RM-77-53

HEALTH CARE RESOURCE ALLOCATION MODELS - A CRITICAL REVIEW

R.J. Gibbs

October 1977

Preface

The aim of the IIASA Modeling Health Care Systems Task is to build a National Health Care System model and apply it in collaboration with national research centers as an aid to Health Service planners. The research envisaged is described in the IIASA Research Plan 1977. It involves initially the construction of four linked sub-models dealing with population, disease prevalence, resource allocation, and resource supply. This paper is concerned with resource allocation. It reviews and classifies the literature on resource allocation models in the Health Care field and suggests which type of model is appropriate to the IIASA Task.

Abstract

The purpose of a resource allocation sub-model, within the IIASA National Health Care System (HCS) model, is to represent the interaction between the demand for and supply of health care resources. The overall model can then be used to examine the consequences of alternative health care policies, particularly policies concerned with the production of new resources. The HCS resource allocation models in the literature can be classified into three types: macro-econometric, behavior simulation, and system optimization. For each type some examples are described and the strengths and weaknesses of the approach are assessed. The macro-econometric approach has many advantages; it is well-tried and has standard methods, terminology, and computer programs. But it has the important limitation that its results are only strictly valid over the ranges of the variables that exist in the data from which the econometric equations are estimated. Thus the domain of its valid application is normally limited to small variations around the status quo. By contrast, the behavior simulation approach lacks many of the advantages of the macro-econometric approach but is, in principle, capable of exploring a wider range of situations; *provided that the behavioral hypotheses are sound*, such a model can be used to explore situations which are radically different from the status quo and in which variables may lie outside the ranges observed to date. In theory the optimization approach produces the ideal solution but in practice, if applied to the HCS, the approach is likely to founder on the difficulty of defining an objective function that both expresses reasonable objectives for the HCS as a whole and, at the same time, takes adequate account of the practice of the actors in the HCS. Accordingly the behavior simulation approach is recommended for the IIASA Task.

Health Care Resource Allocation Models - A Critical Review

1. THE ROLE OF A RESOURCE ALLOCATION SUB-MODEL WITHIN THE IIASA NATIONAL HEALTH CARE SYSTEM MODEL

1.1 Background

The main task of the IIASA Health Care System (HCS) Modeling team is to build a National HCS model and to apply it as an aid to HCS planning in collaboration with National Centers. The long-term strategy for this task, as set out in earlier papers by Venedictov [1] and Kiselev [2], envisages the construction of a mathematical simulation model relating activities both within the HCS and between the HCS and other interacting systems (e.g. the population, environment, and socio-economic systems). The purpose of the simulation model is to illuminate the future consequences of alternative policies both for the HCS and the interacting systems and thus assist planners to examine strategic options.

Within this framework the current short-term plan for the IIASA HCS Modeling Task, as set out in the IIASA Research Plan 1977 [3] and in a recent paper by Shigan [4], is to concentrate effort initially on modeling the HCS itself and its interaction with one external system - the population system. The plan envisages the construction of four connected sub-models dealing with population, the estimation of disease prevalence, the supply of resources (for health services), and the allocation of these resources to competing demands (see Figure 1). The demands for health care resources comprise treatment activities for the sick population and screening and preventive activities for the general population. The output of the resource allocation sub-model describes the immediate consequences of the interaction between demand and supply. This output will be of interest to the user in its own right and will be the first product of the integrated HCS model. (There are of course longer term effects of the resource allocation process on population, prevalence, and resource supply, e.g. the effect of a treatment program on mortality, and it is hoped to eventually incorporate these effects into the integrated HCS model by means of feedback loops from the resource allocation sub-model to the other three).

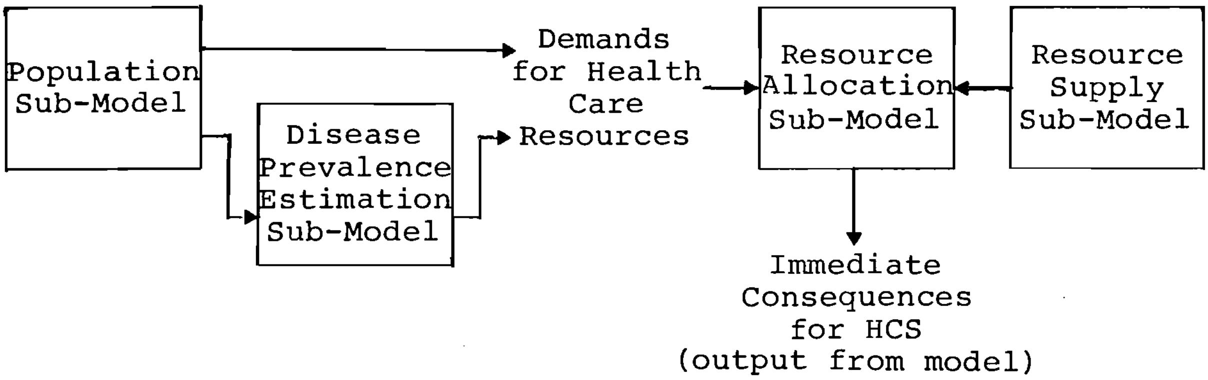

Figure 1. Schema for a National HCS Model containing four connected sub-models.

The first stage of implementing this plan resulted in the design of a prototype model, along the lines of the schema in Figure 1, by Klementiev [5]. A working version of this prototype model, concerned with degenerative disease only, was built and run with hypothetical test data, as described by Olshansky [6]. However, the Research Plan [3] recognizes that a much more elaborate version of the model will need to be developed to represent the activities of the HCS as a whole. In the elaboration of the model, work is required on all four of the sub-models of the schema in Figure 1. This Research Memorandum is concerned with one of these: the resource allocation sub-model.

1.2 The Purpose and Form of this Memorandum

The purpose of this Memorandum is to review the literature on HCS resource allocation models, to classify the models into different types, to consider their advantages and disadvantages, and to assess the type that is most appropriate for the IIASA HCS Modeling Task. It is hoped that the paper will stimulate discussion and comment among scientists working in this field.

The next stage of the work will be to construct a more elaborate resource allocation sub-model.

The literature review here builds upon the start made by the review-analysis of National HCS models by Fleissner and Klementiev [7]. Their review-analysis was concerned with reviewing all types of National HCS model whereas this Memorandum is concerned only with models dealing with resource allocation. This Memorandum considers certain models in more detail, and attempts to go beyond reviewing into drawing conclusions about the appropriateness of different types of model for different tasks. The previous review-analysis [7] suggested a classification of National HCS models into three types: econometric, simulation, and optimization. A modified version of this classification is employed in this paper where the following three types of resource allocation model are defined:

macro-econometric: models consisting of linear equations (or transforms of linear equations) relating aggregate variables such as consumption, supply and price of health services, and population attributes, whose parameters are estimated by multiple regression analysis of current or historic aggregate data;

behavior simulation: models based on hypotheses concerning the behavior of physicians, patients, and other decentralized decision makers in the HCS; and

system optimization: models designed to identify the set of resource allocations that optimize a defined objective function of the HCS.

Each type of model is reviewed, in turn, in Sections 2, 3, and 4, and their advantages and disadvantages are assessed. In Section 5 the way in which each type of model could fit into the HCS modeling schema of Figure 1 is considered and the appropriateness of each type of model for the IIASA HCS Modeling Task is assessed.

2. MACRO-ECONOMETRIC MODELS

2.1 Introduction

Many examples of macro-econometric models can be found in the literature on health system resource allocation; see for example the bibliography assembled in reference [8]. The common characteristic of these models is that their hypotheses are expressed in terms of linear relationships (or transformations of linear relationships) between aggregate amounts of quantities such as consumption, supply, price of health care services, and factors describing attributes of the population and the environment. The other distinguishing feature of the approach is that the coefficients of the equations expressing these relationships are estimated by multiple regression analysis using cross-section and/or time series data. This estimation process also allows for testing of the hypotheses in the sense that those equations which fail tests of statistical significance are rejected. This means that a large number of alternative equations (hypotheses) can be tested and the eventual econometric model consists of the sub-set of equations (hypotheses) which have survived the tests.

Three examples of the macro-econometric approach will now be considered in detail: Yett, et al. [9], Feldstein [10], and Harris, D. [11]. These examples illustrate the wide range of relationships which have been examined using this approach. Following this the advantages and disadvantages of the approach will be assessed.

2.2 The Model of Yett, et al.

Some authors, particularly those from countries where the HCS has a predominant private sector, offer hypotheses in which the consumption of health care is mainly dependent upon supply and price variables. Consider, for example, the model of Yett, et al. [9], which describes the HCS with a set of 47 equations. The model can be illustrated by examining the four equations describing one particular sector of the HCS, the in-patient activity of short-term voluntary and proprietary (STVP) hospitals; these equations are displayed in Table 1.

Table 1. Four illustrative equations from the model of Yett, et al. [9], describing the in-patient activity of STVP hospitals.

$$(1)\quad \text{PD-P} = \underset{(-2.38)}{-36.2394}\ \text{P-HP/P-OP} + \underset{(5.67)}{61.1378}\ \text{\%OLD} + \underset{(4.91)}{656.3433}\ \text{HBEN/P-HP}$$

$$- \underset{(-9.32)}{0.9308}\ \text{PDGA} + 317.8782\ , \qquad \bar{R}^2 = .81\ , \ \text{S.E.} = 119.4853\ .$$

$$(2)\quad \text{P-HP} = \underset{(22.40)}{0.8971}\ \text{P-HP}_{t-1} - \underset{(2.04)}{0.1418}\ \text{OCCP} + \underset{(3.48)}{2.1657}(.25\ \text{W-RN} + .07\ \text{W-AH} + .13\ \text{W-PN} + .55\ \text{W-NM})$$

$$+ \underset{(4.05)}{7.0271}\ \text{EMBD} + \underset{(0.56)}{0.0525}\ \text{KPBD} - 0.8362\ , \qquad \bar{R}^2 = .96\ , \ \text{S.E.} = 3.4737\ .$$

$$(3)\quad \text{OCCP} = \frac{\text{PD-P}}{0.00365\ \text{BEDP}}\ .$$

$$(4)\quad \text{BEDP} = \underset{(15.25)}{0.7958}\ \text{BEDP}_{t-1} + \underset{(3.74)}{0.7645}\ \text{PD-P} + \underset{(1.26)}{0.0226}\ \text{HBFF}_{t-2} - 0.0554\ , \qquad R^2 = .99\ ,$$

$$\text{S.E.} = 0.1092\ .$$

where PD-P = Annual number of in-patient days provided by STVP hospitals (millions);
PH-P = Average daily service charge in STVP hospitals (dollars);
P-OP = Average revenue per out-patient visit at STVP hospitals (dollars);
%OLD = Percentage of population aged 65 and over;
HBEN = Benefits per capita for hospital care paid by private and public insurance programs (dollars);
PDGA = Weighted average of in-patient days provided by STVP hospitals per thousand population;
OCCP = Average percentage occupancy rate in STVP hospitals;
W-RN = Annual wage paid to general duty hospital RNs (thousands of dollars);
W-AH = Annual wage paid to health professionals (thousands of dollars);
W-PN = Annual wage paid to practical nurses (thousands of dollars);
W-NM = Annual wage paid to non-medical hospital employees (thousands of dollars);
EMBD = Number of personnel per bed in STVP hospitals;
KPBD = Value of plant assets per bed in STVP hospitals (thousands of dollars);
HBFF = Hill Burton funds for hospital construction (millions of dollars).

Equation (1) represents consumption (P-DP) as a function of price (P-HP), price reimbursement (H-BEN/P-HP), the activity of a "competing" set of hospitals (PDGA), and an attribute of the population (%OLD). Equation (2) represents price as a function of price in the previous year ($P\text{-}HP_{t-1}$), the intensity of bed utilization (OCCP), weighted wage rates (W-RN, etc.), staffing levels (EMBD), and current capital valuation (KPBD). Equation (3) is a definitional equation showing that occupancy is the ratio of consumption to supply. Equation (4) represents supply (BEDP), consumption (PD-P), and the availability of capital building funds in the year before last ($HBFF_{t-2}$). Although some of these variables occur in some other equations in the model, these four equations suffice to illustrate the approach. Essentially the equations represent how the aggregate quantities of consumption, supply and price of hospital services mutually influence each other; only one quantity, (%OLD), describes a feature that is exogenous to the hospital system.

The endogenous variables in the complete model include variables describing in-patient and out-patient care, including not only STVP hospitals but also state and local government ones, and a number of categories of health manpower (e.g. surgical specialists in private practice, registered nurses). The exogenous variables consist mainly of variables describing insurance parameters (e.g. percentages of the population enrolled in different types of scheme) and aspects of federal or other government intervention such as parameters of the Medicare scheme, availability of federal funds for new hospital construction, and the output of medical schools.

The coefficients of the equations in the model were estimated separately by multiple regression analysis of cross-section data. The performance of the model was then tested by initializing the endogenous variables to the 1975 levels for the State of California, solving the equations simultaneously year by year, and comparing the results with historical data for California for the period 1968 to 1977. Yett, et al. illustrate how the model can be used as an aid to planning by forecasting the year by year consequences for the endogenous variables of policy generated changes in the exogenous variables.

2.3 The Models of Feldstein

Although the macro-econometric approach has been most widely used in countries whose HCS have strong private sectors, it has also been used in countries with predominant public sectors. For example one of the most well-known studies is that by Feldstein [10] of the U.K. National Health Service. In the U.K. the consumer usually pays no direct price for health care services. Thus price variables are irrelevant. Feldstein chose not to include population and environmental variables (except for population as the denominator in certain consumption and supply variables). So his equations represent consumption variables as functions of supply variables alone.

Most of Feldstein's equations treat aggregate consumption as a function of various groups of supply variables at differing levels of aggregation. Consider, for example, one of the production functions he suggests for U.K. acute hospitals (equation 4.15, p. 98 of Feldstein [10]):

$$W = AM^{0.387}B^{0.465}N^{0.047}S^{0.069} \quad ,$$

where W = weighted number of in-patients treated p.a.,

M = supply of doctors,

B = supply of beds,

N = supply of nurses,

S = other supplies.

Such an equation allows one to predict the aggregate consumption of hospital care in terms of the supplies of the different inputs of a hospital. With equations of this type Feldstein analyzes the likely consequences for hospital output (i.e. aggregate consumption) of changes in input ratios and proceeds to compute input ratios that are optimum under stated conditions. One of his most interesting findings is that an increase in the ratio of doctor supply to the other hospital inputs, within the same total expenditure, would be expected to elicit an increase in production, i.e. an increase in the number of in-patients treated.

In another part of his analysis (Chapter 7), Feldstein represents disaggregated consumption variables as functions of a single aggregate supply variable. The equations are of the form:

$$\log C_i = a_i \log S + b_i \quad ,$$

where C_i = consumption by disease type i,
S = aggregate supply of acute hospital beds,
a_i, b_i = coefficients to be estimated.

Thus the coefficient, a_i, is the elasticity of consumption by disease type i with respect to supply. Feldstein used three types of consumption variable for each disease type: beds used, and its two components, cases treated, and mean length of stay. An illustrative set of his results, for certain disease types, is shown in Table 2 (from Table 7.10 in Feldstein). Consider

Table 2. Elasticities of consumption, by selected diagnoses, to aggregate bed supply* for English acute hospitals in 1960, from Feldstein [10]. (Standard errors in brackets.)

Disease (1)	Beds Used** (2)	Admissions** (3)	Mean Stay (4)
Acute appendicitis	0.15(0.36)	-0.16(0.33)	0.31(0.17)
Acute upper respiratory infections	2.57(1.00)	1.53(0.52)	1.04(0.74)
Peptic ulcer	0.85(0.52)	0.29(0.40)	0.56(0.51)
Abdominal hernia (female)	1.39(0.44)	0.52(0.22)	0.87(0.44)
Haemorrhoids	1.14(0.62)	0.70(0.48)	0.44(0.24)
Tonsils and adenoids	0.55(0.46)	0.23(0.38)	0.33(0.38)
Arteriosclerotic heart disease	2.22(0.70)	1.14(0.51)	1.08(0.99)
Malignant neoplasms	0.58(0.30)	0.68(0.29)	-0.10(0.20)
Varicose veins (female)	1.40(0.70)	0.78(0.41)	0.62(0.67)
Males	1.03(0.14)	0.66(0.13)	0.37(0.15)
Females	0.97(0.11)	0.63(0.17)	0.34(0.21)
All Persons	—	0.65(0.15)	0.35(0.15)

* Elasticities calculated with respect to total number of beds per 1 000 population.

** Per 1 000 population, 1960.

for example the results for haemorrhoids. These can be interpreted as implying that a 1% increase in aggregate bed supply is associated with a 1.14% increase in beds used for haemorrhoids patients, which in turn is composed of a 0.70% increase in the number of haemorrhoids cases and a 0.44% increase in their mean length of stay.

In interpreting these results and considering their relevance to this type of analysis we can start by quoting from Feldstein's text (pp. 220-221):

> The elasticity values for acute appendicitis and acute upper respiratory infections are quite satisfactory. Appendectomies are very inelastic to bed scarcity while the less serious respiratory infections are highly elastic; in addition, the former's elasticity is properly concentrated on the mean stay while the latter's is almost completely in the number of admissions. The values for peptic ulcer also seem appropriate, showing a low overall elasticity which is concentrated on the mean stay.
>
> The results for some other disease groups are not as satisfactory. Cases of abdominal hernia and of haemorrhoids show greater than average overall elasticity. It seems inappropriate that important surgical repair procedures such as these should occupy a smaller proportion of beds in regions of greater relative scarcity and that the number of cases treated should be no less elastic than average.
>
> In contrast, it would seem desirable that admissions for tonsillectomy and adenoidectomy should be highly elastic; in regions where beds are relatively more scarce, tonsillectomy and adenoidectomy cases should occupy a smaller proportion of available beds. The decision to operate for tonsillitis has often been cited as an example of medical fashion that is not founded on medical knowledge. Nevertheless, one in twenty hospitalizations in Britain are for tonsillectomy and adenoidectomy. And, more important for our current discussion, the number of these procedures and the length of stay show very low elasticity to bed scarcity.
>
> Arteriosclerotic heart disease, including coronary, usually presents very serious medical cases; nearly a third of hospital admissions in this category die in hospital. Despite this, the number of cases admitted is extremely elastic to bed scarcity. Patients with malignant neoplasms (cancer) also have a high hospital fatality rate; more than a fifth of all admissions die in hospital. This is about 50 per cent higher than in

> the United States, reflecting in part the greater tendency in Britain to keep terminal cancer cases in hospital after there is no longer any hope of helping them. As the table shows, this length of stay is unaffected by relative bed scarcity.
>
> Women with varicose veins enter hospital for a surgical operation. The high elasticity of bed use for varicose veins, with the greater proportion of this due to the number of cases, probably reflects a greater reliance on alternative methods of outpatient treatment, as well as a generally lower rate of care, in those regions in which beds are relatively more scarce.
>
> It is difficult to understand why such striking examples of inappropriate elasticities should have been found. It may be possible that some of the results can be explained by differences among the regions in the actual incidence of the diseases. Since the hospital statistics on which this study has been based provide the only sound measure of morbidity for these conditions, the influence of area-specific incidence rates could only be studied for those diseases in which mortality is a good indicator of morbidity. Doing so for heart disease and cancer does not suggest that the geographical pattern of mortality would explain our findings.
>
> We shall not try to offer any behavioral explanations of the individual elasticity values. To do so properly, we should have to develop a complex theory of medical admission and treatment decisions based on the factors that motivate patients to seek care and the way in which doctors diagnose and treat each type of disease. But again it is likely that any explanation should begin by recognizing that the doctors themselves are not aware of the allocation patterns that they have established.

Although Feldstein considers that some of his elasticity estimates are inappropriate there is good reason to believe that they correctly represent hospital practice, or at least hospital practice as it was in 1960. At that time tonsillectomy and adenoidectomy were common treatments for cases of infections of the tonsils and adenoids--although this practice is much less common today--and the low elasticity values are therefore plausible. Similarly the high elasticity values for hernia and haemorrhoids are consistent with the fact that there were (and often still are), long waiting lists for these conditions in areas of relative bed scarcity. Thus, given the prevailing medical practice and within the ranges of his data, Feldstein could reasonably claim that his elasticity values represent how the pattern of admissions and length of stay would be likely to respond to changes in the levels of total bed availability.

2.4 The Model of Harris, D.

Most econometric studies in this field, like those of Yett and Feldstein, have placed relatively little importance on factors exogenous to the HCS. There are however some notable exceptions, e.g. Harris, D. [11], Fleissner [12], Newhouse [13]. For example in the study by Harris, D. [11], variables describing consumption are represented as functions of both supply variables and variables describing certain population characteristics. The variables describing population characteristics were derived from a factor analysis of 21 population characteristics in the 56 New York State counties (see Table 3). Four factors were identified and Table 3 shows the loadings of each of the original 21 characteristics on each factor, and the commonality estimates (h^2) which state the proportion of variance in each characteristic explained by the four factors. From inspection of the loadings Harris interprets the factors as follows:

Factor 1: "metropolitan/middle class" - predominant in counties with a metropolitan character and high socio-economic status.

Factor 2: "age/illness" - predominant in counties with large elderly populations and chronic morbidity.

Factor 3: "fertility/family" - predominant in counties with high fertility, large young population, and low educational level.

Factor 4: "city/inner suburb" - predominant in counties containing a central city and a high proportion of unmarried women.

Harris then fitted a number of equations relating these factors and supply and consumption variables by multiple regression analysis of cross section data from 56 New York State counties for 1970. Harris's equations were based on a three stage causal model. The stages of causation are studied by path analysis, an elaboration of multiple regression analysis, which reveals

Table 3. Factor analysis of population characteristics for 56 New York State counties (definitions from 1970 Census), from Harris [11].

Variables in each factor are indicated in italics.

Variable	Factor 1	Factor 2	Factor 3	Factor 4	h^2
1. Median family income........	*0.927*	-0.243	0.034	0.071	0.924
2. Personal income per capita..	*0.876*	0.096	-0.210	0.099	0.830
3. Population per sq. mile.....	*0.867*	0.014	-0.179	-0.077	0.789
4. Size of population..........	*0.827*	-0.050	-0.089	0.160	0.720
5. Percent of families below poverty level..............	*-0.774*	0.205	-0.059	-0.129	0.661
6. Percent of work force in white-collar jobs...........	*0.771*	-0.308	-0.258	0.158	0.780
7. Percent of pop. living in urban areas..............	*0.738*	-0.135	-0.107	*0.483*	0.807
8. Percent nonwhite............	*-0.674*	-0.018	-0.092	0.149	0.486
9. County located in SMSA......	*0.560*	-0.230	0.241	*0.514*	0.689
10. Percent of pop. over 25 who have completed 13 or more years of school............	*0.551*	*-0.548*	*-0.391*	-0.127	0.773
11. Crude death rate............	*-0.497*	*0.782*	-0.194	0.083	0.903
12. Percent of pop. age 65 and over....................	*-0.428*	*0.811*	-0.287	0.083	0.930
13. Median age..................	0.252	*0.894*	-0.174	0.094	0.902
14. Percent reproductive age females (15-19)............	0.251	*-0.797*	*0.438*	0.121	0.904
15. Percent of pop. under age 5......................	-0.121	*-0.441*	*0.846*	0.044	0.926
16. Crude birth rate............	*-0.477*	*-0.407*	*0.576*	0.165	0.752
17. Fertility ratio*............	-0.314	0.018	*0.905*	-0.099	0.927
18. Percent females (14 and older) married.............	0.213	0.256	*0.795*	*-0.346*	0.862
19. County has central city.....	0.249	-0.109	-0.056	*-0.681*	0.540
20. Sex ratio..................	-0.176	-0.170	0.302	*-0.056*	0.531
21. Infant mortality rate.......	-0.059	*0.335*	0.038	*0.494*	0.361

*Number of children age 5 and under per female between 15 and 54 years of age.

the causal effect of one variable on variables at later stages in the model both directly and via variables at intermediate stages. The model structure and results are shown in Figure 2. The numbers on the linking arrows are the standardized partial regression coefficients; they measure the effect of each causally prior variable on each subsequent one. The numbers on the arrows without origins are the square roots of the unexplained variance.

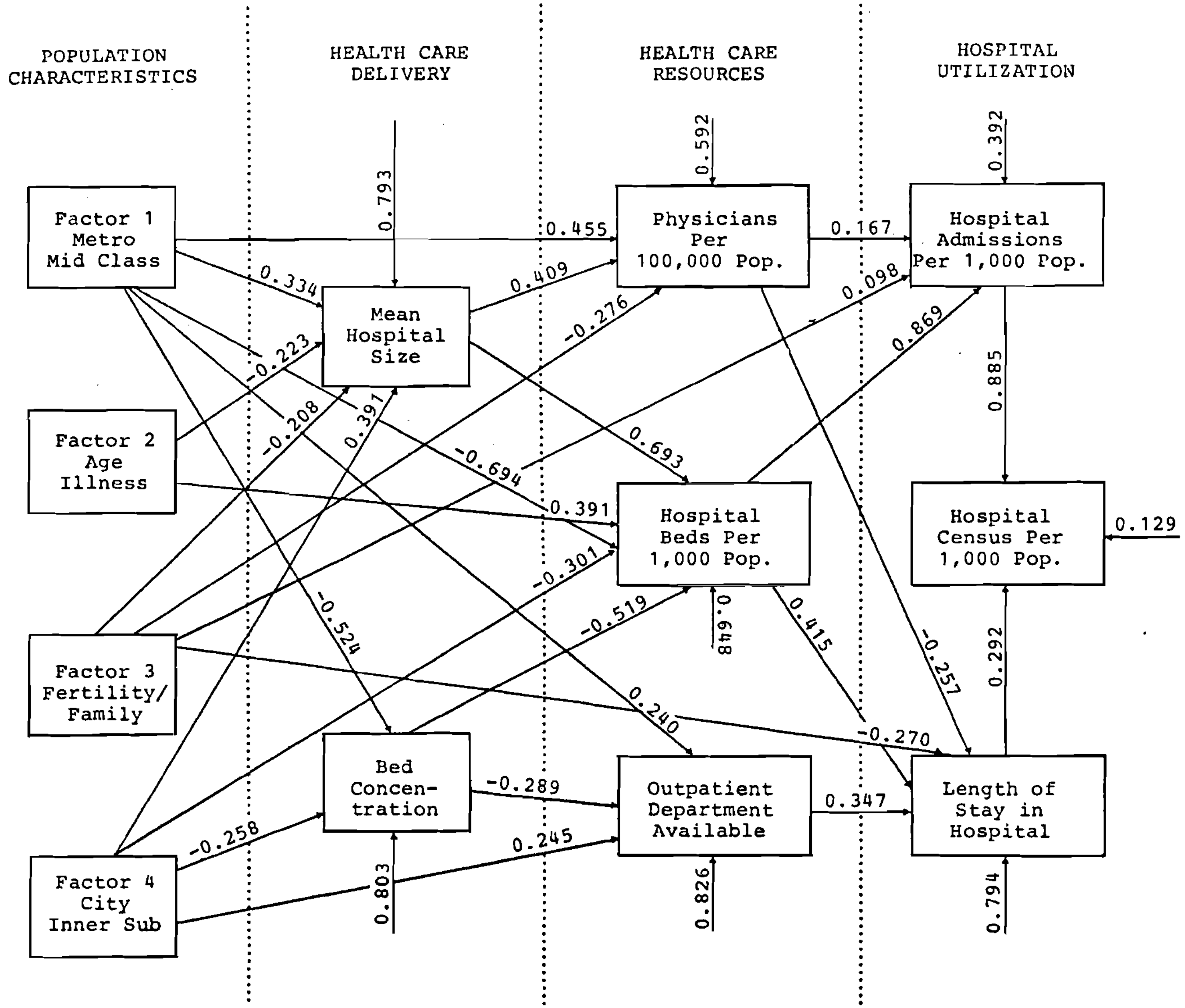

Figure 2. The results of the study of Harris [11], expressed in terms of path analysis.

The results in Figure 2, coupled with the zero order matrix (not shown here), allow one to trace causation. For example the population factor 1, "metro/middle class", has a weak net negative effect on the hospital admissions variable, as revealed by a zero order correlation coefficient of -0.13, but the path analysis reveals that this is the net sum of two strong countervailing indirect effects. The first, positive, indirect effect is via the positive causal effects from (a) factor 1 to physician supply and (b) from physician supply to hospital admissions. The second, stronger, negative, indirect effect is via (a) the negative effect of factor 1 on hospital bed supply and (b) the positive effect of hospital bed supply on admissions.

From his results Harris concludes that for the most part hospital consumption variables depend causally on supply variables and that the effects of population characteristics on consumption are mainly transmitted indirectly via supply variables and only to a small extent directly. This, according to Harris, leads to the following implications.

> The results of the present study, coupled with those of a companion study, using longitudinal data, should help end the controversy surrounding Roemer's thesis*, at least in urbanized areas. It should now be clear that supply can create its own demand rather than that demand leads to congruent levels of supply. To supply additional beds to areas with current high demand (utilization) in an attempt to "satisfy" the "unmet need" in these areas is thus seen as a futile exercise. Additional beds will always lead to additional use, and health care expenditures will continue to rise. To base hospital construction priorities primarily on past use and current demand is hopeless since areas will never have "enough" beds.
>
> Excessive hospitalization and its high costs can be reduced by controlling the number of beds available to a local population and its physicians, as well as by controlling the supply of physicians. Money that would have been spent on hospital stays can then be used to help finance and develop more rational health care delivery systems that make more extensive use of ambulatory care and other substitutes for inpatient care.

---

*Roemer's thesis is that hospital utilization is primarily dependent upon supply factors; see reference [14].

Harris suggests a sequence of causation, as shown in Figure 2, of the following type: population characteristics → health care delivery → health care resources → hospital utilization. If one accepts this sequence, which is based purely on a priori considerations, then one can also accept the important result that hospital utilization depends mainly on supply factors. This result can then be used in planning, as Harris suggests, by adjusting planned levels of supply so as to achieve some desired levels of utilization.

2.5 Review

Having looked at three examples of the macro-econometric approach to modeling HCS resource allocation let us assess its advantages and disadvantages.

First the advantages. The approach has been widely used. The three examples described above illustrate both the wide range of relationships that have been studied and the importance of the relationships which show how the pattern of consumption of health services is influenced mainly by the pattern of supply. The approach is relatively easy to apply since standard methods and computer programs exist and results are relatively easy to compare using a standard terminology.

The data requirements are not usually excessive. The approach has been successfully used in both private and public HCS and there are grounds for believing that certain types of macro-econometric model might be universally applicable, at least among the IIASA member countries.

There are however some important disadvantages and limitations. Macro-econometric models relate aggregate amounts of quantities in linear (or transformations of linear) equations and the coefficients are estimated from regression analysis of cross section or time series data describing current or historic values of these quantities. It follows that the application of such models in a predictive or planning mode is only valid, strictly speaking, for situations in which

(1) the value of each variable is within the range observed in the data;

(2) there is no structural change in the HCS or related systems, e.g. a large-scale introduction of screening or a massive pollution of the water supply; and

(3) there is no major change in the way in which key resources are used, e.g. telephonic medical consultation, the use of hospitals more for nursing care and concentrating medical care more on ambulatory and domiciliary settings.

Thus the approach is only strictly valid for describing incremental changes around the status quo. The hypotheses underlying the equations describe how total quantities of the variables relate to each other in the present and/or recent past.

The approach is essentially descriptive - it allows one to efficiently summarize the current or recent historic relations between variations in the variables. This is a serious limitation if the main reason for building a model of the HCS is to assist decision makers consider non-incremental changes.

We now turn to the second type of resource allocation model--behavior simulation models.

3. BEHAVIOR SIMULATION MODELS

3.1 Introduction

In this type of model there are hypotheses concerning the behavior of the consumers and the suppliers of health care, i.e. patients and HCS personnel. These models can directly represent how scarce resources are rationed between competing demands. Three examples will be described: Rousseau [15], McDonald, et al. [16-18], and Klementiev [5].

3.2 The Model of Rousseau

Rousseau [15] starts by dismissing the approach in which resource requirements are calculated directly from population characteristics. He supplies evidence showing "...that on the contrary the demanded resources or consumption of resources is directly related to the available resources and that in practice and in the global perspective one has to accept the hypothesis that the demand could never be saturated, or if we can envisage a saturation, it is not at a level society can afford". Here he echoes the conclusion of Harris, D. [11] and many other authors. Indeed no author of an HCS resource allocation model has argued against this conclusion.

Rousseau tackles the problem of allocating known quantities, A_i, of medical services (treatments) by category to different categories of physicians (specialists), each with a given capacity, M_s, of work. The variables in the model are the allocations X_{is} of services to physicians. The behavioral hypothesis is based on two assumptions:

1. There exists an ideal scheme of practice that would be attained if enough resources were available.
2. The observed practice is as close as possible to the ideal scheme, given the resource constraints.

Clearly Rousseau faced a serious difficulty in quantifying the parameters of ideal practice. In the lack of other information he chose (1) to set these parameters equal to the observed mean values of X_{is} over the nine districts of Quebec and (2) to assume that the objective function of the HCS in the different districts was to minimize the sum of squared deviations from the "ideal" practice. Rousseau found that under these conditions his model produced a set of values of X_{is} that not only corresponded closely with observed practice but also corresponded significantly closer than a second set of values derived (without the model) by assuming that a district allocates its resources in the same proportions as the Quebec average. Thus there is some evidence that the model can satisfactorily represent how the HCS allocates its resources. Having tested the model in a descriptive setting Rousseau plans to apply the model to predicting how resource allocation would change in the future if additional resources are made available. In this way the model could be used by planners to examine the consequences of alternative strategies for resource investment.

Rousseau concludes in his paper that, since "the control variables for the planning agency are very limited", resource allocation models should, like his, "be of the type 'user optimization' rather than 'system optimization' in the sense that the models should predict how the different actors in the system would react given certain constraints rather than try to model how they should react for the well-being of the whole system". This can be viewed as a positive response to the problem posed by Feldstein (see Section 2.3) of developing "a complex theory of medical admission and treatment decisions". It might be argued that in countries with predominant public sectors in the HCS "the control variables for the planning agency" are not so limited as in Canada and system optimization may be more feasible. However even in these countries one suspects that many of the decisions on resource allocation are taken by individual clinicians and other staff in the HCS at the point of delivery of care rather than by the central planning agencies and so we may reasonably assume that Rousseau's conclusion holds for these countries too.

3.3 The Model of McDonald, et al.

Another behavior simulation model is the Inferred Worth Model of the U.K. National Health Service, by McDonald, et al. [16-18]. In this model there are three groups of variables which describe the allocation of scarce resources:

Cover - the numbers of patients, categorized by disease type and other factors, who receive treatment (d_i),

Modes - the usages of alternative forms of treatment that are permitted for each patient category (x_{il}),

Standards - the average amounts of resources consumed per patient in a given category in a given mode (u_{ilk}),

where i denotes patient category, l denotes mode, and k denotes resource. Thus the model represents the response of the HCS to the scarcity of a resource in terms of

- treating fewer patients, in one or more categories (less cover), or
- treating them in different ways (shift in the balance of mode use), or
- treating them less intensively (lower standards), or
- some combination of these.

The central hypothesis of the model is that the HCS attempts to maximize a function of the cover, modes, and standards variables. This function is a de facto utility function or, in Rousseau's terms, a "user optimization" function. It is termed the inferred worth function because its parameters are estimated from observations of the effects of the prevailing value system in the HCS. In other words the model is fitted to historical data on resource allocation. In running the model the inferred worth function is maximized subject to constraints on resource availability. In applications the model is used to estimate the consequences, in terms of cover, modes, and standards, of setting the vector of resource availability to different values. It is therefore used with planners as a device for examining options in long-term strategic planning of resource investment in the HCS.

The underlying theories of this model and that of Rousseau are somewhat similar (although there are considerable differences in the types of variables in the two models). Both models envisage local decision makers in the HCS attempting to optimize their own objective function subject to overall resource constraints. A further similarity is that both models postulate an ideal scheme of practice. In McDonald's model this is expressed in terms of:

- total numbers, D_i, of sick individuals by category (i.e. the maximum numbers that ought to be treated),
- the ideal standards, U_{ilk}, for patients of a given category in a given mode.

For several instances the current performance of the HCS in England is below these ideals; (in other words, demands for health care are not saturated). Experience with using the model suggests that budgetary and resource constraints prevent the HCS attaining these ideals and are likely to continue to so prevent it for the foreseeable future, i.e. the demands for health care will never be fully saturated. In effect the inferred worth model represents the HCS striving to achieve these ideals, within resource constraints. It also represents the different degrees of priority which the HCS places on the attainment of the different ideals; these differing priorities are incorporated as parameters in the inferred worth function. Thus the model represents resource allocation as the outcome of a rationing process in which these differing priorities of demand are balanced against constraints on resource supply.

The main features of the McDonald model that are not present in the current version of the Rousseau model are:

1. the complete range of HCS resources are represented, not merely physician time;
2. the model represents the consumption of combinations, or packages, of resources by patients rather than a single resource per patient;
3. the estimation of ideal performance is based not on an average of current performance but on information from surveys, medical literature, and professional opinion; and
4. the model is more concerned with alternative locations (modes) of care, e.g. hospitalization vs domiciliary care, rather than care by one medical specialist vs another.

3.4 The Model of Klementiev

A third, and rather different, type of behavior simulation model is contained within the HCS model design suggested by Klementiev [5]. The overall model describes certain aspects of morbidity, resource supply, and resource allocation. The morbidity aspect of the model, which has been further developed since, considers the processes by which individuals transfer between the states healthy, latent sick, revealed sick, treated sick, dead, and between different stages of sickness. Three stages of sickness are defined for degenerative disease corresponding to out-patient treatment, acute in-patient treatment, and terminal care. Latent sick individuals can become revealed sick either by self-referral to a physician or via the screening process. The resource allocation aspect of the model is based on a queue discipline hypothesis which can be roughly expressed as follows:

1. In a given time period the patients in each stage of the disease who are in the state of being treated have prior claims on physician resource for that stage. The remaining availability of physician resource is then calculated.
2. The number of revealed sick (but not being treated) in each disease stage who can transfer in the given time period to the state of receiving treatment will be determined by <u>either</u> the remaining availability of the physician resource (supply), <u>or</u> the number of revealed sick not yet receiving treatment (demand), whichever is smaller.

Although this model is still at an early stage of development it has already been run for one class of disease, degenerative disease, but on hypothetical data (see Olshansky [6]). Thus there has not yet been an opportunity to test the ability of the model to fit data on the real performance of the HCS.

The model is designed to enable the user to examine the consequences of policy options for (a) the proportion of physician resource devoted to screening rather than treatment and (b) the proportions of the physician treatment resources devoted to each of the three stages of the disease. "Inefficient" decisions on these proportions result in either patients accumulating and waiting at one or more stages in the treatment process or under utilization of one or more of the treatment resources.

A somewhat similar approach, though with more dimensions, has been used to simulate the care of elderly handicapped individuals in the U.K.; see for example Harris, R. [19]. In this simulation model a number of categories of elderly handicapped are defined, in terms of severity of handicap, home situation, etc., and a priority ordering of categories is stated. For each category a number of alternative forms of care are defined, e.g. hospital, residential, domiciliary care, with a preference ranking. The model simulates how the HCS allocates clients to the alternative forms of care by a queue discipline mechanism. That is to say that if the first preference form of care is not available for clients in a given category, the model allocates them to the next highest ranked form of care that is available. In the model, higher priority client categories are allocated to a package of care before the lower priority categories so that, in general, the high priority categories are more likely to be allocated to a high preference form of care. The model can be used to simulate the consequences of providing different mixes of resources (services) for the care of the elderly.

The queue discipline mechanism has been suggested in other models - for example in the model of the effects of establishing a health screening program described by Atsumi and Kaihara [20].

3.5 Review

In considering the relatively sparse literature on behavior simulation models it is interesting to note the variety of hypotheses. Rousseau describes specialist physicians striving to attain an ideal distribution of activities between themselves. Klementiev and Harris, R. suggest a system of patient selection based on a queuing discipline. McDonald suggests that the participants in the HCS are striving after ideals in the numbers and types of treatment and that the degrees to which the ideals are approached in practice is a result of striking a balance in priorities. The adequacy of the hypothesis is crucial in a behavior simulation model since, by contrast with a macro-econometric model, it is not possible to test rapidly a large number of alternative hypotheses. If this approach is to be pursued it is evident that much thought will need to be given to hypothesis construction.

There are some important advantages in the behavior simulation approach to resource allocation. It can operate at a disaggregated level, if necessary, since the basic hypotheses describe the behavior of physicians and other personnel at the local level. More importantly, and in contrast to the econometric approach, it can in principle be validly applied to exploring situations in which there is a structural, rather than a merely incremental, change in HCS, since, through its hypotheses, it contains a type of information about the behavior mechanisms of the actors in the HCS that is not usually contained in macroeconometric models. An analogy from the physical science may serve to illustrate this point. Boyle's Law and Charles' Law describe the relationships between the pressure, volume, and temperature of a fixed mass of gas and were established by observing the behavior of samples of different gases within certain ranges of the three variables; in some respects the equations of these Laws are analogous to the equations of macroeconometric models of the HCS. By contrast the kinetic theory of gases is based on hypotheses concerning the behavior of gas molecules and was found not only to correctly predict the macro relationships of Boyle's and Charles' Law, within the ranges where they apply, but also to correctly predict the deviations from the Laws outside these ranges, for example near the liquefaction points of gases. Provided that their hypotheses are sound behavior simulation models of the HCS have analogous properties. Thus they can, in principle, be used to examine the consequences of (a) major changes in the balance of resources in the HCS, outside the range of current or recent historic variation, (b) changes in the ways in which resources are used, (c) changes in medical technology (provided that the changed technology coefficients can be forecast), and (d) changes in the pattern of morbidity. Another advantage is that by attempting to represent the real-life process of resource allocation the model may be more transparent to the HCS planner than the more abstract representation of most econometric models; thus the planner may be more willing to use a behavior simulation model since its mechanism is more likely to correspond to the mental model which he already possesses and which has been built up through his personal experience of the HCS.

Thus the behavior simulation approach is strong in areas where the econometric approach is weak. However the reverse is also true. Firstly there has been comparatively little experience in the application of behavior simulation models to the HCS as a whole and the few models that have been used are somewhat different, one from the other. Thus there are no standard methods, programs, and terminology for this approach and the results of different models are not so readily comparable. More importantly behavior simulation models are relatively difficult to build since they require an intimate understanding of the workings of the HCS and relatively difficult to apply since they usually require data (for parameter estimation) on the preferences and priorities within the HCS that may not readily be available. Lastly it may be that a universal behavior simulation model does not exist, even among the IIASA member countries. That is to say there may be such differences in the factors governing the behavior of patients and HCS personnel in different countries that no one behavioral model would be valid for all.

Let us now turn to the third type of resource allocation models - system optimization models.

4. SYSTEM OPTIMIZATION MODELS

4.1 Introduction

We come now to the group of models concerned with optimizing some assumed objective function for the HCS. We term this approach "system optimization" after Rousseau [15], who distinguishes it from "user optimization" - see page 17. There are many examples in the literature of optimizing models applied to Health Care problems, but only a few are concerned with the HCS as a whole. The most commonly used technique for this type of model is mathematical programming. A useful review of this field of work has been made by Boldy [21]. He concludes that, in the field of strategic planning of the HCS, classical system optimization models are unlikely to be as useful as models of a more exploratory nature:

> In the strategic planning area, there have been a number of mathematical programming models developed for planning the prevention and control of disease or population growth and a start has been made towards their implementation. However, perhaps the most potentially valuable mathematical programming models are those, such as are being developed by O.R. Service of the DHSS, which are concerned with exploring the wider aspects of allocating resources both between the different health and social services care sectors and between the various patient/client

groups. Because of the general lack of detailed knowledge concerning the relative effectiveness of given forms of care provided to given types of patient/client, a situation which is likely to continue for the foreseeable future, such mathematical programming models are likely to be used in a "what-if" rather than an optimizing manner. In other words, their use is likely to lie in the exploration of the resource consequence and other effects of different policy options so that a more well-informed decision can be made (Boldy [21], pp. 446-447).

Boldy regards "the general lack of detailed knowledge concerning the relative effectiveness of different forms of care" as the main objection to the application of classical system optimizing models whereas Rousseau's main objection, [15], is that system optimization is inconsistent with the behavior of physicians at the point of delivery of health care. To assess these two objections and to consider also the advantages of the system optimization approach we will consider two examples of the approach being applied in a real planning situation.

4.2 The Model of Feldstein, et al.

The first example is the model of Feldstein, Piot, and Sundaresan [22] which was applied to the planning of the control of tuberculosis in the Republic of Korea. The authors propose an objective function of the form:

$$\sum_j V_k B_{kj} x_j \quad ,$$

where V_k = social benefit of type k;

x_j = amount of activity j; and

B_{kj} = amount of benefit per unit of activity.

The function is to be maximized subject to resource constraints of the form:

$$\sum_j A_{ij} x_j \leq m_i \qquad \forall\, i \quad ,$$

where m_i = availability of resource i, and

A_{ij} = amount of resource consumed per unit of activity.

The model was applied to the control of tuberculosis in the Republic of Korea. The technology coefficients A_{ij} were estimated from experience with TB programs elsewhere; the benefit coefficients B_{kj} were calculated on the basis of an economic analysis of research findings and physicians' judgements concerning the clinical outcomes of each activity. The output from the model showed the optimum set of counter-tuberculosis activities for different categories of population defined in terms of age and urban/rural split. The model was run for four different forms of the objective function in which social benefit was expressed respectively in terms of temporary disability, permanent impairment, excess mortality, economic loss. Sensitivity analyses involving changes in parameters in the objective function were performed. The results enabled the authors to propose some robust conclusions for the design of a tuberculosis control program in Korea.

The sensitivity analysis is particularly interesting since it is argued in this Memorandum that uncertainty in quantifying the objective function is one of the major disadvantages in using optimizing models. The robustness of the results of Feldstein, et al. is in contrast to the sensitivity found by Ashford, et al. [23], in an application of linear programming to maternity services. Ashford, et al. found that there was "a fundamental disagreement about the relative merits of different procedures" between the various experts consulted. This led Ashford, et al. to run their model with three different objective functions to reflect these differing views. They found that the use of one of these three led to "a radically different solution" to that obtained with the other two.

On the basis of these contrasting experiences we can only conclude that the scientist who decides to build a system optimizing model for the HCS can have no confidence in advance that his model will not founder on uncertainty in the objective function. Indeed we must fear that the more comprehensive the scope of the model, i.e. the wider the range of HCS activities covered in the model, the greater is the risk of encountering areas where no single valued objective function can be satisfactorily applied.

4.3 The Cost Minimization Version of the Model of McDonald, et al.

Another optimization model is the early version [24] (see also [16]) of the McDonald model [16-18]. For this version of the model the authors chose as the objective function the minimization of the total resource cost. Thus they avoided the problems, described above, of quantifying a maximization of benefits function. This version of the model was designed to identify the set of resource allocations and associated resource availabilities that minimizes total resource costs subject to upper bounds on the availability of individual resources, lower bounds on the number of patients to be treated and specifications of the alternative types of treatment permitted for each category of patient:

$$\text{Minimize} \sum_k \sum_i \sum_l C_k x_{il} U_{ilk}$$

subject to

$$\sum_l x_{il} \geq D_i \qquad \forall\, i \quad ,$$

$$\sum_i \sum_l x_{il} U_{ilk} \leq b_k \qquad \forall\, k \quad ,$$

$$x_{il} \geq 0 \qquad \forall\, i,l \quad ,$$

where i = patient category,

k = resource type,

l = mode type,

C_k = unit resource cost,

x_{il} = number of patients i treated by mode l,

U_{ilk} = amount of resource k consumed per patient i in mode l,

b_k = maximum availability of resource k, and

D_i = minimum number of patients of type i to be treated.

The only variables in this formulation are the x_{il}. The D_i and the U_{ilk} are constants representing the ideal numbers of treated patients and standards of treatment. The problem as formulated above was quickly shown to be infeasible. In other words if realistic constraints on resource availability were assumed there was no allocation that could treat all D_i patients at the required standards, U_{ilk}. Feasible solutions could only be obtained by relaxing some of the resource constraints to an extent that would permit very large growths of these resources from their current levels, e.g. a trebling of the number of home nurses. Such solutions were unrealistic. Thus although the solutions had a theoretical validity they were of little practical value to planners.

In order to produce solutions that were both feasible and realistic, the data input was modified. The values of first the patient demands, D_i, and then later the standards of treatment, U_{ilk}, were lowered to levels that corresponded more nearly with the prevailing levels in the HCS. Feasible solutions were indeed obtained in this exercise and the results were of some limited practical value. However the behavior of the model was unsatisfactory in one important respect. The model tended to select modes of care where the prevailing standards were low rather than those where the prevailing standards were close to the ideals, U_{ilk}. An example will clarify this phenomenon.

Consider a typical patient category in the model from the group of categories of elderly chronically disabled patients. One of the permitted modes of care for this category is long-term care in a geriatric hospital. An alternative mode is a package of community-based care services, such as home nurses, home helps, and day centers. The ideal standards for this mode of care (showing the main services only) are displayed in Table 4. This mode is typical of the domiciliary and community-based modes of care defined for other categories of elderly disabled patient in the model but the precise mix of services and the ideal standards vary from one category to another; (for example the categories with greater physical disability have, in general, higher ideal standards). Estimates of the prevailing standards for this mode of care are also shown in Table 4; in general they are significantly lower than the ideals.

Table 4. Example of a community-based mode of care for an elderly disabled patient with ideal and prevailing standards.

Service	Ideal Standard	Estimate of Prevailing Standard
Home Nurse (visits p.a.)	35	7
Home Help (visits p.a.)	135	82
Day Center (attendances p.a.)	150	13

Now the behavior of the model in this situation is to tend to select the domiciliary mode of care, for all patient categories for which it is permitted, in preference to the geriatric hospital mode for two reasons:

(1) the objective function of the model is cost minimization and, at prevailing standards, the community-based mode is cheaper for all categories; (whereas at ideal standards for the more disabled categories it would be more expensive);

(2) the community-based mode, at prevailing standards, makes relatively low demands on scarce resources; (whereas, at ideal standards, the constraints on resource availabilities would more tightly limit the use of these modes).

Thus the main result from running their version of the model consisted of a decrease in hospital-based modes of care and an increase in community-based care relative to the prevailing actual situation in England. Such a shift in the balance of care would, in theory, elicit large financial savings in the provision of hospital services, and relatively small increases in expenditure on additional community-based services.

This aspect of the model's behavior is at variance with the behavior of the HCS in the real world. It is well-known in England that many elderly disabled patients are hospitalized precisely because the alternative domiciliary care could only be offered at unsatisfactorily low standards. In other words decision makers at the point of care delivery reject one form of care at low standards, despite its cheapness, in favor of another one at high standards. This type of difference between the behavior of the model and the real HCS also occurs for most other categories of patient defined in the model, e.g. the mentally ill and the acute ill. It arises because the objective function of the model, cost minimization, is different to that of the HCS. This difference is important and the results obtained from running this formulation of the model are of limited practical value. This is an example of the danger foreseen by Rousseau of developing models for "system optimization" rather than "user optimization". Even if the central authority of the HCS, in this case the Department of Health and Social Security, had subscribed to the "system objective" of cost minimization, the model's results would still have been of little practical value since this objective is inconsistent with the "user objective" being pursued by the actors in the HCS. Following this experience McDonald's team proceeded to modify their model into a "user optimization" version, the inferred worth model, which has been described earlier as an example of behavior simulation.

4.4 Review

Two examples of system optimizing models have been given here and by considering the different ways in which they were applied we may draw some conclusions about the appropriateness of this type of model. The model of Feldstein, Piot, and Sundaresan was applied to a very specific sector of the HCS, the counter-tuberculosis program. In this particular sector it appears to be possible to estimate the benefits of alternative activities on a number of different scales and that the main conclusions from the model are not crucially affected by the choice of scale. Furthermore the paper gives the impression that decision makers in Korea were both willing and capable of designing and implementing this program at the level of detail specified in the model (e.g. by specifying the population groups at whom mass screening is to be directed). To the extent that this was true the optimizing model was appropriate. By contrast the cost minimization version of the McDonald model was concerned with the resource allocation in the HCS as a whole and it became evident that even if the central decision makers for whom the model was developed wished to allocate resources so as to minimize costs they were not in a position to do so. They had, and still have, the main say in deciding the aggregate availability of many of the key resources in the HCS but the decisions to allocate these resources between competing patient groups rest mainly with individual personnel at the point of care delivery.

Let us assume, as mentioned earlier (end of 3.2), that there is a significant degree of decentralized decision making in the HCS of all countries. Then we may conclude that "system optimizing" models are likely to be appropriate only for certain individual sectors of the HCS where there is a strong influence from the center and where the clinical outcomes of the alternative procedures are reasonably well-known, but that system optimization is unlikely to be appropriate for planning the HCS as a whole.

There are many advantages in using optimizing models of resource allocation. There are a number of well-known techniques, such as linear programming, which can readily be applied and for which computer programs are available. The models are relatively easy to build and most of the data is relatively easy to obtain (with the exception of parameters for the objective function in benefit maximization models - see below). Optimizing models are capable of exploring situations in which major structural changes are envisaged rather than mere incremental changes, provided the technology coefficients can be safely assumed to remain unchanged. By their very nature they are capable of incorporating a planner's goals into the objective function and so they hold out the promise, in principle, of leading the planner to the desired solution in one step rather than by a series of model manipulations and runs which is required with macro-econometric and behavior simulation models.

On the other hand there are the two serious disadvantages mentioned earlier. Firstly there is the difficulty of defining an objective function which corresponds to some formal quantitative statement of the objectives of the HCS. Objectives like "maximize the health of the population" are easy to state qualitatively but notoriously difficult to express in an acceptable quantitative form. HCS planners themselves are particularly aware of this difficulty and, in the experience of this author, are not aware of employing any single readily quantifiable conteptual objective function when they are planning services for the HCS as a whole.

A second disadvantage is that the solutions produced by the model are likely to be of theoretical rather than practical interest since there is no guarantee that the real behavior of the HCS will ever follow the "system optimizing" behavior of the model. As Boldy [21] has observed for system optimizing models of hospital location, "...these models tend to ignore aspects of patient behavior and as such are unlikely to be implemented, consequently more complex models have been developed involving the simulation of such aspects".

5. ASSESSMENT OF THE APPROPRIATE TYPE OF RESOURCE ALLOCATION SUB-MODEL FOR THE IIASA HCS MODEL

5.1 Introduction

In this section the ways in which the three types of resource allocation sub-model reviewed in the previous sections would fit into the overall HCS model are considered (5.2) and the appropriateness of each type is assessed (5.3).

5.2 The Role of Each Type of Resource Allocation Sub-Model in the Overall HCS Model

If there were no constraints on the supply of HCS resources then planning the HCS would be a relatively simple matter. The appropriate model schema would be a simple variant of the schema illustrated in Figure 1, which included four sub-models concerned with population, disease prevalence estimation, resource allocation, and resource supply. In this variant, shown in Figure 3, the resource allocation sub-model is really no more than a list of standard, ideal, resource requirements per unit of prevalence or, for screening resources, per unit of population; the required supply of each resource can then be calculated by combining these figures with the outputs of the population and disease prevalence sub-models.

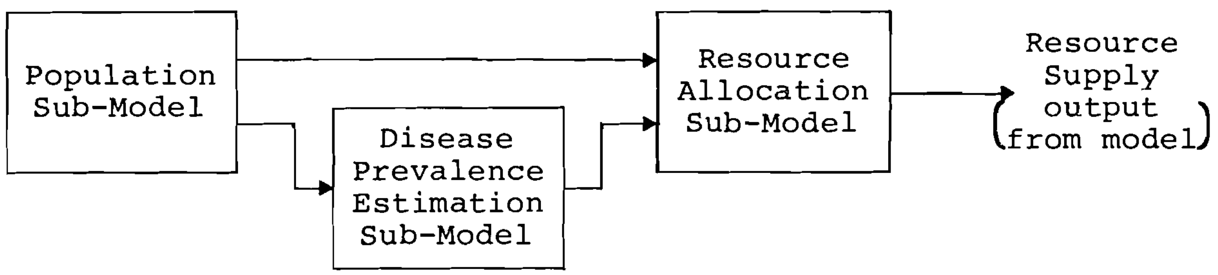

Figure 3. Model schema for the case of unconstrained resource production.

It is assumed here, following Harris, D. [11], Rousseau [15], and many others, that this case, of unconstrained HCS resource supply, does not exist in real life and that in all countries the total potential demand for health care exceeds the capacity of the HCS to provide it, i.e. the demands for health care cannot be fully saturated by the delivery system. Thus the simple schema in Figure 3 does not apply and the process of resource allocation in the HCS is one of allocating scarce resources between competing demands. Let us now consider how the three types of model considered in this paper represent this resource allocation process and how the HCS model schema of Figure 1 would apply for each type.

With the macro-econometric type of model information is usually supplied to the model on one or more of the following variables: supply and price of resources and population attributes. (In some cases, such as the model of Feldstein, population features only as a denominator in the supply variables.) The output usually includes the consumption of health care services, e.g. numbers of patients treated in various categories. In no instance is information on demand or disease prevalence supplied as an input. The fraction of disease prevalence that receives treatment (i.e. consumption divided by prevalence) would have to be calculated outside the model. Thus the model schema has to be adapted as in Figure 4.

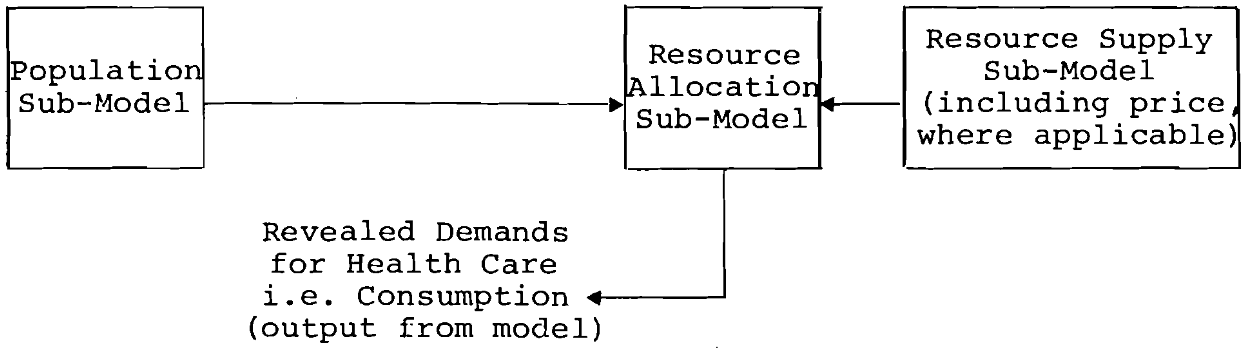

Figure 4. Model schema for using the macro-econometric resource allocation sub-model.

With this type of model the planner can submit options on policies for resource supply and discover the model's estimates of the consequent pattern of consumption.

With the behavior simulation type of model both the demands for health care and the supply of resources can be supplied as input data and the model can estimate the outcome in terms of variables such as the fraction of demand or prevalence that is met by the service, i.e. the fraction of patients who receive treatment and the fraction who do not, and the types of treatment they receive. The model schema for this case is shown in Figure 5. Planners' options for resource supply can be tested out with the model as with the macro-econometric case of Figure 4, but in this case the structure of demand and disease prevalence is part of the input of the resource allocation sub-model; this would be an important advantage in a situation where the structure of disease prevalence is expected to change within the planning horizon.

With the system optimizing type of model information can be supplied on the demands for health care but for resource supply all that is required is a set of upper bounds on the maximum supply of the main resources. The model then calculates an optimum allocation of resources to demands, within the constraints, and thus produces the optimum pattern of resource supply. The model schema for this case is shown in Figure 6.

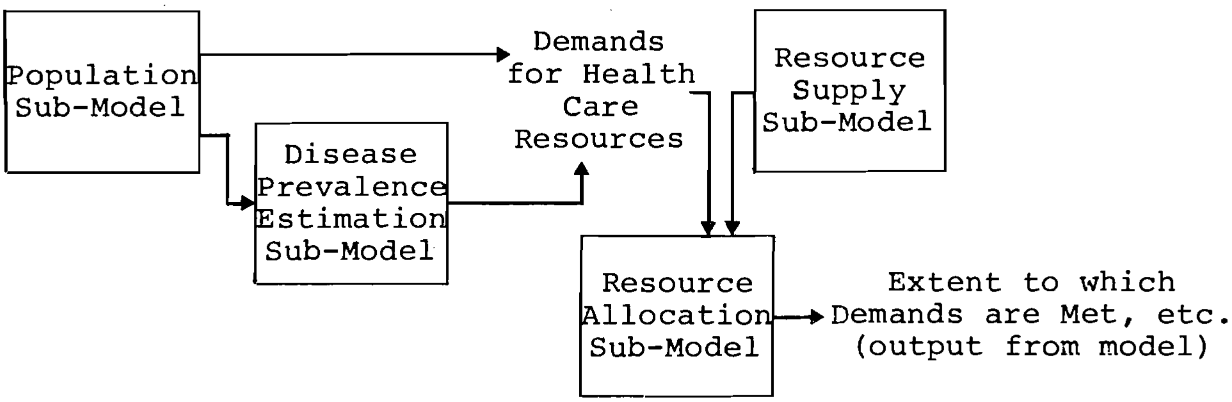

Figure 5. Model schema for using the behavior simulation type of resource allocation sub-model.

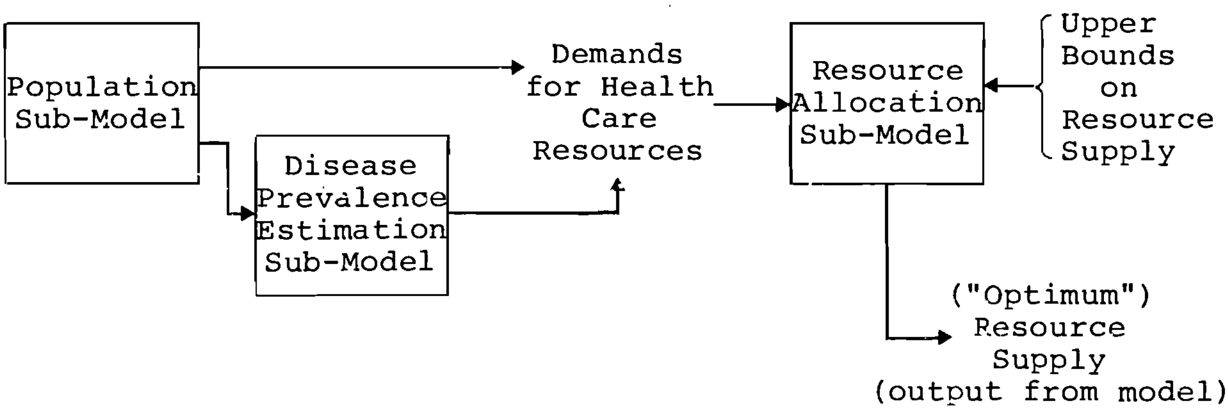

Figure 6. Model schema for using the system optimizing type of resource allocation sub-model.

5.3 Assessment

Having considered the different ways in which the three types of resource allocation sub-model would fit into the overall HCS model schema let us assess the appropriateness of each type for the IIASA HCS Modeling Task. The advantages and disadvantages of each type of model have been described in the previous three sections and are summarized in Table 5.

Table 5. Summary of advantages and disadvantages of three types of resource allocation model.

	Macro-Econometric Models	Behavior Simulation Models	System Optimization Models
Advantages			
	1. Well tried and successful.	1. Easy to use at disaggregated levels.	1. Standard techniques & terminology.
	2. Standard techniques and terminology.	2. In principle can be applied to situations of structural or major change from the status quo.	2. Easy to compare results.
	3. Easy to compare results.	3. Relatively acceptable to HCS planner since hypotheses likely to correspond with planner's experience.	3. Most of the data is relatively easy to obtain (apart from coefficients of objective function).
	4. Data requirements reasonable.		4. In principle can be applied to situations of structural or major change from the status quo.
	5. Large number of tried hypotheses can be tested, therefore easy to build model.		5. In principle supplies the HCS planner with the "correct" solution in one step.
	6. Probably applicable universally.		
Disadvantages			
	1. Hypothesis limited to linear relations (or transformations of linear relations) between aggregate quantities	1. Relatively little experience at strategic level.	1. Impossible to define satisfactory objective function for HCS system as a whole.
	2. Results not valid outside the range of variables in the data used for estimation of coefficients and therefore model is not appropriate to situations of structural or major change from the status quo.	2. Standard techniques and terminology not available.	2. Unrealistic since real behavior of HCS participants will not correspond to "system optimum" allocations.
		3. Results difficult to compare.	
		4. Relatively difficult to build since deep understanding of HCS behavior is required.	
		5. Some of the data may be difficult to obtain.	
		6. May not be applicable universally.	

The only disadvantages or limitations of the econometric approach are its high level of aggregation, its use of linear equations, and the limitation of its validity to situations of only incremental change. Apart from these the advantages of the approach, as compared with the other two approaches, are very strong. Thus, if the IIASA HCS Modeling Task is to be concerned with the study of incremental changes in the HCS, there is a clear case for concentrating on the econometric approach.

If however, as seems more likely, the HCS Modeling Task is to be concerned with situations of major structural changes in the HCS then the behavior simulation approach is the most strongly indicated ab initio. The main doubts about using this approach concern (1) the reliability of the behavioral hypotheses under conditions of structural change and (2) their universality. It appears that the only way to resolve these doubts is by the classical scientific procedure of building a model and testing it under a range of different situations.

It is therefore recommended that the IIASA HCS Modeling Team should embark on the construction of an HCS resource allocation sub-model of the behavior simulation type. The point of departure would be the existing behavior simulation model of Klementiev [5] which is based on a queue discipline hypothesis. However the work would need to enlarge considerably upon this model and draw upon the experience of Rousseau [15] and McDonald [16-18], who have shown that the behavior of the HCS, particularly the way it adapts to resource scarcity, cannot be adequately represented solely in terms of queuing mechanisms and that some account has to be taken of the value system of the actors in the HCS.

Although it is recommended that the HCS Modeling Task should concentrate on the behavior simulation approach this is not meant to imply that nothing can usefully be learnt from experience with macro-econometric and system optimizing models. In any case the three types of approach, as applied in practice, are not entirely mutually exclusive; (for example Feldstein used his <u>econometric</u> hospital production function as part of a procedure to suggest the <u>optimum</u> mix of hospital inputs). Thus it is to be expected that in building a behavior simulation model of resource allocation in the HCS there will be some recourse both to econometric methods (e.g. for those aspects of the HCS which are not likely to be subject to structural change), and to optimization methods (e.g. for representing "user optimization" behavior). Nevertheless the basic philosophy of the behavior simulation approach is distinct from those of the other two and the main purposes of this paper are (1) to clarify the distinction and (2) to recommend that IIASA concentrate on the behavior simulation approach.

References

List 1: General

[1] Venedictov, D.D., Modeling of Health Care Systems, in *IIASA Conference '76*, Vol. 2, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1976.

[2] Kiselev, A., *A Systems Approach to Health Care*, RM-75-31, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1975.

[3] *Research Plan 1977*, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1977.

[4] Shigan, E.N., Systems Analysis in Health Care, in E.N. Shigan and R. Gibbs, eds., *Modeling Health Care Systems - Proceedings of an IIASA Workshop*, Conference Proceedings, International Institute for Applied Systems Analysis, Laxenburg, Austria (forthcoming).

[5] Klementiev, A., *Mathematical Approach to Developing a Simulation Model of a Health Care System*, RM-76-65, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1976.

[6] Olshansky, W., *Basic Model of Health Care Systems*, Internal Paper, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1976.

[7] Fleissner, P., and A. Klementiev, *Health Care System Models: A Review*, Research Memorandum, International Institute for Applied Systems Analysis, Laxenburg, Austria, RM-77-49.

List 2: Macro-Econometric Models

[8] U.S. Department of Health, Education, and Welfare, *An Analysis of Health Manpower Models*, Monographs and Data for Health Resources Planning Series, DHEW Publication No. (HRA) 75-19, Washington, D.C., 1974.

[9] Yett, D.E., et al., A Macroeconometric Model of the Production and Distribution of Physician, Hospital, and Other Health Care Services, in D.D. Venedictov, ed., *Health System Modeling and the Information System for the Coordination of Research in Oncology*, CP-77-4, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1977.

[10] Feldstein, M.S., *Economic Analysis for Health Service Efficiency*, North-Holland, Amsterdam, 1967.

[11] Harris, D.H., Effect of Population and Health Care Environment on Hospital Utilisation, *Health Services Research*, 10, 229 (1975).

[12] Fleissner, P., An Integrated Model of the Austrian Health Care System, in N.T.J. Bailey and M. Thompson, eds., *Systems Aspects of Health Planning*, North-Holland, Amsterdam, 1975.

[13] Newhouse, J.P., Forecasting Demand and Planning of Health Services, in N.T.J. Bailey and M. Thompson, eds., *Systems Aspects of Health Planning*, North-Holland, Amsterdam, 1975.

[14] Roemer, M.I., and M. Shain, *Hospital Utilization Under Insurance*, Hospital Monograph Series, No. 6, American Hospital Association, Chicago, 1959.

List 3: Behavior Simulation Models

[15] Rousseau, J.-M., The Need for an Equilibrium Model for Health Care System Planning, in E.N. Shigan and R. Gibbs, eds., *Modeling Health Care Systems - Proceedings of an IIASA Workshop*, Conference Proceedings, International Institute for Applied Systems Analysis, Laxenburg, Austria (forthcoming).

[16] McDonald, A.G., G.C. Cuddeford, and E.M.L. Beale, Mathematical Models of the Balance of Care, *British Medical Bulletin*, 30, 3 (1974), 262-270.

[17] McDonald, A.G., and R.J. Gibbs, Some Requirements for Strategic Models of Health Services Illustrated by Examples from the United Kingdom, in D.D. Venedictov, ed., *Health System Modeling and the Information System for the Coordination of Research in Oncology*, CP-77-4, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1977.

[18] Gibbs, R.J., and I.L. Coverdale, *Some Compatible Inter-Related Sub-Models of the National Health Service*, paper presented at the International Cybernetics Conference, Vienna, Austria, April 1976.

[19] Harris, R., The Development of Resources and Services for the Long Term Care of the Elderly, in *The Use of Operational Research in European Health Services - Report of a Working Group*, WHO, Copenhagen, 1975.

[20] Atsumi, K., and S. Kaihara, Planning a National Medical Information System: A Systems Approach, in N.T.J. Bailey and M. Thompson, eds., *Systems Aspects of Health Planning*, North-Holland, Amsterdam, 1975.

<u>List 4: System Optimization Models</u>

[21] Boldy, D., A Review of the Application of Mathematical Programming to Tactical Strategic Health and Social Service Problems, *Operational Research Quarterly*, <u>27</u>, 2, ii (1976), 439-448.

[22] Feldstein, M.S., M.A. Piot, and T.K. Sundaresan, *Resource Allocation Model for Public Health Planning - A Case Study of Tuberculosis Control*, WHO, Geneva, 1973.

[23] Ashford, J.R., G. Ferster, and D. Makuc, An Approach to Resource Allocation in the Re-Organised National Health Service, in G. McLachlan, ed., *Problems and Progress in Medical Care: Essays on Current Research* (Ninth Series: *The Future—and Present Indicatives*), Oxford University Press for The Nuffield Provincial Hospitals Trust, London, 1973.

[24] Jackson, R.R.P., Contribution to the Delivery of Health Care, in D. Howland, ed., *Cybernetic Modelling of Adaptive Organizations: Proceedings of a Conference Held in Oporto, 27-31 August 1973*, NATO, Advisory Panel on Operational Research, 1974.