

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/13038301>

An Iterative Estimation and Validation Procedure for Specification of Semi-Markov Models with Application to Hospital...

Article in *Operations Research* · December 1982

DOI: 10.1287/opre.30.6.1082 · Source: PubMed

CITATIONS

42

READS

78

3 authors:



[Elliott Weiss](#)

University of Virginia

112 PUBLICATIONS 764 CITATIONS

[SEE PROFILE](#)



[Morris A. Cohen](#)

University of Pennsylvania

102 PUBLICATIONS 5,766 CITATIONS

[SEE PROFILE](#)



[John C Hershey](#)

University of Pennsylvania

101 PUBLICATIONS 5,015 CITATIONS

[SEE PROFILE](#)

An Iterative Estimation and Validation Procedure for Specification of Semi-Markov Models
with Application to Hospital Patient Flow

Author(s): Elliott N. Weiss, Morris A. Cohen and John C. Hershey

Reviewed work(s):

Source: *Operations Research*, Vol. 30, No. 6 (Nov. – Dec., 1982), pp. 1082-1104

Published by: [INFORMS](#)

Stable URL: <http://www.jstor.org/stable/170201>

Accessed: 09/08/2012 11:36

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



INFORMS is collaborating with JSTOR to digitize, preserve and extend access to *Operations Research*.

<http://www.jstor.org>

An Iterative Estimation and Validation Procedure for Specification of Semi-Markov Models with Application to Hospital Patient Flow

ELLIOTT N. WEISS

Cornell University, Ithaca, New York

MORRIS A. COHEN and JOHN C. HERSHEY

University of Pennsylvania, Philadelphia, Pennsylvania

(Received May 1980; accepted July 1982)

This article presents a methodology to identify and specify a continuous time semi-Markov model of population flow within a network of service facilities. An iterative procedure of state space definition, population disaggregation, and parameter estimation leads to the specification of a model which satisfies the underlying semi-Markov assumptions. We also present a test of the impact of occupancy upon realizations of population flows. The procedure is applied to data describing the movements of obstetric patients in a large university teaching hospital. We use the model to predict length-of-stay distributions. Finally, we compare these results with those that would have been obtained without the procedure, and show the modified model to be superior.

SEMI-MARKOV MODELS of customer flows through a network of service facilities are particularly useful due to their ability to provide a rich model structure within a relatively simple context. The models capture two types of randomness; both the next destination of the individual and the length of stay in any facility are treated as random variables. They also permit the detailed analysis of the resources utilized in delivering service to a population. These models have been applied to diverse contexts such as the flow of patients within a hospital (Hershey et al. [1981]), the movement of orders through a job shop (Jackson [1957]), and the transmission of information within a computer network (Kleinrock [1976]). Semi-Markov models are an improvement over single facility queueing models (e.g. Shonick and Jackson [1973]) which analyze only one stage of service or aggregate all stages into one stage.

The purpose of this article is to present a methodology that can be used to develop a continuous time semi-Markov model of customer flows. This methodology uses an iterative scheme to identify model structure,

Subject classification: 271 hospital patient flow model, 567 estimation and validation procedures, 799 semi-Markov model specification.

estimate model parameters, validate the model assumptions, and specify a final model where the *a priori* hypothesis is that an underlying semi-Markov process describes the system. In particular, our emphasis is on providing a systematic approach to an often neglected portion of this model development process, namely statistical validation of the underlying assumptions. After applying this methodology and obtaining a final model, one should ideally test the model against an independent data base. If the model is confirmed, it can subsequently be used to analyze resource allocation decisions.

Our application is to the flow of patients within a hospital. We discuss the results of using the specification methodology for data collected from an obstetric service. Since the analysis was carried out in the context of this health care system, henceforth we will discuss the procedure and results in terms of the specific case of the patient flow model. It is important to note that the methodology can be adapted to other application areas as well.

Health care delivery within a hospital facility consists of the provision of a sequence of services to each patient as his/her health condition and medical requirements change over time. In "progressive patient care facilities," such as a hospital's maternity section or coronary care facility, patients are assigned to units of the facility according to their care requirements, and are reassigned as their care requirements change.

A number of semi-Markov models have been developed to describe the stochastic process of patient movements in order to analyze bed planning decisions in progressive patient care facilities. Kao [1972, 1973, 1974], and Hershey et al. examine such resource allocation decisions via analytic methods, while Cohen et al. [1980] simulate the semi-Markov model. The semi-Markov models are useful for both personnel scheduling and capital investment planning within hospitals and are important because of their potential for identifying relationships among capacities (number of beds) of the units providing the health services, quality of delivered services, and utilization of the units. For example, while a reduction in capacity of the maternity ward can lead to a higher level of average utilization, it can also increase the likelihood that this unit is fully occupied. An increase in the capacity of the unit will have the opposite effect. Variations in patient mix caused by changing demographics or alternative treatment patterns can also be modeled using the semi-Markov model. The effect of these changes on resource utilization can be predicted by varying the appropriate input parameters, i.e., patient arrival rates. In cases where patient types are not homogeneous with respect to treatment patterns, these changes may be significant.

A key assumption of semi-Markov models is that the probability of transfer to each of the units in the network of facilities or to absorbing states outside of the network depends only on the identity of the currently

occupied unit. Similarly, it is assumed that the length of stay in each unit is dependent only on the current location and next destination. The model is therefore specified by the transition probabilities among locations and the parameters of the particular length of stay distribution for each possible transfer. These inputs are appropriate under "ideal" conditions where unit capacities do not serve as limiting constraints on the flow of individuals within the network so that individuals' "paths" are assumed to be independent. The models reported to date all require input of the ideal model parameters. When capacity is limiting, ideal transfers and lengths of stay may not actually be realized since desired transfers may be blocked. Cohen et al. have investigated policies and procedures for redirecting or bumping individuals which require estimates of the ideal parameters. Hershey et al. present a technique for recovering data affected by these blocking policies.

As pointed out by Fetter and Thompson [1969], changes in a patient's state of health may not be Markovian for many diseases, but rather may depend on the patient's entire past history (the sequence of locations occupied by the patient and the treatment delivered to date). Thomas [1968] developed a Markov chain model that was adequate to model certain summary measures of the movement of coronary patients through a coronary care facility. However, this model did not actually describe the stochastic flows of patients in such systems in terms of a semi-Markov process. Kao [1974], expanding Thomas' work, also showed that a semi-Markov model could generate summary measures that were similar to observations from a coronary care facility. These summary measures were the distribution of total length of stay and probabilities of successive patient state locations predicted by the imbedded Markov chain. However, like Thomas, Kao did not explicitly demonstrate that the state-to-state flows of patients generated by the semi-Markov process were statistically equivalent to the actual flows. In particular, neither Thomas nor Kao tested whether transition probabilities or holding time distributions were independent of previous locations of patients.

Shachtman et al. [1981] tested the goodness of fit of a Markov chain model for fecund (capable of becoming pregnant) women. They validated the homogeneity of transition matrices for different cohorts of women as well as the independence of transitions and previous states. Kastner et al. [1979], and Kastner and Shachtman [1981] used a Markov chain model to analyze the incidence of nosocomial (hospital acquired) infections. Aside from these studies, there have been no detailed validations of Markovian models applied to health care analysis. When models are to be used for planning and resource allocation purposes, it is clear that such validation is required and must go beyond tests of aggregate output statistics.

This article is organized as follows. In the next section we present the formulation of a semi-Markov model for patient flow within a hospital facility. Section 2 is an overview of the analysis, while Section 3 is a detailed presentation of the various estimation and validation steps. In this latter section the specifics of the various statistical tests employed are outlined. The data collected from a large teaching hospital's obstetric service to test our procedure are discussed in Section 4. In Section 5, we present the results of applying the methodology. Although we do not confirm the final, revised model using a separate data base, we discuss in Section 6 the improvement gained from using the revised model to estimate the underlying length of stay distribution. Conclusions are presented in the final section.

1. THE PATIENT FLOW MODEL

Consider a progressive patient care hospital facility composed of m units labeled $1, \dots, m$, each of which has finite capacity. In addition, "units" $m + 1, \dots, r$ are absorbing states external to the hospital (e.g., home or death). Let

- $p_{ij}^{(l)}$ = the probability that a patient of type l will require a transfer to unit j for the next transfer given that the patient is in unit i , i.e. the transition probability;
- $\tau_{ij}^{(l)}$ = the random variable denoting the time a patient of type l should stay in unit i before the transfer to unit j is called for, i.e. the holding time;
- $h_{ij}^{(l)}(\cdot)$ = the holding time probability distribution.

The $r \times r$ transition probability matrix is

$$P^{(l)} = \begin{bmatrix} m & r-m \\ \hline Q^{(l)} & S^{(l)} \\ \hline 0 & I \end{bmatrix} \begin{matrix} m \\ r-m \end{matrix} \quad l = 1, \dots, s \quad (1)$$

where the submatrix 0 consists of zeros, the submatrix Q describes the transfers of patients within the hospital, the submatrix S describes the transfers of patients from within the hospital to the absorbing states outside the hospital, and I is the identity matrix. The number of subpopulations is denoted by s . The subscripts denoting all of the states are assumed to be arranged appropriately. Reentry to the hospital is considered to be a new admission. Assumptions implicit in this formulation are described in the next section. This model has been used previously by Kao [1974], Cohen et al., and Hershey et al.

2. ANALYSIS OVERVIEW

Figure 1 summarizes an iterative procedure for the specification of the semi-Markov model. Hypotheses regarding the movements of patients and the semi-Markov assumptions are first tested on the initial definitions of patient types and of state space. Depending on the results, the patient type and state space definitions may be modified. The hypotheses are then retested on the modified model. This iterative procedure continues until a model is found that satisfies the semi-Markov assumptions and predicts the actual movement of patients. The final model is one which is not inconsistent with an underlying semi-Markov process. This final model is also not unique, in that many definitions of state spaces and subpopulations that are consistent with model assumptions are possible.

All patients visiting the facility can be described in terms of a number of factors (e.g., age, socioeconomic status, initial diagnosis, past medical history, etc.). The first stage in the procedure determines if the initial classes of patients should be either combined or further subdivided. Two obvious criteria for initial specification of subpopulations, i.e., patient type, are medical procedure and level of a priori medical risk, since transition probabilities and holding times may depend upon a patient's classification with respect to these two criteria. If so, it is important to specify different subpopulations because of the ultimate use of the model for policy analysis. For example, the analyst may desire to use the model to assess the impact of changes in patient mix on the utilization of units of the facility. Subpopulations exhibiting different flows must therefore be identified along with their current and projected relative size in order to measure expected resource use.

The second testing stage examines the Markov assumptions. Recall that transfers should depend only upon the current location of the patient and that length of stay should depend only upon the current location and the next destination of the patient. Thus, transfers and holding times must be shown to be conditionally independent of previously visited locations given the current location. In general, when the state definitions in a particular specification of the model do not yield the necessary independence, certain states can be subdivided to incorporate past history. The redefinition of the state space may then satisfy the Markov assumptions. An alternative approach to satisfying the Markov assumptions would be to test a complete second-order state space, i.e., one where all states were further specified by the previous location and transfers and holding times were assumed to be conditionally independent of previously visited locations given the current and previous location. The approach adopted here will generally result in a more parsimonious state space.

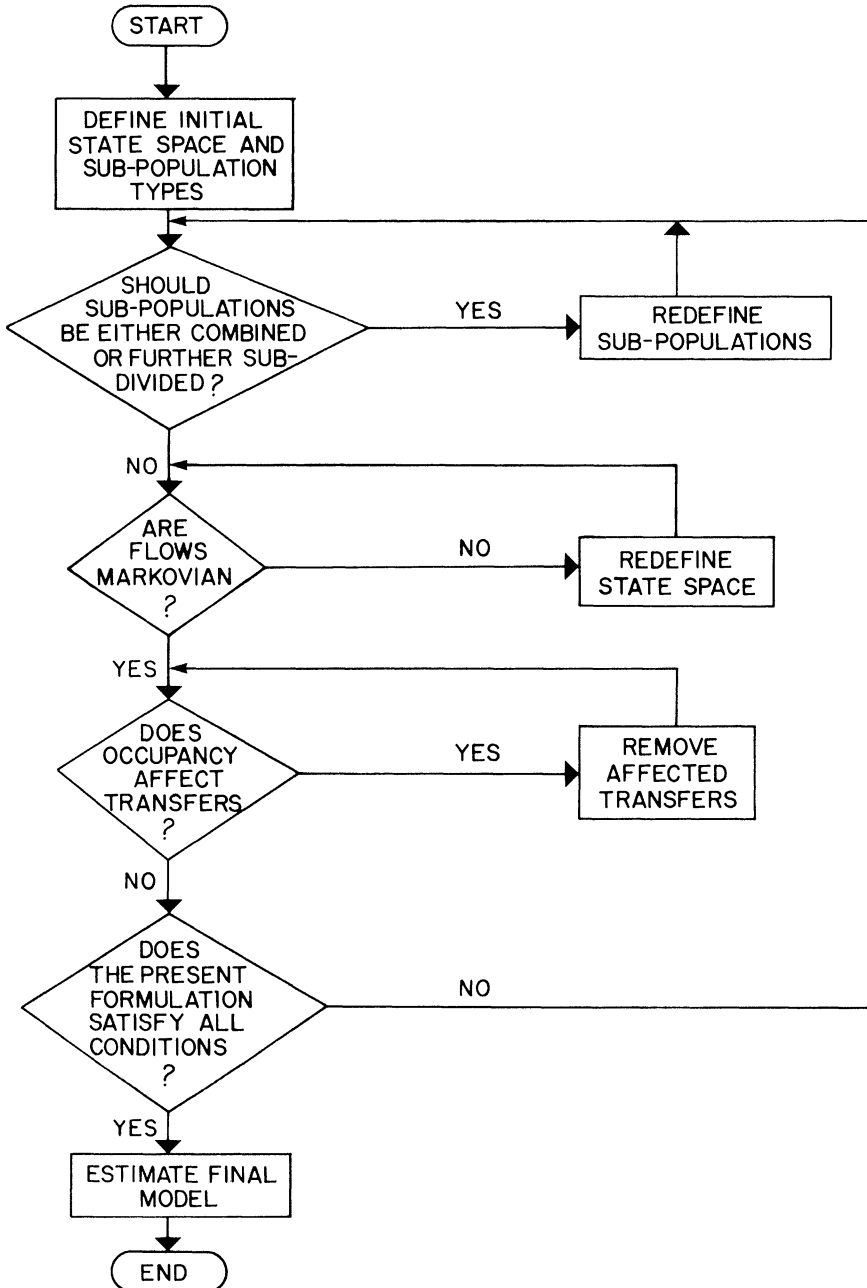


Figure 1. Specification procedure overview.

The semi-Markov model requires estimates of parameters of “ideal” transfers and lengths of stay in the units, i.e., the transfers and lengths of stay that would be observed if patient movements were unaffected by capacity constraints. In gathering actual data, some of the observed transfers may deviate from the ideal movements due to the interactions of patient flows caused by full or nearly full units (see Hershey et al.). At the third testing stage, we investigate whether transition probabilities or holding times are related to occupancy rates. In cases where high occupancy rates appear to affect the observed parameters, we recommend removal of data from high occupancy periods for estimation of the ideal model (as in Hershey et al.).

The nature of this sequential procedure is such that any modification of subpopulation or state-space definitions or any removal of occupancy-affected transfers may result in the invalidation of previously validated assumptions. We therefore recommend iteration through the entire procedure until the final formulation satisfies all three test stages simultaneously without requiring further redefinitions of subpopulations or state spaces, and without requiring further removal of occupancy-related transfers. Of course, any such final formulation is not unique, since it is possible that many formulations could pass all tests. Alternatively, it is possible that this iterative process will not converge with respect to a given level of significance. Nevertheless, if convergence is achieved, the formulation developed by this procedure will have been based on an initial formulation which came, *a priori*, from expert medical opinion and detailed observation of the actual system. The final formulation will satisfy the semi-Markov model requirements and will meet the objective of providing semi-Markov model parameters which can legitimately be used in subsequent policy analysis.

3. DETAILS OF THE SPECIFICATION METHODOLOGY

Table I summarizes the procedures employed in the first three testing stages. We now discuss this table in detail.

Should Subpopulations Be Either Combined or Further Subdivided?

The first tests concern the grouping of subpopulations. These tests must be conducted on two levels—the transition probability matrices and the holding time distributions.

Maximum likelihood estimates of the transition probabilities can be determined as follows (Anderson and Goodman [1957]):

$$p_{ij}^{(l)} = n_{ij}^{(l)} / n_i^{(l)}; \quad i = 1, \dots, m; \quad j = 1, \dots, r; \quad l = 1, \dots, s;$$

TABLE I
SPECIFICATION PROCEDURES

Validation Stage	Statistical Question	Procedure	
		Initial estimation	Statistical test
A. Should subpopulations be either combined or further subdivided?	Are transition probabilities drawn from the same distributions?	Maximum likelihood estimates of transition probabilities (Anderson and Goodman)	Likelihood ratio test of the equality of transition probabilities (Anderson and Goodman)
	Are holding times drawn from the same distributions?	Not necessary; statistical tests use the empirical holding time distributions	Kolmogorov-Smirnov two-sample test (Conover) -or- Kruskal-Wallis test (more than two samples) (Conover, Siegel) -or- Likelihood ratio test (discrete data)
B. Are flows Markovian?	Are transition probabilities independent of previous states?	Maximum likelihood estimates of transition probabilities (Anderson and Goodman)	Likelihood ratio test of the independence of transition probabilities and previous states
	Are holding time distributions independent of previous states?	Not necessary; statistical tests use the empirical holding time distributions	Kolmogorov-Smirnov test, Kruskal-Wallis test, or likelihood ratio test
C. Does occupancy affect transfers?	Are transition probabilities independent of occupancy rates?	Maximum likelihood estimates of transition probabilities (Anderson and Goodman)	Likelihood ratio test of the equality of transition probabilities
	Are holding times independent of occupancy rates?	Not necessary; statistical tests use the empirical holding time distributions	Kolmogorov-Smirnov test, Kruskal-Wallis test, or likelihood ratio test

where

- $p_{ij}^{(l)}$ = the estimated probability of transferring from unit i to unit j for a patient in subpopulation l ,
- $n_{ij}^{(l)}$ = the number of patients in subpopulation l observed transferring from unit i to unit j ,
- $n_{i\cdot}^{(l)}$ = the total number of patients in subpopulation l transferring from unit i .

After estimating the transition matrices for all subpopulations, one can test whether they differ. Anderson and Goodman present a test to determine the equality of transition matrices. The statistic

$$\chi^2 = \sum_{l=1}^s \sum_{j=1}^r \sum_{i=1}^m n_{i\cdot}^{(l)} (p_{ij}^{(l)} - p_{ij}^{(\cdot)})^2 / p_{ij}^{(\cdot)}$$

where

- $p_{ij}^{(\cdot)} = n_{ij}^{(\cdot)} / n_{i\cdot}^{(\cdot)}$,
- $n_{ij}^{(\cdot)}$ = the total number of patients of all types who transfer from unit i to unit j ,
- $n_{i\cdot}^{(\cdot)}$ = the total number of patients of all types who leave unit i ,

asymptotically follows the chi-squared distribution with $((s - 1) \times (m - 1) \times (r - 1) - Z)$ degrees of freedom, where Z is the number of structural zeroes in the transition matrix. Structural zeroes are defined as those transitions which cannot be observed. These are discussed in detail in Bishop et al. [1975]. For example, elements on the diagonal of the transition matrix are zero in our specification of the semi-Markov patient flow model. The transition matrix which we test is the $m \times r$ submatrix $[Q, S]$ from Equation 1, since by definition there is no variability in the submatrix $[0, I]$ of the overall transition matrix.

Before combining any subpopulations, it is also necessary to determine whether the holding time observations are drawn from the same distributions. If not, the subpopulations cannot be combined. Holding time distributions are defined for every possible source-destination transition where transfers are observed. Either a Kolmogorov-Smirnov two-sample test or a Kruskal-Wallis test (Conover [1971]) can be used to test whether observed holding times come from the same distribution for each observed source-destination transition that at least two populations have in common. At this stage we need only the empirical distributions. Estimation of theoretical distributions is necessary only at the final specification stage.

The Kolmogorov-Smirnov two sample test statistic (for $s = 2$) is

$$T_{ij} = \sup_x |S_{ij}^{(1)}(x) - S_{ij}^{(2)}(x)|; \quad i = 1, \dots, m; \quad j = 1, \dots, r;$$

where x represents holding time and $S_{ij}^{(l)}(x)$, $l = 1, 2$, is the empirical

cumulative distribution function based on the sample of transfers from unit i to unit j for patients of type l .

A Kruskal-Wallis test can be used to test the null hypothesis that the holding times come from the same distribution when there are more than two subpopulations to be tested (i.e., $s > 2$). The alternate hypothesis is that at least one of the populations tends to yield larger observations than at least one of the other populations. The Kruskal-Wallis test assumes that the data are continuous. When ties are present, an adjustment to the test statistic is appropriate. The effect of correcting for ties is to make the result more significant, i.e., one would be more likely to reject the null hypothesis (Siegel [1956]). The Kruskal-Wallis statistic is given by

$$T_{ij} = (12/(n_{ij}^{(l)} \cdot (n_{ij}^{(l)} + 1))) \cdot \sum_{l=1}^s R_{ij}^{(l)^2} / n_{ij}^{(l)} - 3(n_{ij}^{(l)} + 1);$$

$$i = 1, \dots, m; \quad j = 1, \dots, r;$$

where

$R_{ij}^{(l)}$ = the sum of the ranks of the observations of holding time for a patient of type l from unit i to unit j after all the observations of holding times have been placed in ascending order,

and the other variables are as previously described. The test statistic is asymptotically distributed as chi-squared. Tables for small samples are available.

An alternative test would be a likelihood ratio test. This test would be necessary where there were a significant number of ties in the length of stay data or the data were assumed to be discrete. The test statistic would be

$$\lambda_{ij} = \prod_{l=1}^s \prod_{t=1}^T (f_{.t}/f_{lt})^{n_{lt}}; \quad i = 1, \dots, m; \quad j = 1, \dots, r;$$

where f_{lt} is the fraction of type l observations in time class t , n_{lt} is the number of type l observations in time class t , $f_{.t}$ is the overall fraction of observations in class t , and T is the number of classes. The specification of break points for the time classes and the number of classes, T , are both arbitrary selections. In this case, $-2\log\lambda_{ij}$ would have a chi-square distribution with $(s - 1) \times (T - 1)$ degrees of freedom.

It is obvious that considerable judgment must be used at this stage. Two subpopulations might be erroneously combined simply because they have very few observed transitions in common. Since the Kolmogorov-Smirnov statistic is computed only for common transitions, the analyst might have very few cases for hypothesis testing. Furthermore, all individuals in the newly defined subpopulations would have a nonzero transition probability between source-destination pairs that had been ob-

served in either of the two original subpopulations. Therefore, caution must be used to insure that very few observations escape the Kolmogorov-Smirnov test, and that if subpopulations are combined, transitions are not permitted which would not be sensible for either of the two subpopulations forming the new subpopulation. We note that the Kolmogorov-Smirnov tests will not be necessary for subpopulation pairs in which the hypothesis regarding the homogeneity of transition matrices is rejected, and this is increasingly likely as the observed transitions shared by the subpopulations decreases.

The test for homogeneity of holding times is conservative due to the fact that populations must pass tests for each of the source-destination pairs. As the number of pairs to be tested increases, the probability of obtaining a spuriously significant difference increases. If there are no *a priori* expectations of differences for specific source-destination pairs, the investigator may wish to adjust the α level to make the test less conservative.

Note also that the procedures presented above only provide tests for the possible combination of subpopulations. In order to determine the necessity of further subdividing the groups, characteristics which might identify subpopulations must be identified prior to any data collection and tests conducted to see if they can be combined.

Are Transitions Markovian?

The next step is to test whether the movements of patients are independent of previous states. If not, a new state space definition that accounts for past history will be necessary. This assumption must also be tested on two levels—transition probabilities and holding time distributions. The Markov hypothesis for transition probabilities can be stated as follows:

$$H_0: p_{1ij} = p_{2ij} = \dots = p_{mij} \quad \text{for } i = 1, \dots, m; \quad j = 1, \dots, r;$$

where

$$p_{kij} = \text{the probability that a patient makes a transition from unit } i \text{ to unit } j \text{ given that the immediate past location was unit } k \text{ for } k = 1, \dots, m; i = 1, \dots, m; j = 1, \dots, r.$$

The likelihood ratio criterion (Anderson and Goodman) can be used to test this hypothesis. To estimate each probability, maximum likelihood ratios are again used for all observed transitions, where

$$p_{kij} = n_{kij}/n_{ki}; \quad k = 1, \dots, m; i = 1, \dots, m; j = 1, \dots, r;$$

$$n_{kij} = \text{the number of patients observed going from unit } k \text{ to unit } i \text{ to unit } j,$$

and

$$n_{ki\cdot} = \sum_{j=1}^r n_{kij}.$$

The transition frequencies can be placed in a contingency table format with each table corresponding to a separate state i . The rows of each table correspond to observed previous locations k , and the columns denote observed destinations j . Let x_i be the number of rows and y_i the number of columns of each of the m tables. The null hypothesis can then be stated as

H_0 : the row percentages of the contingency table are all equal.

The likelihood ratio is

$$\lambda_i = \prod_{k=1}^{x_i} \prod_{j=1}^{y_i} (p_{ij}/p_{kij})^{n_{kij}}, \quad i = 1, \dots, m;$$

where the symbols are as defined above. Under the null hypothesis, $-2\log\lambda_i$ has a χ^2 distribution with $(x_i - 1) \times (y_i - 1)$ degrees of freedom.

The hypothesis that holding time distributions are independent of previous location can be stated as

$$H_0: h_{1ij} = h_{2ij} = \dots = h_{mij}; \quad i = 1, \dots, m; \quad j = 1, \dots, r;$$

where h_{kij} represents the probability distribution describing length of stay in unit i , if the next desired location is unit j given that the patient's immediate past location was unit k . A Kruskal-Wallis test or likelihood ratio test as described previously can be used to test whether the probability distributions are the same for previous locations of patients. Observe again that this is a conservative procedure that will permit the researcher to reject the Markov assumption if any of the source-destination pairs are rejected. Another procedure would be a joint test that tested all transitions simultaneously. Rejection of the joint test would then require individual tests as described above.

In general, the length of time spent in any one state may not only be a function of the previously occupied states but also of the length of time spent in previous states. This dependence may be observed by using regression or correlation analysis in each case where the sample size is sufficient for both identification of the relationship and additional partitions of the state space, and may be considered as an alternative approach to our methodology. We note that one of the strengths of our procedure is that none of the tests is distribution specific, i.e., no assumptions or specifications of probability distributions are necessary in the testing stage.

Does Occupancy Rate Affect Patient Transfers?

Once the state space has been defined for each subpopulation the final

test we recommend is to determine if occupancy rates have affected transfers. Researchers may also use these same tests to determine if other hypothesized factors have affected patient flows. Any observations affected by occupancy rate (or these other factors) should be removed before estimation of the final model since existing models all assume that the parameters represent those that would occur in the absence of capacity constraints.

To identify effects of occupancy rates, likelihood ratio tests and Kolmogorov-Smirnov tests can be performed on transition matrices and holding time distributions, separating the observations into arbitrary categories such as "slow," "normal" and "busy." Medical judgment should be used to categorize these periods.

Does the Present Formulation Satisfy All Conditions?

The procedures of assumption testing and model fitting for the semi-Markov model require changes in state space and subpopulation definitions during the hypothesis testing process. Changes at one stage may result in the invalidation of tests performed during previous stages of the process. We therefore recommend that each stage of the process be performed until a formulation is found that satisfies all of the assumptions simultaneously. Alternatively, each time an assumption is violated, one can go back and repeat all previous tests. The resulting model, although meeting all of the prespecified requirements, will not in general yield a unique formulation. The order of the tests presented in Figure 1 is arbitrary and may be considered a variable which the researcher may modify to fit his/her own requirements.

4. DATA

The maternity service of a large university teaching hospital was selected as an application site. The care units consisted of labor rooms, delivery rooms, recovery rooms, and maternity wards. Since most patients entered the system via the emergency room, and since some required the use of other services outside of the maternity system, the set of care units also included the emergency ward and "other" units. Since neither maternal deaths nor transfers to other hospitals occurred within our sample, home was the only absorbing state required. The patient flows of newborns are not considered in this paper and hence attention is restricted to the movement of the expectant mother from time of entrance to the hospital through subsequent discharge from the hospital.

We collected retrospective data from the medical records of all patients who entered the maternity service during the month of April 1978. This exhaustive sample allowed us to reconstruct the hospital census for each

of the maternity service units. Specific data included age, race, risk score (Goodwin et al. [1969]), and procedures administered while in the hospital, as well as the times and locations of initial arrivals and subsequent transfers for each patient. (The Goodwin-Dunne-Thomas risk score is a score ranging from 0 to 10 which is used to predict mortality and morbidity for pregnant women. It includes factors such as age of the pregnant woman, gestational age, and conditions such as diabetes, hypertension, anemia, and abnormal glucose tolerance in its calculation.)

The medical records indicate the arrival time at a facility. Times were usually entered to the nearest quarter hour, although the time of birth was recorded exactly. Lengths of stay were therefore assumed to represent a continuous random variable. When the same patient entered the hospital more than once during the sample period, a separate visit was recorded.

A total of 339 visits to the hospital maternity service was observed. The major classifications of visits were vaginal deliveries (176), Caesarian sections (33), predelivery screenings (78), and "others" (49). The three remaining visits were for abortions, which are not included in this study. Predelivery screenings include scheduled admissions to the labor floor for procedures such as ultrasound or fetal monitoring. "Others" are unscheduled admissions which usually are the result of complications such as premature labor or bleeding, or of false labor. Summary statistics concerning the maternity service data are available in Weiss [1980].

Semi-Markov assumptions were not tested for the emergency ward since this facility was used as a distribution point for entering patients. All expectant mothers entered the hospital through the emergency ward and were quickly referred to another facility within the obstetric service. We have formulated the model by having the stream of patients who enter the emergency ward subsequently decomposed by next destination. An alternative formulation would have excluded the emergency ward from the state space and would have distinct streams of patients entering the other units. The two formulations are equivalent if the assumption of instantaneous patient distribution is accepted.

5. RESULTS

In this section we describe the results of applying the methodology described in Section 3. Detailed results of the statistical tests and a step by step discussion are available in Weiss et al. [1982].

The seven initial states were Labor, Delivery, Recovery, Maternity, Emergency, Other, and Home. At the outset, four medical procedures—vaginal deliveries, Caesarian sections, predelivery screenings and "others"—were chosen as subpopulation groupings. After rejecting the hypothesis that all four patient types exhibited the same transition

matrix, we tested the hypothesis that patients having a vaginal delivery were the same as transitions of patients having a Caesarian section. We also tested the similarity of transitions of predelivery screenings and others. These two pairs were tested because of their observed similarities. The results indicated that the transition matrices for vaginal deliveries and Caesarian sections were significantly different at the 0.05 level of significance but there was not a significant difference for the predelivery screenings and "others." Therefore, transition matrices were deemed necessary for three subpopulations as defined by medical procedure (vaginal deliveries, Caesarian sections, and nonbirths). Kolmogorov-Smirnov tests regarding the empirical holding time distributions did not lead to rejection of the homogeneity hypothesis for the nonbirth grouping. Further analyses therefore dealt with only these three groupings.

We note that in carrying out the subpopulation identification stage, it is important to avoid either over- or underspecifying the model. Aggregating patients with dissimilar transfers can lead to misleading results. Specification of more subpopulations than required will lead to unnecessary computations and may cause statistical problems associated with small sample sizes. It is necessary to continue redefining subpopulation groups, using conventional medical opinion and observation along with analysis of intermediate results, until a delineation of patient types is found that is parsimonious for computation and understanding yet detailed for analysis.

The next criterion for specification of subpopulations was maternal risk as defined by the Goodwin-Dunne-Thomas risk scores. This delineation is of concern because of the possibility of changes (through regionalization of maternity services or demographic changes) in the risk distribution of admitted patients and their possible differential utilization of hospital resources. Risk scores were collected for vaginal deliveries and Caesarian sections. Low, medium, and high risk patients were defined by risk scores of 0-3, 4-6, and 7-10, respectively. Likelihood ratio tests were performed testing the homogeneity of transition matrices as defined by procedure and risk. The hypothesis of homogeneity was rejected for the patients having vaginal deliveries. Similarly, using the Kruskal-Wallis test adjusted for ties there were cases where we rejected the hypothesis that the observed holding times came from the same distribution. This implies that separate estimation procedures should be used for each class of risk. However, at this point we felt that further subdivision of the population would lead to problems associated with small sample size. New subpopulations were therefore not identified. The result of the tests concerning risk are significant in that they indicate the need for a larger sample size in future studies, particularly if researchers are interested in modeling the effects of changes in the risk distribution at a particular facility.

Likelihood ratio tests rejected the hypothesis that state-to-state transition probabilities were independent of previous location for both the labor room and maternity. It is clear that the creation of new states which incorporate past history can always convert a non-Markov system into a Markov system. This must be done for all states which violate the independence assumption. Medical judgment should be used to construct states which both exhibit the necessary independence and make sense in terms of the delivery of care. All partitions of states must be constructed in a manner that results in new state definitions which are mutually exclusive and collectively exhaustive.

At this stage, the new states that were introduced (after a series of possibilities were examined) involved the partition of the labor room into two classes—those patients who arrive at the labor room from maternity (LFM) and all others (LR). The maternity ward was partitioned into four classes—maternity patients who arrive from the emergency ward (MFE), maternity patients who arrive from labor who had not previously been in maternity (MFL), maternity patients who arrive from the labor room who had previously been in maternity (MLM), and maternity patients who arrive from any other sources (MAT). Similarly, Kruskal-Wallis testing of the holding time distributions indicated the need for partitioning the delivery room into two separate states—delivery from maternity (DFM) and delivery from all other sources (DEL).

The impact of occupancy was measured by partitioning observations into two classes—“busy” and “nonbusy” as determined by the 90th percentile of observed occupancy. The homogeneity tests for transition matrices and holding time distributions could not be rejected. Thus, if any of the transfers were the result of capacity constraints, we were not able to observe them. This result was not surprising since we selected our sample period to be a typical month and specifically avoided busy periods.

As a final step, we repeated all previous stages of the specification procedure using 11 states—Emergency (ER), LFM, LR, Recovery (RR), MFE, MFL, MLM, MAT, DFM, DEL, and Other (OTH)—and three subpopulations. All tests were satisfied.

6. MODEL ACCURACY

Using the model state space definitions specified above, we used maximum likelihood estimates for parameters of Weibull distributions of the holding times. Weibull distributions were chosen because of their widespread use in reliability systems, a process similar to the health status of an individual. We used a grouping procedure similar to Kao's [1974] that resulted in the specification of five theoretical distributions. It was necessary to resort to a discrete multinomial distribution in two cases since the Weibull distributions did not give a satisfactory fit. The

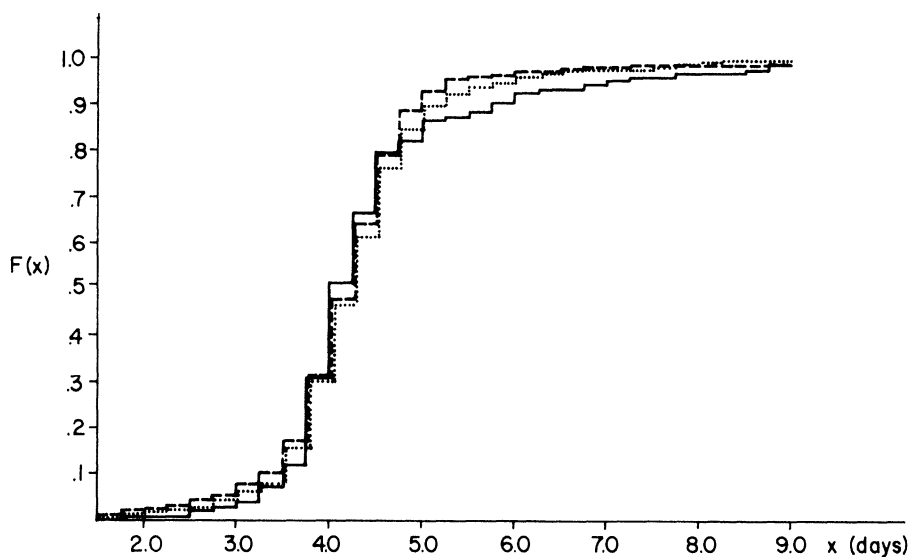


Figure 2. Total length of stay distribution: vaginal deliveries.
 Key: (—) Actual, (---) revised model, (....) original model.

transition matrices and parameters of the holding time distributions are presented in the Appendix.

As a test of the accuracy of the final model, the theoretical total length of stay distributions within the maternity service for each patient type as predicted by the model were calculated and compared to the realized length of stay distributions. The theoretical distribution for length of stay

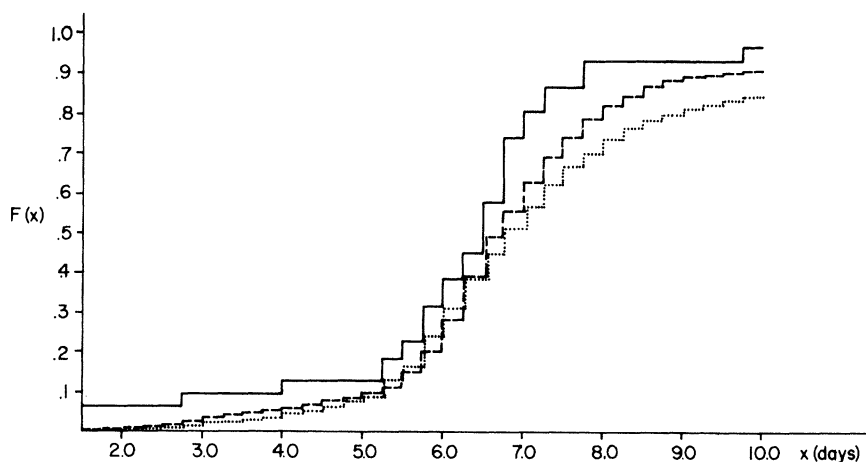


Figure 3. Total length of stay distribution: Caesarian sections.
 Key: (—) Actual, (---) revised model, (....) original model.

in unit i before transfer to an absorbing state j is given by

$$f_{ij}(x) = \sum_{k=1}^m \int_{y=0}^x p_{ik} h_{ik}(y) f_{kj}(x-y) dy + p_{ij} h_{ij}(x);$$

$$i = 1, \dots, m; \quad j = m+1, \dots, r;$$

which is the continuous analog to the discrete recursive relationship given in Kao ([1972], p. 195).

A Monte-Carlo simulation was used to find the total length of stay distributions associated with both the modified state space derived above and the original state space as defined by the physical units of the

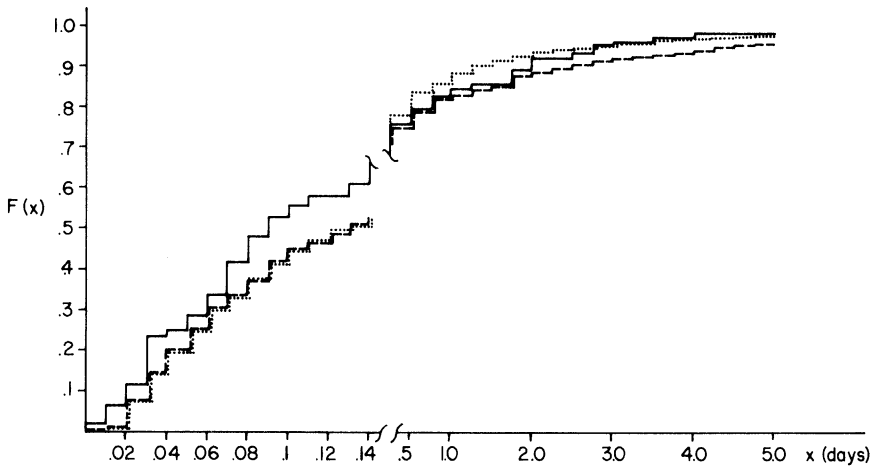


Figure 4. Total length of stay distribution: nonbirths.

Key: (—) Actual, (---) revised model, (....) original model.

maternity service. Discretized graphs presented in Figures 2–4 compare the actual total length of stay distribution with the theoretical distributions as given by the two models. A Kolmogorov-Smirnov test accepts the revised model's total length of stay distribution at the 0.05 level for each of the three groups—vaginal deliveries, Caesarian, and others. The values are 0.086, 0.203, and 0.102, respectively, corresponding to critical values of 0.102, 0.233, and 0.116 at the 0.05 level. When the model is tested using the original state space, the Kolmogorov-Smirnov statistics are 0.075, 0.271, and 0.098, respectively. Thus, with the original model, the distributions for vaginal deliveries and “others” are accepted at the 0.05 level, but the distribution for Caesarian sections is rejected. The results of the revised model are therefore closer to the actual data for Caesarian sections than are the results of the original model.

7. CONCLUSIONS

The methodology presented in this article can be used to specify a semi-Markov flow model through an iterative process of state space definition, population disaggregation, and parameter estimation. The methodology indicates the significance of issues concerning the identification of subpopulations with similar flow characteristics, the existence of dependence of patient movements upon previous locations, and the effect of occupancy rates upon patient flows.

Subpopulation identification is particularly important for the modeling of anticipated changes in the patient mix of a hospital. For example, an increase in the fraction of Caesarian sections, or a redistribution of the

Source	Destination											
	ER	LR	DEL	RR	MAT	OTH	MFL	LFM	MLM	MFE	DFM	HOME
ER		0.949	0.017							0.034		
LR			0.935 (I)				0.065 (I)					
DEL		0.017 (I)		0.966 (II)	0.017 (I)							
RR			0.028 (I)		0.966 (III)	0.006 (I)						
MAT						0.006 (I)		0.011 (I)			0.017 (I)	0.966 (VI)
OTH					1.0 (I)							
MFL			0.091 (V)					0.636 (V)				0.273 (V)
LFM			0.440 (I)	0.080 (I)					0.480 (I)			
MLM								1.0 (V)				
MFE								1.0 (V)				
DFM				0.75 (I)	0.25 (I)							
HOME												1.0

Figure 5. Transition matrix: vaginal deliveries.

risk classes of expectant mothers, may lead to important changes in hospital resource requirements. The dependence of future flows upon previous locations and the creation of new states with which to deal with this dependence gives the model additional information for the prediction of patient movements. Thus, although we must increase the size of the state space to ensure that the Markovian assumptions are not violated, increased accuracy is obtained.

The tests of the impact of occupancy upon patient flows enable estimation of the model describing "ideal" patient movements by identification and removal of observations affected by high occupancy. Where occupancy does affect movements, the "ideal" parameters may no longer be appropriate for analysis of the progressive patient care system. A

Source	Destination											
	ER	LR	DEL	RR	MAT	OTH	MFL	LFM	MLM	MFE	DFM	HOME
ER		0.758	0.030			0.030				0.182		
LR			0.964 (I)				0.036 (I)					
DEL		0.061 (I)		0.879 (I)	0.030 (I)	0.030 (I)						
RR					1.0 (I)							
MAT												1.0 (VII)
OTH		0.250 (I)		0.500 (I)	0.250 (I)							
MFL								1.0 (I)				
LFM			0.417 (I)						0.583 (V)			
MLM						0.286 (V)		0.714 (V)				
MFE		1.0 (I)										
DFM								1.0 (I)				
HOME												1.0

Figure 6. Transition matrix: Caesarian sections.

model that permits different mechanisms describing patient movements is therefore required. (See Cohen et al., and Hershey et al. for approaches to this problem.)

The usefulness and necessity of the procedures developed in this paper depend upon the specific populations chosen and the specific criteria selected. For some populations, the fact that the independence assump-

Source	Destination											
	ER	LR	DEL	RR	MAT	OTH	MFL	LFM	MLM	MFE	DFM	HOME
ER		0.975								0.025		
LR			0.016 (I)			0.008 (I)	0.194 (V)					0.782 (IV)
DEL		1.0 (I)										
RR												
MAT												
OTH		1.0 (I)										
MFL								0.333 (V)				0.667 (V)
LFM									0.846 (V)			0.154 (I)
MLM								0.182 (V)				0.818 (V)
MFE								1.0 (V)				
DFM												
HOME												1.0

Figure 7. Transition matrix: nonbirths.

TABLE II
PARAMETERS OF ESTIMATED WEIBULL DISTRIBUTIONS^a

Distribution	γ	θ	ξ
I	0.9824557	0.2370313	0.0027
II	1.1468499	0.0257421	0.0168
III	1.3435260	0.0121660	0.0345
IV	0.9448905	1.1186774	0.0200
V	0.8071992	1.1675207	0.0138

^a $f(x) = (\gamma/\theta) \cdot (x - \xi)^{\gamma-1} e^{-(x-\xi)^\gamma/\theta}$
 $x \geq \xi$
 $\gamma, \theta > 0.$

TABLE III
DISCRETE MULTINOMIAL APPROXIMATIONS: MATERNITY-HOME TRANSFER

Length of Stay $x(\text{days})$	Distribution	
	VI Vaginal Deliveries $f(x)$	VII Caesarian Sections $f(x)$
1.25	0.006	0.000
1.50	0.006	0.000
1.75	0.006	0.000
2.00	0.000	0.000
2.25	0.012	0.032
2.50	0.012	0.000
2.75	0.036	0.000
3.00	0.018	0.032
3.25	0.030	0.000
3.50	0.199	0.000
3.75	0.193	0.000
4.00	0.187	0.032
4.25	0.175	0.000
4.50	0.024	0.000
4.75	0.024	0.065
5.00	0.000	0.000
5.25	0.006	0.065
5.50	0.024	0.129
5.75	0.006	0.129
6.00	0.012	0.129
6.25	0.000	0.032
6.50	0.006	0.097
6.75	0.006	0.162
7.00	0.006	0.000
7.25	0.000	0.000
7.50	0.000	0.032
7.75	0.000	0.000
8.00	0.000	0.000
8.25	0.006	0.000
10.00	0.000	0.032
23.50	0.000	0.032

tions are violated may not have a substantial impact on a given criterion, such as total length of stay. This was shown for the vaginal deliveries and "others" subpopulations. For other populations, however, the validation and further specification of the semi-Markov model can significantly alter the results of the model. This was demonstrated for the case of Caesarian sections, where the improvements were great enough to transform the unacceptable results from an unrevised model to acceptable results from a revised model. Thus, the proposed methodology can be an important first step before using analytic or simulation models for forecasting resource utilization within a semi-Markov system.

APPENDIX: TRANSITION MATRICES

The transition matrices estimated for the three populations are given in Figures 5 to 7. Roman numerals in parentheses refer to the probability distributions given in Table II. Discrete multinomial approximations for the maternity-home transfer are given in Table III.

ACKNOWLEDGMENTS

This research was supported in part by funds from the National Health Care Management Center (NHCMC) at the University of Pennsylvania. The NHCMC is supported by Grant HS-02577 from the National Center for Health Services Research, Office of the Assistant Secretary of Health, DHHS.

The authors would like to thank Janet Bly who assisted in the design of the data collection instrument and Doreen Scarangelli and Jody Samuels who collected the data. We also thank the medical and administrative staff of the hospital for their assistance.

REFERENCES

- ANDERSON, T. W., AND L. A. GOODMAN. 1957. Statistical Inference about Markov Chains. *Ann. Math. Statist.* **28**, 89-110.
- BISHOP, Y., S. FEINBERG AND P. HOLLAND. 1975. *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, Mass.
- COHEN, M. A., J. C. HERSHEY AND E. N. WEISS. 1980. Analysis of Capacity Decisions for Progressive Patient Care Hospital Facilities. *Health Serv. Res.* **15**, 145-160.
- CONOVER, W. J. 1971. *Practical Nonparametric Statistics*. John Wiley & Sons, New York.
- FETTER, R. B., AND J. D. THOMPSON. 1969. A Decision Model for the Design and Operation of a Progressive Patient Care Hospital. *Med. Care* **7**, 450-462.
- GOODWIN, J., J. DUNNE AND B. THOMAS. 1969. Antepartum Identification of the Fetus at Risk. *Can. Med. Assoc. J.* **101**, 57-65.

- HERSHEY, J. C., E. N. WEISS AND M. A. COHEN. 1981. A Stochastic Service Network Model with Application to Hospital Facilities. *Opns. Res.* **29**, 1-22.
- JACKSON, J. R. 1957. Networks of Waiting Lines. *Opns. Res.* **5**, 518-521.
- KAO, E. P. C. 1972. A Semi-Markov Model to Predict Recovery Progress of Coronary Patients. *Health Serv. Res.* **7**, 191-208.
- KAO, E. P. C. 1973. A Semi-Markovian Population Model with Application to Hospital Planning. *IEEE Trans. Syst. Man Cybernet.* **3**, 327-336.
- KAO, E. P. C. 1974. Modeling the Movement of Coronary Patients within a Hospital by Semi-Markov Processes. *Opns. Res.* **22**, 683-699.
- KASTNER, G. T., AND R. H. SHACHTMAN. 1981. A Stochastic Model to Measure Patient Effects Stemming from Hospital Acquired Infections, Institute of Statistics Mimeo Series 1364, University of North Carolina.
- KASTNER, G. T., R. H. SHACHTMAN AND D. QUADE. 1979. Infection Incidence for Alternative Hospital Infection Surveillance and Control Programs: A Markovian Approach, paper presented at ORSA/TIMS Joint National Meeting, Milwaukee.
- KLEINROCK, L. 1976. *Queueing Systems: Computer Applications*. John Wiley & Sons, New York.
- SHACHTMAN, R., J. SCHOENFELDER AND C. HOGUE. 1981. Using a Stochastic Model to Investigate Time to Absorption Distributions. *Opns. Res.* **29**, 589-603.
- SIEGEL, S. 1956. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York.
- SHONICK, W., AND J. JACKSON. 1973. An Improved Stochastic Model for Occupancy Related Random Variables in General-Acute Hospitals. *Opns. Res.* **21**, 952-965.
- THOMAS, W. H. 1968. A Model for Predicting Recovery Progress of Coronary Patients. *Health Serv. Res.* **3**, 185-213.
- WEISS, E. N. 1980. Semi-Markov Models of Patient Flows for Progressive Patient Care Capacity Planning, unpublished Ph.D. dissertation, University of Pennsylvania.
- WEISS, E. N., M. A. COHEN AND J. C. HERSHEY. 1982. Estimation of a Semi-Markov Model of Obstetric Patient Flows, Graduate School of Business and Public Administration, Cornell University. Working Paper 81-09.