International Institute for Applied Systems Analysis
IIASA
www.iiasa.ac.at

Modeling of National-Regional Health Care Systems: Comments and Proposals for IIASA Biomedical Project

Milsum, J.

IIASA Working Paper

1976

MODELING OF NATIONAL/REGIONAL HEALTH CARE SYSTEMS: COMMENTS AND PROPOSALS FOR IIASA BIO-MEDICAL PROJECT

John H. Milsum*

August 1976 WP-76-24

*Dr. Milsum is Professor of General Systems in the Department of Health Care and Epidemiology, and Director of the Division of Health Systems at the University of British Columbia in Vancouver, Canada. He visited IIASA in May, 1976.

2361
Laxenburg
Austria

International Institute for Applied Systems Analysis

1. OVERVIEW OF HEALTH CARE SYSTEM MODEL OF IIASA BIO-MEDICAL PROJECT

The basic idea behind this model is that it should establish a set of interconnected elements which are essentially invariant in their basic principles, as between different regional and national health care systems. Bearing in mind the models which have been and are already being developed, it is clear that there is a considerable diversity of approaches and of objectives in the particular models. In consequence, I feel that the following subsystems will have to be included as "autonomous" elements or sub-models, although all of them may not necessarily be included in every national model. Nevertheless, it seems to me important to identify the functions of each of these sub-models whenever it is appropriate.

Population Model - by age and sex, with corresponding birth, death and migration rates.

Morbidity Model - perhaps divided into "natural" and "registered"; basic variables are different disease classifications, or/and dynamic health/sickness classifications.

Resource Model - personnel.

Resource Model - facilities.

Resource Utilization Patterns Model - "treatment".

Allocation Model, ideally with Optimization.

Planning Model for Long-term Resource Allocation.

Environmental or External Subsystems.

In turn, some of these may need to be broken down into submodels of more specific aspects; for example, the "treatment" model for utilization of resources may have relatively autonomous activities devoted to prevention, screening, and treatment, perhaps in different modes such as acute care hospitals, ambulatory facilities, nursing homes, etc.

The choice of these "invariant" subsystems does not, of course, limit either the specificity or generality of the detailed modeling which may be performed on them; indeed, it MUST not so limit the modeling if the subsystems are to be, in fact, "invariants". As an example, we note that the morbidity model in some cases might only involve measures of static prevalence, or, on the other hand, it may be modeled to include the dynamics of the transition from health to illness in the population, and the dynamics of the illness process itself.

2. COMMENTS ON INITIAL PROPOSALS FOR IIASA SUB-MODELS

2.1. Population [1]

In general this model seems to follow conventional techniques, with the choice of five-year intervals being appropriate. However, it would not seem justified to increase, in the interval above 60, to a ten-year age group, because of the aging nature of many countries' populations. In other words, the population segments in the five-year groupings from 60-65 and 65-70 would still seem to justify being kept separate. Clearly, the economy in computing of this aggregation is absolutely minimal.

Perhaps more substantially I feel that the aggregating of the two sexes within the age groups 0-4, 5-9, 10-14, 60-69, and 70+, is not acceptable. Presumably the purpose is to economize in computing time, by this removal of five age-sex groups, but it seems totally unjustified in view of the following factors:

(i) Death rates in the young age groups are widely different between males and females; for example, in North America they are approximately in the ratio of 5:2 at age 15. There is a corresponding difference in old age also, although the ratio is not so large. Furthermore, there are significant differences between sexes in the types of illness, or accidents, which are the main causes of death in these various age-sex groups. Thus, any health care model which wishes to explore such aspects would be frustrated, and this would not seem appropriate for a subsystem meant to be universally applicable to all nations.

(ii) The artifice of a sex-ratio is needed at age 15 in order to divide the previously aggregated sexes into separate male and female streams. In general, this will necessarily introduce errors in the simulation of populations, since in the real population this ratio could vary with time. Further, it might vary at either end of the age spectrum as a result of particular preventive or screening programs, which might, in some circumstances, favor one sex over the other.

(iii) With this aggregation of sexes in some age groups, the total male and female sub-populations can never be known accurately, nor therefore any changes in them as time progresses.

2.2. Health Care System Model

The division of the population into the three groups of healthy (HP), latent sick (LD), and registered sick (RP), means essentially taking the population model and dividing it into some further strata. In principle, the model is also going to include three types of illness, with different dynamic trajectories, namely degenerative, vaccinateable and therefore eradicable, and acute or episodic. The initial modeling, however, is only to cover the case of the degenerative diseases. The problem seems to be how to join the two models compatibly, since they are both needed, but on the other hand, for some purposes, may best be kept separate.

A further problem is that of determining the intiial proportions of HP, LD and RP groups in the population, and how they are to be updated. It is to be noted that the equations given in the Working Paper for the three categories, do show loss of population in these groups due to deaths. On the other hand, the inputs due to births are not shown. The equations also seem to show that the same death rate characteristics will be applied to each of the three sub-groups, but this would not seem reasonable since, rather clearly, the death rate would be significantly less among the healthy population than among the sick population. Again, it may be even harder to determine the distribution of the newborn children among these three sub-groups.

It should be noted that there are, in principle, nine states to which each state may possibly transfer in the next time interval, as depicted in Figure 1. A notable factor swelling this number is that of the latent sick group which has been screened and recognized but is, as yet, untreated. Of course, death is an absorbing state and there are no returns from this. Since each of these transitions is theoretically possible in successive time intervals, the total number of transition probabilities to be identified is 72 for each of the 23 (32 proposed in Section 3) age-sex groups, for approximately 1.6×10^3 numerical values. Admittedly, many of the transitions will have low levels of probability, but nevertheless there seems to be a major identification problem here.

There will presumably be considerable difficulty in estimating some of the transition probabilities. For example, since the latent sick population (LS) is really an unknown, how can the transition probabilities for Death be estimated, in each of the I age-sex groups?

It should also be noted that the model, as presently proposed, calls for the latent sick and registered sick to come from two separate streams, namely, from the Screening process and from Self-Request for treatment.

There is the further difficulty that the above description presumes a Markov-type process, in which the probability for any particular transition is independent of the person's history of how he got into the present state, i.e. independent of the history of previous transitions. Theoretically, this is clearly not reasonable, although it has often been found in practice that the theoretical objection does not interfere with practical accuracy very much. A different way of viewing this same problem is that once an individual enters any particular health-state group, say RB, at time T, he then becomes an undifferentiated member of this pool, with the same probability of leaving it at (T + 1) as any other member, even though some of the latter may already have been there for many periods. To "tag", "label", or "track" each individual would, on the other hand, clearly represent an impossible task, indeed essentially involving a vast number of Monte Carlo-type micro-modeling runs. But nevertheless, this points out that the proposed macro-modeling cannot be faithful to the true dynamics of the real-life situation.

Finally, it should be noted that the basic population model, with different health states integrated out, is still needed for various purposes; in particular, for the generation of birth rates, since presumably pregnancy probabilities cannot be estimated to vary with the health status of the females (and males?) concerned.

I would suggest that, for most planning purposes, a ΔT of one month is probably unnecessarily short, in view of the disadvantages also inherent in it. Perhaps three months would represent an appropriate compromise, since this would permit seasonal fluctuations of the year to be included.

2.2.1. Health Care Resources

Physicians are used in this initial modeling as the common variable to represent the need for health care resources. Presumably other resources then are derived through applying certain ratios. It does not seem to me that this is representative of the situation in various countries, particularly for planning models which are looking into the future. Especially in this time horizon, we cannot expect that both multi-professional health care teams, and the use of public health nurses (for example), in an autonomous way will be vastly increased.

It should be noted here that, in some Western countries at least, there is a tendency now to define health care professional as any of those with "professional" or degree training who are in direct "clinical practice" relationship with patients, typically physiotherapist, occupational therapist, clinical psychologist, dentist, and even such new professions as the clinical (bio-medical) engineer.

This increasing substitution of other health care professionals in lieu of physicians means that the health care resource staffing cannot be described in terms of physicians alone, if an adequately broad resource model is to be provided. Indeed it should be noted that, while many countries are still trying to increase the ratio of physicians to population, nevertheless there is increasing background thinking that the level may already be too high, and thus that there may, in fact, already be an excess of physicians. Certainly, in countries such as Canada, it is generally thought that there is at least a distribution problem in that there are certainly too many physicians in the big cities.

A final comment is that the use of other health professionals than physicians will be especially important in screening and preventive programs.

2.2.2. Allocation of Resources to Screened Sick

The present tentative model apparently only proposes to provide treatment to those who have been screened from the latent sick population as "sick", and to those who have requested medical attention by themselves, provided that there is "unutilized workload" of physicians available. Apparently, the first priority in the model is to have the physicians allocated to the screening programs themselves and to the treatment programs, and only after these needs are satisfied are they to be allocated to the "screened-positive" population itself. In view of the recognized importance of preventive/screening practices, it would seem that significant priority should be attached to treating these patients, and indeed I would urge that this should be part of the overall decision-making block on allocation of priorities.

Incidentally, this issue also raises a question about how the model should keep track of those who are diagnosed as sick, but who do not immediately receive treatment. The model equations correctly do not show them as joining the RP, so that by the same token they are now implicitly relegated back into the LD pool, unless a fourth group is to be kept track of, as indicated in Figure 1, namely LRU. It certainly would not seem appropriate to return these persons back into the LD pool, where they would become undifferentiated from the so far undiagnosed LD.

2.2.3. Disease Types

The three disease types proposed for incorporation in the model are, as already noted, degenerative, acute, and vaccine-eliminateable. I am not clear why chronic diseases are not included, since these are widespread, but do not necessarily fall into any of the above three groups. Of course, some chronic disease does eventually degenerate significantly, so that these diseases could be fitted into the degenerative group.

In regard to modeling the degenerative disease population as one homogeneous group, the first big problem would seem to be how to average the tremendously large variation which will be involved in the different stage times, and also in the costs involved, in coming up with one standardized "illness trajectory".

The purpose of defining this type of degenerate disease as a basic element for the model, as I see it, is in order to be able to have a "dynamic" model of the health care system. However, the use of the chosen dynamic trajectory can only have a significant effect upon the operation of the system if:

a. there are very important changes occurring in the health care system, in terms of the morbidity rates in this illness group;

b. the average time chosen for the trajectory is long; and

c. the time increment of computation is very short.

Stated differently, it would seem that if the system is anywhere near the steady-state condition, then one could obtain approximately the same dynamic modeling by allocating costs on an average basis for each episode* of degenerative disease.

Indeed, as a general comment, it seems to me rather inappropriate to disaggregate age-sex into the order of 20 groups, and then only to disaggregate the multitude of illness types to the extent of 3 groups.

2.2.4. Screening

The proposed model categorizes patients revealed by screening according to both age-sex and by phase of illness. This would seem to require that the corresponding prevalence rates be known among the not-RP population. Presumably, these would have to be obtained as sampled estimates from the population, but it would seem that this could be an area where significant error could be introduced into the simulation.

---

*episode; defined here as teh overall duration of sickness and its care for the particular patient with the given diagnostic.

3. PROPOSALS re CONCEPTUAL HEALTH CARE SYSTEM MODEL

A proposal for the conceptual format of the health care system model, which may be useful for discussion purposes, is shown in Figure 2. In many respects, it is an extension of the HSDIM, Medics and other models already in the literature. Its various features are now discussed, especially in relation to suggested changes from HSDIM. The changes suggested are intended to be particularly relevant for the needs of an internationally useful model.

3.1. Population Model, P(I,T)

In order to provide an adequately general population model, it is proposed that:

(i) the sexes be always grouped separately;

(ii) the 60-70 year group be divided into two age groups, of 60-65, and 65-70, of two sex groups each;

(iii) the 70+ years group be divided into two age groups of 70-80, and 80+, of two sex groups each;

(iv) that the perinatal groups be monitored, as two sex groups, but not included in the population totals, for obvious reasons.

Thus, there would result I = 1,...,32 (+2 perinatal) instead of the present 23. The male and female parts of the population could be followed separately as:

$$\left.\begin{array}{ll} P_M = \sum_{k=1}^{16} P_I \ , & I = 2k - 1 \\ & \\ P_F = \sum_{k=1}^{16} P_I \ , & I = 2k \end{array}\right\} \quad 3.1$$

The extra computing effort involved would be more than justified by the model's increased flexibility and informative capacity, in my opinion.

As shown in Figure 2, it seems conceptually helpful to show separately the three basic rates effecting population dynamics, namely,

Birth Rates	BR(I,T)	,
Perinatal Mortality Rates	DR(I = 0,1,T)	,
Mortality Rates	DR(I,T)	.

These specifically will be affected by multiple feedbacks from various other subsystems, including most notably the subsystem of Prevention-Screening-Treatment, but also Environmental Subsystems, Human Life-Styles and Human Genetics.

It should be emphasized that if death rates change as a result of changes in the health care system, for example the treatment of morbidity, then this provides a primary effect to be fed into the population model.

3.2. The Morbidity Model

Basically, this model must identify the different groups of non-healthy people from the population, and subject them to different health care regimens before hopefully transferring them back to healthier groups. However, the particular disaggregation techniques used for this modeling can evidently depend importantly on the viewpoint of the modelers and the intended purpose of the model.

On the one hand, the actual provision of health care treatment may be considered to constitute a single "episode" within the basic time increment on which the model operates, for any one patient. For a time interval of one month or more, this will typically describe the situation of some 90% of the first contacts of a patient with the health care system, which ordinarily do not go on to extensive or expensive hospital treatment, etc. However, these large numbers of cases may only represent some 30% of the amount of money spent in the health care system. For chronic and degenerative diseases, the "morbidity rates" defined for these diseases would have to be appropriate so that the patients were effectively reintroduced into the system in successive time increments. In principle, however, there would be no dynamic following of each patient through the developing ranges of his sickness. This form of "static" modeling of the individual's illness allows for disaggregation into an importantly large number of different disease groupings, so that the changing patterns of disease in the population and the changing needs for different types of specialized health care personnel can by studied.

On the other hand, it can be viewed as more important to model the dynamic nature of illnesses such as the degenerative ones, where the modification of their course by such programs as screening is in mind. However, as argued in Section 2, the large number of probability transitions needed to describe this form of modeling, effectively precludes the disaggregation of diseases into a significant number of groups also.

The matter of the choice of the time increment, ΔT, for the running of the model, as compared to such natural "time periods" as the intervals between different phases of development of a degenerative disease, the periods of episodic treatment for acute illnesses, and the waiting periods between various forms of treatment after contact has been made, is very important in regard to the form of modeling undertaken.

If at one end of the scale the ΔT is chosen to be one year, then the dynamics of illness trajectories become largely unimportant, and the dynamic nature of the model is then focused on how trends develop within the system over such typical periods as a ten-year span. It should be noted that in developed countries, typically the order of 90% of the population will have contacted the health care system for treatment of some sort during the course of a one-year period.

A typical, practical lower limit would seem to be one month, since between one week and one month is typically the natural time for such periods as waiting time between appointments, length of stay in hospital, etc. Typically, something like perhaps 15% of the population will have been treated by the health care system in this period. With such a ΔT, seasonal variations can of course appear significantly, and indeed the morbidity rates and even birth and mortality rates may need to be given seasonal variations in such a modeling. With such a ΔT, also the degenerative disease trajectories now emerge as being relatively long-term, and therefore it is more appropriate to provide for continuity in the successive development of resource consumption by such patients over subsequent time intervals.

At this moment I do not think any hard guidelines can be established for which type of modeling for morbidity is most "universal". Probably both types should be explored within the IIASA project.

3.3. Prevention-Screening-Treatment

As regards what the HSDIM has so far called "preventive" (in comparison with "treatment"), I would propose that this be split into two groups:

(i) prevention - reduction of disease risk for asymptomatic but concerned (interested) well people;

(ii) early diagnosis - screening.

Since screening also results in treatment, the previous use of the word "treatment" should be modified, so that the third group becomes:

(iii) sickness treatment - acute and chronic sick caring and curing.

I have argued elsewhere on the vital need for planners to provide significant resources in developing a <u>better</u> HCS. Unless we wish to argue that there is no way to prevent people from getting sick in the first instance (because "this is what they want to do"!), then any procedures which can be effective ten to twenty years before the diseases become symptomatic, are obviously well worth considering first. In any case, I believe most national planning agencies for health care systems will recognize the importance of including such prevention within the next decade.

In comparison, early disease detection must still apply conventional medical treatment to diseases detected perhaps from one to ten years earlier than when they become normally symptomatic. There is indeed much controversy in the literature about the cost:benefit situation for different diseases, and the analyses seem to produce wildly different conclusions. Of course, any screening program which can reasonably be expected to be worthwhile should be incorporated and hopefully the model will be able to predict at least some cost:benefit aspects. Discounting aspects are pertinent and will be discussed under Performance and Optimization.

It is fairly clear, as shown in Figure 2, that the patient-flows for treatment result directly from the matrix multiplication of the population model and the registered morbidity. On the other hand, early treatment and pure preventive programs cannot be applied only to the part of the population suffering "underlying" morbidity, because these people, by definition, are not known a priori. Therefore, to a greater or lesser extent--depending upon the amount of information known about particular diseases and risks of diseases--the early diagnosis and preventive programs must be applied to the populations "at risk".

It is important to note that, while the acute/chronic treatment plans are at least relatively well-known and finite in number, the possible programs for early detection and prevention are almost infinite in number, with widely different extents of knowledge about their cost effectiveness aspects.

The model also must be able to incorporate other choices than the conventional hospital treatment for acute/chronic treatment of sickness, and this is indicated in Figure 2, for such alternatives as: day surgery and home care; policlinic; community health center; geriatric units; extended care units; and the developing USSR system called "dispensarization".

3.4. Performance Measures; Objective Functions; Improvement or Optimization Techniques

The first basic point to agree on is that the meaningful outputs of the health care system are not represented by the levels of activity within the system. Particularly, for example, the numbers of bed-days provided, the number of physician visits, and the number of X-rays taken are examples of typical and necessary health care activities. Clearly, however, they are not the end result of the process in which we are interested, and do not inherently represent any measure of effective performance in the system. Indeed, with the "technological imperative" providing increasingly powerful but expensive computerized, laboratory and other test equipment, the problem of how to control and, in particular, to limit such activities to the most beneficial level is a matter of great concern. It is also unsatisfactory to utilize such widely available "vital statistics" information about the population as life expectancy, death rates for various ages and groups (including, for example, perinatal) and indeed the proportion of the GNP spent on the health care system. To take life expectancy as an example, it is clearly now increasingly possible to keep people alive longer, but not necessarily with any guarantee of an adequate, let alone improved, quality of life.

The approach to measuring health care system effectiveness must, in principle, relate to the quality of life perceived by each person himself. The Health State Utility concept provides a conceptual approach to this in that this is a measure defined continuously between the values 0 for death and 1 for complete (WHO) health. A complete trajectory for a life may then be shown, as suggested in Figure 3. Presumably, the ideal trajectory would be one of complete health at the value 1 throughout life until a "natural" death occurred "during sleep" at some socially-accepted adequate age. A more typical trajectory for people in a developed country will show a rise into early mid-life, following the frequent acute episodes of childhood, until a slow deterioration sets in, somewhere around late middle-age. The particular problem of our highly scientific and technical sick-care system is that it can prolong this form of life through a slow degeneration, with much disease along the way.

The approach to using Health State Utility as an outcome measure is that, in principle, the ΔH, or improvement of health, due to any particular health care program, can be established as a new trajectory over time, as suggested in the figure. Theoretically then, by integrating programs over the population, a total measure of the effectiveness of this system could be achieved. Indeed, this allows for all sorts of techniques to be developed for improvement or optimization of the system, using interactive and game-playing techniques.

The problem remains of establishing the changes in Health State Utility and their time developments for all the multifold activities of a health care system, most specially the comparative differences between the use of resources in the different modes of preventive, early diagnosis, and acute/chronic sick care.

Discounting is a phenomenon which must also be mentioned here, since we are talking about the effects of health care programs over a significant period of time. There is no general agreement on the rate of discount to be applied, although certain arbitrary numbers have been proposed for particular situations; for example, the British Civil Service recommends 10% for all of its capital projects. The particular problem resulting from discounting for the health care system, and even more emphatically for the phases of prevention and screening, are that the costs are incurred now, while the benefits follow as a stream extended over time, but typically not starting for several years afterwards. In the case of prevention, which may not produce its effects for 20 years in some cases, the result is that the benefits may only receive a weighting of about 10%, that is, the present value of these dollars of benefit is only about 10 cents.

This would seem to represent a significant disharmony between the economic evaluation and our desired social evaluation of the situation. Clearly, after preventive and screening measures have become accepted and been in effect for 20 years or more, there will be, in any current year, a flow of benefits which are available, in principle, to finance the continuing flow of costs for the new and current cohorts to which these programs are now going to be applied. In this social context, it may not be appropriate to discount the benefits, but rather to treat them as simultaneous streams of canceling costs and benefits. In this context, the period of the transient build-up of costs, before the benefits begin to emerge, would be written off to capital investment in the population, just as one invests in other social institutions, such as universities and recreational facilities.

Clearly, this is a matter for decision making at the sociopolitical level, which is indicated in the Figure 2 as "Funds Allocation (Planning)".

3.5. Planning of Allocations

Given the possibilities for the model providing new and important information on cost-effectiveness of various health care system scenarios, the government decision makers must decide upon budgetary allocations. As shown in Figure 2, the main identifiable areas for such allocations are both for the capital investment in development of new resources, and in the allocation for on-going operation of each of the three areas of prevention, early diagnosis, and acute/chronic sickness treatment. Furthermore, there is a seventh area, that of the development of the underlying background medical science. It must also be noted that many of the personnel resources and technical facilities developed under each of the three headings have, in fact, a considerable measure of substitutability, in that they can be transferred from one type of activity to another, if this seems appropriate. For those subsets of the resources for which this comment is applicable, it is necessary to have an intelligent process available to "integrate" the developments and utilization of these resources.

Finally, it should be noted that the health care system, as so modeled, is still only one part of the overall government provision of facilities to the population. Indeed, the health care system is so intimately imbedded in the overall social system that even this extensive system which we have proposed is still only causally responsible for some proportion of the health state of the population. Almost all of the other identifiable systems of government have greater or lesser effects upon health, and here we name only the following:

- education; educated people look after their health better.
- agriculture; nutrition affects health.
- environment; forests, lands, water; the environment affects our health.
- legal and criminal justice system; the rapid decision of divorce cases, etc. can reduce stress-inducing illness.
- industry and transportation; pollution affects health.

3.6. Concluding Comment

An intriguing idea is that "health" is a "universal measure". Specifically, it may be that the overall integrated sickness in the community represents a proxy for the extent to which the population feels its quality of life is unsatisfactory. In this respect, it may provide a performance measure for almost all of government branches of activity, such as those suggested above. Thus, while even in its own right the health care system is a tremendously important system, touching on all our aspects of physiological, psychological, social, ethical and moral problems, it may also represent the nearest we shall ever have to a universal monitor, for government and society as a whole, of its perceived satisfaction with life.

FOR POPULATION AGE-SEX GROUP I

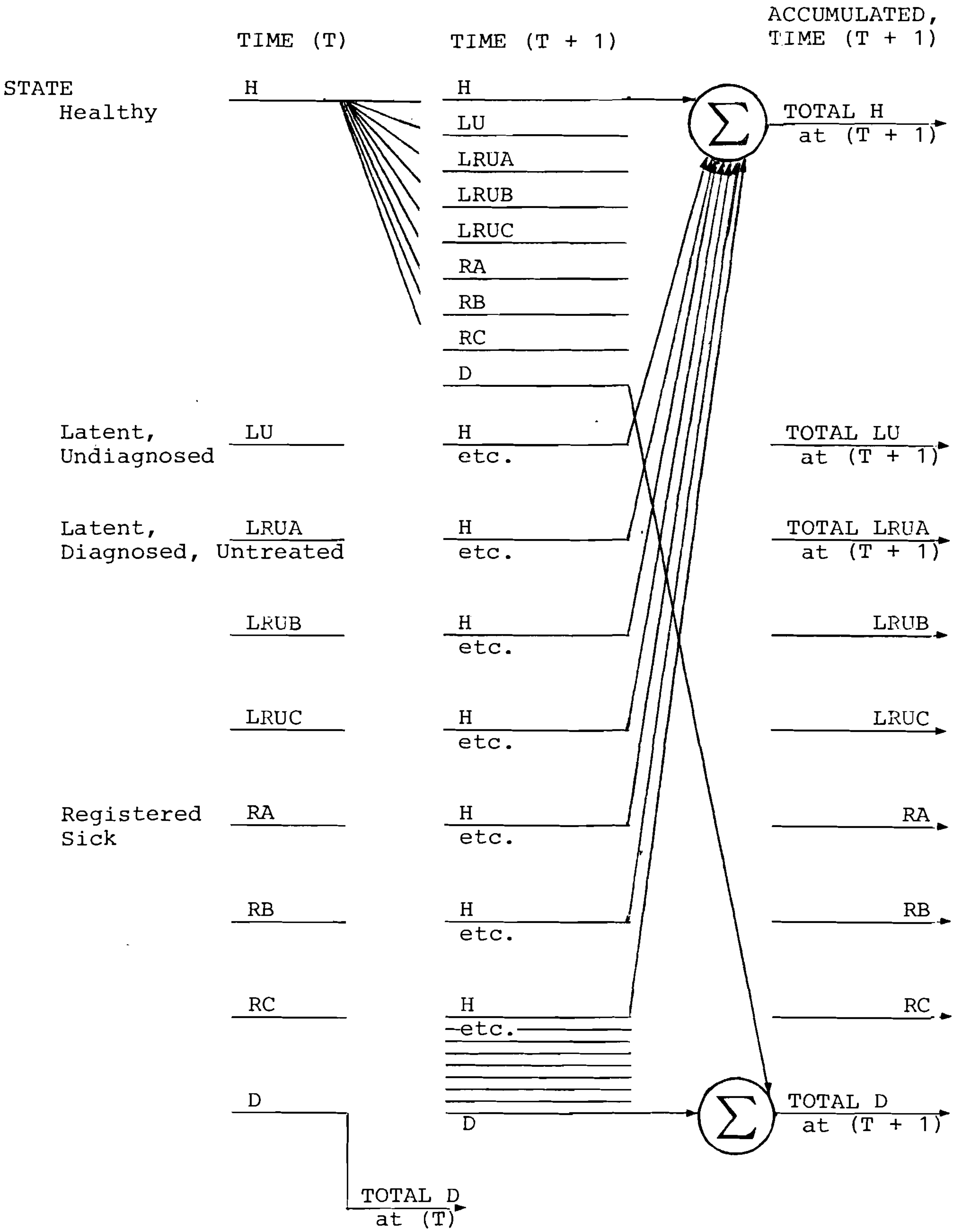

Figure 1. The State Transitions Between HP, LD and RP

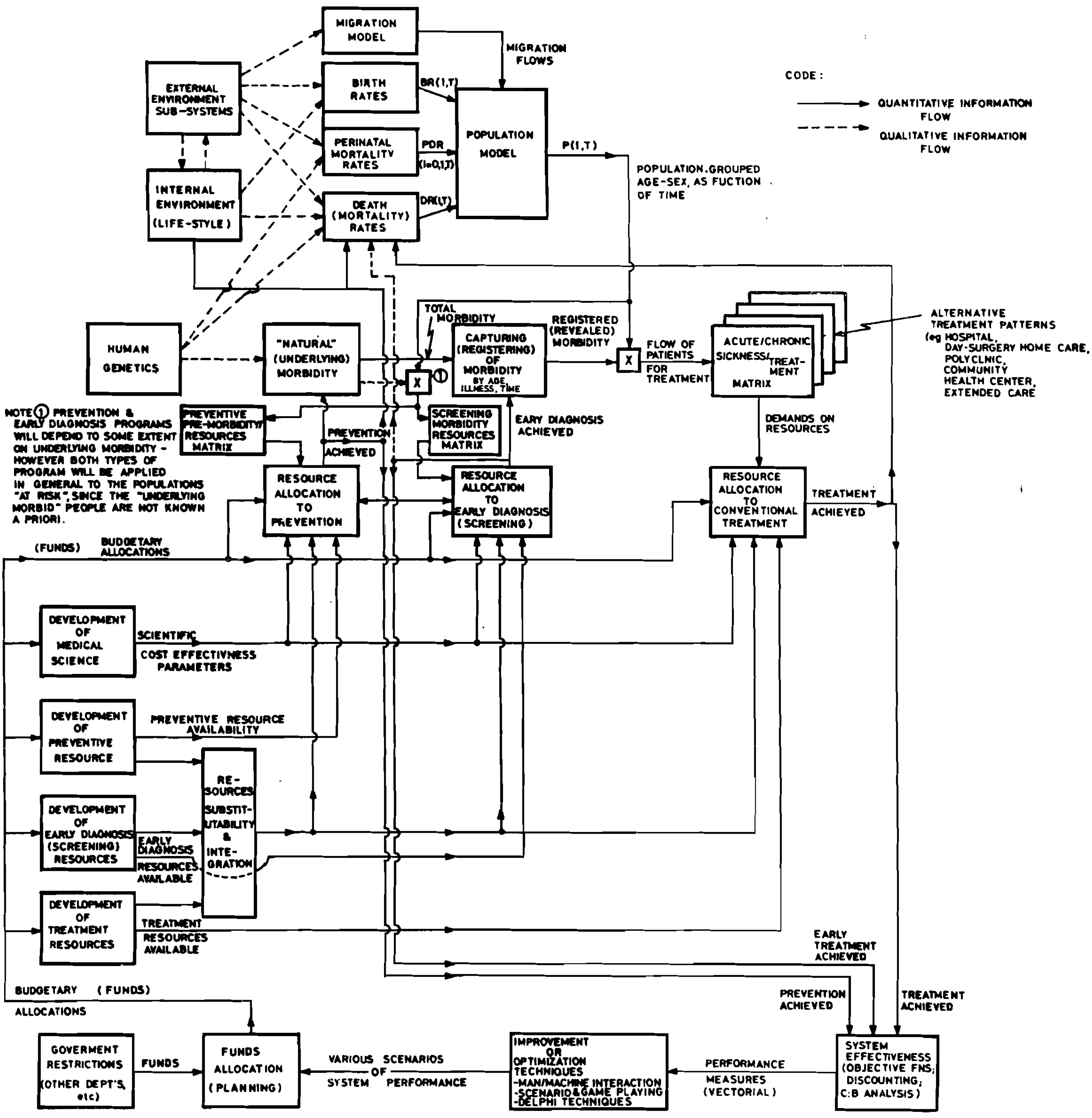

FIG. 2. TENTATIVE FLOW DIAGRAM FOR GENERAL MODEL OF HEALTH CARE SYSTEMS

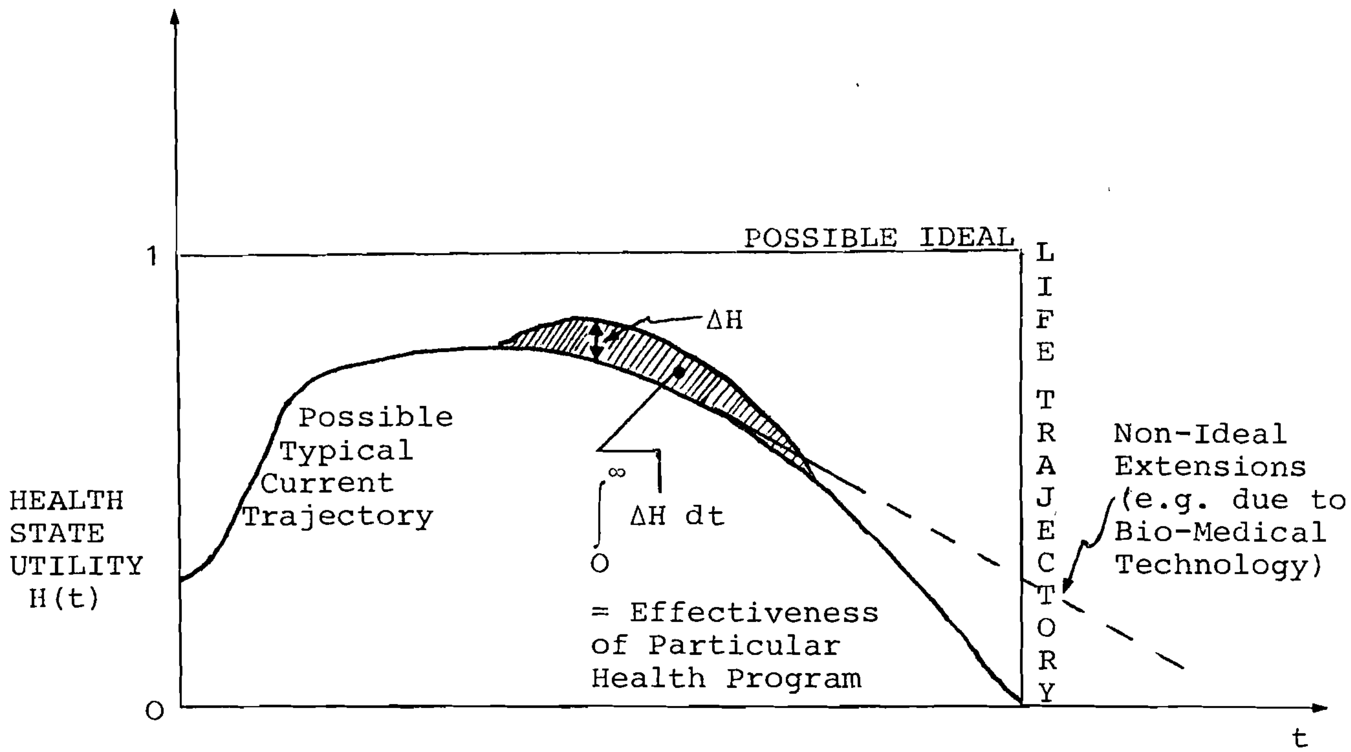

Figure 3. Health State Utility as HCS Performance Measure

4. REFERENCES

[1] Klementiev, A.A. A Computer Method for Projecting a Population's Sex-Age Structure. RM-76-36, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1976.