A STOCHASTIC MODEL TO MEASURE PATIENT EFFECTS STEMMING FROM HOSPITAL ACQUIRED INFECTIONS

by

George T. Kastner
and
Richard H. Shachtman

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1364

August 1981

A STOCHASTIC MODEL TO MEASURE PATIENT EFFECTS STEMMING FROM HOSPITAL ACQUIRED INFECTIONS

by

George T. Kastner [1]

and

Richard H. Shachtman [2]

August 27, 1981

(1) Associate Professor
Instituto de Estudios Superiores de Administracion
Apartado 1640
Caracas, Venezuela 1010A

(2) Professor
Department of Biostatistics
426 Rosenau Hall 201H
University of North Carolina
Chapel Hill, N. C. 27514

A STOCHASTIC MODEL TO MEASURE PATIENT EFFECTS STEMMING FROM HOSPITAL ACQUIRED INFECTIONS

George T. Kastner
Instituto de Estudios Superiores de Administracion, Caracas, Venezuela

Richard H. Shachtman
University of North Carolina, Chapel Hill, N. C.

Abstract

We introduce a Markov chain model to represent a patient's path in terms of the number and type of infections s/he may have acquired during a hospitalization period. The model allows for categories of patient diagnoses, surgery, the four major types of nosocomial (hospital-acquired) infections and discharge or death. Data from a national medical records survey including 58,647 patients enable us to estimate transition probabilities and, ultimately, perform statistical tests of fit, including a validation test. Novel parameterizations (functions of the transition matrix) are introduced to answer research questions on time-dependent infection rates, time to discharge or death as a function of patient characteristics at admission, and conditional infection rates reflecting intervening variables (e.g., surgery).

A STOCHASTIC MODEL TO MEASURE PATIENT EFFECTS STEMMING FROM HOSPITAL ACQUIRED INFECTIONS

by

George T. Kastner
and
Richard H. Shachtman

Researchers modeling hospital activities often require a quantitative representation of a patient's probabilistic health status changes during a hospitalization stay. Frequently, the stochastic model utilized is a Markov model. See for example Thomas (1968), Bush et al. (1971) and Kao (1972). These investigators present disease-specific applications; Bush et al. study adults with primary active tuberculosis while both Thomas and Kao study the recovery process of coronary care patients.

Our model, also a Markov approach, follows a patient's path in terms of type and number of infections s/he may have acquired during a hospitalization period in a particular hospital/patient group, independent of the cause of admission. This model allows us to predict the incidence of nosocomial (hospital acquired) infections (NIs) at the patient-day level. Moreover, our approach suggests novel parameterizations (functions of the transition matrix) to provide decisionmakers with additional criteria for evaluating hospitalization outcomes. We provide answers to several research questions, for example: how can we estimate the time spent in distinct health states? How much longer does a patient with an NI spend in the hospital than a patient with similar characteristics who has not acquired an NI? Does her/his

length of hospitalization (LOH) change significantly if s/he acquires a secondary NI? What is the effect on these measurements of intervening variables, such as surgery? Suppose we could eliminate certain types of NIs; how would this influence LOH?

Generally the answers to these questions are dependent on hospital type and patient group. Some conclusions emerge from aggregate measurements, such as the incidence and prevalence of patients contracting NIs and mean LOH, using traditional statistical and epidemiological methods. Other questions, in particular those requiring time-dependent estimators, cannot be answered with traditional approaches and require model building and new parameterizations. The above questions imply that the model include LOH measurements which reflect the impact of intervening variables, e. g., the time period during which a surgical intervention occurred. Furthermore, the model should allow for simulation exercises providing answers to if-then questions, which are not obtainable in observational studies.

We provide both the model structure and parameterization ideas for analyses with respect to a comprehensive set of hospital/patient categories. Eventually, refinements of this model may be used to reflect differences in Infection Surveillance and Control Programs (ISCPs), Haley and Shachtman (1980), and their impact on outcome measurements, such as those proposed as parameterizations and some corresponding numerical results.

Specific objectives are to demonstrate how the model is used in unique hospital/patient groups, to:

1. produce a daily measure of infection incidence,
2. estimate the relative proportion of infections by site,

3. estimate patient LOH in distinct states during a hospitalization episode,
4. determine conditional infection rates,
5. compute time to absorption as a function of patient characteristics at admission, and
6. determine the influence of "elimination" of certain types of NIs on measures like LOH.

To achieve the first three objectives, we directly exploit the statistical attributes of the proposed Markov model. For the subsequent objectives we develop functional expressions of the transition probabilities (parameterizations).

In the following sections we introduce a 16-state Markov chain, along with corresponding assumptions, to describe and analyze possible relationships with primary diagnosis, infection status, time of surgical procedure and LOH. We also discuss estimation and statistical tests for the various assumptions, compare empirical results with model computed results and suggest parameterizations corresponding to the research questions.

1. MODEL ASSUMPTIONS AND STRUCTURE

During a hospitalization episode a patient may contract infections at various sites, s/he may acquire more than one infection prior to her/his discharge and s/he may die during her/his hospital stay. Each combination of these events describes a possible path through a patient's states of health in terms of NIs. Moreover, given daily observations from the patient's medical record, her/his current health state determines the next health state and we assume this state reflects

the health history of the patient for NI status. This is the underlying Markov assumption; we discuss test results for the Markov property in a later section.

We also make the following assumptions:

1. All patients enter the system uninfected and an algorithm determines their infection status, onset, type (hospital or community acquired), site (below); see Appendix E of the American Journal of Epidemiology, May, 1980.
2. A patient is eligible for our analysis if her/his LOH is four or more days and during these days the patient contracts no infections.
3. We consider four major infection sites: UTI - Urinary Tract Infection; BACT - Bacteremia; LRI - Lower Respiratory Infection; and SWI - Surgical Wound Infection.
4. Cases of multiple infections with the same onset day are labeled UTI before LRI before SWI before BACT. In our patient data base only 131 such cases out of 58647 studied were recorded.
5. Transition probabilities are time homogeneous.
6. We assume unlimited capacity in each state and independent patient-level observations.

To demonstrate the model, we study all patients within a homogeneous hospital group. For patient level analysis we study patient groups characterized by service (medicine or surgery), age, and primary diagnosis, within a homogeneous cohort of hospitals.

Our data within each hospital comes from the Study on the Efficacy of Nosocomial Infection Control (SENIC) sample. Medical chart

reviewers conducted a Medical Record Survey sampling records of discharges from two time periods; see the American Journal of Epidemiology, May, 1980. Each record abstraction includes demographic characteristics of the patient, service, daily records on antibiotics administration, results of cultures and diagnosis of infection.

We describe a patient's health trajectory during a hospitalization period by a 16-state Markov chain model. All patients enter the system through one of four uninfected states, each characterizing a different primary diagnosis group. There are eight states describing infection status, four are primary NIs and four are secondary NIs. We also reflect the occurrence of the first surgical procedure making a distinction between the state describing a surgical intervention which occurs prior to a primary NI, and one occurring after the patient has already been infected. There are two absorbing states, discharge and death.

To complete the state definitions we define the Primary Diagnosis Groups (PDGs). From the 83 Major Diagnostic Categories (MDCs) suggested by Fetter et al. (1978) we selected 18--those which showed the highest frequency of occurrence in our data base. A medical epidemiologist from the Centers for Disease Control (CDC), helped us combine the selected MDCs into four groups which are similar in terms of organ systems. Table 1 summarizes the definition and size of our group selection.

Table 1

TABLE 1

STATE DEFINITION AND CLASSIFICATION OF PRIMARY DIAGNOSIS GROUPS

Hierarchy	State	Definition by Organ System	Number of Patients
Uninfected (Primary Diagnosis Group)	PD1	Vascular System Diseases	16,548
	PD2	Lung and Pleural Diseases	5,835
	PD3	Musculo-Skeletal and Superficial Injuries	9,609
	PD4	Other Diseases included in the 18 MDCs	26,655
	SUR1	First Surgical Procedure (prior to acquiring an NI)	
Primary Infection	S1	First Urinary Tract Infection (UTI)	
	S2	First Bacteremia (BACT)	
	S3	First Lower Respiratory Infection (LRI)	
	S4	First Surgical Wound Infection (SWI)	
	SUR10	First Surgical Procedure (after acquiring an NI)	
Secondary Infection	S10	Second Infection is UTI	
	S20	Second Infection is BACT	
	S30	Second Infection is LRI	
	S40	Second Infection is SWI	
Absorbing	S5	Discharge	
	S6	Death	
		Total	58,647

2. DATA, PATIENT AND HOSPITAL GROUPS

Our data is based on a random, stratified sample of 338 hospitals with approximately 500 randomly selected patients in each hospital, yielding a total of 169,526 patients. Since we consider only patients with LOH of four or more who are uninfected on their fourth day of stay, we are left with 118,221 patients. Of these we select the patients admitted with a primary diagnosis corresponding to the selected MDCs, which leaves us with 58,647 patients.

Stratification of the data is by two hospital level variables, one representing the size of the hospital and whether or not affiliated with a medical school (BEDMED) and the other representing the size of the standard Metropolitan Statistical Area (SMSA). Table 2 summarizes the levels of these variables and displays the number of hospitals and patients contained in each stratum.

TABLE 2

Since our main purpose is to develop and demonstrate a modeling approach, we do not perform separate numerical analyses for each patient group and hospital category. We concentrate on cells 1, 3, 5, 6 and 8 which total 46,479 patients. The numerical results presented in later sections demonstrate how our methodology can provide answers to the research questions posed earlier. Policy making purposes require the comparison of results for several combinations of patient hospital groups; some comparisons and their implications appear in Kastner (1980a).

TABLE 2

STRATIFICATION VARIABLES DEFINITION AND NUMBER OF PATIENTS AND HOSPITALS IN EACH STRATUM

(The upper numbers correspond to the number of patients, the lower numbers to hospitals, and those in parentheses represent the cell numbers.)

SMSA (millions) / BEDMED	SMSA ≤ 0.5	0.5 < SMSA ≤ 2.5	2.5 < SMSA	Total Patients Total Hospitals
Beds < 200	24481 (1) 141	5725 (4) 31	3190 (7) 19	33396 191
200 ≤ beds and affiliated with Medical School	1920 (2) 11	3928 (5) 22	3869 (8) 25	9717 58
200 ≤ beds not affiliated with Medical School	8404 (3) 48	5797 (6) 32	1333 (9) 9	15534 89
Total Patients Total Hospitals	34805 200	15450 85	8392 53	58647 338

3. PARAMETER ESTIMATION AND TESTING

We construct parameter estimates and statistical tests to evaluate the basic assumptions of the model. These tests help assess how accurately and how completely the model represents the empirical process, a procedure seldom employed in the literature.

With a continuous parameter space, the patient's health status can change at any point in time. In practice, of course, we cannot realistically expect to have observations concerned with hundred of patients to be recorded instantaneously. Daily recording of patient health states occurred during the Medical Record Survey; thus we use a discrete parameter space with one day interval lengths.

A. Transition Probability Estimates

Given the state space, time parameter space and independence of patient behavior, we construct maximum likelihood estimators of the transition probabilities for transferring from health state i to state j in hospital group h and patient group p:

$$p_{ij}(h,p) = n_{ij}(h,p) / \sum_{k=1}^{N} n_{ik}(h,p), \quad i,j = 1,\ldots,N,$$

where $n_{ij}(h,p)$ = total number of observed transitions from i to j, in hospital group h and patient group p, and

N = number of states in the chain.

The corresponding variances are given by:

$$V_{ij}(h,p) = p_{ij}(h,p)\,[1 - p_{ij}(h,p)] \sum_{k=1}^{N} n_{ik}(h,p), \quad i,j=1,\ldots,N$$

We present an example of the transition probability estimates for the combination of hospital cells 5 and 8 and weekday admissions in Figure 1. We make a few observations with respect to daily transition prob-

abilities across all hospital groups, although this matrix is typical. Our interests require time-dependent results based on the complete LOH of a patient, thus we refer general statements on patient status to the parameterizations section.

Figure 1

---

1. The daily probability of acquiring a primary or secondary BACT is lower than the corresponding probability for any other NI. Moreover, the daily probability of death after acquiring a BACT is higher in most hospital groups than for other infections.

2. The daily proportion of patients leaving the hospital who die after BACT and LRI is generally higher than after other NIs.

3. The highest infection incidence is of UTI followed by SWI and then LRI. The conditional UTI infection rate is over 50%, while the conditional BACT rate is less than 7%; we discuss the method of calculation in the parameterization section.

4. We do not find similar dominance for second infections. UTI and SWI show the highest incidence, independent of the primary NI. Given that a secondary infection occurred, the UTI incidence ranges from 25 to 65%. (In one case no secondary UTI was observed after primary BACT.) The SWI secondary incidence ranges from as low as 8% to as high as 67%. For some groups no secondary SWI was observed; when a high incidence of secondary SWI is observed, it follows a primary SWI. See Section 5.

5. The daily transition probability within any one primary diagnosis state ranges from 0.78 to 0.90 throughout all hospital groups. In addition, an uninfected patient has about a 93% chance of not acquiring an infection during

	PD1	PD2	PD3	PD4	SUR1	S1	S2	S3	S4	SUR10	S10	S20	S30	S40	S5	S6
PD1	.8919				.0117	.0044	.0004	.0003	.0004						.0865	.0044
PD2		.8866			.0112	.0044	.0005	.0026							.0891	.0055
PD3			.8788		.0112	.0041		.0012	.0013						.1023	.0010
PD4				.8309	.0185	.0045	.0004	.0008	.0018						.1421	.0011
SUR1					.8937	.0079	.0008	.0039	.0067						.0844	.0026
S1						.9182				.0074	.0043	.0027	.0019	.0035	.0581	.0039
S2							.9336			.0041	.0124				.0332	.0166
S3								.8820		.0067	.0178			.0067	.0735	.0134
S4									.9137	.0012	.0047	.0012	.0024	.0106	.0650	.0012
SUR10										.9245	.0063	.0063	.0063	.0094	.0409	.0063
S10											.9617				.0273	.0109
S20												.9777			.0111	.0111
S30													.9408		.0329	.0263
S40														.9596	.0354	.0051
S5															1	
S6																1

FIGURE 1. TRANSITION PROBABILITY ESTIMATES - 16 STATE MODEL
WEEKDAY ADMISSIONS FOR HOSPITAL CELLS 5 AND 8

his/her LOH.

6. Patients in the initial state PD4 generally have a higher daily probability of a surgical intervention than the other patients. The most common NIs (primary or secondary) after a surgical procedure are UTI and SWI.

7. After a surgical procedure, patients have a higher chance of developing an NI. In surgical patients and nonsurgical patients, the overall ratio of the daily transition probabilities from an uninfected state to a state of primary infection is approximately 2.6, i.e., the chance for an NI following surgery is almost three times higher.

8. We computed estimates of the standard deviations in each hospital group and admission day. Approximately 75% of the estimates are smaller than 0.01 and most of the remaining values are between 0.01 and 0.02.

Before using any of the results for statistical inference, we submit the models to several goodness-of-fit tests. The purpose of the tests is three-fold: to assess the extent to which the models portray reality, to verify that the assumptions made while formulating the model are justified by the data and to partition the patient groups into more homogeneous subgroups for these and future analyses. We test and discuss three assumptions: time stationarity, the Markov property, and the geometric holding time distributions. We perform the tests on cells 1,3, 5,6 and 8; details are in Kastner et al. (1980b).

B. Testing for Time Homogeneity (Stationarity)

We implicitly assumed stationarity in estimating the transition matrices. Stationarity is often used in Markov chain analysis and is necessary for steady state results. The characteristics of interest to our models are hospital institutional variables and patient mix by age,

service and diagnosis; these are believed to remain nearly constant over a short time period, say one to three years.

Other possible sources of non-stationarity exist. First, there is a chance that the patient's health status pattern and the LOH in each state is influenced by whether admission occurs on a weekday (Monday through Thursday) or a weekend (Friday through Sunday). Second, the LOH may influence the transition probabilities in such a way that patients with shorter LOH have a lesser chance of acquiring an NI. For this test we define three groups: patients with LOH shorter than the 25% quantile (Q); patients with LOH longer than Q but shorter than the median; and the remaining patients, whose stay is longer than the median. Third, we examine whether a seasonality trend exists by studying four time periods during the year which are consistent with other SENIC analyses: the four quarters from April 1, 1975 through March 31, 1976.

To address the first case we employ a test suggested by Billingsley (1961) for comparing the transition probabilities of two independent chains. To test the other stationarity cases, which involve more than two cohorts, we use the likelihood ratio statistic suggested by Anderson and Goodman (1957). In the selected cell we generate transition probability estimates and test the null hypothesis,

$$H_o : \quad p_{ij}(t) = p_{ij}, \; i, j = 1,\ldots,N, \quad t = \quad 1,\ldots,T,$$

where N is the number of states and T the number of time periods. Under the alternative hypothesis, the estimates for a given hospital cell, at a given time period t, are

$$\hat{p}_{ij}(t) = n_{ij}(t) / \sum_{k=1}^{N} n_{ik}(t) \; , \; i, j = 1,\ldots,N.$$

We test the hypothesis for all states jointly since the random variables $\hat{p}_{ij}(t)$ and $\hat{p}_{ij}$ for two different states i are asymptotically independent; hence, adjusting for the two absorbing states, the joint chi-squared statistic is asymptotically distributed chi-squared with (N-2)(N-3)(T-1) degrees of freedom.

Our test results show that the computed chi-squared statistics for the five cells range from 33,983 to 175,543 for day of admission, while the 0.999 critical value for 210 degrees of freedom is 295. In the LOH case, the statistics values range from 1,899 to 15,497, and the 0.999 percentile with 420 degrees of freedom is 536. On the other hand, the values of the statistics for the seasonal pattern tests range from 9 to 20 versus a 0.999 critical value, for 630 degrees of freedom, which equals 771.

We conclude that both the day of admission and the LOH of the patient are sources of non-stationarity, but that there is no evidence of seasonal pattern. For the first case, we deal with two separate cohorts, patients admitted during weekdays and patients admitted during weekends. In each hospital group, the uninfected states contribute all but a negligible source of time fluctuation creating nonstationarity due to LOH. We expect more instability in uninfected health states as the LOH increases, since the patients are exposed longer to the risk of acquiring an NI. Since all other states do not demonstrate time dependency, we will assume stationarity with respect to LOH; see Section 4.

C. Testing the Markov Property

The key assumption underlying our model is that first-order Markov dependence holds, i.e., that future movement of a patient is stochastically dependent only on his/her present health state. For each hospital group,

we test first-order Markov dependence against second-order Markov dependence, which assumes that a patient's next movement could be probabilistically dependent on both the current health state and the immediately preceding state. See Anderson and Goodman (1957). Assuming time stationarity we have

$$H_o: \text{ for all } 1 \le i,j \le N,\ p_{kij} = p_{ij},\ k = 1,\ldots,N,$$

where

$$p_{kij} = \Pr\{X_n = j | X_{n-1} = i,\ X_{n-2} = k\}, \text{ and}$$

$$p_{ij} = \Pr\{X_n = j | X_{n-1} = i\}, \quad i, j, k = 1,\ldots,N,\ n = 0, 1,\ldots$$

In our case the second order property is true by definition for non-absorbing states and seems acceptable, statistically, for death and discharge. We adopted two tests for the hypothesis: (1) the ordinary chi-squared, see Zahl (1955), and (2) the test suggested by Kullback, Kupperman, and Ku (1962). The latter suggest that the two statistics are asymptotically equivalent, and both have central chi-square limiting distributions with $S_p - S_0$ degrees of freedom, where S_p denotes the number of permissible triplets $(k \rightarrow i \rightarrow j)$, and S_0 is the number of permissible triplets not observed.

We computed the ordinary chi-squared statistic and the Kullback, Kupperman and Ku (KKK) statistic for each of the five hospital cells. The chi-squared values range from 61 to 95. The KKK 0.999 critical values vary, due to a range of degrees of freedom, from 211 to 264. Since the lowest 0.999 critical value, for 140 degrees of freedom, is 211, we accept the equivalence between second-order and first-order dependence for all hospital groups.

Shachtman et al. (1981a) explain that the logarithmic transformation used by Kullback et al. in their χ_k^2 statistic smooths out the relative chi-squared contribution made by less than expected cell frequencies.

They also observe that the χ_k^2 statistic for this test is more robust than the ordinary chi-square (χ_o^2) suggested by Zahl (1955) "with respect to the infrequent occurrence of highly improbable transitions."

Billingsley (1961) suggests a statistic for the hypothesis H_o: $p_{ij}(h) = p_j(h)$, i.e., that the transition probability $p_{ij}(h)$ is independent of state i. (Under H_o the Markov chain is a sequence of independent events.) A chi-squared statistic for each hospital group, ranging from 2,531 to 12,766, with $(N-1)(N-3)$ degrees of freedom (195 in our case), indicates, at the 0.999 significance level, that we can reject the hypothesis of an independent sequence of health states.

These tests imply that the (first-order) Markov property is satisfied.

D. Testing the Holding Time Distribution

A necessary condition for a Markov chain is that the holding times be geometrically distributed where these times are the number of days the process returns to the same state, prior to moving to any other state.

The probability function of the geometric distribution is given by

$$\Pr\{X = k\} = \begin{cases} p^k(1-p), & k = 0, 1, 2, \ldots \quad , 0 \le p \le 1 \\ 0 & , \text{ otherwise.} \end{cases}$$

The null hypothesis for the underlying test is H_o: the probability that a patient stays k consecutive time units (days) in health state i, given the patient is in state i in hospital group h, equals $\hat{p}_{ii}^k(h)\,[1-\hat{p}_{ii}(h)]$, $k = 1, 2, \ldots,$ and the $p_{ii}(h)$ are the diagonal elements of the corresponding transition probability matrix. We observe frequencies of the holding times from the path-coded patient level data.

We employ the Pearson chi-square goodness-of-fit test, see Conover (1971), and the Kolmogorov-Smirnov test, see Conover (1971, 1972). Due to the amount of computation needed, we discuss numerical results only for hospital cell 1. The Pearson chi-squared statistic ranges from 131 to 5,808,168; the corresponding degrees of freedom are 3 or 4. Thus, we reject the hypothesis that the holding time distributions are geometric. The individual contributions to the summary chi-squared statistic imply that the model overestimates frequencies for short holding times and underestimates them for long times. The large contributions of these extreme cases result in large summary chi-squared values. The Kolmogorov-Smirnow tests results are similar.

E. Summary of Validation Tests

The validation tests lead to the following conclusions:

1. The basic Markov property is satisfied.

2. Day of admission and LOH of patients are sources of non-stationarity.

3. The holding time distributions are not geometric.

Thus, some of the underlying assumptions for the Markov chain model do not hold. With respect to stationarity, we proceed with our analysis considering two patient cohorts defined by day of admission. The probable reason for LOH nonstationarity is the large contribution to the chi-squared value exhibited by the uninfected states. Patients enter the system through a set of uninfected health states characterized as primary diagnosis groups which, implicitly, represent an a priori risk level of contracting an NI. We believe that a more disaggregate set of the initial states, which better account for a priori susceptibility to NI, will eliminate the non-stationarity due to LOH. Results of patient risk

analyses by Hooton et al. (1980) support this supposition. Further research on patient risk categories is in progress at CDC and an index will soon be available to categorize patients. Since the transitions among all other states are stationary, we will continue assuming stationarity with respect to LOH.

Most Markov chain models discussed in the literature do not test for geometric holding times. Does this render the untested models invalid? In some cases analysts sensed difficulties and increased the cardinality of the state space, see Thomas (1968); this increases the number of test cells and substantially increases computer time. More significantly, there is a certain trade-off involved: do we redefine some states, add new states and iterate this procedure until we find a state space that better satisfies the assumptions, even if the ultimate model fails to describe reality as accurately as the original model? Here, reality may mean simply the ability to predict LOH. Or, instead of reshaping the state space, and thus the model, do we correct for non-geometric lengths of time in states by adjusting when constructing parameterizations? Insofar as one major purpose in this modeling effort is to predict overall LOH, we deem it sufficient to present results that provide answers to the research questions posed earlier. In the following section we present comparisons between empirical LOH and model developed LOH, which suggest that our model is indeed consistent and responds to the stated purpose. Nonetheless, the model should be used with caution.

4. VALIDATION - EMPIRICAL RESULTS VS. MODEL GENERATED RESULTS

In this section we compare empirical LOH with comparable model results: model based mean time to discharge, or death, (i.e., time to

absorption), versus observed LOH values for certain hospital/patient groups.

To account for nonstationarity caused by admission day, we consider separately patients admitted on weekdays and weekends. We combine cells 3 and 6 (see Table 2) yielding a group of large hospitals, not affiliated with medical schools, in relatively small metropolitan areas. We also combine cells 5 and 8 which consist of large hospitals, affiliated with medical schools, and located in relatively large metropolitan areas.

We present the mean LOH of patients in the resulting four cohorts for each primary diagnosis group, the corresponding standard deviations and the 95% confidence intervals, in Table 3. The model derived computations come from the standard forumla for mean times to absorption; see Section 5.

Table 3

There are no apparent differences in LOH for patients admitted on weekdays or on weekends. Note, however, that based on model estimates, patients admitted to medical school affiliated hospitals stay from two to four days longer (depending on the PDG and the admission cohort) than patients admitted to non-medical school affiliated hospitals. The large standard deviations associated with each mean LOH estimate suggest caution in comparing LOH values.

As anticipated, the values shown for theoretically derived means are similar to those corresponding to the empirical LOH means. In all but two cases, the absolute difference between the empirical mean and the

TABLE 3

MODEL-DERIVED AND EMPIRICAL ESTIMATES OF LOH MEANS, STANDARD DEVIATIONS AND 95% CONFIDENCE INTERVALS FOR THE FOUR PDG'S BY MEDICAL SCHOOL AFFILIATION AND ADMISSION DAY

		Large Non-Medical School Affiliated		Large Medical School Affiliated	
Source	Primary Diagnosis	Weekday Admissions	Weekend Admissions	Weekday Admissions	Weekend Admissions
Model Derived	PD1	13.47 (9.51)	13.33 (9.27)	15.56 (12.30)	15.55 (12.64)
	PD2	12.84 (8.85)	11.48 (7.58)	15.25 (12.21)	15.60 (12.63)
	PD3	12.27 (8.42)	12.94 (9.00)	14.39 (11.41)	15.63 (12.46)
	PD4	10.13 (7.00)	10.19 (6.90)	12.10 (9.92)	12.29 (10.42)
Empirical Estimates	PD1	13.83 (11.67)	13.77 (11.08)	15.73 (14.56)	15.79 (12.79)
		(13.36, 14.13)	(13.22, 14.32)	(14.98, 16.48)	(14.94, 16.64)
	PD2	12.64 (9.18)	11.47 (7.72)	14.96 (15.00)	15.17 (12.89)
		(12.00, 13.28)	(10.81, 12.13)	(13.54, 15.70)	(13.57, 16.77)
	PD3	12.86 (11.66)	13.25 (10.55)	14.74 (13.82)	15.95 (17.68)
		(12.26, 13.46)	(12.56, 13.94)	(13.78, 15.70)	(14.43, 17.47)
	PD4	9.74 (5.43)	9.76 (5.56)	11.56 (14.05)	11.79 (11.66)
		(9.56, 9.92)	(9.54, 9.98)	(10.99, 12.13)	(11.12, 12.46)

derived model result is only one-half day or less. Substantial standard deviations are displayed in all cases. A comparison of empirical standard deviations with model derived results shows that the latter tend to be smaller.

The 95% confidence intervals presented are used as a descriptive indicator of how close the model based mean LOH is to the empirical mean LOH. The confidence intervals include the model mean for 14 of the 16 comparison groups; the fourth diagnosis group shows a confidence interval which excludes the model derived mean for the two non-medical school affiliated cohorts.

The above results attest to the usefulness of the model in estimating the overall LOH for patient cohorts. Moreover, it would appear that having violated the LOH stationarity assumption does not lead to significant deviation in the values of aggregate parameter estimates.

5. PARAMETERIZATIONS

Investigation of the influence of intervening variables on NI incidence and the corresponding LOH requires several functional expressions. The LOH and times spent in distinct health states are important for both policy makers and physicians for resource allocation purposes. We estimate time dependent and absorbing NI rates, which are also of particular interest to hospital epidemiologists. All of the comparisons use only parameters from the transition matrix. Time-dependent expressions necessitate derivation of parameterizations--functions of the transition probabilities. Our estimates for patient LOH in distinct health states (holding times), also used for testing the geometric distribution in Section 3, are presented in Kastner (1980a).

A. Relative Infection Rates

The transition probabilities are measures of daily

transitions; p_{ij} is the probability of a patient moving from $X_{n-1} = i$ to $X_n = j$ on any given day n. The relative daily infection rate at site j for (uninfected) patients from PD_i is

$$R_{ij} = \tilde{p}_{ij} / \sum_{k=1}^{4} \tilde{p}_{ik}$$

where $\tilde{p}_{ij} = p_{PDi,Sj}$, $i,j = 1,2,3,4.$

The relative daily death rate for patients from PDi is

$$D_i = \tilde{p}_{i5}/(\tilde{p}_{i5} + \tilde{p}_{i6}).$$

Similar expressions obtain for surgical patients. Values for these formulae appear in Table 4 for joint hospital groups 3,6 and 5,8 by admission day.

Table 4

B. Probability of Acquiring an NI

To estimate the probability that a patient (ever) acquires an infection during his LOH we consider the first passage time probability; for $n \geq 1$

$$f_{ij}(n) = Pr[X_n = j; X_m \neq j, 1 \leq m \leq n-1 | X_0 = i]$$

$$= p_{ii}^{n-1} p_{ij} + \sum_{m=1}^{n-1} p_{ii}^{n-m-1} p_{is} p_{ss}^{m-1} p_{sj}.$$

Where i represents PDi, $i = 1,2,3,4$; j represents Sj, $j=1,2,3,4$; and s represents SUR1. The result holds for $n \geq 2$ and $f_{ij}(1) = p_{ij}$.

This result obtains due to the upper triangular structure of the transition matrix. To assess the probability of ever entering an NI state,

TABLE 4

RELATIVE DAILY NI RATES AND DEATH RATES AMONG UNINFECTED PATIENTS

Hospital Group	Admission Day	Primary Diagnosis Group	Relative Daily Infection Rates (%) UTI	BACT	LRI	SWI	Death Rate (%)
3 and 6 Moderate Size Non-Medical School Affiliated	Weekday	PD1	72.41	3.45	8.62	15.52	5.72
		PD2	35.71	4.76	59.52	-	6.46
		PD3	61.54	3.85	3.85	30.77	0.86
		PD4	49.23	3.08	10.77	36.92	0.70
		SUR1	43.04	-	24.05	32.91	2.45
	Weekend	PD1	84.85	3.08	7.69	4.62	5.24
		PD2	51.06	6.38	36.17	6.38	5.34
		PD3	69.74	2.63	11.84	15.79	0.94
		PD4	40.74	4.94	9.88	44.44	10.38
		SUR1	36.71	-	20.25	43.04	3.61
5 and 8 Large, Medical School Affiliated	Weekday	PD1	80.00	7.27	5.45	7.27	4.84
		PD2	58.67	6.67	34.67	-	5.4
		PD3	62.12	-	18.18	19.70	9.68
		PD4	60.00	5.33	10.67	24.00	0.77
		SUR1	40.93	4.15	20.21	34.72	2.99
	Weekend	PD1	66.67	12.34	9.88	11.11	5.34
		PD2	61.96	4.35	33.70	-	8.06
		PD3	47.76	5.97	34.33	11.94	1.63
		PD4	66.67	3.45	4.60	25.28	6.88
		SUR1	46.67	3.33	21.67	28.33	1.48

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}(n) = p_{ij}/(1-p_{ii}) + p_{is}p_{sj}/(1-p_{ss})(1-p_{ii}) \ , \ i,j = 1,2,3,4 \ .$$

We present these probabilities in Table 5.

Table 5

The last row of this table presents the probability of a randomly selected patient ever acquiring an NI during her/his LOH. We compute these probabilities from

$$\Pr[\text{ever contracting NI } j] = \sum_{i=1}^{4} a_i f_{ij} \ .$$

Where a_i is the probability of entering the system into state PDi from Table 1. A similar approach yields relative daily secondary infection rates conditional on a primary NI, and the relative daily death rate among patients with a primary NI.

We construct the cumulative daily infection incidence function by using first passage time distributions for $n \geq 2$:

$$F_{ij}(n) = \sum_{m=1}^{n} f_{ij}(m)$$

$$= \begin{cases} \dfrac{p_{ij}(1 - p_{ii}^{n})}{(1 - p_{ii})} + \dfrac{p_{is}p_{sj}}{(1 - p_{ii})}\left[\dfrac{1 - p_{ss}^{n-1}}{1 - p_{ss}} - \dfrac{p_{ii}(p_{ii}^{n-1} - p_{ss}^{n-1})}{(p_{ii} - p_{ss})}\right], & p_{ii} \neq p_{ss} \\ \dfrac{p_{ij}(1 - p_{ii}^{n})}{(1 - p_{ii})} + \dfrac{p_{is}p_{sj}}{(1 - p_{ii})}\left[\dfrac{1 - p_{ii}^{n-1}}{1 - p_{ii}} - (n - 1)p_{ii}^{n-1}\right], & p_{ii} = p_{ss} \end{cases}$$

Let i be PD4 and compare $F_{ij}(n)$ for $j = 1,2,3,4$, $n = 2,5(5)30$ and $n \longrightarrow \infty$, as well as $F_{ij}(n)/f_{ij}$, which yields the difference in the

TABLE 5

Probability of Acquiring an NI, Weekday, Hospital Cells 5 and 8

Entering State	NI S1	S2	S3	S4	Total
PD1	.0487	.0045	.0067	.0105	.0704
PD2	.0461	.0052	.0265	.0062	.0840
PD3	.0407	.0007	.0133	.0166	.0713
PD4	.0347	.0032	.0087	.0175	.0641
Randomly Selected from PD1-PD4	.0407	.0034	.0106	.0142	.0689

rate at which the cumulative incidence approaches the probability of (ever) acquiring infection j. The transition probabilities (from Figure 1) are: $p_{ii} = 0.8919$, $p_{ss} = 0.8937$, $p_{is} = 0.0117$, and p_{ij}, p_{sj}, respectively, (0.0044, 0.0004, 0.0003, 0.0004), (0.0079, 0.0008, 0.0039, 0.0067). Results are in Table 6 and the last row agrees with the fourth row of Table 5.

Table 6

A fascinating outcome of this analysis is that the rate at which the cumulative incidence approaches the probability of (ever) acquiring an infection, $F_{ij}(n)/f_{ij}$, does not differ significantly as a function of the infection type. Note, also, that even by day 5 a patient has already achieved 40% of her/his chance of contracting an NI and 80% by day 15. Tables for PD2, PD3 and PD4 differ somewhat in the rapidity of convergence but have the same property for $F_{ij}(n)/f_{ij}$.

C. Time to Absorption and Corresponding Ramifications

The mean time to absorption in S5 or S6 from any state is the equivalent of the mean remaining LOH for a patient. To compute this quantity we consider the submatrix of P for the transient states denoted by Q; Q is a 14 x 14 upper triangular matrix with components q_{ij}. Let $N = (I-Q)^{-1}$ be the fundamental matrix which we employ to compute the mean time to absorption, see Lemmas 1 and 2 in the Appendix.

Let m be the (row) vector of mean times to absorption; $m^T = N E = (m_1, m_2, \ldots, m_{14})^T$, where E is a (column) vector of ones. By Lemma 1 the ith entry in m is given by

$$m_i = \sum_{j=1}^{14} \bar{q}_{ij} = \sum_{j=i}^{14} \bar{q}_{ij} = 1/p_i + \sum_{j=i+1}^{14} \sum_{k=i+1}^{j} (q_{ik}/p_{ii})\bar{q}_{kj} .$$

Results appear in Table 7.

TABLE 6

CUMULATIVE DAILY INFECTION INCIDENCE
WEEKDAY ADMISSIONS FOR HOSPITAL CELLS 5 AND 8
WHERE PRIMARY DIAGNOSIS GROUP IS PD1

n	$F_{ij}(n)$ UTI	BACT	LRI	SWI	$F_{ij}(n)/f_{ij}$ (%) UTI	BACT	LRI	SWI
2	.0091	.0008	.0009	.0014	.1868	.1856	.1416	.1345
5	.0204	.0019	.0025	.0039	.4191	.4185	.3777	.3698
10	.0326	.0030	.0043	.0066	.6685	.6688	.6373	.6294
15	.0395	.0037	.0053	.0082	.8110	.8120	.7901	.7827
20	.0435	.0040	.0059	.0092	.8925	.8939	.8798	.8731
25	.0457	.0042	.0062	.0097	.9390	.9407	.9325	.9264
30	.0470	.0044	.0065	.0101	.9655	.9675	.9634	.9577
∞	.0487	.0045	.0067	.0105				

Table 7

As expected, among hospitals not affiliated with medical schools mean times to absorption for weekend admissions are usually the same or longer than the corresponding means for weekday admissions. Exceptions occur with secondary infections and surgery subsequent to an infection. Surprisingly the mean times to absorption for patients in medical school affiliated hospitals are mixed for weekend versus weekday admissions; they are longer than those for the other larger hospitals (not in cells 5 and 8). Patients with no infections in PD4 stay the least time, while patients in PD1, i.e., vascular diseases, stay longer than the other uninfected patients.

Among patients with a primary infection, it appears that those with BACT exhibit the longest mean time to absorption, followed by those who acquire a UTI. In the case of secondary infections, bacteremia is associated with longest mean time to absorption, followed by LRI and UTI. Also, patients who go through surgery after acquiring an NI have a mean absorption time which is 2.5 to 11 days longer than the mean time of patients who undergo surgery when uninfected.

We consider two absorbing states, discharge and death; let d denote either of them. We compute the probability of absorption in state d by

$Y_d = NZ_d = (1-Q)^{-1}Z_d$, with the ith element being

TABLE 7

MEAN TIME TO ABSORPTION (DAYS) FOR HOSPITAL GROUPS 5 AND 8
AND ADMISSION DAYS

Starting Transient State	Large Medical School Affiliated Weekday Admission	Large Non-Medical School Affiliated Weekday Admission	Large Medical School Affiliated Weekend Admission	Large Non-Medical School Affiliated Weekend Admission
PD1	9.47	11.56	9.33	11.55
PD2	8.84	11.25	7.48	11.60
PD3	8.27	10.39	8.94	11.63
PD4	6.13	8.10	6.19	8.29
SUR1	11.03	12.61	9.72	13.38
S1	13.80	18.69	13.76	18.99
S2	15.36	21.44	9.65	23.35
S3	12.75	15.16	14.90	16.16
S4	13.00	17.47	11.42	17.59
SUR10	13.86	23.67	12.22	17.70
S10	12.96	26.14	17.59	25.76
S20	27.33	44.90	14.00	37.50
S30	27.00	16.89	10.33	28.75
S40	16.24	24.75	16.43	23.00

$$(Y_d)_i = \begin{cases} \sum_j q_{ij} (Z_d)_j & i = 1 \\ (Z_d)_i / p_{ii} & i > 1 \end{cases}$$

where $N = (I-Q)^{-1}$ for a given Q matrix; Z_d is a column vector of transition probabilities for any of the states included in the Q matrix to the absorbing state d, with $(Z_d)_i$ the ith element of Z_d. For example, let Q include states S1, SUR10, S10, S20, S30 and S40 and Z_d represent the transition probabilities from these states to absorbing state S5, discharge. In Table 8 we summarize the corresponding absorption probabilities.

Absorption probabilities in state S6 (death) are generally higher in hospitals affiliated with medical schools. Also, these probabilities are higher among patients admitted over the weekend as compared to weekday admissions. Among patients admitted on weekdays the average of these probabilities is 0.1731 in large medical school affiliated hospitals vs. 0.0939 in the non-affiliated; among patients admitted on weekends the corresponding mean probabilities are 0.2266 vs. 0.1596. One explanation for the latter may be that weekend admissions are not always scheduled, are often sequelae of emergencies, and evidence critical health conditions. Also, absorption probabilities in state S6 are higher among (a) patients who undergo surgery after acquiring a primary NI; and (b) patients with a

---

Table 8

---

secondary NI as compared to primary NI patients. These observations do not necessarily imply that an NI leads to a higher probability of death;

TABLE 8

THE PROBABILITIES OF ENDING IN ABSORBING STATE S(j)
HAVING STARTED IN TRANSIENT STATE S(i)

Starting Transient State	Cells 3 and 6 Weekday Non-Medical School Affiliated Absorbing States		Cells 5 and 8 Weekday Medical School Affiliated Absorbing States		Cells 3 and 6 Weekend Non-Medical School Affiliated Absorbing States		Cells 5 and 8 Weekend Medical School Affiliated Absorbing States	
	S5	S6	S5	S6	S5	S6	S5	S6
PD1	0.944	0.056	0.948	0.052	0.946	0.056	0.937	0.063
PD2	0.937	0.063	0.937	0.063	0.944	0.056	0.914	0.086
PD3	0.988	0.012	0.981	0.019	0.983	0.017	0.970	0.030
PD4	0.989	0.011	0.983	0.017	0.982	0.018	0.980	0.020
SURL	0.969	0.031	0.955	0.045	0.956	0.044	0.958	0.042
S1	0.920	0.080	0.887	0.113	0.920	0.080	0.822	0.178
S2	0.861	0.139	0.684	0.316	0.371	0.629	0.536	0.464
S3	0.903	0.097	0.825	0.175	0.817	0.183	0.828	0.172
S4	0.955	0.045	0.933	0.067	0.931	0.069	0.888	0.112
SUR10	0.926	0.074	0.799	0.201	0.971	0.029	0.873	0.127
S10	0.800	0.200	0.714	0.286	0.824	0.176	0.680	0.320
S20*	1.000	-	0.500	0.500	0.667	0.333	0.250	0.750
S30	0.667	0.333	0.556	0.444	0.667	0.333	0.500	0.500
S40	0.920	0.080	0.875	0.125	0.857	0.143	0.692	0.308

*More observations may be needed to verify probabilities associated with low frequency events.

patient with complications and severe illness may stay longer in the hospital and thus become more susceptible to acquiring an NI. Such a patient's chances of mortality are larger to begin with and the NI may be just an additional factor. Therefore, it is reasonable to refine the model with a patient risk index, to test whether there really is a significant difference in the probability of death, mean time to absorption, and other measures of interest. See, for example, Hooton et al. (1980).

Among uninfected patients, the mortality rate is larger for patients in the PD2 primary diagnosis group, i.e., respiratory and lung diseases, followed by PD1, vascular diseases. Our primary diagnosis grouping is too coarse to permit major policy recommendations, but the results suggest significant potential for further research and policy analysis.

6. POST-CONDITIONED TABOO PROBABILITIES

In traditional cohort comparisons for epidemiologic analyses, the data requirements for conditional rate estimations may be significant. To maximize the utility of existing data, in this section we extend our analysis beyond known limiting results and derive probabilities of absorption conditioned on constraints on visits to one or more prespecified states. See Shachtman et al. (1981b) for the methodology.

Analyses using our model yield estimates of differences in the cumulative distributions, or proportions, of patients acquiring secondary infections. Furthermore, by computing cumulative probabilities with conditions defining distinct primary infections as taboo states, we can compare the effect of each primary infection, or a combination of primary infections, on the occurrence of secondary infections, as well

as on the time to absorption.

The probability that the process is in state k at time n without having entered state h, given the process started at state j, is

$$ {}_hp_{jk}(n) = Pr\{X_n = k;\ X_m \neq h,\ 1 \leq m \leq n-1 | X_o = j\},\ h \neq k $$

and is called a post-conditioned taboo probability. The corresponding post-conditioned first passage taboo probability is

$$ {}_hf_{jk}(n) = Pr\{X_n = k;\ X_m \notin \{h, k\},\ 1 \leq m \leq n-1 | X_o = j\},\ h \neq k. $$

Let P be the transition probability matrix, and let P_h be a modified matrix defined by: $P_m = I_h P$, where I_h is an identity matrix with the diagonal element corresponding to taboo state h set to zero. Thus the matrix P_h will have a row of zeros corresponding to the taboo state. If $R^1 = P$, then the n-step post-conditioned taboo transition probability matrix is given by $R^n = R^{n-1}P_h = P(P_h)^{n-1} = ({}_hp_{jk}(n))$.

Let k correspond to a secondary UTI (state S10), the n-step post conditioned first passage taboo probability is

$$ {}_hf_{SO,S10}(n) = Pr\{X_n = S10;\ X_n \notin \{h, S10\},\ 1 \leq m \leq n-1 | X_o = SO\},\ h \neq S10 $$

and h may be S1, S2, S3, S4 or SUR1. Note that SO represents a selected initial uninfected state. This yields

$$ {}_hp_{SO,S10}(n) = \sum_{m=1}^{n} {}_hf_{SO,S10}(m)\ {}_hp_{S10,S10}(n-m), \quad n \geq 2, $$

where h = S1, S2, S3, S4 or SUR1.

Let S10 be absorbing for computational purposes; then ${}_hp_{S10,S10}(n-m) = 1$ for all $n \geq 2$. Thus ${}_hp_{SO,S10}(n) = \sum_{m=1}^{n} {}_hf_{SO,S10}(m)$.

Let h = S1 (primary UTI) be taboo, and consider ${}_{S1}p_{SO,S10}(n)$, n=1,...,40.

The results shown in Table 9 indicate an initial increase and then a decrease in the post-conditioned probabilities as a function of n. (All the numerical examples are for large, medical school affiliated

hospitals, cells 5 and 8.)

Table 9

In this table ${}_{S1}p_{SUR1,S10}(n)$ is consistently largest for all n and ${}_{S1}p_{PD4,S10}(n)$ is second. This indicates that surgery patients, without a primary UTI, have a higher probability of acquiring a secondary UTI than patients who do not have surgery. Also, the patients who enter the system through the uninfected state PD4 have a higher probability for a secondary UTI after any non UTI primary NI. Note that PD4 is less specific in terms of the organ system definition than the other initial states.

We now consider a more complex post-conditioning structure:

$$Hi\,f_{SO,S10}(n) = \Pr\{X_n = S10,\ X_m \notin Hi\ ,\ 1 \le m \le n-1 \mid X_o = SO\};$$

with Hi defined as $Hi = \{S1, S2, S3, S4\} \sim \{Si\}$, $i = 1, 2, 3, 4$. This means that $X_m \notin Hi$ if and only if $X_m = SO$ or Si. From Shachtman et al. (1980b) we have the following results, $n \ge 2$ and $h \ne k$:

(1) $$h p_{jk}(n) = \sum_{r \ne h} h p_{jr}(n-1) p_{rk} \ ;$$

(2) $$h p_{jk}(n) = \sum_{m=1}^{n} h p_{jk}(m)\, h p_{kk}(n-m) = h f_{jk}(n) + \sum_{m=1}^{n-1} h f_{jk}(m)\, p_{kk}(n-m)$$

which leads to the following iterative expression for n-step post-conditioned first-passage taboo probabilities:

$$h f_{jk}(n) = h p_{jk}(n) - \sum_{m=1}^{n-1} h f_{jk}(m)$$

with

$$h f_{jk}(1) = f_{jk} = p_{jk} \text{ and } k \text{ absorbing.}$$ Since

$$Hi\,p_{SO,S10}(n) = p_{SO,Si}(n-1)\ p_{Si,S10}\ ,$$

TABLE 9

THE n-STEP POST-CONDITIONED TABOO PROBABILITIES

$_{S1}P_{S0,S10}^{(n)}$

Where S0 is PD1, PD2, PD3, PD4 or SUR1.

(Each entry is $10^5 \times {}_{S1}P_{S0,S10}^{(n)}$)

	S0				
n	PD1	PD2	PD3	PD4	SUR1
2	0.77	0.96	1.00	1.56	3.7
5	5.98	7.21	7.33	10.90	27.6
10	17.45	20.18	20.31	27.58	75.6
20	30.36	32.90	32.57	38.48	119.0
30	28.17	29.10	28.40	30.95	102.3
40	20.26	20.19	19.46	19.97	69.2

for each primary diagnosis initial uninfected state (S0) we have,

$n \geq 2$,

$$p_{S0,Si}^{(n-1)} = \sum_{m=1}^{n-1} f_{S0,Si}^{(m)} \; p_{Si,Si}^{(n-m-1)} =$$

$$= \begin{cases} p_{S0,Si} \; (p_{Si,Si}^{n-1} - p_{S0,S0}^{n-1})/(p_{Si,Si} - p_{S0,S0}), & p_{Si,Si} \neq p_{S0,S0} \\ (n-1)p_{S0,Si} \; p_{Si,Si}^{n-2} , & p_{Si,Si} = p_{S0,S0} \end{cases}$$

Thus, by combining the two previous expressions, we get

$$Hi p_{S0,S10}^{(n)} =$$

$$= \begin{cases} p_{S0,Si} \; p_{Si,S10} \; (p_{Si,Si}^{n-1} - p_{S0,S0}^{n-1})/(p_{Si,Si} - p_{S0,S0}), & p_{Si,Si} \neq p_{S0,S0} \\ (n-1) \; p_{S0,Si} \; p_{Si,S10} \; p_{Si,Si}^{n-2} , & p_{Si,Si} = p_{S0,S0} \end{cases}$$

Let Hi = HS1 , i.e., primary UTI; Table 10 summarizes the n-step post conditioned probability computations corresponding to all primary NI states, with UTI (S1) as the taboo state and secondary UTI (S10) as the "absorbing" state; the computations are based on observations of weekday admission patients in large medical school affiliated hospitals.

Table 10

In all but the last case presented in the Table above, the n-step probability function reaches its maximum at n=10. This means that after 10 days the patient's probability of a secondary UTI, even though s/he acquired a primary UTI, decreases as her/his LOH increases.

TABLE 10

The *n*-Step Post-Conditioned Taboo Probabilities

$_{HS1}P_{S0,S10}^{(n)}$

Where S0 is PD1, PD2, PD3, PD4 or SUR1

(Each entry is $10^5 \times {}_{HS1}P_{S0,S10}^{(n)}$)

	S0					
n	PD1	PD2	PD3	PD4	SUR1	$_{S1}P_{SUR1,S10}^{(n)}$
2	2.27	0.81	1.73	1.73	3.67	3.72
5	6.43	2.28	4.78	4.25	10.72	27.60
10	8.17	2.87	5.85	4.40	14.28	75.65
20	5.57	1.92	3.73	2.21	10.60	119.04
30	2.78	0.95	1.77	0.91	5.71	102.25

Not surprisingly, these probabilities are highest among surgical patients. One possible reason for their higher UTI incidence is that for some period after surgery a urinary catheter is often used. Catheters, believed to be a major cause of UTIs when insufficient precautions are taken, are often associated with a primary UTI patient acquiring a secondary UTI. Patients in the PD1 state of primary diagnosis, vascular diseases, have the second highest incidence of acquiring a secondary UTI after a primary UTI.

It is interesting to compare ${}_{HS1}P_{SUR1,S10}(n)$ and ${}_{S1}P_{SUR1,S10}(n)$; in the latter case the only taboo state is primary UTI, while in the first expression all states of primary infection, except primary UTI, are taboo states. See the last two columns of Table 10. The probabilities in which S1 is the only taboo state increase with n, even among surgery patients, to a maximum at n=21, and then decrease; this supports the above argument that, even among surgery patients after a certain LOH, the probability of acquiring a secondary UTI is larger when a primary UTI is already present.

From a theorem developed by Shachtman et al. (1981b), we have that the probability distribution $\{{}_hF_{jk}(n): n = 0, 1,\ldots,\}$ where ${}_hF_{ij}(n) = \sum_{m=1}^{n} {}_hf_{ij}(m)$, is the post-conditioned time to absorption from state i distribution. By using the expression for ${}_{Hi}P_{S0,S10}(n)$, and definitions from Shachtman et al.(1981b), in a general form, we develop an expression for the pre-conditioned first passage taboo probability which applies to our models. We define the pre-conditioned first passage taboo probability by

$$ {}_hq_{ij}(n) = Pr\{X_n = j \mid X_m \notin \{h,j\} , 1 \le m \le n - 1 ; X_0 = i\} . $$

From a theorem, Ibid.,

$$ {}_hq_{ij}(n) = {}_hf_{ij}(n)/[1 - {}_jF_{ih}(n-1) - {}_hF_{ij}(n-1)] , \quad n \ge 2 \text{ and } h \ne j . $$

We extend this result to our collection, Hi, of taboo states. Let Hi be defined as before and consider $Hi^{q}_{S0,S10}(n)$. Then, following the above results we get

$$Hi^{q}_{S0,S10}(n) = Pr\{X_n = S10 \mid X_m \notin Hi \cup \{S10\},\ 1 \le m \le n-1 \mid X_0 = S0\},$$

which is explicitly given by

$$Hi^{q}_{S0,S10}(n) = \sum_{r \neq S0,Si} Hi^{q}_{S0,r}(n-1)\ Hi^{q}_{r,S10}(1) =$$

$$= Hi^{q}_{S0,S0}(n-1)\ p_{S0,S10} + Hi^{q}_{S0,Si}(n-1)\ p_{Si,S10}.$$

But for our model, $Hi^{q}_{S0,Si}(m) = p_{S0,Si}\ p^{m-1}_{S0,S0}$, and $p_{S0,S10} \equiv 0$, hence,

$$Hi^{q}_{S0,S10}(n) = p^{n-2}_{S0,S0}\ p_{S0,Si}\ p_{Si,S10}$$

The corresponding cumulative pre-conditioned first passage taboo probability is

$$\sum_{n=2}^{M} Hi^{q}_{S0,S10}(n) = \sum_{n=2}^{M} p^{n-2}_{S0,S0}\ p_{S0,Si}\ p_{Si,S10} =$$

$$= p_{S0,Si}\ p_{Si,S10}\ (1 - p^{M-1}_{S0,S0})/(1 - p_{S0,S0}), \quad M \ge n \ge 2.$$

The above result, with Si = S1, is the probability that a patient moves from state S0 (uninfected) to state S10 (secondary UTI) for the first time at time $n \ge 2$, without visiting the states S2 (BACT), S3 (LRI), or S4 (SWI), by time $M \ge n$. Table 11 summarizes the numerical results for M = 2, 5, 10, 20 and 30. As with the n-step post conditioned taboo probabilities, these are highest when the uninfected state is the surgical state (SUR1) or when the patients enter the system from the vascular disease primary diagnosis group (PD1).

Table 11

TABLE 11

THE CUMULATIVE PRE-CONDITIONED FIRST PASSAGE TIME PROBABILITIES

$_{HS1}q_{S0,S10}^{(M)}$

where S0 is PD1, PD2, PD3, PD4 or SUR1

(Each entry is 10^5 x $_{HS1}q_{S0,S10}^{(M)}$)

	S0				
M	PD1	PD2	PD3	PD4	SUR1
2	2.268	0.810	1.728	1.728	3.672
5	7.482	2.656	5.573	5.007	12.445
10	12.585	4.431	9.099	7.191	21.718
20	16.476	5.744	11.499	8.056	29.812
30	17.473	6.065	12.018	8.135	32.347

We now propose a parameterization to assess the effect of relative LOH in a state of secondary infection on probabilities of discharge or death. Let n be the LOH from the onset day of a primary infection for a patient contracting multiple infections, and r be the onset day of her/his secondary infection. We define

$$g(n,r) = Pr\{X_n = d;\ X_{n-1} = S_j;\ X_r = S_j\ ;\ X_{r-1} = Si | X_o = Si\}\ ,\ 1 \le r \le n-1\ ,$$

where d is either absorbing state S5 or S6; S_j , $j = 10, 20, 30, 40$, is a secondary infection state, and Si, $i = 1, 2, 3, 4$, is a primary infection state. Then

$$g(n,r) = a_{Si}\, p^{r-1}_{Si,Si}\ p_{Si,Sj}\ p^{n-r-1}_{Sj,Sj}\ p_{Sj,d} / a_{Si}$$

$$= \frac{p_{Si,Sj}\ p_{Sj,d}}{p_{Si,Si}\ p_{Sj,Sj}} \left(\frac{p_{Si,Sj}}{p_{Sj,Sj}}\right)^r p^n_{Sj,Sj} \quad , n \ge 2,\ 1 \le r \le n-1,$$

and a_{Si} is the prior proability for state Si.

Suppose the overall LOH is fixed, $n = n_o$; then, for fixed transition probabilities:

$$g(n_o,r) = C\ (p_{Si,Sj}/p_{Sj,Sj})^r,\quad 1 \le r \le n-1,$$

where C depends only on the transition probabilities and n_o .

For fixed n_o , i, j and d, $g(n_0,r)$ is an increasing function for $(p_{Si,Si}/p_{Sj,Sj}) > 1$, and a decreasing function of $(p_{Si,Si}/p_{Sj,Sj}) < 1$.

If the LOH in the primary infection state is fixed at r_o , then $g(n,r_o)$ tends to 0 as n increases. Thus, if the probability of retaining a primary infection is higher than that of retaining a secondary infection, the probability of absorption increases as the relative LOH, r, increases.

For illustration, we again consider patients admitted on weekdays in large hospitals not affiliated with medical schools (cells 5 and 8). Let Si be a primary UTI (S1) and Sj a secondary UTI (S10); we consider both absorbing states. The expression for $g(n_o,r)$, with n_o = 31, yields the results in Table 12. In this table C(d) denotes the constant term C, where d can be either discharge (S5) or death (S6); $g(n_o,r,d)$ is the absorption probability conditioned on the relative duration, n - r, of the secondary infection. In this case $p_{Si,Si} < p_{S10,S10}$ and $g(n_o,r,d)$ decreases as r increases for fixed n_o = 31, and d = S5 or S6. The expression $g(n_o,r,S6)/g(n_o,r,S5) = p_{S10,S6}/p_{S10,S5} = 0.250$ is the ratio of conditioned absorption probabilities, of death to discharge, for multiple UTIs.

TABLE 12

---

8. SUMMARY AND CONCLUSIONS

We posed four research questions: (1) How can we estimate the hospitalization time spent in distinct health states? (2) How much longer does a patient with an NI spend in the hospital than a patient with similar characteristics who has not acquired an NI? (3) What is the effect on these measurements of intervening variables, e.g., surgery? and (4) If we could eliminate certain types of NIs, how would this influence length of hospitalization? Our findings provide models, tests, parameterizations and numerical results for these and related questions. Furthermore, our methodological framework permits applications beyond

TABLE 12

DISCHARGE AND DEATH PROBABILITIES CONDITIONED ON RELATIVE DURATIONS OF SECONDARY INFECTIONS, $n_o = 31$

(Each entry is 10^4 times the original expression.)

r	C(S5)	$g(n_o,r,S5)$	C(S6)	$g(n_o,r,S6)$
1	0.329	0.320	0.082	0.0798
5		0.303		0.0755
10		0.283		0.0705
20		0.246		0.0615
30		0.215		0.0536

those employed in this paper.

Perhaps the most significant aspect of this study is its development of a control and prediction tool for hospital medical epidemiologists and planners. Flexibility is embodied in the potential for modifying health state definitions and the design of the model. By allowing for more specific and refined initial, i.e., uninfected diagnostic coding, we may achieve more homogeneous patient cohorts for analyses. Currently under development at CDC, see Hooton et al. (1980), is a patient NI risk index which accounts for a large number of variables including age, sex, purpose of hospitalization, types of underlying or previously acquired illnesses, duration of surgery, duration of treatment with urinary catheters, and others. When initial health states are defined in terms of such a risk index, our models and their sequelae provide quantitative measures of the differences among patient risk groups.

We also recommend the development of an algorithm which identifies the NI resolution time ("offset" or end of the NI episode). Then we can generalize the Markov chain network to be non-hierarchical, allowing the patient's return to an uninfected health state prior to discharge, death or acquisition of additional NIs. Both tasks require further research.

APPENDIX I

Lemma 1:

Let Q be an nxn upper triangular matrix with entries (q_{ij}); thus $q_{ij} = 0$ for all $i > j$, and $0 < q_{ii} < 1$. Let $p_{ij} = 1 - q_{ij}$. Then, (i) $(I-Q)^{-1}$ always exists. (ii) Let $\bar{q}_{ij}$ be the (i,j) element in $(I-Q)^{-1}$. Then

$$\bar{q}_{ij} = \begin{cases} 1/p_{ii} & , \ i = j \\ \sum_{k=i+1}^{j} (q_{ik}/p_{ii})\ \bar{q}_{kj} & , \ i < j \\ 0 & , \ i > j \end{cases}$$

Note, the Lemma allows an iterative determination of $\bar{q}_{ij}$; e.g., we first compute

$$\bar{q}_{j-1,j} = (q_{j-1,j}/p_{jj})\ \bar{q}_{j,j} = (q_{j-1,i}/p_{jj}^2)$$, then we compute

$\bar{q}_{j-2,j} = f(\bar{q}_{j-1,j}$ and known transition probability terms), etc.

Proof: (i) Let $(I-Q) = \tilde{q}_{ij}$. By the assumptions we have that

$$\tilde{q}_{ij} = \begin{cases} \leq 0 & , \ i < j \\ = 0 & , \ i > j \\ \varepsilon\ (0,1) & , \ i = j \ . \end{cases}$$

Thus, (I-Q) is also upper triangular and its determinant is the product of the diagonal elements.

$$|| \ I-Q \ || = \prod_{j=1}^{n} \tilde{q}_{ij} > 0,$$ nonsingular, and $(I-Q)^{-1}$ exists. *

(ii) Proof by induction. Let n=1, then $(I-Q)^{-1} = 1/p_{11}$; for

$n = 2$ we have by (1) that

$$\bar{q}_{11} = 1/p_{11}\ , \quad \bar{q}_{22} = 1/p_{22}\ ,$$

$$\bar{q}_{12} = (q_{12}/p_{11})\ \bar{q}_{22} = (q_{12}/p_{11}p_{22}),\ \text{and}\ \bar{q}_{21} = 0.$$

We get,

$$(I-Q)\ (I-Q)^{-1} = \begin{bmatrix} p_{11} & -q_{12} \\ 0 & p_{22} \end{bmatrix} \begin{bmatrix} 1/p_{11} & q_{12}/p_{11}p_{22} \\ 0 & 1/p_{22} \end{bmatrix} = I.$$

We now assume the relationship holds for $(n-1)$ and show it holds for n.

The result follows immediately from Lemma 2.

Lemma 2.

Let Q be an nxn ($n \geq 2$) upper triangular matrix with entries (q_{ij}) such that $q_{ij} = 0$ for all $i > j$, $p_{ij} = 1-q_{ij}$, and $0 < q_{ij} < 1$. Let K be an $(n-1) \times (n-1)$ matrix with entries $k_{ij} = q_{ij}$, $i, j = 1,\ldots,n-1$. Then if $\bar{q}_{ij}$ and $\bar{k}_{ij}$ are, respectively, the (i, j) the element in $(I-Q)^{-1}$ and $(I-K)^{-1}$,

(i) $\bar{q}_{ij} = \bar{k}_{ij}$, $i, j = 1,\ldots,n-1$;

(ii)

$$(2)\quad \bar{q}_{in} = \begin{cases} 1/p_{nn} & , \ i=n \\ \sum_{\ell=i+1}^{n-1} (k_{i\ell}/p_{ii})\ \bar{q}_{\ell n} + (q_{in}/p_{ii})\ \bar{q}_{nn} & , \ i=1,\ldots,n-1 \end{cases}$$

Proof: Let n=2; then

$$(I-Q)^{-1} = \begin{bmatrix} \frac{1}{p_{11}} & \frac{q_{12}}{p_{12}\,p_{22}} \\ 0 & \frac{1}{p_{22}} \end{bmatrix} \quad ; \quad (I-K)^{-1} = 1/p_{11} \text{ , and}$$

(2) holds. Since the first (i-1) elements of the ith row in (I-Q) are 0, we have

$$\sum_{j=i}^{n} \tilde{q}_{ij}\ \bar{q}_{in} = \begin{cases} 0, & i \neq n \\ 1, & i = n \end{cases}$$

We now consider the inner product of the nth column of $(I-Q)^{-1}$ and the ith row of (I-Q). This yields the ith entry in the nth column of the product $(I-Q)\ (I-Q)^{-1}$. This element is given by

$$\sum_{j=1}^{n} \tilde{q}_{ij}\ \bar{q}_{in} = \sum_{j=i}^{n} \tilde{q}_{ij}\ \bar{q}_{in} = \tilde{q}_{ii}\ \bar{q}_{in} + \sum_{j=i+1}^{n} \tilde{q}_{ij}\ \bar{q}_{in} =$$

$$= p_{ii}/p_{nn} + \sum_{j=i+1}^{n} (-q_{ij})\ \bar{q}_{in} = \begin{cases} 0, & i \neq n \\ 1, & i = n \end{cases}$$

Suppose i=n; then

$$\sum_{j=1}^{n} \tilde{q}_{ij}\ \bar{q}_{jn} = \tilde{q}_{nn}\ \bar{q}_{nn} \text{ , which equals 1}$$

if and only if $\bar{q}_{nn} = 1/\tilde{q}_{nn} = 1/p_{nn}$

Continuing by construction we let i=n-1; then

$$\sum_{j=1}^{n} \tilde{q}_{n-1,j}\ \bar{q}_{jn} = \sum_{j=n-1}^{n} \tilde{q}_{n-1,j}\ \bar{q}_{jn} = \tilde{q}_{n-1,n-1}\ \bar{q}_{n-1,n} + \tilde{q}_{n-1,n}\ \bar{q}_{n,n}$$

which using the result from the previous step, equals 0 if and only if

$$\bar{q}_{n-1,n} = \frac{-\tilde{q}_{n-1,n}\ \bar{q}_{nn}}{\tilde{q}_{n-1,\,n-1}} = \frac{q_{n-1,n}}{p_{n-1,n-1}\ p_{nn}} \quad . \quad \text{Now let } i = n-2;$$

then

$$\sum_{j=1}^{n} \tilde{q}_{n-2,j}\ \bar{q}_{jn} = \tilde{q}_{n-2,n-2}\ \bar{q}_{n-2,n} + \tilde{q}_{n-2,n-1}\ \bar{q}_{n-1,n} + \tilde{q}_{n-2,n}\ \bar{q}_{n,n} = 0.$$

if and only if

$$\bar{q}_{n-2,n} = -\frac{\tilde{q}_{n-2,n}\ \bar{q}_{n,n}}{\tilde{q}_{n-2,n-2}} - \frac{\tilde{q}_{n-2,n-1}\ q_{n-1,n}}{\tilde{q}_{n-2,n-2}} =$$

$$= \frac{q_{n-2,n-1}\ q_{n-1,n}}{p_{n-2,n-2}\ p_{n-1,n-1}\ p_{nn}} + \frac{q_{n-2,n}}{p_{n-2,n-2}\ p_{nn}}$$

similarly, we have that

$$\sum_{j=1}^{n} \tilde{q}_{ij}\ \bar{q}_{jn} = 0 \text{ if and only if } \bar{q}_{jn} = -\frac{\sum_{j=i+1}^{n} \tilde{q}_{ij}\ \bar{q}_{jn}}{\tilde{q}_{ii}} =$$

$$= \frac{\sum_{j=i+1}^{n} q_{ii}\ q_{jn}}{p_{ii}} = \sum_{j=i+1}^{n-1} \frac{q_{ij}}{p_{ii}}\ \bar{q}_{jn} + \frac{q_{in}}{p_{ii}}\ \bar{q}_{nn} \ ;$$

this proves the Lemma. *

REFERENCES

American Journal of Epidemiology. 1980. Special Issue: The SENIC Project, 111, no. 5.

Anderson, T.W., and L. A. Goodman. 1957. Statistical Inference about Markov Chains. Annals of Mathematical Statistics. 28, 89-109.

Billingsley, P. 1961. Statistical Methods in Markov Chains. Annals of Mathematical Statistics. 32, 12-40.

Bush, J. W., M. M. Chen, J. Zaremba. 1971. Estimating Health Program Outcomes Using a Markov Equilibrium Analysis of Disease Development. American Journal Public Health. 61, 2362-75.

Conover, W. J. 1971. Practical Nonparametric Statistics. New York: John Wiley and Sons, Inc.

Conover, W. J. 1972. A Kolmogorow Goodness of Fit Test for Discontinuous Distributions. Journal of the American Statistical Association. 67, 591-596.

Fetter, R.B., Y. Shin, J. L. Freeman, R. F. Averill and J. D. Thompson. 1980. Case Mix Definition by Diagnosis-Related Groups. Medical Care, 18, no. 2, Supplement, pp. 1-53.

Haley, R. W. and R. H. Shachtman. 1980. The Emergence of Infection Surveillance and Control Programs in U. S. Hospitals: An Assessment, 1976. American Journal of Epidemiology. 111, no. 5, 574-591.

Hooton, T. M., R. W. Haley, D. H. Culver. 1980. A Method for Classifying Patients According to the Nosocomial Infection Risks Associated with Diagnoses and Surgical Procedures. American Journal of Epidemiology, 111, no. 5.

Kao, E.P.C. 1972. A Semi-Markov Model to Predict Recovery Progress of Coronary Patients. Health Services Research. 7, 191-208.

Kastner, G. 1980a. Some Models for Cost Effectiveness Analysis of Infection Surveillance and Control Programs in U. S. Hospitals. Doctoral Dissertation and NIPNOTE 122, The SENIC Project, Dept. of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, N. C.

Kastner, G.,R. H. Shachtman, D. Quade, I. McCanless, R. Langston. 1980b. Validation for the 16-State Markov Chain Model. NIPNOTE 128, The SENIC Project, School of Public Health, Dept. of Biostatistics, University of North Carolina, Chapel Hill, N. C.

Kullback, S., M. Kupperman, H. U. Ku. 1962. Tests for Contingency Tables and Markov Chains. Technometrics. 4, 573-608.

Shachtman, R. H., J. R. Schoenfelder, C. J. Hogue. 1981a. Using a Stochastic Model to Investigate Time to Absorption Distributions. Operations Research. 29,589-603.

Shachtman, R. H., J. R. Schoenfelder, C. J. Hogue. 1981b. Conditional Rate Derivation in the Presence of Intervening Variables Using a Stochastic Model. Operations Research. 29, to appear.

Thomas, W. H. 1968. A Model for Predicting Recovery Progress of Coronary Patients. Health Services Research. 5,185-212.

Zahl, S. 1955. A Markov Process Model for Follow-up Studies. Human Biology. 27, 90-120.