased. As a result, the total time the ambulance is available to answer calls decreases. Consequently, response times appreciate significantly.

The North Vancouver ambulance also has more pressure on it with a depot at 38 or 39. It has to travel further into West Vancouver, time to travel back to the hospital increases,

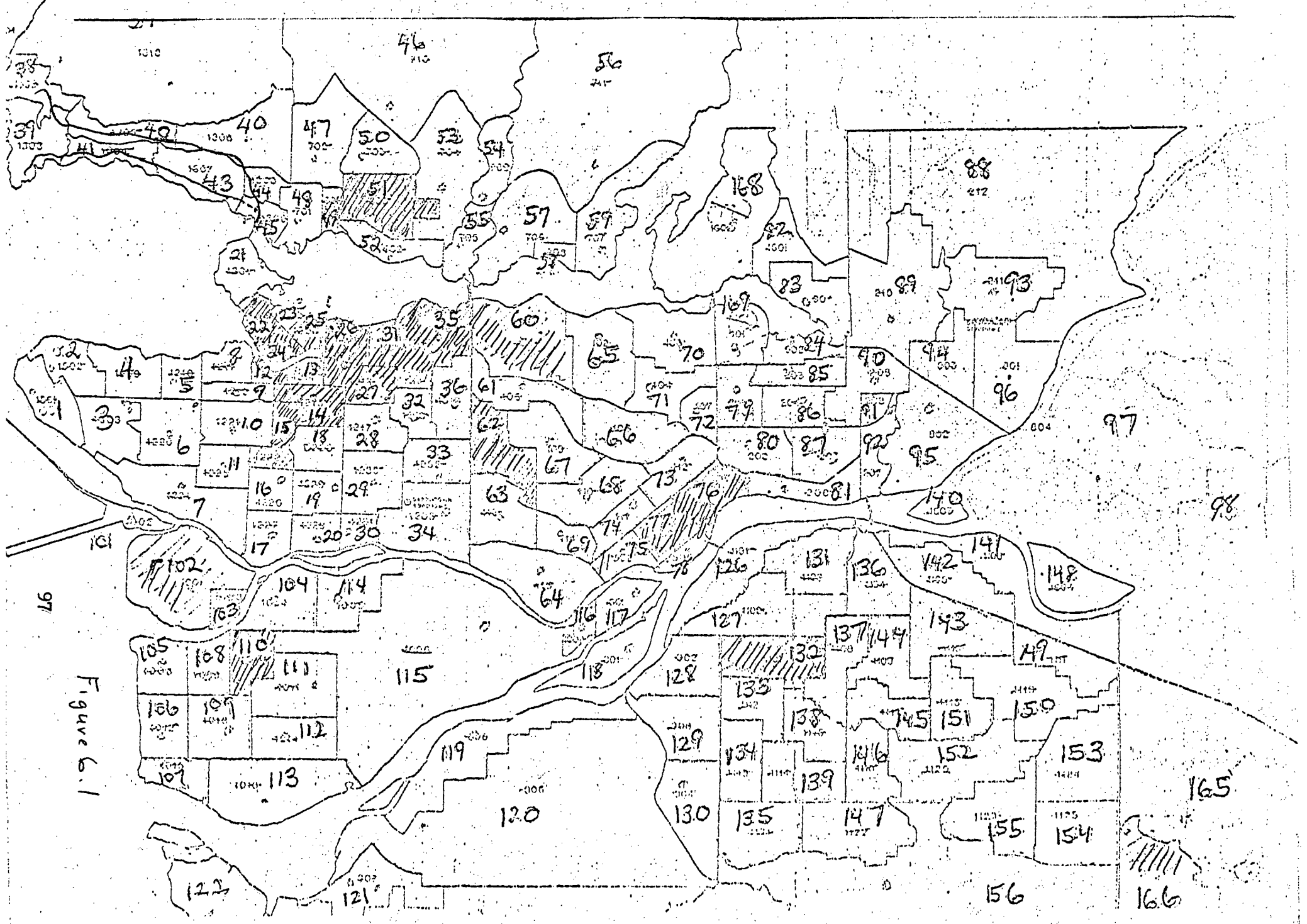


Figure 6.1

and, thus, total service time increases. Total availability is depreciated. Overall response and service times increase. Only Horseshoe Bay residents profit from the movement of the West Vancouver ambulance. Response and service times for the total G.V.R.D. do not alter. The best policy for all West and North Vancouver ambulances is not to move the West Vancouver ambulance. Of course, experiments can be done relocating the North Vancouver ambulance or adding more ambulances which may alter the results obtained.

EFFECT OF MOVING WEST VANCOUVER AMBULANCE WESTWARD

(Results in Minutes)

	Current Depot Node 43	Depot Node 41	Depot Node 40	Depot Node 38	Horseshoe Bay Depot Node 39
West Vancouver emergency call response time	7.39	8.98	9.14	10.75	9.55
North Vancouver emergency call response time	5.91	6.11	6.14	6.39	6.27
North Vancouver ambulance service time	51	51	51	51	51.6
North Vancouver number of calls serviced	269	269	269	269	269
West Vancouver ambulance service time	66	75.03	76.2	94.83	94
West Vancouver number of calls serviced	57	57	57	57	57
Horseshoe Bay subregion response time for all calls	23.08	21.68	22.01	20.98	19.97
North Vancouver ambulance response time for transfer calls	16.9	16.9	16.9	16.9	16.9
North Vancouver ambulance response time for normal calls	15.89	15.89	15.89	15.89	15.89
West Vancouver ambulance response time for transfer calls	31.03	31.53	31.54	32.99	32.35
West Vancouver ambulance response time for normal calls	15.15	15.95	16.28	19.01	17.47

TABLE 6.1

6.3 THE EFFECTS OF REMOVING THREE NIGHT CREWS

Experiment #2 consisted of two separate simulation runs each with three night crews removed. The evening shift runs from 12 midnight until 7:30 a.m. There are 15 ordinary ambulances and two paramedic crews on duty during the night shift. The depot locations and corresponding shifts are stated in Table 6.2. Paramedic crews were not removed in the experiment. The depots where ambulances are removed are in nodes 6, 23 and 166 in the first experiment and nodes 15, 31 and 166 in the second run. (See figure 6.1 for node locations.) The results gathered are listed in Table 6.3.

The removal of three night crews causes dramatic results. All average response times for all calls and in all regions of the G.V.R.D. increase significantly.

Let us begin examination of these results by stating that all nodes where ambulances were removed did have a high demand. Node 23 is a downtown node and is near St. Paul Hospital. Node 6 is a residence region with the V.G.H. and Shaunagsy nearby. Node 166 is where the Langley ambulance is. This ambulance has a very large spatial territory to patrol. Node 15 is residential with the Shaunagsy and V.G.H. nearby and is less than ten minutes from downtown Vancouver. At node 31, the demand rate is also high as it is in the heart of the "red light" district. Thus the removal of an ambulance readily

puts a strain on nearby crews.

With other crews having to pick up the slack, their availability to service other calls decreases. Consequently, more calls need to wait for an ambulance to become free. The increased queuing time results in increased response time.

The most significant increase in average response time is for ordinary (priority II) calls. Vancouver and Burnaby are hardest hit with increases. Since the majority of calls are in these two regions, this result is not surprising. The amount of increase is. Vancouver has the greatest proportion of ordinary calls. Further, the largest percentage of all calls are non-emergency calls. Two Vancouver night crews are removed in each simulation run, leaving only five crews in Vancouver on the night shift. This is a 28% decrease in ambulance crews for Vancouver. These remaining five crews answer emergency calls first. Then if no other emergency calls are on queue, the ordinary call is responded to.

With only five crews available each crew answers more calls decreasing its availability. Combining this fact with the fact that most calls are ordinary calls, one can readily appreciate the significant increase in Vancouver's average response time.

Burnaby's average response time for ordinary calls also increased due mainly to only five crews available in Vancouver. Obviously Vancouver needed help and got some from Burnaby.

Increased travel time for Burnaby increases its response time. Also, Burnaby has one paramedic serving in the evening and itself has a fair number of ordinary calls. Calls in Burnaby suffer as a result of the Burnaby ambulance travelling to Vancouver. This includes both non-paramedic and paramedic calls.

The paramedic unit in Burnaby helps out on ordinary calls when it is not busy. The removal of a night crew in Vancouver makes the Burnaby crew work more, creating more work for the paramedic unit. As a result, the average response time for the paramedic unit is augmented. The chain reaction can be extended to include the ordinary and paramedic unit in New Westminster. Also with less crews in Vancouver, a paramedic crew may have to travel there to answer a paramedic call. With the paramedics answering more ordinary calls, the paramedic call in Vancouver may get queued resulting in a long wait.

The increased work for all units causes an increase in emergency response time, but not as significantly. This is due to the high priority and low number of emergency calls. Response time for rural regions increases significantly. The large increase in rural emergency response time can be attributed to the removal of a unit at node 166. It was the only night crew in Langley. Now the closest ambulances are in White Rock and Surrey who, while covering their own large regions, are required to travel the long haul to cover for the absence

of the Langley ambulance. Further, Surrey and White Rock no longer get help from the ambulance at Langley. If they get any help it would be from the ambulance at New Westminster. The large spatial area required to cover coupled with the loss of 1/3 of the ambulance crews explains the significant increase in average rural emergency response times. It goes without saying that the average ordinary response time increased.

This experiment strongly indicates that the removal of crews from the night shift is not viable. The reaction is far reaching. If anything possible crews need to be added and experiments can be easily done with the simulation program.

The second result of the experiment is that concerning paramedic crews. In order for these crews to be most effective in saving lives in cases such as cardiac arrests, they must be available as much as possible. Removal of night crews in Vancouver or Surrey results in less availability.

EFFECTS OF REMOVING THREE NIGHT CREWS

(Results in Minutes)	Present	Remove Nodes 6,23,166	Remove Nodes 15,31,166
Overall emergency response time	6.21	7.61	7.78
Urban emergency response time	6.08	7.26	7.43
Rural emergency response time	6.77	9.21	9.35
Vancouver emergency response time	5.80	7.29	7.50
Burnaby emergency response time	6.04	6.87	7.74
New Westminster emergency response time	6.19	6.58	6.58
Overall normal* response time	13.00	17.99	18.64
Vancouver normal response time	10.99	16.44	17.44
Burnaby	11.85	18.48	18.15
New Westminster	16.56	17.48	20.60
Overall paramedic response time	6.09	7.41	7.11
Vancouver paramedic response time	5.26	6.59	6.43
Burnaby	6.38	17.02	8.05
New Westminster	6.29	7.84	8.43

TABLE 6.3

* Priority 2 calls - non-emergency

6.4 MOVING A PARAMEDIC AMBULANCE FROM BURNABY TO VANCOUVER

In experiment #3 a Burnaby paramedic crew, node 62, is relocated at the Vancouver General Hospital (V.G.H.), node 14. Presently there are no paramedic cars in Vancouver. The paramedic unit was to help out on other calls when it was not busy and was given a 15 minute radius maximum for answering paramedic calls. Paramedic calls outside the 15 minute maximum would be treated as emergency calls, answered by other vehicles.

Currently, the Vancouver district has the highest percentage of all types of calls, including emergency. Some of these emergency calls are paramedic type but were treated as emergency since no paramedic ambulances were available to answer the call. (An experiment done by the dispatchers labelling these calls bore out this fact.) Any paramedic vehicles travelling to Vancouver came from far away Burnaby. Victims of cardiac arrests in Vancouver cannot wait for a paramedic crew to come from Burnaby. For paramedic crews to be effective, response time must be as low as possible, ideally within three minutes.

Adding a paramedic ambulance at the V.G.H. gives Vancouver one more all day crew, four daytime and two night time at node 14. Burnaby, node 62, would be reduced to two day crews and no night crews. All other factors remain constant. The results have been tabulated in Table 6.4.

Two significant results arise. First, both the average emergency and paramedic response times increase. Second, the average response times for ordinary and transfer calls are reduced.

Response time for emergency calls increases due to the loss of an all day ambulance in Burnaby. Without the paramedic unit at node 62, average response time for emergency and paramedic calls is increased significantly in Burnaby. This indicates how much work the paramedic unit was doing in Burnaby. Now the remaining cars have to pick up for the loss and are having difficulty keeping up with the increased workload. Further, the fact there is no car in Burnaby at night probably significantly affects Burnaby's average response time. During the day, this ambulance can be covered for.

At night, ambulances must come from Vancouver to answer a call in Burnaby. The increased travel time directly augments response and total service time. Thus Vancouver's average emergency response time suffers. Also, North Vancouver suffers an increase in its emergency response time. This is possibly due to the ambulance in North Vancouver crossing the second narrows bridge to answer a call in Burnaby during either the busiest part of the day when most ambulances are occupied, or at night when no ambulance is available.

The increase in Burnaby's and North Vancouver's emergency response time is not fully balanced by the reduced response

time experienced in Vancouver resulting in an overall appreciation in the average emergency response time.

Vancouver is benefited by the additional unit at the V.G.H. Both its emergency and paramedic response time are reduced, although marginally. Most of this result can be attributed to the very high demand for all calls in Vancouver. All ambulances in this municipality are busy and their availability is low. It must be kept in mind that the paramedic unit helps out on ordinary calls.

A reduction in the average paramedic response time took place in New Westminster. Most likely the paramedic unit was no longer required to make any long trips to Vancouver. The increase in the West Vancouver paramedic response time is mainly due to the 15 minute radius for the Vancouver paramedic. Previously paramedic calls in West Vancouver may have been answered by the Burnaby unit. Since that unit no longer exists there either the West or North Vancouver ambulance answers the paramedic calls. Since there are few paramedic calls in that region, it is not likely that the paramedic call got tied up in the queuing list causing the increased response time.

In order to improve the paramedic average response time, it appears that more of these types of vehicles are necessary.

It is also seen in Table 6.4 that the average response time for ordinary and transfer calls is reduced. With the high number of ordinary calls and transfer calls in Vancouver and the

additional vehicle helping out on ordinary calls, this result is not surprising. The majority of calls are ordinary calls and most ordinary calls occur in Vancouver. An additional ambulance takes some of the workload off of the other cars. As a result, all cars are more available to service calls and consequently, less calls need be queued. Thus average response time for ordinary and transfer calls is reduced.

In summary the addition of a paramedic unit at the V.G.H. helps Vancouver residents but Burnaby suffers somewhat. It may be worthwhile to have an all day vehicle in Burnaby. This would help out both Vancouver and Burnaby, especially on the evening shift.

TABLE 6.4 Moving paramedic vehicle from Burnaby (node 62) - VGH (node 14)

	Base Run	Move Para at (62) - Para at (14)
Average Transfer Response Time	20.46	20.36
Avg. Ordinary Response time	12.73	12.71
Avg. Emergency Response Time	6.33	6.41
Avg. Paramedic Response Time	6.08	6.17
Vancouver Emergency R. T.	5.86	5.84
North Van. Emerg. R. T.	5.91	6.82
West Van. Emerg. R. T.	7.39	7.39
Burnaby Emerg. R. T.	6.44	7.29
New West. Emerg. R. T.	5.81	5.81
Vancouver Para. R. T.	5.18	5.12
North Van. Para. R. T.	5.15	5.15
West Van. Para. R. T.	4.14	6.28
Burnaby Para. R. T.	6.91	7.62
New West. Para. R. T.	6.29	6.23

6.5 REPLACING AT V.G.H. AN ORDINARY (EMA-2) AMBULANCE WITH A PARAMEDIC (EMA-3) AMBULANCE UNDER TWO CONDITIONS

For the fourth experiment a paramedic unit replaces an all day ordinary unit at node 14. Presently there are no paramedic cars in Vancouver. The loss of an EMA-2 vehicle in Vancouver leaves the municipality with ten EMA-2 cars serving during the day shift and only six EMA-2 units on the night shift. No other municipalities experience any changes in the volume of vehicles.

The EMA-2 ambulance at node 14 answered all types of calls. For this experiment the paramedic ambulance has two dispatch policies:

1. the paramedic crew only answers emergency and paramedic calls, and

2. the paramedic crew helps with ordinary (priority II) calls when it is not busy. It does not answer any transfer calls.

Two simulation runs were required, one with policy (1) and the other under policy (2). With a paramedic in Vancouver, more emergency type calls will be deemed as requiring a paramedic crew. Previously, since Vancouver had no paramedic, these calls were serviced as emergency calls. A study done by the dispatchers gave the proportion of paramedic calls per day in Vancouver. Thus in the simulation, a certain percentage of emergencies were given paramedic status to more accurately represent the real situation. The resulting statistics are

stated in Table 6.5.

Under condition (1), all call types, except emergency, show an increase in average response time. Emergency type calls show a marginal decrease in average response time. The reason for this is apparent. Vancouver has lost one all day vehicle to answer non-emergency calls. Having an EMA-2 unit which handles all types of calls relieves the pressure on other ambulances, freeing them up to help out on paramedic and emergency calls more. Vancouver has the majority of ordinary and transfer calls. Even with the EMA-2 unit, the vehicles are quite busy, especially during the middle part of the day when the greatest majority of transfer calls occur. Now with one less EMA-2 vehicle the ordinary and transfer calls are being queued up even more. Further, on the night shift the six EMA-2 vehicles have to cover up for the lost vehicle. The higher utilization of the remaining EMA-2 cars reduces their availability and as a result, average response time for non-emergency calls appreciates.

On a community by community basis for emergency response time, most municipalities are unaffected. Vancouver, Burnaby and New Westminster experience marginal improvements in their emergency response time. The explanation here is that the paramedic unit at node 14 only answers emergency and paramedic calls. Its availability is thus greater than that of an EMA-2 unit. Less emergency calls as a result are queued and average

response time for emergency calls improves in Vancouver.

Ambulances in Burnaby increase their availability slightly since long trips into Vancouver are no longer necessary. The vehicles have a smaller radius to cover and this directly improves their response time. New Westminster's vehicle now no longer have to travel into Burnaby to help out there. Thus the policy employed by the EMA-3 vehicle at V.G.H. results in three municipalities lowering their average response times for emergency calls.

Paramedic response time, on the contrary, is increased. By having the simulation create more paramedic calls, the number of these types of calls queued is increased. With an increase in paramedic calls in Vancouver, paramedic response time is increased. In New Westminster, the paramedic response time is also increased. With a higher number of paramedic calls in Burnaby, the EMA-3 unit in New Westminster is helping out there. Travel time is increased. With help from New Westminster, Burnaby's average paramedic response time is reduced slightly.

This policy has some implications. First, the high number of transfer calls in Vancouver creates a big problem when a vehicle is taken away. Perhaps some other form of transportation is needed for these types of calls. Special cars may be added to handle transfer calls during the 11 a.m. to 2 p.m. shift when most transfers are released from the hospitals. Second, as more paramedic units are added, more paramedic calls

occur. The public requests for paramedics increases. The screening for these types of calls is critical in order for the EMA-3 vehicles to be of utmost effectiveness.

For the second condition, the new EMA-3 unit gives first priority to emergency and paramedic calls and only answers ordinary (priority II) calls when all other nearby ambulances are busy. The situation is especially criticized during the day shift. The paramedic unit must help out due to the high frequency of calls during the day and there is still a loss of an ordinary unit. The EMA-3 does not answer any transfer calls. With the paramedic unit helping on ordinary calls (the majority of calls) the emergency response time in Vancouver increases, from the base run. Burnaby has a very minute increase in its emergency response time. All other areas are scarcely or not at all affected. The result is an increase in the overall emergency response time. The majority of emergency calls occur in the Vancouver region.

The average response time for ordinary calls is decreased with the help of the paramedic unit. Ordinary ambulances in Vancouver are free to answer ordinary calls since the paramedic is mainly servicing emergency calls. Other areas are not affected.

Paramedic response time in Vancouver, Burnaby and New Westminster increases. North Vancouver experiences a drop in its paramedic response time due to a paramedic unit being

close by in Vancouver. The increase in average paramedic response time for the three high demand municipalities can be attributed to the increase in the number of paramedic calls. In Vancouver, with the EMA-3 unit participating in servicing ordinary calls, the increase is more apparent. Further implications of the Vancouver EMA-3 unit helping out on non-emergency calls are that Vancouver may now be getting help from the Burnaby EMA-3 crew. Burnaby's response time increases. As a result, the paramedic in New Westminster may have to travel in Burnaby. Overall, everybody's average response time for paramedic calls increases.

These results indicate that

1. more EMA-2 vehicles are required in Vancouver, especially during the day shift,

2. a study should be conducted to see if some other form of transportation for transfer patients can be obtained,

3. the effective use of paramedic vehicles requires effective screening of calls by dispatchers,

4. Vancouver does require paramedic vehicles, and

5. with the adaptation of policies (2) and (3), the most effective use of paramedics would be for them to only answer emergency and paramedic type calls.

TABLE 6.5

Replace VGH EMA-2 w/EMA-3 under two conditions	Base Run	Change VGH All-Day Ordinary Ambulance into Paramedic Ambulance	Change VGH All-Day Ordinary Ambulance into Paramedic Ambulance
		Paramedic Amb. Answers Paramedic & Emergency Calls Only	Paramedic Ambulance Helps w/Ordinary Calls When Not Busy
Average Transfer Response Time	20.46	21.48	20.73
Average Ordinary R. T.	12.73	14.15	12.61
Average Emergency R. T.	6.33	6.27	6.36
Average Paramedic R. T.	6.08	6.59	6.70
Vancouver Emerg. R.T.	5.86	5.83	5.91
North Vancouver Emerg. R.T.	5.91	5.91	5.91
West Vancouver Emerg. R. T.	7.39	7.39	7.39
Burnaby Emerg. R.T.	6.44	5.99	6.46
New West. Emerg. R. T.	5.81	5.60	5.81
Vancouver Para. R. T.	5.18	6.15	6.27
North Van. Para. R. T.	5.15	5.21	4.90
West Van. Para. R. T.	4.14	4.14	4.14
Burnaby Para. R. T.	6.91	6.75	7.08
New West. Para. R. T.	6.29	6.68	6.58
No. of Para. Calls Answered by EMA-3	176	277	280

6.6 EXPERIMENT #5 - OPTIMAL NUMBER OF EMA-2 UNITS REQUIRED TO ACHIEVE A GIVEN FRACTILE OF EMERGENCY CALLS ANSWERED IN LESS THAN FIVE MINUTES

This experiment proceeded in two stages. First a number of experiments were run to optimally locate ambulances to service night shift demands. These locations were initially obtained using the p-median model. The solution set from the p-median was then implemented in the simulation. Experiments with the simulation lead to some reallocation of the depots taking the impact of congestion into account. Initially the idea was that only one ambulance per base would be required for the night shift. But, very heavy demand rates in some urban nodes made it more efficient to stack ambulances at some bases. (See figure 6.11) Alternatively one could subdivide the nodes, but this could result in ambulances a few blocks apart. In rural areas one ambulance per base remained appropriate (figure 6.100. While stacking was necessary, ambulances were added to make the number of cars at a base roughly proportionate to demand. Figures 6.8 and 6.9 provide the fractile response surfaces for emergency and paramedic alls, respectively. In figure 6.8, for instance, the 60th fractile line has been estimated by interpolation. For any combination of bases and ambulances on this line one would expect about 60% of the calls to be answered in five or less minutes, provided the bases were optimally located and the ambulances optimally allocated.

Only a limited range of alternatives were examined due to the heavy expense of computation.

Examination of figure 6.8 reveals some interesting facts. The present number of ambulances and bases, 19, answer 42.2% of all emergency calls in five minutes or less at night.

TABLE 6.6

EMERGENCY AMBULANCE CALLS ANSWERED IN 5 OR LESS MINUTES (NIGHTS ONLY)

# of Depots	# of Ambulances	Added Ambulances	% Increases	Added	% Increase
19	19	4 (23)	7.1%	12(31)	6.1
23	23	4 (27)	12.6%	12(35)	3.2
27	27	4 (27)	7.4%		
31	31	4 (31)	8.1%		

TABLE 6.7

# of Ambulances	Bases	Added Bases	% Increase
23	23	4 (27)	- 2.3%
27	23	4 (27)	- 4.5%
31	27	4 (31)	- 1.9%
35	31	4 (35)	- 5.2%

Adding four ambulances to the 19 depots, 19 ambulances combination results in an 7.1% increase in the number of emergency calls answered. When another eight ambulances are added the marginal increase is 6.1%. This result is explained by the fact that when the first four ambulances were added they took

a good deal of pressure off the first original 19 ambulances. With the addition of eight more ambulances most calls nearby the depots were taken care of quickly but there still were not enough ambulances in rural areas when travel times are longer. Looking at the case where one starts with 23 depots and 23 ambulances and increases the number of cars to 27, one gets an increase in the number of calls answered within five minutes of 12.6%. If eight more cars are added the percentage increase is only 3.2%. Here again the effect of diminishing marginal returns is apparent. The addition of four cars helps immensely in the urban areas. To reduce response time further one needs the placement of ambulances in rural areas only a few miles apart. The number of calls in rural areas, however, does not justify having ambulances in close proximity to each other. All figures in Table 6.6 and Table 6.7 support stacking of ambulances in urban areas. The demand is heaviest in urban areas and stacking of ambulances there results in at least a 7.1% increase in the number of emergency calls answered in five minutes or less. Figure 6.8 also reveals that going from the 50th fractile curve to the 65th fractile the curve is becoming flatter. Perhaps an alternative measurement would be to split the urban emergency and rural emergency response times. Service Standards for both of these areas would have to be developed by a person from the Emergency Health Services.

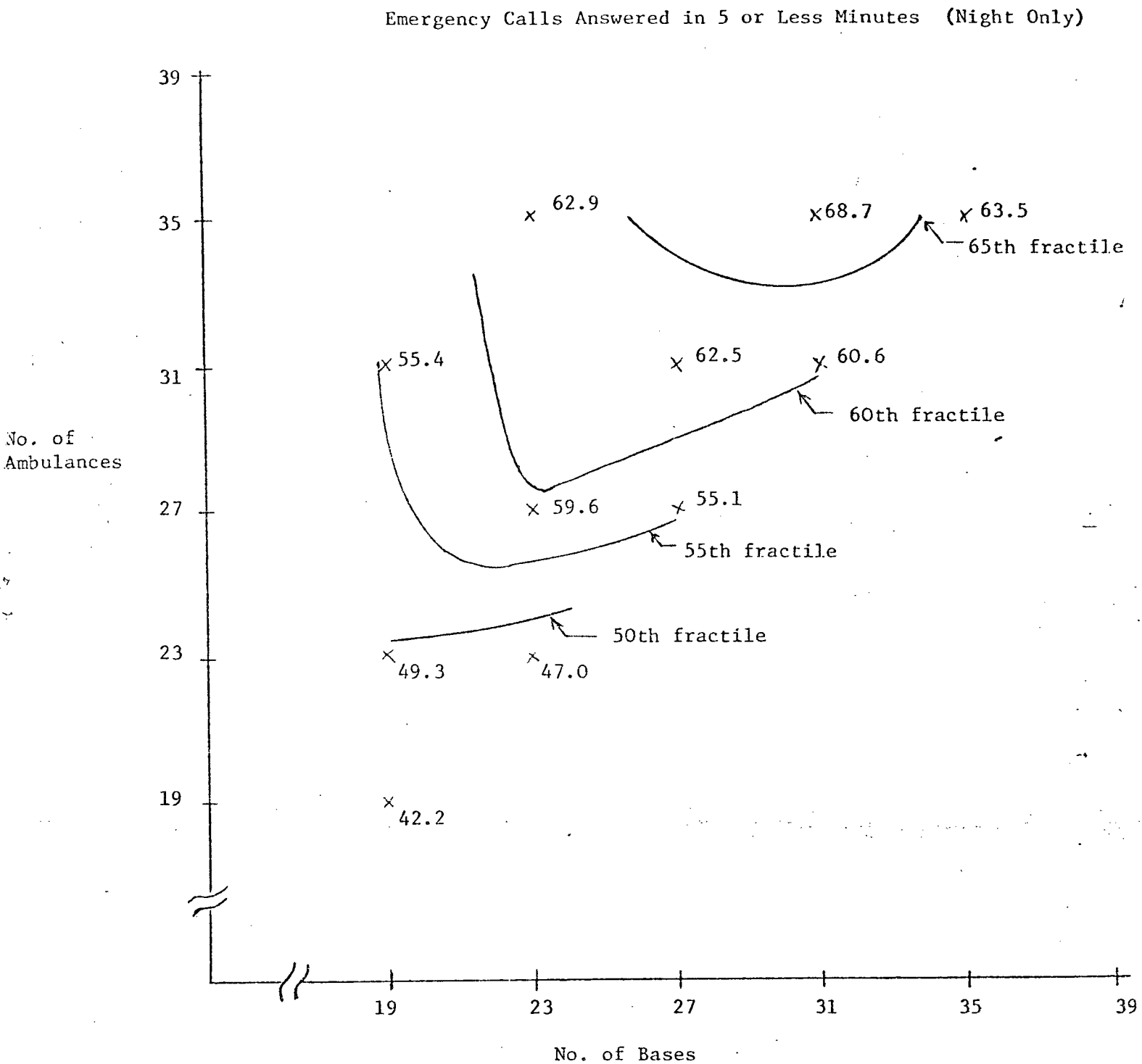

Emergency Calls Answered in 5 or Less Minutes (Night Only)
39
35
31
27
23
19
No. of Ambulances
62.9
68.7
63.5
65th fractile
55.4
62.5
60.6
60th fractile
59.6
55.1
55th fractile
50th fractile
49.3
47.0
42.2
19
23
27
31
35
39
No. of Bases

FIGURE 6.8

In figure 6.9 the fractiles for paramedic calls is depicted. Again stacking of ambulances in urban areas has a big effect. With 23 ambulances and 19 bases only 41.6% of all emergency calls are answered in five minutes or less. Adding eight cars increases the percentage to 49.9%. For the 23 cars and 23 bases the percentage of calls answered within five minutes is 52.4%. Adding 12 ambulances increases the number of calls answered by 12.5%.

In figure 6.9 it can be seen that the 70th fractile curve is flat compared to the other fractiles. At this fractile it is becoming evident that to get any significant increase in paramedic calls answered within five minutes a great number of cars and bases would need to be added.

Again it can be said that perhaps urban and rural regions should be examined separately. Another alternative can be that certain cars be designated to answer only emergency calls and/or only paramedic calls.

Table 6.8 reveals that with nine rural ambulances only 53.7% of rural calls are answered in less than five minutes. Further increases of 3, 4 and 2 ambulances show diminishing marginal increases although doubling of the number of ambulances has an overall effect of 18.5%.

In figure 6.10 it can be seen that increasing the number of ambulances from 40 to 60 has no effect on the rural response time for emergency calls. No effect is also observed when

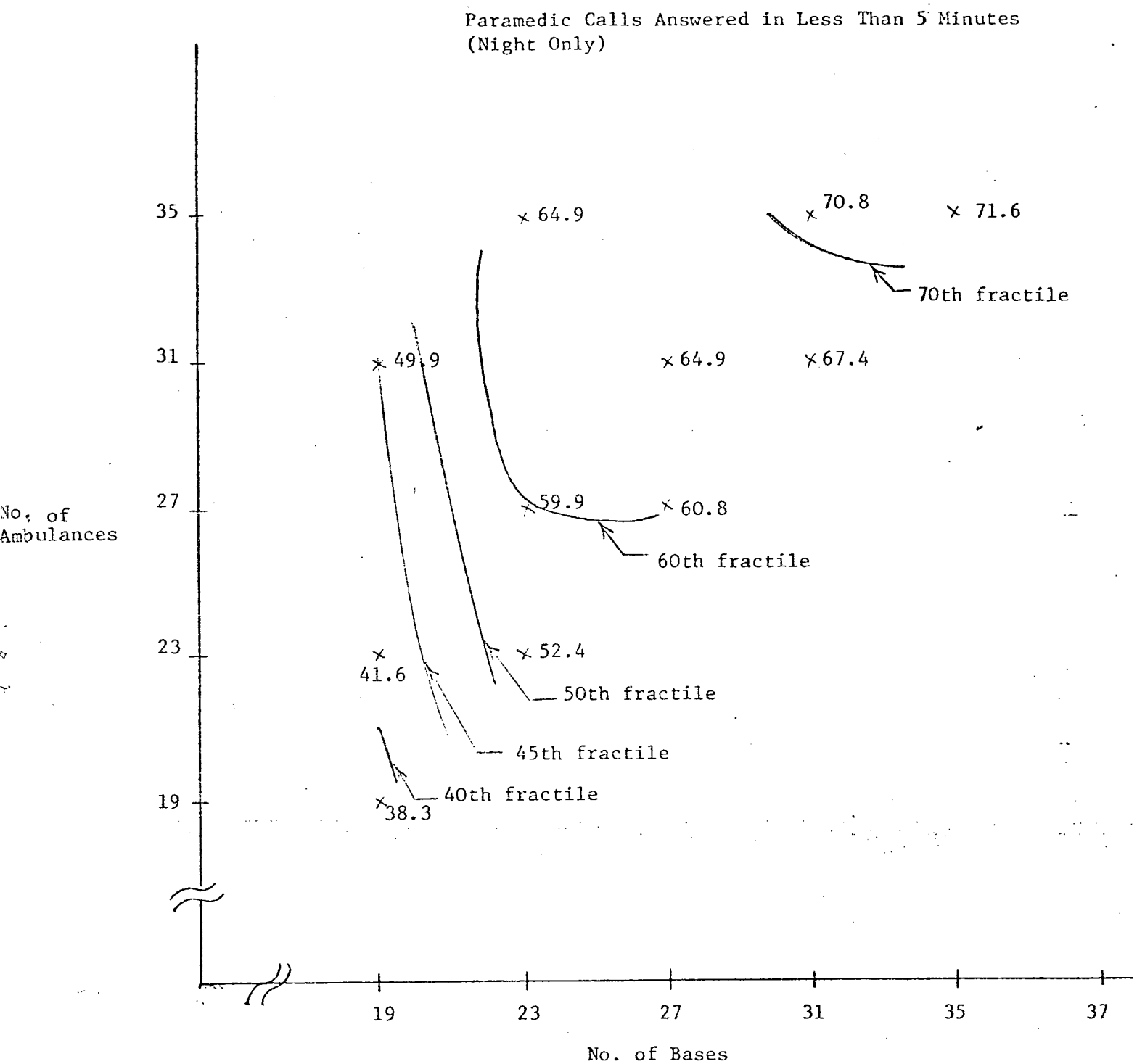

Paramedic Calls Answered in Less Than 5 Minutes
(Night Only)
No. of Ambulances
35
31
27
23
19
64.9
70.8
71.6
70th fractile
49.9
64.9
67.4
59.9
60.8
60th fractile
52.4
41.6
50th fractile
45th fractile
40th fractile
38.3
19
23
27
31
35
37
No. of Bases

FIGURE 6.9

FRACTILE RESPONSE IN RURAL AREAS (NIGHT ONLY)

Number of Rural Ambulances	Fractile Under 5 Minutes	% Increase
9	53.7	7.4
12	61.1	5.5
16	66.6	5.6
18	72.2	

TABLE 6.8

Fractile Response Under 5 Minutes for Rural Emergency Calls
(Night Only)

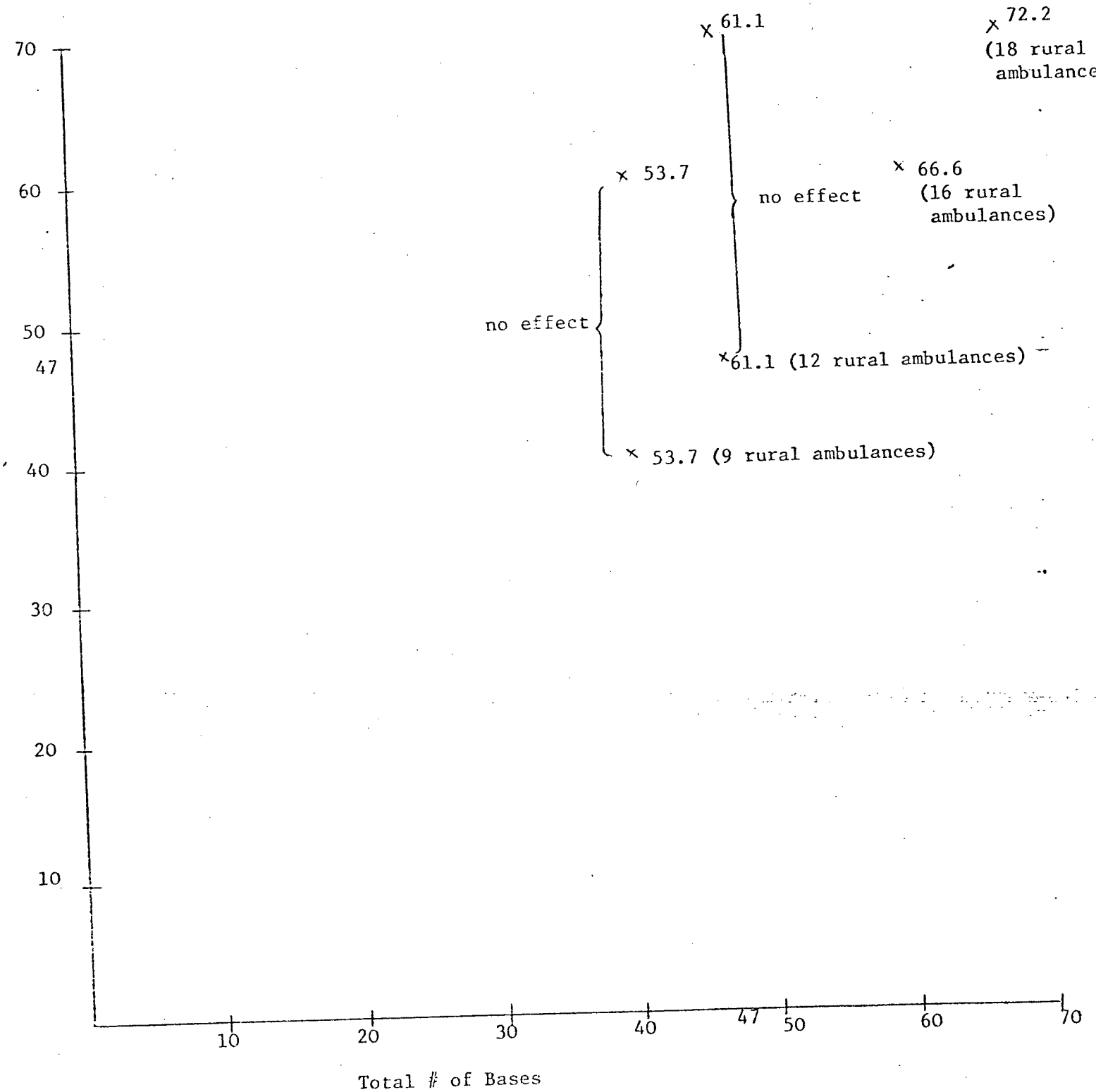


FIGURE 6.10

with 47 bases the number of ambulances is increased from 47 to 70. The only alternative to reducing rural response time is to increase the number of rural ambulances as indicated in figure 6.10. With 70 ambulances in the region and only 12 in rural areas the number of calls answered is 61.1%. Transferring six of the urban cars to rural areas increases the number of rural emergency calls answered to 72.2%. Whereas buses and ambulances are required in rural areas only cars help in urban areas.

Figure 6.11 illustrates the stacking effect of ambulances in urban areas. A 20 car increase at the 40 base level results in a 10.9% increase in emergency calls answered and a 23 car increase at the 47 base level results in a 12.9% increase in the number of emergency calls answered within five minutes. Increasing the number of bases at the 60 and 70 ambulance level results in a decrease in emergency calls answered.

Comparing the 47 base and 70 ambulance level between emergency rural calls and emergency urban calls one sees that only 61.1% of emergency rural calls versus 84.3 emergency calls are answered. The reason is obvious. Large distances in rural areas cause large response times. Urban areas have a heavy demand problem. Stacking is an effective means of reducing emergency response time in urban areas.

Having obtained the response surface in 6.8, it is now possible to pick likely areas for further investigation of the

Fractile Response Under 5 Minutes for Urban Emergency (Night Only)

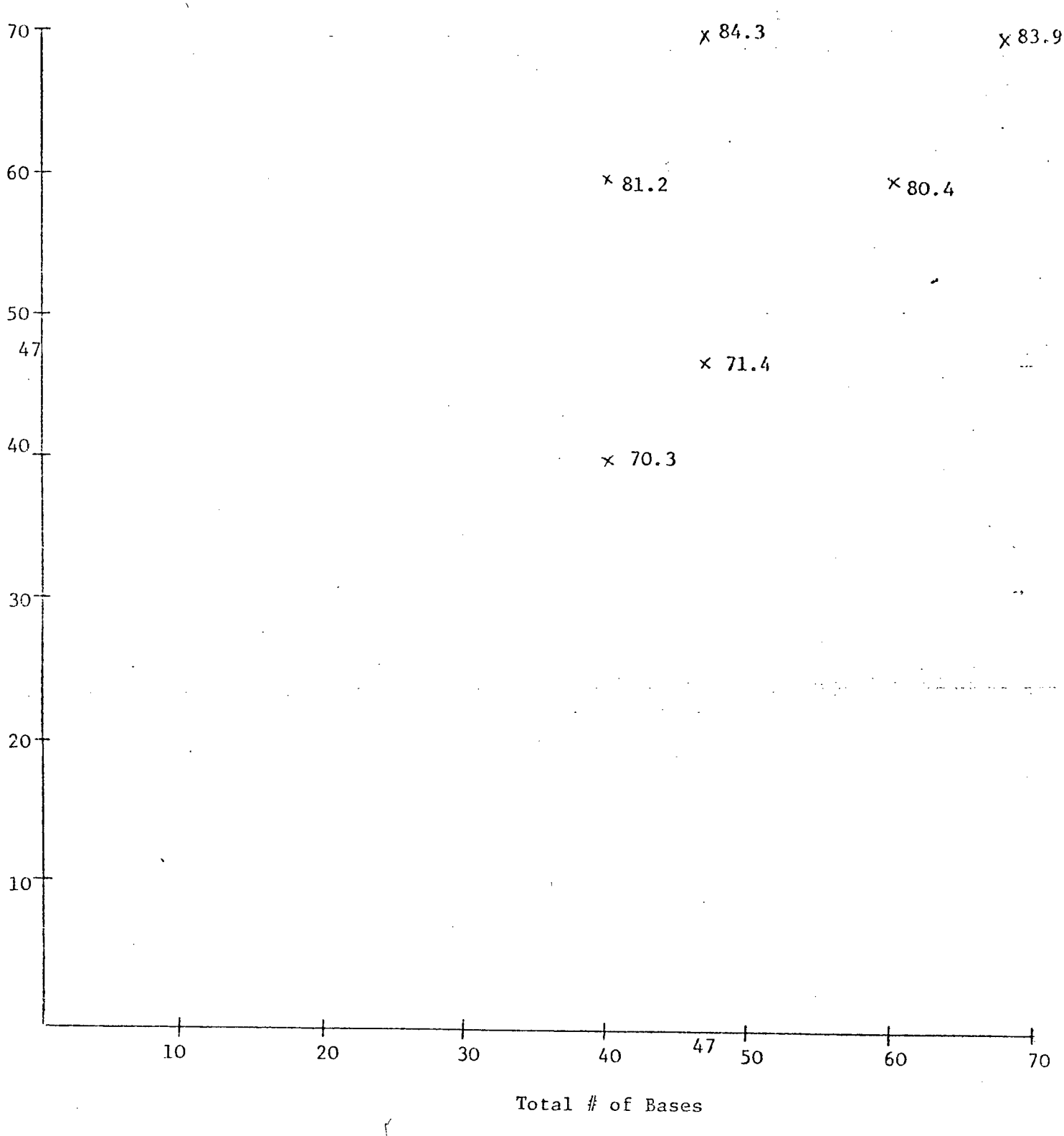


FIGURE 6.11

number of ambulances to assign to gain equivalent fractile responses over the entire day. The appropriate testing point, that which balances the number of bases and the number of ambulances on any given fractile contour line, must be decided by the ambulance management. For the purposes of this experiment, several combinations were chosen as starting points for all-day runs. Starting from the 23-base 27-ambulance combination, the addition of 14 ambulances resulted in an all-day five-minute fractile for emergency calls of 54.8%, a drop of 4.8% over the night shift. With a start of 27 bases and 27 ambulances, 14 more ambulances caused a drop of 2.4%, with a start of 27 bases and 31 ambulances, the drop was 0.1%.

The important observation is that the night shift response surface and the all-day response surface have similar properties. During the day shift the frequency of calls increases but the proportion of emergency calls is not large. There is a significant increase in the number of transfer calls but the frequency of other types of calls do not increase significantly.

It is becoming evident from this experiment and others that either more ambulances or specialized ambulances are needed in urban areas. With the high frequency of urban calls certainly more ambulances are needed. To further reduce response time for emergency type calls (including paramedic calls) perhaps a number of cars should only be used to answer emergency calls. Another alternative is to obtain other forms of trans-

portation for transfer patients. As the population in the G.V.R.D. grows and demand increases the ambulance service must examine innovative alternative policies in order to remain effective. Some alternative policies are discussed in the following sections.

6.7 HIGH URBAN DENSITY

Congestion is the major problem of the present ambulance system. There are not enough ambulances in either the day shift or night shift to meet demand. As a result ambulance availability is low and the number of queued calls is significant. As population increases demand will increase. Some policies to further investigate are:

1. Set separate goals for urban and rural response times. By dividing up the region one can focus more effectively on the problems of one area.

2. Increase the number of vehicles as indicated by the results of experiment #5 for the day shift. The optimal number will have to be decided by ambulance service management.

3. Use of other than ambulance personnel to transport transfer patients. This would alleviate much congestion between the hospital discharge hours from 11 a.m. to 2 p.m.

6.8 PARAMEDICS - FUTURE USE

From the experiments conducted and data gathered it is evident that there is strong demand for more paramedic units.

Presently in the G.V.R.D. there are two paramedic units. These units are being used for other types of calls when other vehicles are busy. Another factor which results in loss of time for paramedics is the high number of cancelled calls these vehicles answered.

In order to efficiently use the paramedic units effective screening of calls must first take place. This may take place when dispatches become more experienced. A second policy is to use paramedics only for designated paramedic calls. In order to facilitate obtaining this goal more ambulances are needed. Finally more paramedic ambulances are required. By gathering data from dispatchers on where and when paramedics are required one can determine through the simulation where these vehicles should be located. Placements of paramedic crews at hospitals is a useful policy since paramedics are well qualified to help out in emergency rooms.

6.9 SUMMARY

The ambulance system is a very complex system. In order to study it effectively many factors must be taken into account. These factors include: the frequency of calls, the spatial and time distribution of calls, the types of calls and the operational procedures of ambulances. Simulation is a means of taking into account these factors and the inter-relationships involved.

In the G.V.R.D. congestion is a major problem. Using simulation one is able to investigate the problem effectively and try out different strategies to see how problems can be dealt with. From the five experiments done in Chapter 6 of this study, it was demonstrated how simulation can be used to examine individual regions, various periods, and the total system.

Areas of concern revealed by the simulation are the high frequency of urban calls, the high number of transfer calls during the day shift and the increasing demand for paramedic vehicles.

BIBLIOGRAPHY

1. Berlin, G.N., Leibman, J.C. "Mathematical Analysis of Emergency Ambulance Location" Socio-Economic Planning Sciences Vol. 8, Dec. 1974.

2. Berlin, G.N., Revelle, C., Elzinga, J. "Determining Ambulance Locations for On-Scene and Hospital Care" Environmental Planning Association Vol. 3, Aug. 1976.

3. Carbone, R. "Public Facilities Location Under Stochastic Demand" Infor Vol. 12 No. 3, Aug. 1976.

4. Chaiken, J., Larson, R. "Methods for Allocating Urban Emergency Units: A Survey" Management Science Vol. 19, Dec. 1972.

5. Church, R., Revelle, C. "The Maximal Covering Location Problem" Papers of the Regional Science Association Vol. 32, Fall 1974.

6. Cooper, L. "Location-Allocation Problems" Operations Research, Vol. 11, No. 3 (May-June 1963) pp. 333-343.

7. Daverkow, S.G. "Location and Cost of Ambulances Serving a Rural Area" Health Services Research Vol. 12 No. 7, Fall 1977.

8. Fitzsimmons, J. "A Methodology for Emergency Ambulance Deployment" Management Science Vol. 19, Feb. 1973.

9. Garfinkel, R.S., Neebe, A.W., Rao, M.R. "An Algorithm for the M-Median Plant Location Problem" Transportation Science Vol. 8, Aug. 1974.

10. Groom, K.N. "Planning Emergency Ambulance Services" Operational Research Quarterly Vol. 28 No. 3 II.

11. Hakimi, S. "Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph" Operations Research Vol. 12, May-June 1974.

12. Hakimi, S. "Optimum Distribution of Switching Centers in a Communications Network and Some Related Graph Theoretic Problems" Operation Research Vol. 13, June 1965.

13. Khumawala, B.M. "An Efficient Algorithm for the P-Median Problem with Maximum Distance Constraints" Geographical Analysis Vol. 5, Oct. 1973.

14. Khumawala, B.M., Neebe, A.N., Dannenbring, D.G. "A Note on Elshaieb's New Algorithm for Locating Sources Among Destinations"

15. Kleinman, J.C., Wilson, R.W. "Are 'Medically Underserved Areas' Medically Underserved?" Health Services Research, Summer 1977.

16. Kolesar, P., Blum, E.H. "Square Root Laws for Fire Engine Response Distances" Management Science Vol. 19 No. 12, Aug. 1973.

17. Krarup, J., Pruzun, P.M. "Selected Families of Discreet Location Problems Part IIIK: The Plant Location Family" Working Paper No. WP-12-77, Faculty of Business, University of Calgary, Aug. 1977.

18. Kuhn, H.W., Kuenne, R.E. "An Efficient Algorithm for the Numerical Solution of the Generalized Weber Problem in Spatial Economics" Journal of Regional Science, Vol. 4 (1962) p. 21-34.

19. Revelle, C.S., Marks, D., Leibman, J. "An Analysis of Private and Public Sector Location Problems" Management Science Vol. 16.

20. Revelle, C.S., Swain, R. "Central Facilities Location" Geographical Analysis Vol. 2, Jan. 1970.

21. Revelle, C.S., Toregas, C., Falkson, L. "Applications of the Location Set-Covering Problem" Geographical Analysis Vol. 8, Jan. 1976.

22. Rojeski, P., Revelle, C.S. "Central Facilities Location Under an Investment Constraint" Geographical Analysis Vol. 2, 1970.

23. Rushton, G., Goodchild, M., Ostresh, L. "Computer Programs for Location-Allocation Problems" Monograph No. 6, Department of Geography, University of Iowa, July 1973.

24. Savas, E. "Simulation and Cost-Effectiveness Analysis of New York's Emergency Ambulance Service" Management Science Vol. 15, Aug. 1969.

25. Schüler, R.E., Holaman, W.L. "Optimal Size and Spacing of Public Facilities in Metropolitan Areas: The Maximal Covering Location Problem Revisited" Papers of the Regional Science Association Vol. 39, Summer 1978.

26. Scott, A.J. "Location-Allocation Systems: A Review" Geographical Analysis, 1970.

27. Siler, K.F. "Level-Load Retrieval Time: A New Criterion for EMS Facility Sites" Health Services Research, Winter 1977.

28. Swain, R.W. "A Parametric Decomposition Approach for the Solution of Uncapacitated Location Problems" Management Science Vol. 21, Oct. 1974.

29. Swoveland, C., Uyeno, D., Vertinsky, I., Vickson, R. "Ambulance Location: A Probabilistic Enumeration Approach" Management Science Vol. 20 No. 4, Dec. 1973.

30. Tan, E. "A Strategy for Ambulance System Designs: An Investigation of the Ambulance System in the Greater Vancouver Regional District" Master's Thesis, University of British Columbia, 1974.

31. Teitz, M., Bart, P. "Heuristic Methods for Estimating the Generalized Vertix Median of a Weighted Graph" Operations Research Vol. 16, Sept.-Oct., 1968.

32. Toregas, C., Revelle, C.S. "Location Under Time or Distance Constraints" Papers of the Regional Science Association Vol. 28, Fall 1972.

33. Toregas, C., Revelle, C.S. "Binary Logic Solutions to a Class of Location Problems" Geographical Analysis Vol. 5, April 1973.

34. Toregas, C., Swain, R.W., Revelle, C.S., Bergman, L. "The Location of Emergency Service Facilities" Operations Research Vol. 19, Oct. 1971.

35. Weber, A. translated as "Alfred Weber's Theory of Location of Industries" by C.I. Friedrich, Chicago, 1929.

36. Whitaker, R.A. "An Algorithm for Estimating the Medians of a Weighted Graph Subject to Side Constraints, and an Application to Rural Hospital Locations in B.C." PHD Thesis, University of British Columbia, March 1971.

APPENDIX

1. Description of Regression on Travel Times

In a separate study,[1] a regression equation was developed that could be used to predict response times.

The data used was gathered from data forms filled out by the ambulance drivers. A sample size of one hundred and twenty-one travel times was obtained.

For each ambulance call, there is a separate response time and a time for travelling from the scene to the final destination (i.e. hospital). The factors considered were: (i) distance travelled, (ii) type of call, (iii) time of day, and (iv) where the travel took place.

The regression equations obtained were:

(T for time and D for distance)

response time	T =	5.8754	if non-emergency call
		2.9894	if emergency call

$R^2 = 0.69563$

scene to final destination	T =	3.7002	if not on 7 a.m.-6 p.m. shift
		0.23498D + 3.7002 D	if on 7 a.m.-6 p.m. shift

$R^2 \doteq 0.726Z$

These equations are graphed on figures A and B.

The regression equations were used to compare the actual travel times (travel times to be used by the simulation model) and the predicted travel times. This could be done since the G.V.R.D. travel time matrix had a corresponding distance matrix. Nodes were chosen from various regions of the G.V.R.D.

(the dispatch area). Obtained for each of these nodes was a predicted travel time to every other node in the G.V.R.D. The predicted travel times were then compared to the actual travel times. The routine "ACTFIT" from the U.B.C. T.S.P. package was used to compare the travel times.

It was observed that the actual travel times were greater than the predicted travel times. For nodes within a few miles of the source node the mean error was almost zero. As one moved further away from the source node the actual travel times exceeded predicted travel times by an overall mean error of 9.8 minutes for downtown source nodes.

The regression indicates that the travel time data being used for the simulation model is conservation. This results from the fact that the data is reflecting morning rush hour traffic. Thus the net effect of using this data is that response times outputted by the simulation model should be conservative.

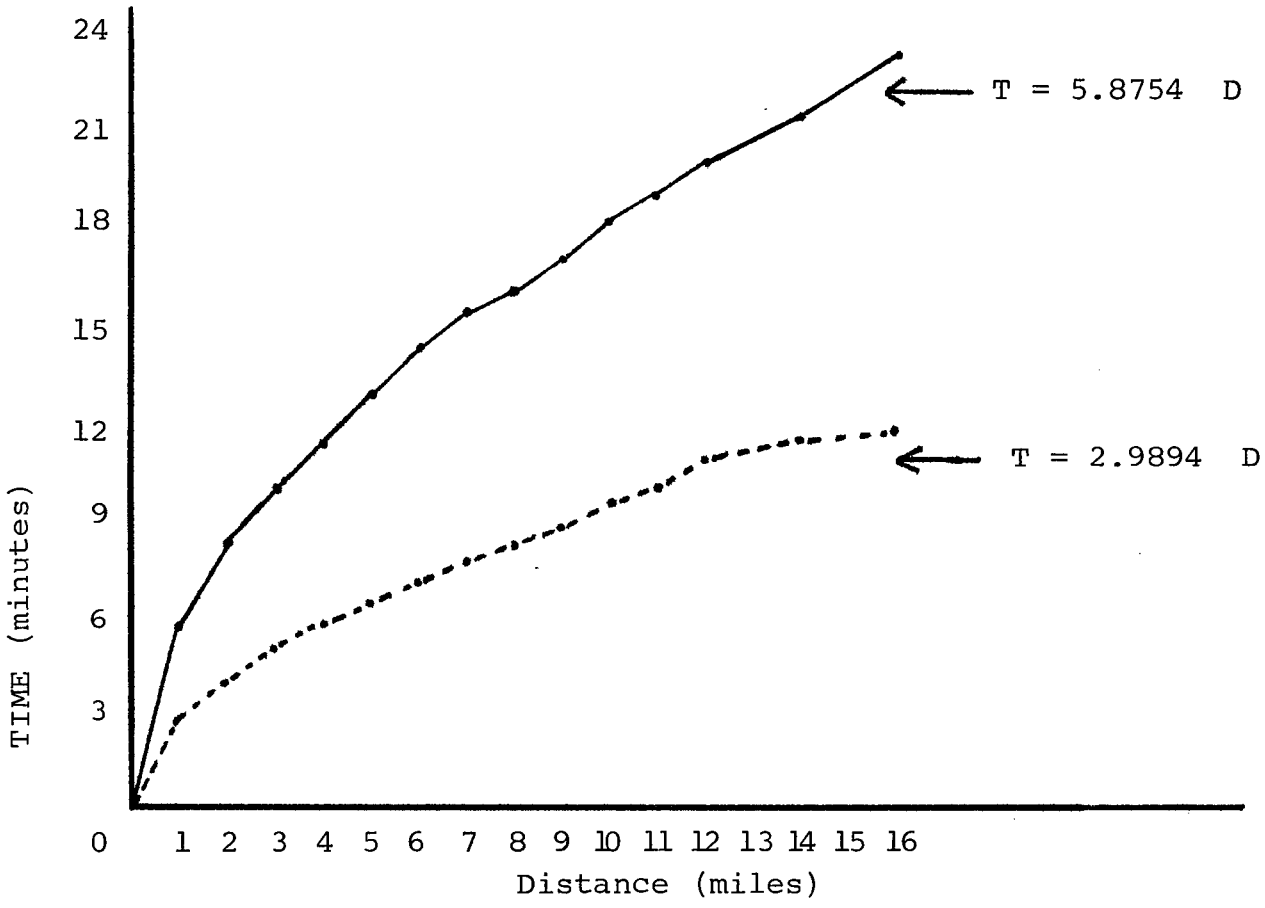


FIGURE A

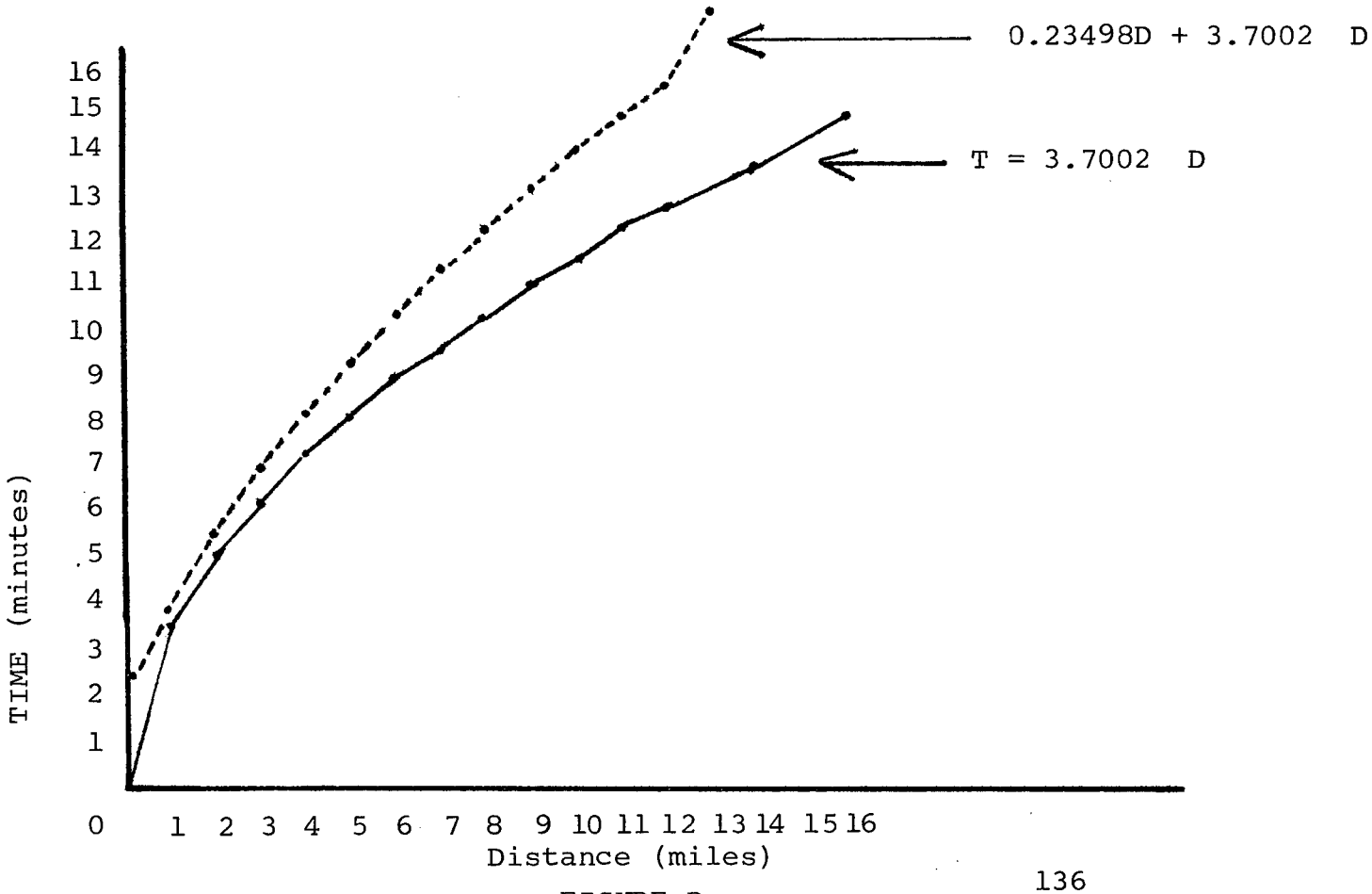
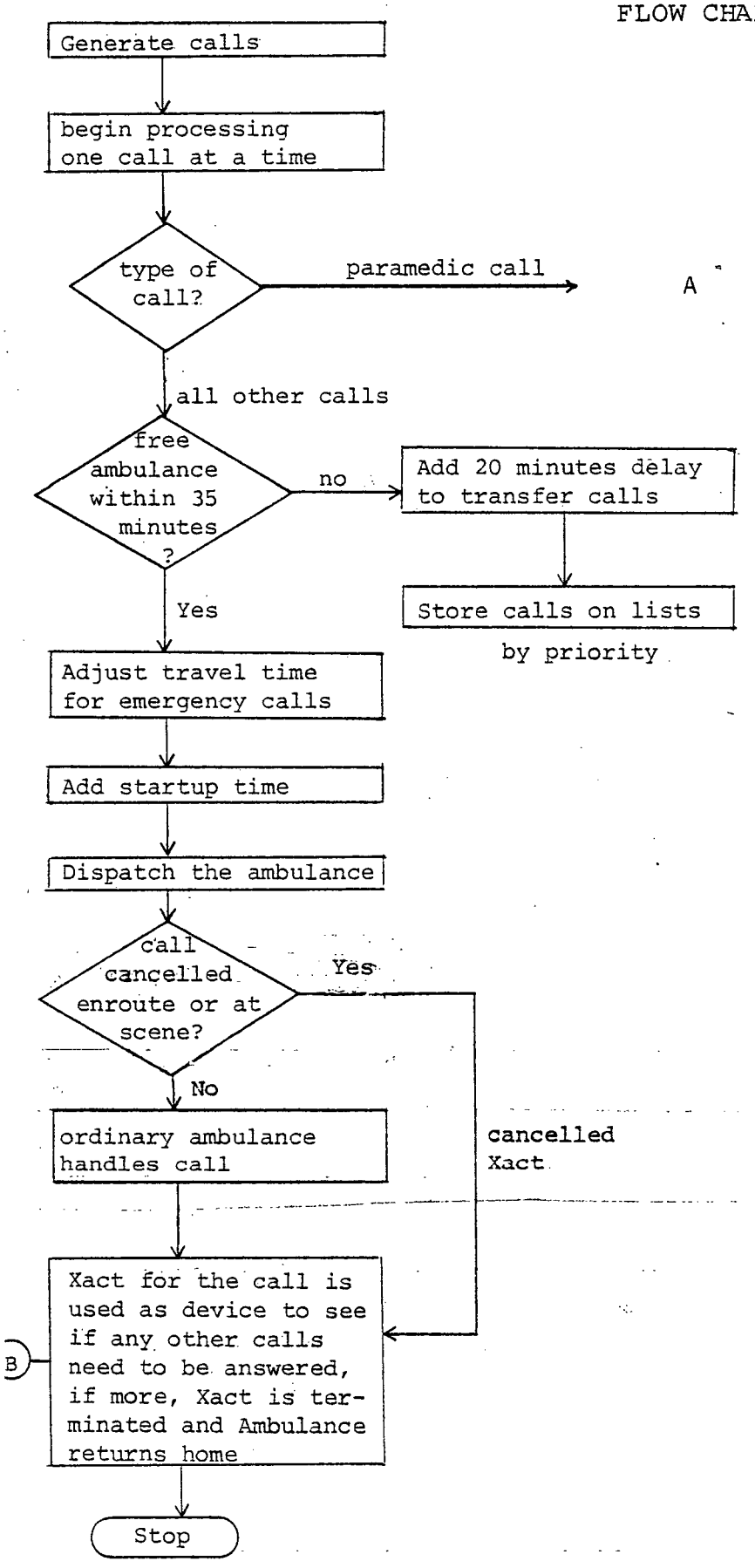
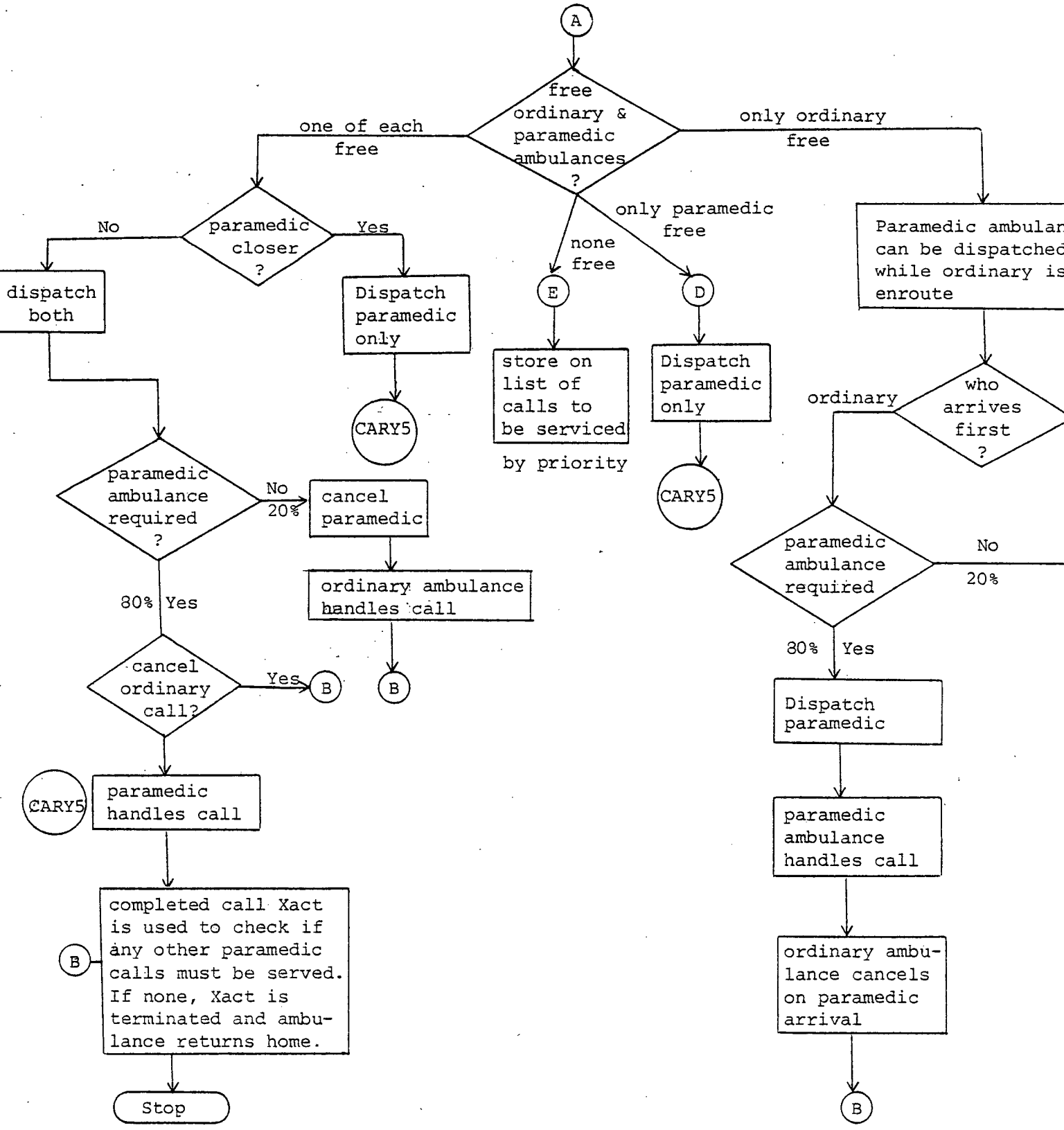
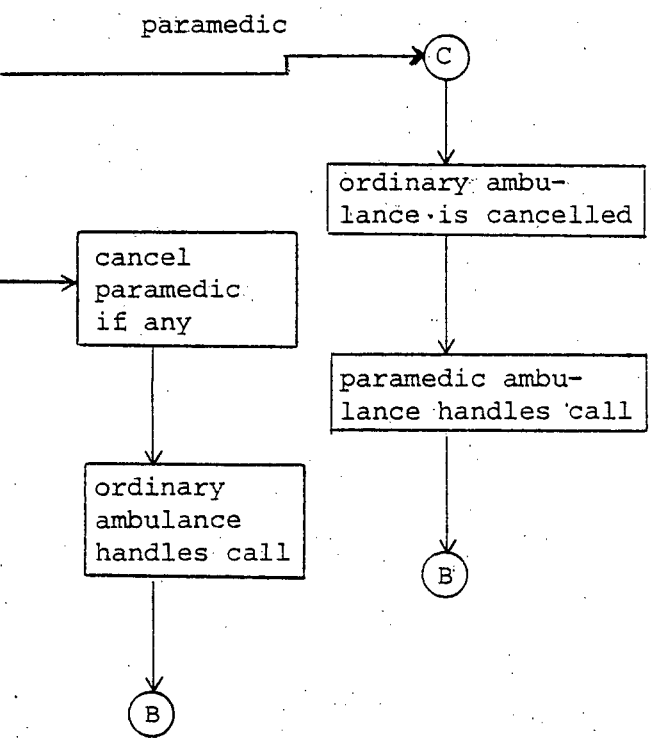


FIGURE B

FLOW CHART OF THE SIMULATION




A
free ordinary & paramedic ambulances ?
one of each free
only ordinary free
only paramedic free
none free
paramedic closer ?
No
Yes
dispatch both
Dispatch paramedic only
CARY5
E
D
store on list of calls to be serviced by priority
Dispatch paramedic only
CARY5
Paramedic ambulan can be dispatched while ordinary is enroute
who arrives first ?
ordinary
paramedic ambulance required ?
No
20%
cancel paramedic
ordinary ambulance handles call
B
80%
Yes
cancel ordinary call?
Yes
B
CARY5
paramedic handles call
completed call Xact is used to check if any other paramedic calls must be served. If none, Xact is terminated and ambulance returns home.
B
Stop
paramedic ambulance required
No
20%
80%
Yes
Dispatch paramedic
paramedic ambulance handles call
ordinary ambulance cancels on paramedic arrival
B


paramedic
C
ordinary ambu-
lance is cancelled
cancel
paramedic
if any
paramedic ambu-
lance handles call
ordinary
ambulance
handles call
B
B