# ALTERNATIVE ANALYSIS OF DIFFERENT METHODS FOR ESTIMATING PREVALENCE RATE

E.N. Shigan

August 1977

Research Memoranda are interim reports on research being conducted by the International Institute for Applied Systems Analysis, and as such receive only limited scientific review. Views or opinions contained herein do not necessarily represent those of the Institute or of the National Member Organizations supporting the Institute. .

·

.

.

# Preface

The aim of the IIASA Health Care Systems Modeling Task is to build a National Health Care Systems 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 submodels dealing with population estimation, disease prevalence estimation, resource allocation, and resource supply. This paper is concerned with disease prevalence estimation. It reviews the different types of morbidity data that are available in different countries and suggests how the prevalence estimation modeling activity needs to adapt to these different situations. .

-

# Abstract

The estimation of disease prevalence is a crucial part of the national planning of Health Services since it is a necessary step in the calculation of resource requirements. There are a number of different types and sources of morbidity data which prevalence estimation models can use. These data can be categorized both by disease type (e.g., terminal degenerative disease, non-terminal disease, infectious disease and trauma), and by the source of the data, (e.g., inpatient data, out-patient data, screening records, temporary disability information). The main problem with much of the morbidity data that is available in disaggregated form is that it comes from one part of the service, such as in-patient data, and will therefore give an incorrect picture of the morbidity of the It therefore needs to be augmented general population. by data relating to the general population, e.g. data from general samples or interviews. Since this general population data is usually expensive to collect, the best strategy is to collect it in aggregate form and combine it with the disaggregated data from particular parts of the service, e.g., in-patient data (which is available in a very disaggregated form in many countries).

# Alternative Analysis of Different Methods for Estimating Prevalence Rate

E.N. Shigan

# 1. INTRODUCTION

The solution of the "health-demands-resources" problem is very important both for developed as well as developing countries.

In the most developed countries some unsatisfactory trends in the health indices in medical resources can be observed. The resources themselves are sometimes not utilized in full scope.

In the developing countries there is also the problem of estimating medical resource demands, considering the real situation in these countries (economics, family planning, education, etc.).

In order to solve the "health-demands-resources" problem, health care organizers, together with other specialists, must perform experiments on the real objects (medical establishment, district, national levels). Under the methodological supervision of WHO specialists, such investigations are being carried out within a common program in several countries. However, all these experiments are very expensive, take much time and are unable to test many alternatives for a planning policy. That is why many national and international organizations are paying so much attention to the development of different mathematical models for simulation activity and trends of health centers, hospitals, health care systems on the district, national, regional and global levels.

The IIASA biomedical group, under the supervision of Dr. D. Venedictov, has been working on the development of a national health care system model since 1975. Such a model will help national level decision makers to consider different versions of planning decisions and to choose the best alternative for them. Considering the health care system alone, i.e. apart from other interacting systems, the appropriate model consists of the following submodels:

population  $\rightarrow$  health  $\rightarrow$  need/demand  $\rightarrow$  resource supply  $\rightarrow$  resource allocation

These submodels are considered to be interrelated and dynamic. A step-by-step approach was used in the elaboration of the NHCS model:

- identification of the common national health care problems;
- verbal description of the national health care system;
- review-analysis of the different health care system models;
- population submodel;
- sickness-prevalence submodels;
- health submodels involving indices of ill-health (mortality, morbidity, invalidity, etc.).

The main component of ill-health which needs the most medical resources is morbidity. That is why the biomedical group has been working on the development of morbidity estimation models since last year. In order to better understand the problems of morbidity modeling, it is necessary to know about problems of morbidity.

# 2. MORBIDITY

In the twelfth report of the WHO Expert Committee on Health Statistics, TRS N 389 (1970), morbidity is understood to be any departure, subjective or objective, from a state of physiological well-being.

Some other terminology, generally accepted to distinguish several types of morbidity and rates, is as follows. General morbidity, or prevalence, is the number of all cases, new and old, of disease and any pathological conditions in a specified population. For estimating the general morbidity level among the population of a town, district, or nation, there is a prevalence rate, or general morbidity rate, denoted GMR.

Besides the concept of general morbidity (prevalence), it is also possible to distinguish between its different parts:

- incidence morbidity (new diseases occur);
- others (known cases, chronic diseases, etc.).

### 3. SOURCES OF INFORMATION ABOUT MORBIDITY

There are two main channels of information about general morbidity data:

- routine statistics, and
- special investigations.

Routine statistics include a set of different official certificates, forms, and medical reports. In some developed countries, such as the Scandinavian countries, Great Britain, and the USSR, routine statistics contain important information about the morbidity of a population. In other developed countries, official statistics are concerned only with economical data and include little information about morbidity.

Special investigations can be local and national, aggregative or sectoral (specific group of disease) clinical studies, interviews, etc.

Information about the general morbidity of a population can be taken from hospitals (*hospital morbidity*). Certainly data about in-patient cases cover the most serious part of diseases.

Also very useful is the analysis of general morbidity data taken from *out-patient visits* to general practitioners, health centers, polyclinics, etc. The spectrum of these data includes more cases about infectious and acute conditions. There exists separately information about the working population - *morbidity with temporary disability*.

In some countries there are very strict rules concerning the confidentiality of health information and it is rather difficult to obtain data about out-patient visits, hospitalized cases, etc. In these countries, different kinds of *interviews* with people are being used as the means of collecting morbidity data.

But all these sources of information about general morbidity of the population do not cover all cases. According to investigations conducted in different countries, annually some 20-30% of the population do not visit medical establishments; this does not imply, however, that all these people are healthy. For this reason, *screening* is now a generally accepted method of examining the population in developed countries. Mainly unknown degenerative chronic diseases are discovered during medical examination of the population. Great help for the estimation of terminal degenerative diseases is given by the use of data concerning the cause of death, and clinical data about the survival of patients of the degenerative group.

Completeness and reliability of sources of information about general morbidity depend very seriously on many factors. These factors can be divided into several groups, according to:

- individual (age, sex, educational level, income, etc.);
- type of health care system (insurance, state, private);
- situation of medical resources (bed/doctor/population ratio, medical equipment);
- accessibility of health care (distance, roads, transport, communication, etc.);
- condition of statistical registration and data processing;
- doctors' possibilities (education level, general or narrow specialist, provision by diagnostic technique, etc.).

All these factors vary greatly from country to country. Therefore, in order to get data on general morbidity in a specific country, different statistical or medical computer centers are collecting data from different sources.

In collecting different morbidity data, some very serious difficulties may arise. Since the same people can visit general practitioners, be hospitalized, and be examined during the screening procedure, it becomes difficult to collect all personal information (identification of person, linkage record study, alphabetization, etc.). Considering these organizational difficulties, computer possibilities, etc., there exists no country at present where all these problems have been solved and where there is a complete computer bank of personal health information from which statisticians and planners could take general morbidity data. In this situation, mathematical models for estimating morbidity must use different types of data in different countries. In Figure 1, different sources of information, which can be used for modeling general morbidity, are introduced. In order to develop a set of mathematical models for calculating an adjusted general morbidity rate, it is necessary to become familiar with different indices, that are used in different countries, and which can be taken from routine statistics, special samplings, or clinical investigations, for the construction of the models.



Figure 1. Scheme of building a general morbidity model.

As one can see from this figure, there exist two ways for estimating general morbidity: natural and mathematical.

- Natural procedure of estimation: obtaining real general morbidity data on the basis of all sources of information.
- Mathematical procedure of estimation: obtaining theoretical, adjusted, hypothetical general morbidity data on the basis of mathematical models.

#### 4. LIST OF MAIN MORBIDITY INDICES

General morbidity (prevalence) rate (GMR): GMR = all cases during one year among population • 1000 ; population Incidence rate (IR): IR = all new cases during one year among population • 1000 ; population Screening morbidity rate (SMR): SMR = <u>number of cases revealed during screening</u> • 1000 ; number of screened population Practically healthy people rate (PHPR): PHPR = <u>number of healthy people</u> • 100% ; population (or its parts) Average number of out-patient visits per capita (AOPV): number of all out-patient visits AOPV = population Frequency of hospitalization (H%):  $H\% = \frac{\text{number of in-patients}}{100\%}$ ; population Infectious morbidity rate (IMR):  $IMR = \frac{number of infectious diseases}{100000}$ ; population Terminal degenerative disease rate (TDMR):  $TDMR = \frac{number of terminal degenerative diseases}{100000}$ , population Nonterminal degenerative disease rate (NTMR): NTMR = number of nonterminal degenerative diseases · 100000 population Rate of trauma frequency (RTF):  $RTF = \frac{number of traumas}{number of traumas} \cdot 100000 ;$ population Temporary disability rate (TDR):  $TDR = number of temporary disability cases <math>\cdot 100\%$ number of workers

It has been accepted to divide these rates into two kinds: general, and specific. General rates are calculated on the general totality, without division according to age, sex, disease, etc. Specific rates are calculated mainly according to the following factors:

- age;
- sex;

- urban/rural;
- disease;
- occupation.

These rates can be calculated for the population of a country, district, or town; temporary disability rates - for one plant, a whole industrial branch, etc.

As usual, the most complete and reliable list of morbidity rates can be taken from specially organized scientific investigations.

#### 5. LIST OF ADDITIONAL INDICES AND DATA

This list of material includes data without which it would be impossible to find the main indices of morbidity:

- population: size, age, sex, urban/rural, administrative division, working, etc.;
- ratio of screened population to entire population;
- ratio of working population to entire population;
- structure of out-patient visits according to age, sex, urban/rural, disease, etc.;
- structure of in-patient visits according to age, sex, urban/rural, disease, etc.;
- structure of causes of death according to age, sex, urban/rural, disease, etc.;
- mortality rate according to age, sex, urban/rural, disease, etc.;
- additional scientific material.

#### 6. DIFFERENT ALTERNATIVES FOR MODELING GENERAL MORBIDITY

Depending on the kind of information available, its completeness and reliability, different alternatives for constructing general morbidity models can be suggested for each particular country. The following variants of modeling can be suggested for the most frequent situations.



	Fi	qure	2
--	----	------	---

This situation exists in many developed countries, where population data (size, age/sex structure), and data on morbidity rates from different causes for each age/sex group are available. All of these data, of course, are available in official annual statistical reports. But in order to calculate terminal degenerative disease morbidity, it is necessary to find data about survival of patients of different age/sex groups, who are suffering from these degenerative diseases (cancer, cardiovascular, etc.). These data can be obtained from different scientific This mathematical model for estimating terminal investigations. degenerative morbidity was elaborated and tested on real material of different countries by the IIASA biomedical team (A. Klementiev, K. Atsumi, S. Kaihara, I. Fujimasa). This method is described in detail in an IIASA publication (Kaihara, S., et al., An Approach to Building a Universal Health Care Model: Morbidity Model of Degenerative Diseases, RM-77-6, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1977).

# Limitations

- The accuracy of this method depends on the completeness of the initial material, particularly on the ratios of autopsies, physician diagnoses, etc., to total deaths.

- If the data on the survival of degenerative disease patients (cancer, heart disease) are taken from special hospitals, or from any selected group of patients, how is it possible to consider this material as being representative of all cases of degenerative diseases in the country?
- Using this method, it is possible to estimate only terminal degenerative groups of diseases, which account for only one part of the total.

# 6.2. Variant 2

In some cases, during the modeling process, researchers can use population data and data available from routine statistics. In Figure 1, all these possible sources are mentioned-data about hospitalization, out-patient visits, screening and temporary disability. But, as usual, in many countries' reports these data are presented only in an aggregative form, without detailization according to age/sex group or disease distribution. Therefore, in order to link the morbidity estimation method with the population submodel, and in order to have an idea about the ratio of hospital cases, screening cases, etc., to general morbidity, it is necessary to have a special sampling investigation.



# Figure 3

The following real situations exist. a).



Figure 4

For this situation, the following data is necessary:

- population size, age/sex (static, dynamic), from census;
- distribution of hospitalized cases according to disease (List A or B of ICD), from annual report;

and from sampling study:

- % of hospital cases according to age, sex, and disease
  (List A or B);
- % of hospital morbidity from sampling general morbidity.

b).



Figure 5

This situation exists very often when the following data are available:

- population, size, age/sex structure from census;
- total number of out-patients (very often);
- distribution of out-patient visits according to age (very rare);

- distribution of out-patient visits according to disease (very rare).

In order to elaborate the general morbidity rate, it is necessary to take the following data from sampling studies:

- distribution of out-patient visits according to age/ sex, disease;
- % out-patient morbidity from general morbidity.



Figure 6

First of all, it is necessary to distinguish among several types of screening:

- Single test screening application of one single test to people.
- Multiphasic screening (multiple) applying a combination of different screening tests to people.
- Mass screening screening the entire population.
- Selective screening screening selective groups of the population.

Of course, the most useful method for calculating the general morbidity rate is mass multiphasic screening; that is, the application of many tests to the entire population. Other screening methods are oriented either to specific diseases, or to specific groups of the population (high-risk groups). But even mass multiphasic screening is effective only for degenerative chronic disease and gives little information about acute conditions and infectious diseases. That is why for calculating the general morbidity rate we need additional information from special investigations, for example percent cases detected during mass multiphasic screening from general morbidity cases (in sampling), according to age/sex, disease.





In some countries, very developed statistics on the temporary disability of employed persons are available. These data may help the mathematical modeler to build models for the estimation of general morbidity. But in addition, he also needs the following data:

- population size, age/sex, urban/rural;

e.)

- distribution of population according to profession, industrial branches;
- from sampling surveys the data on the prevalence rate of this sample of people, and estimation of the proportion of the employed to the unemployed part of the population.



Figure 8

In many developed countries, it is not easy to obtain data on the health of the population through official "routine" statistics or through different research surveys. This is due to two main problems:

- the confidentiality of personal health information;
- the absence (or shortage) of data about the quality of health care in official statistics.

That is why in those countries the method of interviews with people is very often used as the source of information about individual's health.

The interviews cover selected groups of the population; selected districts can be organized with or without all the necessary demands of the representative sampling method. In the questionnaires there are different groups of medical questions--about out-patient visits, hospitalization, screening procedures, etc. Besides the medical questions which are usually asked, there are many others--housing, income, education, etc.

# Limitations

The quality of information obtained by means of interviews depends very seriously on many factors:

- person (his educational level, social level, etc.);
- content of questionnaire;
- who is answering the questions--the person whom we need, or his relatives;
- which time-period is the questionnaire interested in.

In many developed countries, there exist investigations devoted to the comparison between interview data and hospital data, interview data and general practitioner information. These data could be used for modeling the general morbidity in these countries.

In other countries, where the individual's health is not considered as confidential information, this is included in official statistics data and usually accompanies all research surveys.



Figure 9

In many developed countries, there exist well-organized routine statistics, including information about all cases of hospitalization, out-patient visits, screening, and temporary disability data. In some countries, there can exist two or three sources. Therefore, it is very useful to create models for different combinations, like:

- hospitalization + out-patient visits;
- hospitalization + screening;
- hospitalization + temporary disability;

But nevertheless, in order to link-up with the population block (age/sex distribution, etc.), to estimate how completely the combination of these channels reflects the general morbidity of the whole population, we need material from special investigations about all possible sampling rates previously mentioned.

#### Limitations

The accuracy of the models, which could be suggested for estimating the general morbidity of the population on the basis of routine statistics, depends very seriously on how completely these statistics cover all cases of hospitalization, out-patient visits, etc., because sometimes these statistics cover only part of the medical establishments, large plants and factories, and so on. Therefore, this group of the medical establishment cannot be representative of the whole totality. The same people during the year can visit many out-patient centers, can be hospitalized, and can be screened. On the other hand, some people will not visit any medical establishment either because they are healthy or they consider themselves such. Certainly, the accuracy depends also on the reliability of the statistics, including the use of definitions, classification, methods of calculation of different rates, etc. For this reason, in addition to routine statistics, modelers need data from special sampling investigations.

# 6.3. Variant 3

In some countries, there are good statistics on infectious diseases. In this situation, this very important component of general morbidity can easily be taken from official statistics. But usually these data are introduced in countries' annual reports in an aggregative form without division according to age/sex groups. For this purpose, in order to link up with the population submodel it is necessary to use additional research data from sampling surveys. This method is being elaborated by IIASA's biomedical group on the basis of Japanese data.

As to official national statistics regarding all accidents (see Figure 1), not many developed countries have good data covering the whole country. But if data exist about this component of general morbidity, they can be used by mathematical modelers. While building the model, one can face just the same problem--how to link accident data, taken from official statistics in an aggregate form, with the population submodel. For this purpose, it would be very useful to use a sampling survey.

# Limitations

As in Variant 1, all these approaches are used to help estimate different components of general morbidity:

- infectious diseases;
- accidents.

In order to obtain approximate data about general morbidity, we need all components:

-15-

- terminal degenerative;
- nonterminal degenerative;
- infectious diseases;
- accidents; etc.

Occasionally, one can encounter people who had, during the year, different kinds of diseases (for example, arterioschlerosis + influenza). This means that an overlap among all these components exists. Also, this classification is oriented to the ill population and is not concerned with healthy persons or groups in specific physiological conditions (e.g. delivery, etc.), which also need a significant part of the medical resources.

# 6.4. Variant 4



Figure 10

Mathematical modelers can face this situation in developing countries, where there is either no, or only poor, statistical information about health. As usual in these countries national centers, with the help of WHO or other international organizations, carry out special investigations. These data about the health of a sampling of the population, including the distribution of the population according to age/sex, disease, and visits to medical establishments, can be calculated on the basis of all population of the countries.

# Limitations

The accuracy of this alternative depends very seriously on the quality of the sampling study. It is concerned with the representativeness of the study, what kind of sources of information were included (hospital, out-patient data, etc.), the organization of the screening process, data processing, etc.

# 6.5. Variant 5



In some developing countries where no, or only very poor, routine statistics exist on different aspects of health and where a special health survey was either never organized or was concerned with a nonrepresentative group of the population or medical establishments (e.g. included only large hospitals, private sectors, capital cities, etc.), modelers can use only data about population.

In order to estimate the general morbidity rate it is necessary to take data from other countries. In this situation, the most difficult for modelers, it is very important to select the country which will be closest with regard to the socioeconomic situation, climate, demography, type of health care system, etc. The application of one from a group of methods of mathematical taxonomy can be useful in this case.

#### 7. UNIVERSAL AND SPECIFIC APPROACHES

All these methods for modeling general morbidity have their own negative aspects and limitations. That is why the application of a combination of all, or several, of these alternatives can be very useful and can more exactly reflect the real general morbidity. In the case when a combination of these alternatives is used on the basis of real computer data about population, morbidity, mortality, etc., operation researchers will get different adjusted general morbidity rates --in a range from  $GMR_{min}$  to  $GMR_{max}$ .

For this reason, for the prognosis of resources it will be very useful to get forecasting data in three alternatives:  $\begin{array}{c} \text{GMR} \\ \text{min}, & \text{GMR} \\ \text{medium}, & \text{GMR} \\ \text{max}. & \text{Sometimes, especially when modeling} \end{array}$  the infectious disease group, accidents, etc., it will be useful to organize expert estimations before proceeding to the construction of resource models.

In most developed countries, the mathematical modeler can find good data about population, routine statistics, mortality and morbidity, and different research data. That is why for this large group of countries, modeling of general morbidity on the basis of all sources of information will give the best approximation to the real result.

# 8. MEANING OF THE SPECIAL INVESTIGATIONS FOR BUILDING AND TESTING MODELS

Returning to Figure 1, there exist two approaches for modeling the general morbidity of the population. The first proceeds from different groups of diseases:

- terminal degenerative;
- nonterminal degenerative;
- infectious;
- accidents.

The second proceeds from the organization of the health care service and existing real medical sources of information in most countries:

- hospital morbidity;
- out-patients;
- screening;
- temporary disability.

Both these approaches have their own specific problems mentioned above. But besides the specific, there also exist nonspecific, common problems. One such problem is that the individual can suffer from several diseases of different groups, and can visit different medical establishments many times during the same year. That is why the organization of special investigations--collecting all medical events for each person in combination with the depth of information, good registration, and data processing--helps the researcher to avoid these problems (see Figure 12).



Figure 12

If we have, from the sampling, general morbidity rates for each age/sex group it is possible, by excluding the different sources of information, to estimate how much the morbidity rate received differs from the general morbidity rate; for example, what is the quantitative difference between the general morbidity rate and the morbidity rate without hospitalized cases, for each age/sex group, disease group, etc. This method helps us to better estimate the degree of approximation of each adjusted rate to the general morbidity rate (as a fixed point, standard).

The use of data taken from natural experiments will help to estimate how completely different combinations of alternatives reflect the exact general morbidity rate of the population. We suspect that these corrective coefficients will be different in different countries. It depends on the factors of many specific countries. In other words, the qualitative side of this methodology will be the same, but the quantitative coefficients could be different.

For the elaboration of a universal model, the IIASA biomedical group is trying to generalize the results of different national experiments--Bulgaria, the U.K., Japan, the U.S.S.R, Canada, and Czechoslovakia. The elaboration of such a universal model also allows decision makers from different countries to compare their own adjusted general morbidity rate with that of others. Such a universal approach will be very useful for WHO during the comparison of different countries and the estimation of their own needs.

### 9. OTHER RELATED PROBLEMS

The development of mathematical modeling of morbidity is very closely connected with the successful solution of different problems.

# 9.1. Scientific Problems

- Generalization of different studies about the influence of socio-economical, environmental, and other external systems on the level of morbidity;
- Elaboration of methods for integrating morbidity indices with mortality rates, physical development, invalidity, and economical rates (integrative indices);
- Elaboration of different information systems on the basis of the linkage record study, both for the whole population of different localities, as well as for selected groups of the population (for example, high risk groups);

# 9.2. Organizational Problems

- Improving the official statistics as a main channel of morbidity information, and as a source for model building;
- Standardization of definitions, classifications, methods of calculation, etc.;

 Collection of all quantitative results (rates, indices, coefficients, models, etc.) obtained in the course of different scientific investigations (samplings, clinical investigations, etc.) in one national center.

# 10. CONCLUSIONS

- In order to estimate medical resource needs, it is necessary to know the health of the population at present, and to forecast its changes in the future.
- All countries can be classified into several groups, depending on the presence and the degree of development of different sources of information about the general morbidity of the population.
- 3. In modeling general morbidity, the researcher must take into account all available sources of information, because all of them allow for only partial estimation of morbidity and from different points of view.
- 4. The use of experiences of the developed countries, especially concerning sampling investigations of morbidity, will help the modeler to estimate the degree of approximation of different sources of information in quantitative indices, and to build dynamic models for predicting health of the population and resources needed.
- 5. The elaboration of a universal general morbidity model will allow decision makers from different countries to forecast the level of morbidity rates and resource needs, on the basis of national sources of information.
- 6. This universal model allows for the comparison of the present and future situations of the health care systems in different countries, which is very important for international organizations such as WHO, UNDP, etc.

# Bibliography

- Grundy, F., and W.A. Reinke, *Health Practice Research*, PHP 51, WHO, Geneva, 1973, pp. 41-42.
- Kaihara, S., et al., An Approach to Building a Universal Health Care Model: Morbidity Model of Degenerative Diseases, RM-77-6, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1977.
- Kohn, R., and K.L. White, eds., *Health Care: An International Study*, Oxford University Press, London, 1976.
- Sluchanko, I.S., and G.F. Tsercovnyi, Statistical Information in Health Care Management, Meditsina, Moscow, 1972.
- Stimson, D., and R.H. Stimson, Operations Research in Hospitals, Hospital Research and Educational Trust, N.Y., 1972.
- Venedictov, D.D., Modeling of Health Care Systems, in IIASA Conference '76, CP-76-7, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1976.
- WHO Regional Office for Europe, *Health Planning and Organisation* of Medical Care, EURO 4102, Copenhagen, 1972.
- WHO Regional Office for Europe, Health Planning in National Development, EURO 4104, Stockholm, 1972, p. 7.
- WHO Regional Office for Europe, *Health Information Systems*, EURO 4914, Copenhagen, 1973.