THE SPATIAL ALLOCATION OF MEDICAL CARE RESOURCES IN MASSACHUSETTS: AN APPLICATION OF RAMOS

E.J. Rising
L.D. Mayhew

March 1983
WP-83-38

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

The principal aim of the former Health Care Task at IIASA was to develop a family of submodels of national health care systems for use by health service planners. This paper, written as a part of that research activity, applies 1978 acute general hospital discharge data for Massachusetts, USA, to the Resource Allocation Model Over Space (RAMOS). Its publication was delayed, and it is therefore being issued a few months after the dissolution of the Task and the Human Settlements and Services Area.

Andrei Rogers
former Chairman
of the Human Settlements
and Services Area

CONTENTS

THE SPATIAL ALLOCATION OF MEDICAL CARE RESOURCES IN MASSACHUSETTS: AN APPLICATION OF RAMOS

1. INTRODUCTION

This paper describes the use of the model RAMOS (Resource Allocation Model Over Space) with data from the United States, specifically the 1978 acute general hospital discharge data from the State of Massachusetts. The RAMOS model was developed in England and at the International Institute for Applied Systems Analysis (IIASA), as one of a family of models describing the delivery of in-patient medical care. The first versions of this model were designed for medical care systems that are either centrally or regionally planned such as in East European countries and England. Later these models were applied to countries whose systems are based on national insurance schemes (e.g., Italy and Canada), but have less strict planning procedures. Until now, this model has not been used with data from the USA, whose medical care system is regarded as being much more market oriented than any of the previous countries examined.

RAMOS is a type of gravity model related to the Newtonian principle in which the attractive force of two bodies is equated to the product of two masses, multiplied by a deterrence factor (equal to the inverse of the square of the distance between them), and by a coefficient of proportionality. In RAMOS the attraction of patients from one area to another is hypothesized to be proportional to the product of the ability of an origin area to generate patients and caseload capacity of a destination area, and inversely proportional to the cost of access.

The RAMOS model, and others in the IIASA family of models, as well as several applications, are described in a number of papers.* The mathematical statement presented by Mayhew and Taket (1980) for RAMOS only is as follows:

$$T_{ij} = B_j D_j W_i e^{(-\beta c_{ij})} \qquad i = \overline{1,I} \quad j = \overline{1,J} \tag{1}$$

where

T_{ij} = the number of patients from origin area i who receive care in destination area j, in a particular specialty or group of specialties

D_j = the caseload capacity of destination j in the same specialties

W_j = the propensity of origin i to generate patients in these specialties (called the patient generating factor)

c_{ij} = the accessibility cost of treatment for a patient from origin i receiving treatment in destination j

$$B_j = \left[\sum_i W_i{}^{(-\beta c_{ij})}\right]^{-1} \tag{2}$$

which is a constraint assuring that

$$\sum_i T_{ij} = D_j \tag{3}$$

---

*These are predominately IIASA working papers or research reports that have been produced by several "generations" of investigators: Gibbs 1978a,b; Rousseau and Gibbs 1980; Aspden and Rusnak 1980; Hughes 1978; Hughes and Wierzbicki 1980; Mayhew and Taket 1980, 1981; Mayhew and Leonardi 1981.

Gravity models of this type are described as "attraction constrained", or in this case "destination constrained" because the destination capacity must be utilized, but all patients will not necessarily be seen.

As one would expect, the diagonal of the T_{ij} matrix is "strong", that is, most people receive care in the same area in which they live. The patient generating factor W_i and the caseload capacity D_j are the primary independent variables. Their product is analogous to the product of the two masses in the Newtonian model, and they are determined from hospital discharges and local population data.

The deterrence function, $e^{(-\beta c_{ij})}$, is related to the tendency of a patient living in i to be treated in j. This function contains the parameter, β, which is evaluated during calibration, and c_{ij}, a measure of the accessibility costs (e.g., distance or travel time). As one might expect, the diagonal of the accessibility cost matrix, and therefore of the deterrence factor, is usually "weak": that is, it is more convenient to obtain care in the area where one lives.

The proportionality term B_j is designed so that caseload capacities are fully utilized, yet not exceeded. This seems realistic in situations where there may be a queue of patients waiting for (elective) care, but it would appear to disqualify the model for applications in many communities in the United States, for example, when unused capacity occurs. However, this is not necessarily the case. This constraint simply requires that the model be used with a client population and set of resources that match. In fact, under some circumstances this property of the model can be used advantageously; once calibrated this model can be used to determine the quantity of facilities of various types that are needed for particular populations.

RAMOS was designed, and has been used, to solve problems of evaluating facilities location and predicting patient flows within the framework of a single algorithm. It is unique in that previous models have generally focused on single facilities or problems at the local scale of enquiry rather than over large areas, such as an entire state or nation. In addition, earlier models failed to take into account variations in morbidity or the disaggregation of the population by age and sex, medical specialty, or method of payment. These characteristics permit the RAMOS model to be used for planning for various services, for specific locations, and for particular populations.

RAMOS was originally intended for use in medical systems with at least some measure of central control. The model is now used in several countries to produce scenarios based on future population and resource availability in assisting the process of allocating hospital operating funds by region and by area.

In the United States the medical system is neither centrally planned nor centrally funded. This difference has a profound effect on the way the model might (or might not) be used. A brief description of the United States medical system is therefore helpful. There are three issues central to the use of the model that must be touched upon: the methods by which a patient chooses a hospital, the system of payment for hospitalization, and the so-called Certificate of Need Law.

Except for emergencies, most patients receiving inpatient care in hospitals are referred by their local doctor, who is either on the staff of that hospital or has staff privileges. This physician will usually be responsible for the in-hospital care himself, or consult with a specialist. In general, the local physician is the dominant factor in hospital choice, but except in certain cases* there is no compulsion or financial advantage for the patient to go to the hospital selected by his physician.

*Exceptions would be such groups as veterans, members of the armed services and their dependents, and members of a growing number of health maintenance organizations.

In an emergency, however, a patient goes (or is taken) to the hospital that is closest at the time. If admission is required for recuperation or treatment the patient may remain there or be transferred nearer home, depending on his or her wishes.

Patients requiring highly specialized care are referred to the most appropriate hospital that caters to the particular illness, often at a considerable distance from home.

Another exception would occur if the patient had a strong opinion for (or against) a hospital because of something (real or imagined) that may have happened to someone he or she knows. The patient might then ask the doctor for referral somewhere else. This is usually not a problem with physicians who have privileges at several hospitals. In other cases, the patient is perfectly free to consult another doctor for an evaluation for referral, with the idea that this physician will choose a hospital more to the liking of the patient.

With few exceptions, patients receiving care in the United States are responsible for their own hospital costs. They are billed for this, and the money is retained by the hospital as income, although the hospital may also receive supplementary income from endowments, grants, etc. Most patients, however, do not pay bills directly because they carry insurance covering all (or most of) the costs. Persons who fall below certain income criteria and those over 65 years of age are eligible for medicaid or medicare, a governmentally funded system of medical insurance (whose eligibility and benefits vary from state to state).

Most hospitals are non-profit organizations, and although they compete with each other, they do not compete in a free market. The patient-customer of the hospital generally does not know what facilities various hospitals have, or even what facilities are required for the treatment of his or her case. Consequently, the patient must rely on the advice of his or her physician (who is also a supplier of services to the patient-customer). Thus, hospitals compete with each other by attracting a large and outstanding staff of physicians. An important element in attracting this staff is the provision, by the hospital, of the facilities these physicians need (or want) to treat their patients.

The cost of providing the facilities that physicians find attractive is sometimes partially subsidized by government grants, but is mostly paid for by the hospital. The hospital then passes the costs on to the patients on a fee-for-service basis. Until recently the only constraints on the acquisition of facilities and equipment was the availability of capital and the reluctance of Boards of Trustees to approve expansions they felt unjustified. A consequence was that costs of hospitalization increased so rapidly that there was mounting public concern, eventually requiring legislation to control this growth. This resulted in the current "Certificate of Need" legislation, which requires providers of health and medical care to secure a Certificate of Need before any substantial increase is made in facilities, capital equipment, or the services that they offer. The State of Massachusetts was one of the earliest states to enact such legislation, and similar legislation was later passed by Congress. Although under pressure for repeal and/or reduced funding from the current administration, it is still in force.

An individual hospital in Massachusetts that wishes to expand or secure substantial new equipment needs to prepare an application describing and justifying the desired addition. This application is then reviewed twice before the Certificate of Need is issued by the State Department of Public Health. The first review is held at the local level by the staff and advisory board of the local health planning agency; the second is held at the state level by the staff and advisory board of the Department of Public Health, with public hearings in both cases. If the decision is negative, the hospital may seek remedy in the courts or with the state legislature.

Given this operating environment, it is obvious that the opportunity for direct control of the delivery of hospital care in the United States is limited as compared with other, more centrally planned and funded, medical systems. However, a strong influence on the future behavior of the system can be exerted by control at the "margin of growth" through enlightened management of the issuing of Certificates of Need.

Many issues that are subject to approval under the Certificate of Need Law are those that can be investigated within the RAMOS framework. In particular, we have found that despite the many complexities involved in the way patients seek and obtain medical treatment, the structure of the RAMOS model permits us to describe *and* predict the origins and destinations of patients. Investigations can then be carried out for broad categories of facilities, such as medical-surgical beds in Massachusetts, or for specific facilities such as "burn units" or intensive care units. Investigations can encompass broad populations (e.g., all Massachusetts residents), or sub-populations stratified by method of payment, age, sex, ethnic origin, etc. (provided the data are available).

The results of these analyses can be provided as another input to the political process. The results could be placed on record at public hearings through the testimonies of independent analysts, or these results could be used directly by various groups within the system. A hospital could, for example, use the results of this sort of analysis as evidence of a need for expansion, the staff of local health system agencies could use these results to compare submissions from different hospitals, or the State Department of Public Health could use the results from these models to examine the implications of various changes in public policy.

2. OBJECTIVE

The eventual objective of this line of research is the development of models describing the delivery of health care that are consistent with the current system in the United States, and are based to the extent possible on the existing family of IIASA health care models. This development process will follow the classical procedures of model development:

a) construction (or adaptation) of the model

b) calibration

c) validation

d) application

The purpose of the following sections is to provide a detailed record of the first step in this process—the adaptation and calibration of the RAMOS model—and to suggest lines of future research. The various methods of calibrating RAMOS are described elsewhere (Mayhew and Taket 1980); for convenience, however, they are presented briefly below. This description is followed by a discussion of the variables used by the model and how the data were obtained in Massachusetts. The results of the calibrations are then given.

3. METHODOLOGY

The two methods used for calibration are those described by Mayhew and Taket (1980). Each uses a different criterion for the selection of the "best" value for β, the model parameter. The first is by slope calibration, and the second uses maximum likelihood.

The criterion for the best solution using the slope calibration method is obtained from a regression of the *predicted* against the *observed* values of T_{ij}; when the value of the slope of this regression equals one and the intercept is zero, it is clear that predicted and observed flows are (on the average) the same, and the model is calibrated. The method uses an iterative procedure based on systematically incrementing β and estimating the regression coefficients until the results come as close as possible to the prescribed conditions. A measure of goodness-of-fit may be obtained by correlating the *predicted* and *observed* patient flow elements. However, it is the square of the correlation coefficient, the "proportion of explained variance" (by the regression) and symbolized by R^2, that is usually reported. Incidentally, an examination of the actual plot of these pairs of values is useful because points close to the regression line, as well as the outliers, can be identified.

From this result it is but a short step to consider a calibration procedure based on *maximizing* R^2 itself. This approach is theoretically attractive and a suitable technique was developed and tested by Mayhew and Taket (1980). They found, however, that R^2 was relatively insensitive to the parameter β when very close to 1.0, its theoretical maximum, thus making it unsuitable.

The maximum likelihood calibration method has a different basis and requires another procedure. Mayhew and Taket (1980), using the work of Batty and Mackie (1972), noted that, if the deterrence function is based on the negative exponential (see equation 1), predicted flows are most likely to be correct when the *predicted* mean travel cost equals the *observed* mean travel cost. In this case, the observed mean travel cost is first calculated from the travel distance matrix and the observed patient flows. An iteration procedure is then used, which systematically varies the value of β until the travel cost averages to the desired value. The best value of β is the one that produces a match between the criterion value of the observed average and the predicted average (to within the desired tolerance).

Although modifications can be made to accommodate other forms of deterrence functions, Mayhew and Taket (1980) found this unnecessary in the case they examined. It was found, however, that the maximum likelihood method of determining β was sensitive both to the number of zones over which the calibration took place and to how and where the centroids of these zones were located, which is a disadvantage. On the other hand, they noted that this method was capable of very rapid convergence. They concluded therefore, that the maximum likelihood method is very useful for producing initial estimates for β in the early stages of development, after which they recommend the use of the slope calibration procedure to obtain a final estimate.

Although both calibration methods use different criteria and procedures, they are similar with respect to their data requirements. Both require that consistent observations be made on the dependent and independent variables in order to determine the parameter, β.

The data needed for calibration can be described in terms of the model as follows:

1. the observed patient flow $\overline{T}_{ij}$, which is the number of patients from origin i, who actually obtain care in destination j. (Note that this becomes the dependent variable in prediction runs; for calibration it is an independent variable, therefore, a bar is added to distinguish the observational data from the predictions when it is the dependent variable.)

2. the patient generating factor W_i, which is the propensity of each origin area to generate patients who receive care. This is calculated using age- and sex-specific data on morbidity and population. The state-wide average numbers of patients in various specialties (taken from hospital discharge data) are used to produce a weighted average demand for each origin zone based on local population characteristics. The more detailed the morbidity data (i.e., disaggregated by age, sex, and location), the more accurate the result.

3. the caseload capacity D_j, which is the number of patients treated in each destination area in a particular set of clinical specialties. Since the structure of the model does not permit unused or excessive capacity, careful interpretation of this variable is necessary, particularly for disaggregation purposes (see also the subsequent, disaggregated models, DRAMOS (Mayhew 1981)).

4. the accessibility cost of treatment c_{ij}, which represents the relative cost to patients living in area i and receiving care in area j. For current purposes, no account was taken of possible differences in treatment costs between destinations, only travel costs. (Mileage, time, and opportunity costs were considered.) If data on differences in costs for similar treatment in different locations are available, they can be included if desired, although more prior data analysis would be required. The section on results will show that the calibration of the model is very sensitive to the accessibility cost, for this reason it will be discussed later in more detail.

The detail available on the origins and destinations of the patients, plus the fact that the model does not require that the boundaries of the origin and destination zones be the same, has made it possible to consider several alternatives for zonal definition. It was decided to use health service area (HSA) sub-areas, although other categorizations (e.g., towns, countries, etc.) will be used for sensitivity studies in the future. HSA sub-areas seemed an appropriate place to start since they were already used for health planning.

4. DATA

There are six HSAs in Massachusetts broken down into 23 sub-areas. These are shown on a map of Massachusetts in Figure 1, and information about these sub-areas is presented in Table 1. Each HSA contains between 8 and 50 hospitals, and each sub-area contains from 2 to 25 hospitals. The populations of the HSAs range from about 490 thousand to 2.1 million, and the populations of the sub-areas range from about 110,000 to just over 700,000. In the densely populated eastern portion of the state the population is made up of sub-areas containing as few as 3 or 4 towns; in the western portion one sub-area needed as many as 43 towns to reach this population level.

The hospitalization data used for this calibration were taken from the records of patients who were discharged in the calendar year 1978 from 122 licensed, short-stay hospitals in Massachusetts. The hospitals in Massachusetts *not* included were 33 chronic care and rehabilitation hospitals, 31 psychiatric hospitals, 2 hospitals for the mentally retarded, 3 "long stay" specialty hospitals (burns, orthopedic, etc.), 1 army hospital, 1 US public health hospital, and 1 acute general hospital in HSA 6 (sub-region 3) for which data were not available.

The 1978 patient discharge data from Massachusetts are highly regarded by the medical and health professionals for consistency and accuracy. The published data from the year 1979 are not internally consistent for technical reasons, and therefore could not be used. The 1980 data, which have just been put in final form, are highly regarded but were unavailable at the time this paper was written.

The patient discharge data were disaggregated in the following ways:

1. by the hospital of discharge
2. by the home address (town) of the patient
3. by the category of the primary mode of care received by the patient. Such categories were defined as:

HEALTH SERVICE AREA BOUNDARIES

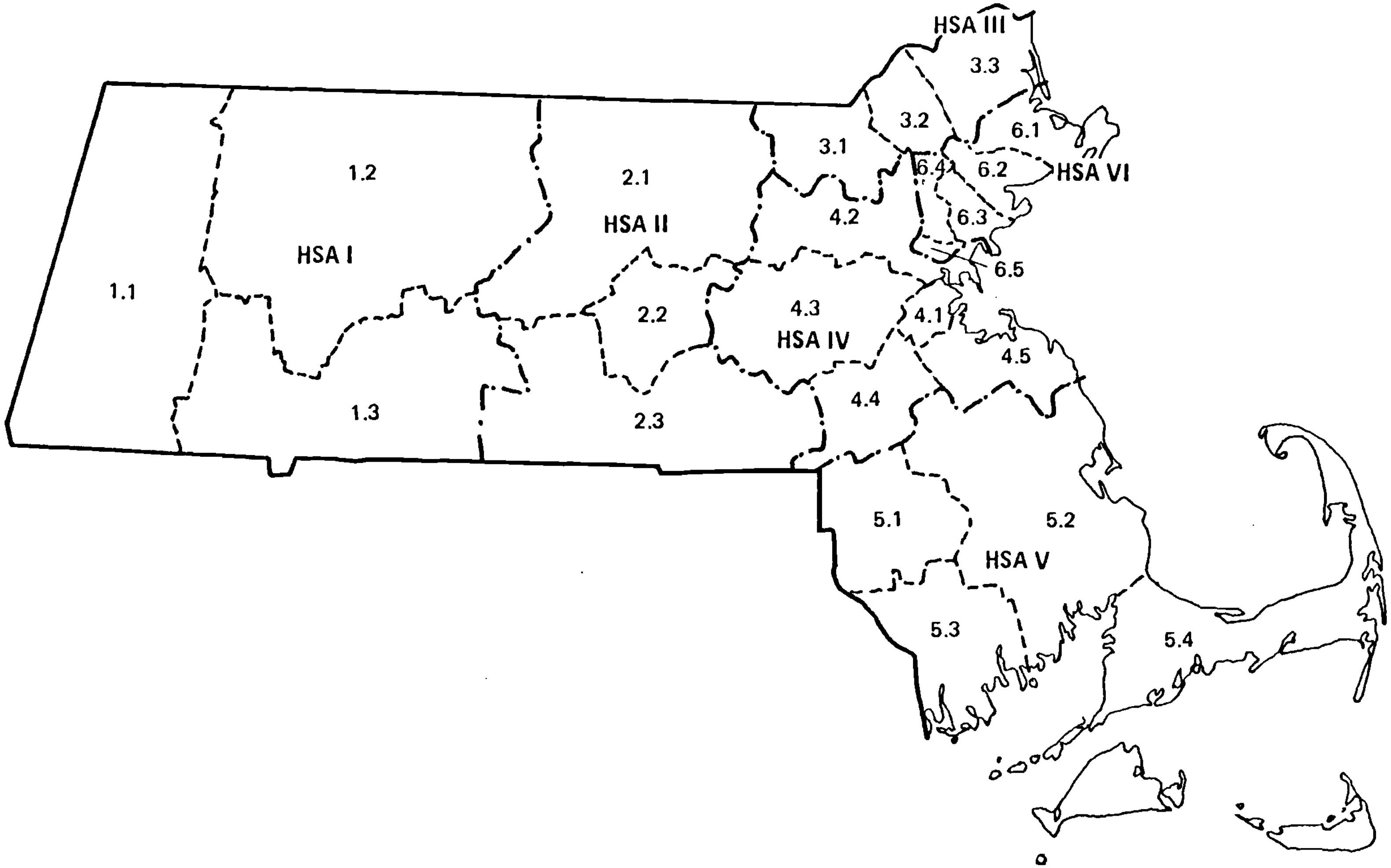

Figure 1. Boundaries of health service areas and sub-areas in the State of Massachusetts.

Table 1. Description of health service areas and sub-areas.

Health service area and sub-areas	Number of towns	Number of hospitals in calibration*	Population (1978 est)
REGION I			
1.1 Berkshire	33	16	145,674
1.2 Northern Valley	43		175,257
1.3 Southern Valley	26		482,586
REGION II		17	
2.1 North Worcester	25		214,380
2.2 Central Worcester	8		237,751
2.3 South Worcester	29		228,390
REGION III		8	
3.1 Lowell-Tewksbury	8		226,991
3.2 Lawrence-Andover	4		146,375
3.3 Greater Newbury	11		114,518
REGION IV		50	
4.1 Central Metro	4		705,204
4.2 Northwest Metro	21		456,914
4.3 West Metro	20		478,314
4.4 Southwest Metro	11		164,082
4.5 South Metro	10		305,625
REGION V		16	
5.1 Attleboro	14		191,634
5.2 Brockton-Plymouth	20		343,718
5.3 Fall River	12		307,373
5.4 Cape Cod	23		180,349
REGION VI		15	
6.1 Cape Ann	9		107,914
6.2 Danvers-Salern	5		132,557
6.3 Greater Lynn	5		132,268
6.4 Eastern Middlesex	5		110,507
6.5 Tri-Cities	3		148,657
OUT-OF-STATE (OS)			
OS-1 New Hampshire (NH)			887,000
OS-2 Rhode Island (RI)			929,000
OS-3 Connecticut (CT)			3,115,000
OS-4 Vermont (VT)			493,000
OS-5 New York (NY)			17,648,000

*Three hospitals have closed since this origin study was made and several have merged.

a) all patients (a total of the following categories)
b) medical-surgical patients
c) obstetrical patients (in-patient abortions and gynaecological patients were counted as medical-surgical patients)
d) pediatric patients
e) psychiatric patients discharged from acute general hospitals

These categories are arbitrary and are an artifact of available data. Further aggregation or disaggregation is possible as interest warrants and data becomes available. This issue of categorization also arises in the discussion of the RAMOS model in the section of this paper on future research.

Two values were used for the accessibility cost in the calibration: first, the actual road mileage, and second, an "adjusted" mileage, which attempts to take into account driving time and difficulty. The adjusted mileages were obtained by increasing the actual road mileage values in the urban areas to account for the fact that one mile of driving is more difficult and time consuming in the city than in rural areas.

A mileage chart used by trucking companies provided estimates of the actual road mileages between the towns that are the centroids of population and centroids of in-patient care. Where these centroids are in the same town, the travel distance is less than would otherwise be the case, but it is not zero; i.e., even if the centroids are exactly coincident, one person traveling 5 miles north would not be "cancelled" by another patient traveling 5 miles south. Therefore, mileages for the accessibility cost in this case were calculated somewhat differently: the area of the town was divided by the number of hospitals, producing an "average area served" by each hospital. This area was then assumed to be round, and its radius was calculated. Two-thirds of this radius was taken as the average for the "actual road mileage" traveled by these patients.

The immediately adjacent 5 states are the origin of approximately 50,000 out-of-state patients who are treated annually in Massachusetts hospitals. Road mileage estimates for these patients were based on their actual numbers and residence of record at the time of discharge.

The adjusted mileages were obtained from the actual road mileages. The values for actual mileages in and around Boston, Worcester, and Springfield were increased from one (Springfield) to nine (Central Boston) units. The adjustment was based on a judgment of the relative increased driving time and the general physical inaccessibility of these locations; rivers (and bridges), and interstate highways (and their access).

Twelve of the 322 origin-destination pairs in the accessibility cost matrix were adjusted, and the average distance traveled by in-state patients was thereby increased from 7.583 *actual* miles to 9.581 *adjusted* miles. Although the adjustment was applied to only 4 percent of the origin-destination pairs, it increased the average distance the patients traveled by almost 21 percent. This percentage increase is large because of the very large number of urban patients who had their access distance adjusted upward.

5. DISCUSSION OF RESULTS

Three sets of results are presented and discussed in this section; the first set, and by far the most important, is the calibration results for the RAMOS model. They show both the slope calibration and maximum likelihood methods using actual road mileages and adjusted mileages. Since Boston is a nationally known referral center, it is also useful to compare the calibration obtained for all patients who receive care in Massachusetts to a calibration that is limited to Massachusetts residents only.

A substantial difference was obtained between the first and second set of results, i.e., the calibration using the actual road mileages and the calibration using adjusted mileages. This difference was one of the most interesting results obtained and is discussed further in the section on accessibility costs.

The third set of results are unrelated to the calibration of RAMOS. They are based on $RAMOS^{-1}$ (RAMOS inverse), a model variant developed with the strategic planning of health care services in mind (Mayhew 1980).

5.1 RAMOS Calibration

The results of the calibration of RAMOS using 1978 Massachusetts hospital discharge data are shown in Tables 2 and 3. Table 2 presents information for all patients treated in Massachusetts, and Table 3 is limited to patients whose home address is in Massachusetts at the time of discharge. The data base for these two tables differ by about 50,000 discharges. Each of these tables presents data separately for four categories of care: medical-surgical, obstetric, pediatric, psychiatric, and the total of all patients in these categories.

These tables show first, the numbers of discharges available for the data base for each category; second, the value obtained for the parameter β from the calibration; and finally, a measure of the goodness-of-fit, R^2. Since the maximum likelihood calibration uses average mileage (or adjusted mileage) traveled by patients as the criterion for the calibration, these values are also included in the tables.

A comparison of the results using actual road mileage to those using adjusted mileage is shown in both tables. They show that relatively small changes in mileages (an adjustment of only 12 of 322 pairs of mileage values) has a significant effect on the goodness-of-fit statistic, R^2. Figures 2 and 3 display this result graphically for one example (slope calibration for medical-surgical, in-state patients). A comparison of Figures 2 and 3 shows the difference in values of R^2 as a difference in the grouping of the points around the "ideal" regression line. Particular attention is called to the point for "Central

Table 2. Calibration of RAMOS using 1978 in-patient discharge data from Massachusetts for all patients treated in that state.

Category of patient care	Number of patient discharges in data base	Average mileage traveled by patient: Actual miles	Average mileage traveled by patient: Adjusted miles	SLOPE CALIBRATION, Actual miles: Calibration coef β	SLOPE CALIBRATION, Actual miles: Proportion of explained variance R^2	SLOPE CALIBRATION, Adjusted miles: Calibration coef β	SLOPE CALIBRATION, Adjusted miles: Proportion of explained variance R^2	MAXIMUM LIKELIHOOD CALIBRATION, Actual miles: Calibration coef β	MAXIMUM LIKELIHOOD CALIBRATION, Actual miles: Proportion of explained variance R^2	MAXIMUM LIKELIHOOD CALIBRATION, Adjusted miles: Calibration coef β	MAXIMUM LIKELIHOOD CALIBRATION, Adjusted miles: Proportion of explained variance R^2
Total (all patients)	851760	10.024	11.984	.1600	.8407	.2300	.9531	.11749	.8577	.13900	.8585
Medical-surgical	658942	10.151	12.150	.1600	.8395	.2300	.9510	.11505	.7998	.13613	.8510
Obstetric-maternity	88192	8.484	10.297	.1900	.8920	.2700	.9489	.14082	.8588	.16368	.8956
Pediatric	84391	10.868	12.695	.1500	.7678	.2100	.9190	.11116	.7245	.13309	.8138
Psychiatric	20182	9.035	10.906	.1900	.8635	.2500	.9578	.13924	.8664	.16412	.8851

Table 3. Calibration of RAMOS using 1978 in-patient discharge data from Massachusetts for in-state patients only.[a]

				SLOPE CALIBRATION				MAXIMUM LIKELIHOOD CALIBRATION			
				Actual miles		Adjusted miles		Actual miles		Adjusted miles	
Category of patient care	Number of patient discharges in data base	Average mileage traveled by patient: Actual miles	Adjusted miles	Calibration coef β	Proportion of explained variance R^2	Calibration coef β	Proportion of explained variance R^2	Calibration coef β	Proportion of explained variance R^2	Calibration coef β	Proportion of explained variance R^2
Total (all patients)	817892	7.553	9.581	.1600	.8524	.2300	.9559	.13316	.8385	.17450	.9202
Medical-surgical	632183	7.490	9.561	.1600	.8503	.2300	.9539	.13321	.8386	.17583	.9186
Obstetric-maternity	85249	7.343	9.202	.1900	.8988	.2700	.9501	.14337	.8769	.17577	.9130
Pediatric	80757	8.256	10.152	.1500	.7856	.2100	.9245	.12432	.7689	.16478	.8868
Psychiatric	19650	7.544	9.458	.1900	.8757	.2500	.9609	.13380	.8427	.17145	.9170

[a]In-state patients are patients who are treated in Massachusetts having an address in Massachusetts at the time of discharge.

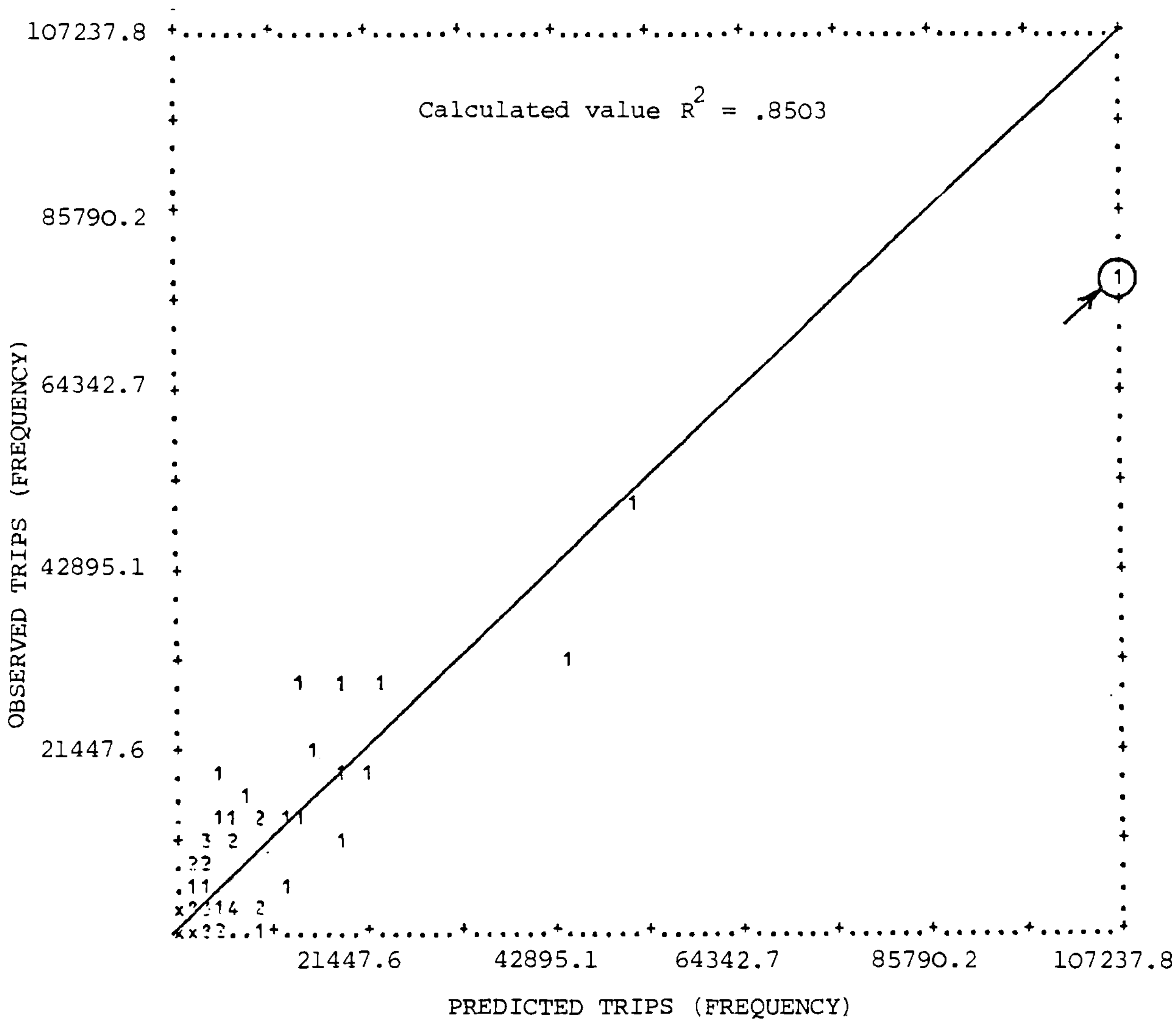

Figure 2. Comparison of predicted and actual patient flow matrix.

(The actual number of patients from origin i who are treated in destination j compared with the predicted number [23 x 28]. Medical-surgical discharges, Massachusetts, 1978 data, road mileage used for accessibility cost.)

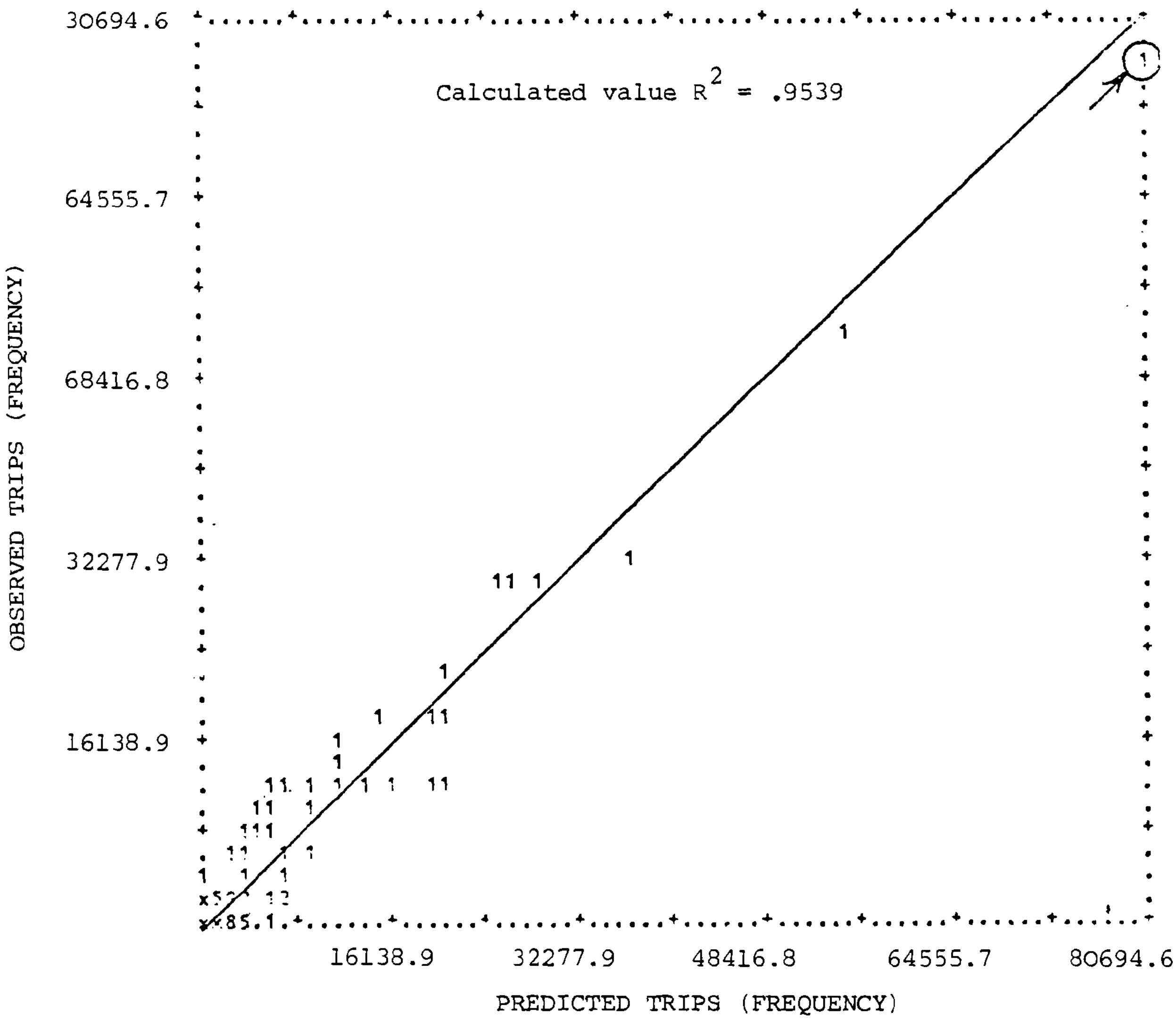

Figure 3. Comparison of predicted and acutal patient flow matrix.

(The actual number of patients from origin i who are treated in destination j compared with the predicted number [23 x 28 matrices]. Medical-surgical discharges, Massachusetts, 1978 data, road mileage adjusted for travel time used for accessibility cost.)

Metro", the sub-area encompassing central Boston (circled and marked with an arrow on these two figures). As expected, this sub-area has the largest value for both the *predicted* and the *observed* values for patient flow. It is interesting to note that this point falls far below the regression line when actual mileages are used (i.e., much less hospitalization is observed to occur in this sub-area than is predicted by the model); however, when the adjusted values of mileage are used, the point for Central Metro is close to the regression line (i.e., the predicted number of cases in the Central Metro sub-area come close to the observed values).

The use of adjusted mileages instead of actual road mileages is responsible for increasing the value of R^2 in the slope calibration, yielding values around 0.95. The same tendency is noted in the maximum likelihood calibration, although not as large.

An examination of the changes in the values of β for the various calibrations is also interesting. Comparing results of the calibration for all patients treated in Massachusetts to the calibration of the model for just in-state patients (from Tables 2 and 3) shows no difference between corresponding values of β for slope calibration but does show differences between about 0.003 and 0.03 for maximum likelihood.

The 5 out-of-state origins are not sufficiently different from the 23 in-state origins in their yield of patients, relative to their distance from the treatment destinations, to affect the slope calibration. However, the situation is somewhat different when using maximum likelihood, because it is based on the average mileage traveled. The 5 out-of-state origins add about two miles (a bit over 20 percent) to the average distance calculated for in-state and all patients categories in this method.

The differences in the values of β and R^2 for each category of care are also of interest. Both Tables 2 and 3 show that the values for medical-surgical patients and the values for total of all patients are very similar (often identical). This is because a broad definition was used for the medical-surgical category, and thereby consists of about 75 percent of the total. On the other hand, restricted definitions were used for obstetric, pediatric, and psychiatric patients (who were discharged from acute general hospitals); therefore, the number of patients in these categories is much smaller. The model fits the pattern of patient flow for pediatric patients with the lowest value of β, while psychiatric and obstetric categories have higher values that are approximately equal. Although the results from the slope calibration are uniformly higher than the values obtained from maximum likelihood, this same pattern persists.

Given the same accessibility cost matrix, a higher value of β will produce a higher value for the deterrence factor. Since the values for β are higher for psychiatric and obstetric patients than for pediatric patients, this means that patients are more likely to seek obstetric and psychiatric care at acute hospitals sited close to home, and are willing to travel further for pediatric care. The difference in β must be interpreted cautiously because it is in the exponent of the deterrence factor.

5.2 Deterrence Factor, Accessibility Costs, and Error

The discussion of the deterrence factor has been based on accessibility costs measured in terms of mileages—actual road mileages in one case, and mileages adjusted for driving time and difficulty in the other. The data in Tables 2 and 3 show substantial differences between the results of using both measures, which suggests that calibration is sensitive to these choices, which is to be expected, since the accessibility cost appears as an exponent in the model.

A more systematic study of driving time and difficulty should further improve the calibration, but in view of the sensitivity of the model, it would cause at least a small modification in the result. This sensitivity is reason to look closely not only at the measurement of distance, but also at the mechanism of the deterrence itself.

Geographic distance is often used in gravity models because it is easy to measure and is an important variable in the process of selecting a hospital (and in marketing in general). When one goes beyond the idea of just geographically defined distance concepts for accessibility costs, there is an opportunity to further refine the model. There is a rich literature on the reasons why patients are treated in particular hospitals, and important work based on factors such as patients' religion or income exists (Morrill and Kelly 1970). These are used to modify geographical distance into a new variable called "social distance" to explain the use of individual hospitals by particular populations. Besides the addition of such explanatory factors as income and religion, investigators have experimented with a range of different types of functions for the deterrence factors. Shannon (1969, 1975) examined four deterrence factors, each based on an existing medical system with slightly different characteristics, and he was able to identify the form of the function for each different system. He also introduced a deterrence function that contained the product of two exponentials, which he suggested would combine the characteristics of two of the systems. Mayhew and Taket (1980) also investigated four different functions for the deterrence factor using the Greater London data, and Mayhew (1982) suggested the inclusion of a "prestige" factor based on the added attractiveness of certain facilities.

The expression used hitherto for the deterrence factor, f_{ij}, is as follows:

$$f_{ij} = e^{-\beta(c_{ij})}$$

If calibration errors are attributable to the accessibility cost, an error term, ε_{ij}, can be added to the exponent so that the observed and predicted flows match exactly

$$f_{ij} = e^{-\beta(c_{ij}+\varepsilon_{ij})}$$

With the error stated explicitly it can be calculated and analyzed; however, before this can be done any additional information available should also be included in the accessibility cost; for example, prestige can be thought of as a factor that enhances the attractiveness of certain destinations. From the point of view of the model, this prestige effect would tend to reduce the importance patients attached to the distance they travel to some destinations; if we call this factor p_j, then the deterrence factor could be rewritten as:

$$f_{ij} = e^{-\beta(c_{ij}-p_j+\varepsilon_{ij})}$$

Further refinement is obtained by adding a term q_i for any known effects related to origins (e.g., socio-economic factors) and a term s_{ij} that is selective between origins and destinations. Positive values for these terms would indicate that the effective distances are increased, as in the case above of adjusted mileages with c_{ij}, and negative values would decrease the effective distance as when prestige factors are active. In its most elaborate form the deterrence function could be written as:

$$f_{ij} = e^{-\beta(c_{ij}\pm p_j\pm q_i\pm s_{ij}\pm\varepsilon_{ij})}$$

where f_{ij}, β, c_{ij} are defined as before, and

p_j = effects known, by destination

q_i = effects known, by origin

s_{ij} = effects known, selectively by origin and destination pair

ε_{ij} = residual error

The complete expression for the deterrence factor is useful to help visualize all the ways that it can be effected, but when this complete expression is substituted into the model one term will drop out. In this destination-constrained model it is the term p_j; it drops out because the deterrence factor appears in the model twice, once directly in the model, and once in the expression for B_j, which is calculated by a single summation over i. (In the origin-constrained case the q_i term would drop out, and the p_j would be retained, by the same reasoning.)

If the estimate of β is presumed to be reliable, the distribution of the residual error term, ε_{ij}, can be obtained for any calibration since $\bar{T}_{ij}$, D_j, W_i, c_{ij} (and any other terms being used in the accessibility cost) are known values. The model can then be solved iteratively for the values of ε_{ij} although the process is complicated by the fact that f_{ij} appears twice in the model. Indirect methods are available for this solution (see Mayhew and Taket 1981).

5.3 RAMOS^{-1}

RAMOS^{-1} is a set of non-linear optimizing models with linear constraints, whereas RAMOS estimates T_{ij} given different configurations of supply and demand. The RAMOS^{-1} finds D_j, the caseload capacity, for each destination consistent with the residence pattern of the total availability of resources and the ease of travel on the region.

A number of different criteria for selecting P_j have been extensively tested (Mayhew and Leonardi 1981) and are currently available. One of these, the equity criterion, uses an objective function that allocates resources such that the relative needs in each area of residence are satisfied. If this criterion were unconstrained a majority of patients would be treated locally, perhaps calling for an unrealistic level of construction of new facilities in some areas, and excessive elimination or reduction in size of facilities in others. Therefore, the model has the capability of constraining the solution so as to hold the amount of change in any area, addition or elimination, to any range the analyst may feel is desirable or feasible. The ability to use variable upper and lower bounds on the amount of change also permits an investigation of the implications of what can be accomplished in various planning periods. Control over the rate of change allows time for conflict between the objective used by the model and any other objectives of the health care system, to become reconciled as the system evolves.

The results of two runs with the RAMOS^{-1} model are shown in Table 4. The same 1978 Massachusetts in-patient discharge and population data are used as before. The accessibility costs are based on actual road mileages, and the value of the parameter β for this data set as 0.16. (See Table 2 for the RAMOS calibration for this data set.) Both runs use no constraints on the amount of increase permitted in the facilities in any area so that areas having particular short falls can be identified; the first run, however, constrains the reduction in resources allocated to any one area to an arbitrary value of 10 percent, and the second run constrains it to 25 percent.

The results produced from these two runs, shown in Table 4, are intuitively reasonable; they show reduced caseloads in areas that appear to be oversupplied and increased caseloads in those that are undersupplied. The first run, which constrains the solution to the elimination of not more than 10 percent of the caseload capacity in any area, showed that 13 of the 23 HSA sub-regions were decreased by the full 10 percent permitted, 3 more areas had their caseload capacities decreased by less than 10 percent, and the caseload capacity of 7 areas were increased. The second run, constrained to a maximum decrease of 25 percent in any area, showed that 9 areas should have their caseloads decreased by the full 25 percent, 2 should be decreased by less than 25 percent, and 11 zones should be increased by some amount. Interestingly enough, 5 of the zones that in the first run indicated they should lose resources, have their

Table 4. Results from RAMOS^{-1} constrained for 10 percent and 25 percent decrease in caseload capacity, no constraint on percent increase.

Zone designation		CASELOAD CAPACITIES				
			Predicted Values			
		Initial value	Constrained to -10% change (actual)	(% change)	Constrained to -25% change (actual)	(% change)
1.1	Berk	21191.00	19072.00	(-10.00)	16074.11	(-24.15)
1.2	No V	15347.00	20554.94	(33.93)	25156.32	(63.92)
1.3	So V	57514.00	51763.00	(-10.00)	43135.00	(-25.00)
2.1	No W	19913.00	18839.43	(-5.39)	21415.28	(7.54)
2.2	Ce W	48661.00	43795.00	(-10.00)	36496.00	(-25.00)
2.3	So W	11339.00	24180.99	(113.26)	31576.37	(178.48)
3.1	L/Tk	20098.00	18088.00	(-10.00)	26279.83	(30.76)
3.2	L/Ad	15918.00	25857.75	(62.44)	27696.14	(73.99)
3.3	Newb	11516.00	24638.21	(113.95)	29687.10	(157.79)
4.1	Ca M	175991.00	158392.00	(-10.00)	131993.00	(-25.00)
4.2	NW M	44237.00	39813.00	(-10.00)	33178.00	(-25.00)
4.3	We M	37859.00	34073.00	(-10.00)	28394.00	(-25.00)
4.4	SW M	10429.00	9386.00	(-10.00)	18370.13	(76.14)
4.5	So M	21675.00	19705.00	(-9.09)	16256.00	(-25.00)
5.1	Attl	14038.00	20925.49	(49.06)	22035.80	(56.97)
5.2	Br/P	27053.00	24348.00	(-10.00)	31705.09	(17.20)
5.3	Fa R	35281.00	32910.92	(-6.72)	42965.82	(21.78)
5.4	C Cd	13902.00	19270.49	(38.62)	21257.49	(52.91)
6.1	C An	10266.00	12639.77	(23.12)	23617.41	(130.05)
6.2	C/Sa	17049.00	15344.00	(-10.00)	12789.00	(-24.99)
6.3	Lynn	13161.00	11845.00	(-10.00)	9871.00	(-25.00)
6.4	E Mx	12934.00	11686.00	(-10.00)	9738.00	(-25.00)
6.5	TriC	17060.00	15354.00	(-10.00)	12795.00	(-25.00)

^{a}1978 in-patient discharge data from Massachusetts, using actual road mileages, calibration parameter β = 0.160.

allocations increased in the second run. This is an artifact of the constraints on change, and the ripple effect caused by trading off allocations between different areas. When large changes are permitted, the areas that are badly out of line will have large changes, and those areas that are at very nearly the right level will have little or no change. However, when constraints are tightened and the amount of changes permitted is limited to small values, those areas that are badly out of line can only be changed by small values, and the remaining change must be made up by re-shuffling allocations elsewhere.

6. CONCLUSIONS

On the basis of this work three general conclusions seem justified:

1. The 1978 Massachusetts data does fit the RAMOS model, and calibrations can be obtained with good agreement between *actual* and *predicted* patient flow provided that the actual number of admissions that occur is used as the caseload capacities (D_j).
2. The calibration of the RAMOS model shows that patients tend to travel farther for pediatric care than for general medical-surgical care, and further for general medical-surgical care than for obstetric and maternity care. This result is both intuitive and generally compatible with earlier work with the model in England.
3. This initial success calibrating RAMOS with the Massachusetts data shows that it is possible to use this model with the market-oriented health system in the United States. However, there is still a long way to go before RAMOS can be considered a valid tool for planning health and medical care delivery in Massachusetts. The more immediate steps in the process of transforming this model into a usable tool are described in the next section.

7. FUTURE RESEARCH

There are many avenues along which research could proceed, but the first priority must be an examination of the validity of RAMOS for prediction using Massachusetts data. A second possible extension of this work is to use DRAMOS (Disaggregated Resource Allocations Model Over Space), a model that attempts to incorporate disaggregation into the basic approach of RAMOS by including the relative elasticities of various care specialties at different resource levels and the average standards of treatment received (Mayhew 1981). These issues are briefly examined below.

7.1 Validity of Prediction

The predictive power of the model cannot be tested until data from more than just the calibration year become available so that predictions of the model can be compared with what actually happens. Ordinarily, the time period for this validation would be greater than two years in order to provide an adequate test of the model.* In the case of Massachusetts data, the dependent variable T_{ij} can be analyzed using a 28×33 matrix. The differences between the predicted values and actual values obtained two years later will consist of 644 elements, which can be analyzed as an error distribution.

For planning, any errors in prediction will be a function both of the validity of the model itself and of the ability of the analyst to make accurate projections of the caseload capacity and the patient generating factors. These latter would be derived from an analysis of bed availability.

Health planners sometimes prefer a greater degree of disaggregation than has been discussed so far. Results might be

---

*If the time necessary to process the hospital discharge data and prepare the data elements is considered, it is well to realize that the information available at time t_1 may already be close to one year old, so the span between the period covered by the data and the period of prediction may be two years or more.

desired in terms of specialty groups, age and sex categories, racial or religious characteristics, and so forth. This will be up to the analyst and the types of problems being studied. However, it is recognized that the greater the degree of disaggregation, the greater the possible imprecision of the predictions. Therefore, sensitivity of the model to changes in the basic data elements will also need to be closely checked before the model can be considered a valid planning tool for routine use.

7.2 Substitutability of Facilities

RAMOS is limited to a single category of patient. Here it was run separately for medical-surgical, pediatric, obstetric, psychiatric, and all categories together. Hospital facilities for these types of patients are generally kept separate, but in emergencies or during reorganization there is flexibility, and patients in one category may use facilities formerly reserved for another. DRAMOS was developed in order to take account of interactions between different categories of patients and different standards of treatment (Mayhew 1981).

It incorporates the elasticities of categories of patients relative to resource availability into the basic framework of the RAMOS family of models. The main output is a predicted patient flow matrix $[T_{ijk}]$ that adds the dimension of patient category k to the former dimensions of origin and destination, i and j. It would be a simple extension to include different modes of care m, such as out-patients, in the framework as well. With DRAMOS it would then be possible to analyze the interrelationships between in-patient and out-patient care for various populations defined by category of care (or perhaps diagnosis), location, method of payment, and so forth.

The calibration and validation of the DRAMOS model is more complex than that of RAMOS, and the data requirements more exacting. If different modes of care are considered, research will need to identify the relative differences in cost between them, for example, in the case of out-patient and in-patient care. It is also important to examine whether or not accessibility costs need to be redefined for each mode separately. Also, if out-patient care is included in the model, more research will be needed to check the suitability of the patient generating functions, W_i. There is also the problem of the determining stability of the basic variables (hospitalization rates and lengths of stay) for each of the disaggregated categories of patients in the model. This issue is easier to study with each category individually before the categories are considered interactively in DRAMOS.

For these reasons it is prudent to postpone work on DRAMOS until each category of patient is calibrated and validated individually. When this is complete, however, the insights obtainable should be very useful for planning because DRAMOS allows for the examination of issues beyond the scope of RAMOS, such as the implications both of the choice of mode and of the flexibility in the use of facilities.

REFERENCES

Aspden, P., and M. Rusnak (1980) *The IIASA Health Care Resource Allocation Submodel: Model Calibration for Data from Czechoslovakia*. WP-80-53. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Batty, M., and S. Mackie (1972) The Calibration of Gravity, Entropy and Related Models of Spatial Interaction. *Environment and Planning* 4:205-233.

Gibbs, R.J. (1978a) *The IIASA Health Care Resource Allocation Submodel: Mark 1*. RR-78-8. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Gibbs, R.J. (1978b) *A Disaggregated Health Care Resource Allocation Model*. RM-78-001. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Hughes, D.J. (1978) *The IIASA Health Care Resource Allocation Submodel: Estimation of Parameters*. RM-78-067. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Hughes, J.D., and A. Wierzbicki (1980) *DRAM: A Model of Health Care Resource Allocation*. RR-80-23. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Mayhew, L.D. (1980) *The Regional Planning of Health Care Services: RAMOS and RAMOS^{-1}*. WP-80-166. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Mayhew, L.D. (1981) *DRAMOS: A Multi-Category Spatial Resource Allocation Model for Health Service Management and Planning.* WP-81-039. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Mayhew, L.D. (1982) Correspondence with E.J. Rising. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Mayhew, L.D., and A. Taket (1980) *RAMOS: A Model of Health Care Resource Allocation in Space.* Laxenburg, Austria; International Institute for Applied Systems Analysis.

Mayhew, L.D., and A. Taket (1981) *RAMOS: A Model Validation and Sensitivity Analysis.* WP-81-100. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Mayhew, L.D., and G. Leonardi (1981) *Equity, Efficiency, and Accessibility in Urban and Regional Health Care Systems.* WP-81-102. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Morrill, R.L., and M. Kelley (1970) The Simulation of Hospital Use and the Estimation of Location Efficiency. *Geographical Analysis* 2:283-299.

Rousseau, J.M., and F.J. Gibbs (1980) *A Model to Assist Planning the Provision of Hospital Services.* CP-80-003. Laxenburg, Austria: International Institute for Applied Systems Anlaysis.

Shannon, G.W. (1969) The Concept of Distance as a Factor in Accessibility and Utilization of Health Care. *Medical Care Review* 26: 141-161.

Shannon, G.W. (1975) A Method for Evaluating the Geographic Accessibility of Health Services. *The Professional Geographer* 27(1):30-36.