SIMULATION OF HOSPITAL UTILIZATION*

D. G. Horvitz, T. D. Hartwell and J. R. Batts, Research Triangle Institute

I. INTRODUCTION

The purpose of this paper is to present a dynamic microsimulation model for projecting utilization of general hospitals by the U.S. civilian non-institutional population. The model has been developed to provide essential inputs to a model designed to project hospital manpower requirements [1], but may be used independently. It attempts to meet the need for improved methods of analyzing the effects of changes in the composition of the population and of changing utilization patterns on hospital admission rates and length of stay, as well as for more refined projections of utilization. Further, the model may be used effectively in simulation experiments to provide prior measures of the consequences of alternative programs designed to improve personal health care in the U.S.

The model has two components. The first is POPSIM, a demographic simulation model [2] which creates a population of individuals in the computer and then projects this computer population forward in time, generating births, deaths and changes in marital status for each individual by the Monte Carlo method and on a competing risks basis. The output of POPSIM for a given simulation period (e.g. a year) provides the input to HOSPEP, the second component of the hospital utilization simulation model. HOSPEP is also a stochastic model, generating hospital episode histories for each individual for a selected simulation period.

The principal variables and characteristics generated by POPSIM for each individual in the population file and used by HOSPEP are age (birth date), sex, race, family income, residence (metropolitan or non-metropolitan) and date of death. Using the POPSIM input, HOSPEP classifies each individual by hospital insurance and generates hospital admission and discharge dates. Further, it specifies the diagnosis (18 classes), whether surgery was performed or not, and the bedsize of the hospital.

The hospital utilization model discussed in this paper is essentially descriptive in nature. By varying those parameters (conditional probability distributions) of the model which determine admission rates, diagnosis, surgery status, length of stay and hospital bedsize, the model projects the resultant changes in utilization. The model is not self-adaptive in the sense of having parameters which adjust automatically through interaction with resource constraints in the system. For example, the model assumes there is a hospital bed available for every episode generated by HOSPEP. However, the model does have a definite potential for linkage with resource models of components of the personal health care system and hence for introducing parameter adjustments through feedback. This implies a distinct possibility that the current model could be adapted to project economic demand for hospital services and not just utilization.

The model through appropriate choice of parameters could be used to project need for hospital services rather than utilization. Currently, statistics on need, and particularly unmet need, are available to a considerably lesser degree than utilization statistics. Although the initial interest and emphasis of the model is in projections of utilization at the national level, it can be adapted for local, state or regional projections provided appropriate data are available.

Most models concerned with projecting utilization of health services are aggregate or macro-models. The microsimulation models discussed in this paper appear to offer greater versatility and flexibility in terms of assumptions and output than might be possible with macro-models. Microsimulation models, such as POPSIM and HOSPEP, by their very nature include the inherent variability in the system and can easily generate distributions for the output variables as well as averages. This is a particularly useful feature when assessing alternative health care programs. A more detailed comparison of microsimulation and macrosimulation demographic models suggests there is a definite trade-off in choice of model based on the number of variables and levels of each variable and on the size of the simulated population required to arrive at reasonable conclusions [4].

II. THE DEMOGRAPHIC MODEL (POPSIM)

A. General Description of POPSIM

POPSIM is a dynamic demographic model designed for computer simulation of the principal demographic processes occurring in human populations. It is classed as a microsimulation model because it generates a vital event history, including the dates of birth, marriage, divorce, widowhood, re-marriage and death for each individual in the computer population. Although POPSIM is a two-sex model, it may be used for simulating cohort as well as period data. POPSIM is a stochastic model in the sense that random sampling from probability distributions is used to determine which events occur to an individual and when they occur. It is a dynamic continuous time model, permitting the probabilities to change with time. The model can be made self-adjusting

* Research sponsored by the Bureau of Health Professions Education and Manpower Training, National Institutes of Health under contract number NIH 70-4085.

or recursive through the use of feedback mechanisms. For example, feedback models appropriate for POPSIM to account for the effect of changing marriage patterns on marriage rates are discussed in [5].

The demand in recent years for new techniques for studying complex demographic phenomena with fewer restrictive assumptions than has been possible with analytic models has led to the development of a variety of microsimulation models [6, 7, 8, 9, 10, 11]. POPSIM is a laboratory research tool designed in response to that demand. It can be used to considerable advantage in a wide variety of simulated longitudinal studies. Further, it can be linked easily to other models (e.g. as the demographic component of models of economic, education or health systems), whether they be micro- or macro-models.

1. The Initial Population (Phase I)

The POPSIM computer program consists of two distinct parts or phases. The first is used to create an initial population. The initial population is conceptualized as a random sample of individuals (stratified by age, sex and marital status) from a hypothetical population. The program, through the use of control cards, offers the user considerable flexibility in specifying the nature of the hypothetical population. For example, the user can choose to confine his population to a cohort of women x years of age in 1970, or to all persons 65 years of age and older in 1966, or to the civilian non-institutionalized U.S. population as of 1960.

The initial population can be considered usefully as a random sample of individuals selected from a population register, without regard to familial relationships. Thus, for example, a married female may be selected for the initial population, while her husband and children may not be chosen. Individuals in the computer population (initial sample plus births) are referred to as primary individuals, and marriage partners and children as secondary individuals. Since secondary individuals are not members of the computer population, information concerning them must be carried by the primary individual. In the simulation of vital event histories (second phase of POPSIM), all events which take place are considered as events to primary individuals. All tabulations produced are counts of primary individuals, or of events which happened to primary individuals. Secondary individuals enter the model only in the sense that they influence the vital event risks to which primary individuals are subject. For example, the risk of a married woman (primary individual) becoming a widow is a function of the age of her husband, a secondary individual.

The list of characteristics describing the primary individuals and the associated secondary individuals is shown in Figure 1. The nature of the entries in the list is dictated in part by the fact that POPSIM is a time-oriented simulation model, with continuous time. Consequently, rather than carry age as a characteristic and update periodically, date of birth is used.

Figure 1

List of Characteristics of Each Primary Individual in the Sample Population*

RECØRD(1)	=	sex of individual
RECØRD(2)	=	date of birth
RECØRD(3)	=	current marital status (single, married, widowed, divorced)
RECØRD(4)	=	parity
RECØRD(5)	=	number of living children
RECØRD(6)	=	current contraceptive method**
RECØRD(7)	=	number of marriages (0, 1, 2+)
RECØRD(8)	=	date of current marital status
RECØRD(9)	=	date of next event
RECØRD(10)	=	next event
RECØRD(11)	=	number of children (secondary individuals) for whom information on age, sex and date of death (if appropriate) is available in a separate array for each mother
RECØRD(12)	=	date of birth of spouse
RECØRD(13)	=	previous contraceptive method**
RECØRD(14)	=	date of last birth
RECØRD(15)	=	number of events which happen to the individual during a simulation
RECØRD(16)	=	race (white or non-white)
RECØRD(17)	=	residence (SMSA or non-SMSA)
RECØRD(18)	=	family income deviate
RECØRD(19)	=	family income (dollars)
RECØRD(20)	=	date family income was computed

*RECØRDS(16) through (20) were added for use with hospital utilization simulation model.

**Not used for hospital utilization simulation model.

The computer program to generate initial populations is written in Fortran for use on IBM 360 System/Model 50 computers. The program is designed to generate a specified number of initial sample populations of a designated size in a single run and to write each population on a disk or on magnetic tape. The input to the program includes (in addition to parameters which control various options in the program and class limits used in the output tables) the following variables:

1. The size of the initial population.
2. The proportion of individuals in each of the eight sex-marital status groups.
3. The proportion of individuals in specific (but arbitrary) age groups for each sex and marital status class.
4. The probabilities with which married females are classified as remarried or not.
5. Parameters to assign age of husband given age of wife.
6. Parameters to assign age of wife given age of husband.
7. Parameters to assign the date of current marital status.
8. Monthly birth probabilities by age, parity and marital status (married or single), appropriate to three time points in the 30 year period prior to the date of the initial population. These values are used to generate a birth history, and hence the parity and date of last birth for each married or single female.
9. Monthly death probabilities by age group, sex and marital status to determine number of living children for married or single females.
10. Parameters to assign parity to widowed or divorced females.

The characteristics corresponding to RECØRD(6), RECØRD(9), RECØRD(10), RECØRD(13) and RECØRD(15) are not assigned when the initial population is generated.

POPSIM creates each initial sample population in the computer by means of a series of subroutines which use random sampling of inverse probability distribution functions, for the most part, to assign a consistent set of characteristics to each individual. For example, a stratified random sampling procedure is used to assign age (date of birth), sex and marital status. Using the first three input variables listed above, a density function is fitted by the computer for each age-sex-marital status group. The distribution function and its inverse are then computed for each of these groups. The age assignment routine then sets up the records for the individuals in each age-sex-marital status group and assigns their ages by stratified random sampling of the associated inverse probability function. This is accomplished by first dividing the (0, 1) interval into n sub-intervals or strata of length $1/n$, where n is the number of persons in the particular age-sex-marital status group, and then generating a uniformly distributed random number for each sub-interval to sample the appropriate inverse. This stratified sampling procedure distributes very effectively the ages of individuals over the entire age interval for each sex-marital status group.

The remaining characteristics are assigned to each individual by sampling the appropriate conditional distribution for the specific age, sex and marital status of the individual. For example, the age of husband is assigned by generating a random observation from the conditional distribution of age of husband given age of wife.

2. Generation of Vital Event Histories (Phase II)

After creating an initial population of desired size and characteristics, a second program uses the Monte Carlo method to generate a vital event history or life pattern for each individual. This program advances the population forward through time in a series of time intervals or steps. At the end of each step, the program prints a series of tables and provides the user the option of updating the probabilities of the various events. The user must specify the total length of the simulation period and the time interval for each step. For example, one may simulate the vital events that occur to the initial population for a period of ten years in ten steps of one year each, or for a period of ten years in a single step, or 25 years in five steps of five years each.

The basic set of input data required by POPSIM for generating vital event histories includes:

1. Monthly birth probabilities for females by age group, marital status (married or not married) and parity (or number of living children).
2. Monthly divorce probabilities by interval since marriage (or by age).
3. Monthly death probabilities by age group, sex and marital status.
4. Annual marriage probabilities for females by age and marital status.
5. Parameters to determine the marital status of the groom given that of the bride.
6. Bivariate distributions of ages of brides and grooms for first marriages and remarriages.

At first glance, the list of required input parameters for both phases of POPSIM appears rather formidable. This is not a peculiarity of this or any other microsimulation model. The data required for any model increase rapidly with the degree of complexity of the model. However, POPSIM does not require all the input data for the model to be obtained entirely from one census or sample survey of the population of interest. Estimates of the various parameters used may be obtained from different prior studies on the population of interest and still be used in the model. Occasionally it may even be possible to use data from studies of different but similar populations and obtain valid results. Clearly, the data employed in a microsimulation model will have an important effect on the accuracy of predictions made with the model for particular populations. However, for most applications, the use of parameter values which have been estimated from data which do not pertain exactly to the population of interest will not hamper the usefulness of the model severely. In specific applications, simulation runs using upper and lower bounds for parameters whose values are in the questionable category may provide confidence bounds needed to accept the results from the model.

POPSIM generates marriages, births, divorces and deaths using stored matrices of monthly transition (event) probabilities. A conditional probability approach is used which permits these event probabilities to depend on the current characteristics and prior history of the individual. For example, the monthly birth probabilities vary by parity and marital status within five year age groups for females between 15 and 45 years of age. Births are not permitted unless the interval since the last live birth is at least nine months.

An event-sequenced simulation procedure is used in which an individual is processed only when an event occurs to him. The first step in this procedure is to generate the time interval (and hence the date) of the next vital event for each individual in the initial population. Since the type of event to occur next (i.e., a birth or a change in marital status or a death) is not known, POPSIM generates the time interval (or waiting time) separately for each of the competing events that can happen to the individual, under the assumption that nothing else does happen to him, and the event with the shortest generated time interval becomes the next event for that individual. Only this next event and its time of occurrence are carried in the record for each individual. This procedure for generating the next event is accurate under the assumption that the input parameters are independent probabilities for the competing events (or net rates) rather than crude rates.

The technique for generating the date of an event uses the inverse of the geometric distribution within time periods for which the monthly probabilities of the event remain constant. Consider, for example, the generation of the date of death of a married male who is exactly 42 years old. The monthly chance of dying for married males in the 40 to 44 age group is read from the stored table; this probability, denoted by $P_{mm}(11)$, refers to the 11-th age group for married males. A uniform random number r between zero and one is generated and the quantity

$$t = \frac{\ln(1-r)}{\ln(1-P_{mm}(11))}$$

is computed. This is the randomly generated number of months (from the current month) until the death of the 42 year old male, provided $t \leq 36$. If $t > 36$, then the generated age at death of this man is greater than 45, in which case a new r and t must be generated using $P_{mm}(12)$, the monthly probability of dying for married males in the 45-49 age group. The month of death will have been determined if the new $t \leq 60$. Otherwise, the process is repeated again, using the appropriate monthly death probability, $P_{mm}(13)$, for the next age group (50-54). The process is continued until a time interval to death which does not extend into the next age group is obtained for this married male. The procedure is equivalent to using a sequence of independent, uniform (0, 1) random numbers and making a decision by comparing with the appropriate probability for each month. During intervals having constant probabilities, this assumes in effect, a Poisson process for the event, or a negative exponential distribution for the intervals between events.

An important point to note is that it is assumed that no other events happen to the individual in the interval. For example, it is assumed that the marriage of the 42 year old male does not end, due to divorce or death of his spouse, when calculating the time interval to his death. However, he is subject to having these competing events occur. To determine whether he becomes widowed before he dies, POPSIM computes the date of death of his wife independently, using the monthly death probability for married females of her age class, and then compares the two death dates to determine which of these competing events is to occur first. Since divorce is also a possibility, the date of this event must be computed and compared with the two death dates to determine the next event for this 42 year old married male. Again, the time interval for each of the competing events is computed under the assumption that the other events do not occur.

For a female between the ages of 15 and 45, a competing risk is that of giving birth to a child. Monthly probabilities of a live birth by age group, parity and marital status are used to generate the time interval to the next live birth.

After the point in time for the next event is generated, it is checked to see if it falls within the interval chosen for that simulation step. If it is, time is advanced to that point and the event processed. If not, the individual is stored and not processed again until the beginning of the next simulation interval. Processing consists of recording the essential facts concerning the event and changing the status of individual characteristics affected by the event.

If the event happens to be marriage, some further processing is required. First, a decision must be made with respect to the marital status (single, widowed or divorced) of the partner. This choice is determined by means of 3 x 3 arrays of probabilities for six age groups (age of the primary individual involved). Once this has been done, the age of the marriage partner is obtained from the appropriate (first marriage or remarriage) bivariate distribution of ages of brides and grooms.

The remaining event which requires special treatment is the birth of a child. The date of birth and sex of the child is recorded in the list of secondary children associated with the mother. The sex of the child is determined by using a random number. The parity and count of the number of living children are both updated. Finally a special note is made of the birth in order to supply a sample of births (primary individuals) for the population.

When the event has been processed, a new next event is generated for the updated individual. This is continued until finally an event is obtained which is beyond the time allotted for the step in the simulation or the individual dies.

POPSIM utilizes the series of births which occur to primary females as a random sample of births to add to the computer population. Thus, whenever a birth occurs to a primary female, the infant is entered as a member of the set of her associated children (secondary individuals), and the program also keeps a separate record of all these births. Then, after the female in question has been advanced beyond the time period specified for the simulation step and has been placed in a storage file, the program returns to the list of births and establishes a RECORD for a new primary individual for each birth. Each of these new primary individuals is immediately subjected to the various risks from the date of birth to the end of the simulation step, before being placed in the storage file. In subsequent steps of the simulation, these individuals are not distinguishable from other individuals who were generated for the initial sample population. It is important to note that no contact is maintained between the mother and her child, except as a secondary individual. Future events for the child as a primary individual are generated completely independently of those generated for the mother through the presence of the child as a secondary individual in her RECORD.

Three successive arrays of information are retained in the population file for each individual, when applicable. The first array is the basic set of characteristics given in Figure 1. The second array contains the date of birth, sex and date of death for each of the secondary children associated with a primary female. The third array contains the history of events which have occurred to the individual during the simulation. The entries which are possible in the history of events array are shown in Figure 2. It is significant to note that the population or history file produced in any simulation run can be used as an initial population file in a subsequent run.

A series of tables are printed at the end of each step of the simulation. Those which have been retained for use with the hospital utilization model include:

1. The distribution of the population at the end of the interval by age, sex and marital status.
2. The distribution of the deaths in the population (primary individuals only) by age (at death), sex and marital status.
3. The distribution of births during the interval by age of mother (at the time of the birth) and the marital status of the mother.
4. The distribution of marriages by age, sex and marital status of the primary individual involved.

The data in the population file at the end of a simulation period essentially constitutes a set of vital event histories for each individual in the computer population. The data in the file (stored on a disk or on magnetic tape) are equivalent to data collected from a sample of individuals in a longitudinal survey. This history feature of the POPSIM output is very useful for special tabulation and analysis of the simulation data.

B. Some Modifications of POPSIM

1. Race

The original version of the POPSIM computer program was not designed to distinguish between segments of the population, such as ethnic or racial groups, in a single simulation run. The program has been modified so that two segments or subpopulations can be processed simultaneously and, at the user's option, either separate or combined tables are printed in the standard tabular output. For the hospital utilization simulation the white and non-white subpopulations are recognized. Separate parameters for births, death and marriages are required for POPSIM.

2. Residence

POPSIM has also been modified to classify each individual in the computer population according to whether he resides in a metropolitan area (i.e. a Standard Metropolitan Statistical Area or SMSA) or in a non-metropolitan area (non-SMSA). The residence assignment is made by age, race (white and non-white) and sex on a stochastic basis by the program which creates the initial program. Further, the program which advances the population through time determines those changes of residence which occur between an SMSA and a non-SMSA in either direction. Changes of residence are generated as events using monthly transitional probabilities on a competitive risk basis and in the same manner as the other eligible events are generated. The propensity to move is based on the age, race, sex and present residence of the individual. The event "change of residence" has been added to the history of events array, recording the current residence status, the age of the individual at the time of the move as well as the date of the event.

3. Family Income

Each individual in the initial population is assigned a family income according to his race, residence, type of household and age of the head of household. Four types of households to which individuals over the age of 18 and also ever married individuals 18 and under can belong have

Figure 2

Detailed Illustration of Possible Entries in the History of Events Array

Type of event	Code for event	Date of event	Marital status of primary individual	Descriptive information
Birth of a child	1	date in months	mother's marital status	interval since previous birth
Divorce	2	date in months	married	length of marriage in months
Death	3	date in months	marital status at death	age at death in months
Marriage	4	date in months	marital status prior to marriage	interval since last change in marital status
Widowhood	5	date in months	married	interval since marriage
Death of a child	6	date in months	(not assigned)	(not assigned)

been defined for purposes of determining family income. They are:

a. If an individual is married, that individual belongs to a "husband and wife" household.
b. If a woman is divorced or widowed and has children under the age of 18, she belongs to a "female head" household.
c. If a woman is widowed or divorced with no children under 18 or is single (never-married) and over the age of 18, she belongs to an "unrelated female" household.
d. If a man is widowed or divorced or is single and over the age of 18 he belongs to an "unrelated male" household.

The model assumes that for a given race, residence, type of household and age of head (19-34, 35-44, 45-64, 65 and over) the log-normal distribution adequately represents the distribution of family income. It is further assumed that the mean family income can change over time, but the variance is not assumed to vary with time, an assumption for which there is empirical support and which is consistent with the assumption that the distribution is log-normal. A computer routine is used by the Phase II program to update family income (consistent with the initial family income) whenever it is called and also at the end of each simulation step. The updated family income and the time of the last update are retained in the individuals RECORD (items 19 and 20).

III. THE HOSPITAL EPISODES MODEL (HOSPEP)

A. Introduction

In two previous studies [3, 14], a probability model for generating hospital admissions and durations of stay for the U.S. civilian non-institutional population, based primarily on age and sex, was developed and refined. Using this model, a computer program was written to simulate hospital episodes for each individual in a given population. In the present study the earlier probability model has been extended to include a more detailed hospitalization history than just admissions and duration of stay and to take into account more demographic characteristics of each individual than just his age and sex. In addition, a computer program has been written to simulate hospital episodes using this extended model.

Specifically, the present study has extended the earlier hospital episodes model in the following ways:

1. For each individual in a given population the model will generate in addition to hospital admissions and duration of stay, the reason for hospitalization (diagnosis), whether or not surgery is performed and the bedsize of the hospital the individual is admitted to.
2. The model considers not only the age and sex of each individual in the population but also his race, family income, hospital insurance status and whether

his residence is located in a metropolitan area (SMSA) or in a non-metropolitan area (non-SMSA).

The computer program for the extended hospital episodes model has been set up to run either with input data from a POPSIM history file which gives the demographic characteristics of a simulated population or with data similar to that which POPSIM generates. The output of the program gives detailed tables on the hospital utilization of the simulated population and also a hospital utilization data tape that can be read as input into a hospital manpower model computer program (cf. [1]).

B. Summary of Earlier Model

A detailed description of the original probability model for hospital episodes is given in [3]. This model was developed to generate hospital admissions and durations of stay for a given population. The hospital admissions model assumes that the number of admissions annually, X , for an individual in a particular age-sex class follows the Poisson distribution. That is,

$$f(X|\lambda) = \frac{e^{-\lambda t}(\lambda t)^X}{X!} \quad X = 0,1,2,\ldots \qquad (1)$$

where the parameter λ is the expected number of hospital episodes during time t for the individual. It also assumes that λ varies over the population and has a Gamma distribution. That is, the distribution of λ for all individuals in the population in a particular age-sex class is assumed to be

$$g(\lambda) = \frac{\beta}{\Gamma(\alpha)} (\beta\lambda)^{\alpha-1}e^{-\beta\lambda}, \alpha>0,\ \beta>0 \ . \qquad (2)$$

It follows that

$$f(X) = \sum_{\lambda} f(X|\lambda)g(\lambda)d\lambda$$
$$= \binom{\alpha+X-1}{X}\left(\frac{\beta}{1+\beta}\right)^{\alpha} \left(\frac{1}{1+\beta}\right)^{X}, \ X=0,1,2,\ldots \qquad (3)$$

which is the Negative Binomial distribution. Thus, the model assumes that the Negative Binomial distribution provides a reasonable description of observed data on number of hospital episodes annually for a given age-sex group. The duration of stay model assumes that the distribution of the number of days spent in the hospital by an individual is distributed as a log-normal variate with parameters depending on the age-sex class of the individual.

C. Modification of Earlier Version of HOSPEP

There are six basic steps in the extended model for generating the hospital episodes for each individual during a simulation period. They are:

Step 1. Assign hospital insurance status to the individual depending on his race, age and family income.

Step 2. Generate the next date of admission to a hospital depending on his race, age, sex, family income and hospital insurance status. If this date is beyond the end of the simulation period, the next individual is processed beginning with Step 1.

Step 3. Assign a diagnosis for the generated admission depending on his age and sex.

Step 4. Determine whether surgery was performed or not depending on the age, hospital insurance status, diagnosis and residence of the individual.

Step 5. Assign the bedsize of the hospital depending on the diagnosis and surgery status of the individual.

Step 6. Generate the date of discharge of the individual from the hospital depending on his age, family income, hospital insurance status, diagnosis, surgery status and size of hospital. If this date is still within the simulation period, the program returns to Step 2 and continues to process the same individual. If the generated date of discharge is beyond the end of the simulation period, the next individual is processed beginning with Step 1.

In addition to these steps the HOSPEP program has special subroutines for delivery episodes and for generating the hospital episodes for persons in their last year of life.

Since the particular factors which explain the variation in the conditional probabilities appropriate to the occurrence of each event or to the distribution of each characteristic to be assigned in the six steps of the model were not known, considerable statistical analysis of available data was necessary before the appropriate probability distributions to use in the model could be specified. The factors listed above for each step were derived from a larger set of factors in each instance.

Although data exist from which the needed joint distributions of the dependent and independent factors could have been computed (e.g. the distribution of number of hospital episodes annually by age, race, sex, family income, residence and hospital insurance status), but these data could not be obtained. Since tabulations of these factors were available only in two, three, four and five-way tables, it was necessary to devise a special procedure whereby the critical factors for each step in the model could be determined.

The special procedure consisted of using iterative proportional fitting to construct a k-dimensional contingency table from a set of lower order marginal tables involving k factors [15]. This procedure essentially disaggregates the data in the lower dimensional tables into the single higher dimensional table in a manner which preserves all of the information on main effects and interactions among the factors that are present in the same lower order tables. Further, any higher order interaction which is missing from all of the lower dimensional tables is not introduced into the k-dimensional table by this procedure. It is of interest to note that the marginal tables from several different studies can be used to derive the requisite k-dimensional table. A general computer program, TABLES, has been written to carry out the iterative procedure for constructing a k-dimensional contingency table from a set of tables of lower dimension.

Using data from the Health Interview Survey and the Hospital Discharge Survey, both conducted by the National Center for Health Statistics, and from the Bureau of the Census Current Population Survey, the program TABLES was used to estimate a higher order contingency table for each of the six basic steps. Analyses of variance on the natural logarithms of the cell frequencies in each table were then carried out to determine the factors which best determined the conditional probabilities for each step.

D. Generation of Hospital Episodes

The computer program for the extended HOSPEP model generates hospital episode histories for each individual on the input population file for a specified simulation period. If the population input is from POPSIM, a HOSPEP edit routine selects the appropriate data for each individual pertinent to the time period of interest, carefully checking for changes in age class, family income and residence during the time period. If there are any changes, the simulation period is broken up into subperiods for which the relevant characteristics of the individual remain the same. This is essential for proper use of the conditional probabilities which generate the hospital admissions and characteristics for each admission.

The basic input data required by HOSPEP, other than the information in the population file for each individual to be processed, includes:

1. Conditional probabilities of having hospital insurance by age, race and family income.
2. Parameters to determine hospital admission dates by race, age, family income and hospital insurance status.
3. Conditional probability distribution of diagnosis by age and sex.
4. Conditional probabilities of surgery status by age, hospital insurance status, diagnosis and residence.
5. Conditional probability distribution of bedsize of hospital by diagnosis and surgery status.
6. Parameters to determine length of stay in the hospital based on age, surgery, diagnosis, family income, hospital insurance status, and size of hospital.
7. Daily probabilities for hospital admission for persons in their last year of life.
8. Parameters to update family income by age, race, residence and type of household.

The categories used for each of the variables in the tables of parameters and conditional probabilities are shown below:

<u>Race</u>
White
Non-White

<u>Sex</u>
Male
Female

<u>Age Group</u>
1 <15 years
2 15-44 years
3 45-64 years
4 65+ years

<u>Family Income Group</u>
1 < $4,000
2 $4,000-$6,999
3 $7,000-$9,999
4 $10,000+

<u>Hospital Insurance Status</u>
Insured
Not Insured

<u>Residence</u>
SMSA
Non-SMSA

<u>Surgery</u>
Surgery
Nonsurgery

<u>Size of Hospital (# beds)</u>
1 < 50
2 50-74
3 75-99
4 100-149
5 150-199
6 200-299
7 300-399
8 >400

<u>Diagnosis</u>
1 Infective and parasitic diseases
2 Malignant neoplasms
3 Benign neoplasms and neoplasms of unspecified nature
4 Allergic, endocrine system, metabolic and nutritional diseases
5 Diseases of blood and blood-forming organs
6 Mental, psychoneurotic and personality disorders
7 Diseases of the nervous system and sense organs
8 Diseases of the circulatory system
9 Diseases of the respiratory system
10 Diseases of the digestive system
11 Diseases of the genitourinary system
12 Deliveries and complications of pregnancy, childbirth and puerperium
13 Diseases of the skin and cellular tissue
14 Diseases of the bones and organs of movement
15 Congenital malformations
16 Certain diseases of early infancy
17 Symptoms, senility, ill-defined conditions and special admissions
18 Injuries and adverse effects of chemical and other external causes

Each individual is processed by HOSPEP for the entire simulation period, one at a time. Hospital insurance status, diagnosis, surgery and size of hospital are assigned by generating a uniform random number (0,1) and comparing it with the appropriate conditional probability distribution.

The earlier model for hospital admissions has been simplified slightly by setting the parameter $\alpha = 1$ (Eq. 2 above). Thus the original Gamma distribution of the parameter λ (which reflects the variation among individuals of the same age, race, family income and hospital insurance status in their likelihood of having to be hospitalized) reduces to a Negative Exponential Distribution with parameter β where $1/\beta$ is the expected number of hospital episodes annually per person for the sub-population. In order to determine the next date of admission, HOSPEP first derives the daily admission probability P appropriate to the individual by computing

$$P = \frac{-\ell n\ R}{365\ \beta}$$

where $\ell n\ R$ is the natural logarithm of a uniform random number (0,1). The next date of admission is then generated by sampling the geometric distribution with parameter P . In order to derive the length of stay d for each admission, it is necessary first to compute the mean and standard deviation of the appropriate normal distribution for log d . For this purpose, linear additive models of the significant effects and interactions of the various levels of age, family income, hospital insurance status, diagnosis, surgery status and size of hospital on the means and standard deviations of the distributions of log d were derived (by analysis of variance) from the k-dimensional contingency table of these factors generated by the program TABLES. A random normal deviate, say z , is then generated and the corresponding random variable log d computed.

As each individual is processed, the HOSPEP program tabulates a series of tables. These tables are printed after all individuals have been processed, including:

1. The distribution of the average population for the simulation period by age, race, sex, family income, and hospital insurance status.
2. The distribution of hospital discharges by age (<15, 15-44, 45-64, 65 or older), race, sex and days in hospital (1, 2-4, 5 or more).
3. The distribution of hospital discharges by age, family income, hospital insurance status and days in hospital.
4. The distribution of hospital discharges by age, sex, diagnosis, surgery status and days in hospital.
5. The distribution of hospital discharges by diagnosis, surgery status, and days in hospital for each hospital bedsize class.
6. The distribution of the sample population by number of hospital admissions (0, 1, 2 or more), age, race and sex.
7. The distribution of the sample population by number of hospital admissions, age, family income and insurance status.
8. Hospital admission rates per 1000 persons by age, race and sex.
9. Hospital days rates per 1000 persons by age, race and sex.
10. Hospital days by age, race and sex.
11. Hospital days by age, family income and insurance status.
12. Hospital days by age, sex, diagnosis, and surgery status.
13. Hospital days by size of hospital, diagnosis, and surgery status.
14. Average length of stay by age, race and sex.

At present no record is retained at the end of the simulation period of the hospital episodes experienced by each individual in the population. It is recognized that such a record could be quite useful for further tabulations of the simulation results and hence provision for including this feature is now underway. The HOSPEP program does prepare an output data file for use as input to a Hospital Service Model designed to derive hospital requirements from hospital utilization data [1]. The information on this file is restricted to hospital episodes only and includes the admission and discharge dates, diagnosis, surgery status and size of hospital.

Parameters for POPSIM and HOSPEP for simulation of hospital utilization for the U.S. civilian non-institution populations have been developed for purposes of testing the model.

ACKNOWLEDGEMENTS

The authors appreciate the many useful discussions with Mr. Garrie J. Losee of the Bureau of Health Professions Education and Manpower Training, NIH. A number of other individuals have contributed important ideas to the development of POPSIM and HOSPEP. We wish to acknowledge specifically Dr. F. G. Giesbrecht, Dr. B. V. Shah and Mr. P. V. Piserchia.

REFERENCES

[1] F. D. Kennedy, Simulation of Health Manpower Requirements in a Hospital Setting, presented at the Conference on a Health Manpower Simulation Model, National Institutes of Health, Washington, August 1970.

[2] D. G. Horvitz, F. G. Giesbrecht, B. V. Shah, and P. A. Lachenbruch, POPSIM, A Demographic Micro-Simulation Model. Paper presented at the meetings of the International Union for the Scientific Study of Population, London, September 1969.

[3] National Center for Health Statistics: Computer Simulation of Hospital Discharges, Vital and Health Statistics, PHS Pub. No. 1000-Series 2-No. 13, Public Health Service, Washington, U. S. Government Printing Office, February 1966.

[4] B. V. Shah, On the Mathematics of Population Simulation Models, presented at the annual meetings of the Population Association of America, Atlanta, April 1970.

[5] B. V. Shah and F. G. Giesbrecht (1969), Mathematical Models for Changing Marriage Patterns. Paper presented at the meetings of the International Union for the Scientific Study of Population, London, September 1969.

[6] J. C. Barrett, A Monte Carlo Study of Reproduction. Presented at the Society for Human Biology Symposium, London, England, 1967.

[7] H. Hyrenius and I. Adolffson, A Fertility Simulation Model. Report #2, Demographic Institute, University of Göteborg, Sweden, 1964.

[8] H. Hyrenius, I. Holmberg, and M. Carlsson, Demographic Models: DM 3. Report #5. Demographic Institute, University of Göteborg, Sweden, 1967.

[9] I. Holmberg, Demographic Models: DM 4. Report #8, Demographic Institute, University of Göteborg, Sweden, 1968.

[10] G. H. Orcutt, M. Greenberg, J. Korbel, and A. Rivlin, Micro-analysis of Socioeconomic Systems. Harper, New York, 1961.

[11] J. C. Ridley and M. C. Sheps, An Analytic Simulation Model for Human Reproduction with Demographic and Biological Components. Population Studies, Vol. XIX, pp. 297-310, 1966.

[12] F. G. Giesbrecht, G. Ranney and J. R. Chromy, Demographic Microsimulation Model POPSIM I: Manual for Programs to Generate the Initial Population, Open Core Model. Technical Report #4, Project SU-285, Research Triangle Institute, September 1969.

[13] F. G. Giesbrecht and L. Field, Demographic Microsimulation Model POPSIM II: Manual for Programs to Generate Vital Events, Open Core Model. Technical Report No. 5, Project SU-285, Research Triangle Institute, September 1969.

[14] F. Giesbrecht, Further Development of Computer Simulation of NCHS Hospital Data. Final Report, Project SU-204, Research Triangle Institute, September 1966.

[15] B. V. Shah, On Predicting k-Dimensional Contingency Tables from Lower Dimensional Marginal Tables, presented at the Conference on a Health Manpower Simulation Model, National Institutes of Health, Washington, August 1970.