CONDITIONAL RATE DERIVATION IN THE PRESENCE OF
INTERVENING VARIABLES USING A STOCHASTIC MODEL

by

Richard H. Shachtman
John R. Schoenfelder
Carol J. Hogue

Department of Biostatistics
University of North Carolina at Chapel Hill

CONDITIONAL RATE DERIVATION IN THE PRESENCE OF
INTERVENING VARIABLES USING A STOCHASTIC MODEL

Richard H. Shachtman (1), John R. Schoenfelder (1), Carol J. Hogue (2)

ABSTRACT

When conducting inferential and epidemiologic studies, researchers
are often interested in the distribution of time until the occurrence of
some specified event, a form of incidence calculation. Furthermore, their
interest will often extend to the effects of intervening factors on this
length. In this paper we impose the assumption that the phenomena being
investigated are governed by a stationary Markov chain and review how one
may estimate the above distribution. We then introduce and relate two
interpretatively different methods of investigating the effects of
intervening factors. In particular, we show how an investigator may
evaluate the effect of potential intervention programs. Finally, we
demonstrate the proposed methodology using data from a population study.

(1) University of North Carolina, Chapel Hill, North Carolina

(2) University of Arkansas for Medical Sciences, Little Rock, Arkansas

CONDITIONAL RATE DERIVATION IN THE PRESENCE OF
INTERVENING VARIABLES USING A STOCHASTIC MODEL

Richard H. Shachtman, John R. Schoenfelder, Carol J. Hogue

Biological and epidemiological outcome investigations often assess the relationship between characteristic variables (e.g., personal or demographic variables, often called risk factors) and condition variables (e.g., development of disease, often called response factors) to determine whether the condition occurs more frequently among those individuals exhibiting the characteristic. For example, in a study of the relationship between smoking and lung cancer, the characteristic variable is smoking status and the condition variable denotes whether lung cancer develops. A researcher would compare the incidence (new case) rate of lung cancer for smokers to that of non-smokers.

Frequently, research will investigate more than just the characteristic and condition; that is, one will want to consider intervening variables. In the above example, an intervening variable with epidemiological implications might be occupational exposure to asbestos powder. In this paper, we investigate two ways of incorporating such variables -- as posterior or prior factors. After noting that traditional epidemiological approaches to both methods of intervening variable analysis often exert an unreasonably large data demand, we show how a Markov chain (MC) may be used to significantly reduce the data requirements. First, however, we conceptually define two methodologies for incorporating intervening variables. Subsequently, we will provide mathematical definitions and derive relationships for the two techniques.

Consider a cohort (homogeneous group) of smokers demarcated at time zero. We may consider asbestos powder (the intervening variable) either as a POSTERIOR factor (we consider all smokers at time zero and later compute the lung cancer rate among those who, as an additional constraint, have not been exposed to asbestos powder as time proceeds) or as a PRIOR factor (we consider only the subset of smokers who will not be exposed to asbestos powder and compute their lung cancer rate).

For a specific individual, treating asbestos powder exposure as a posterior factor considers the probability that s/he will develop lung cancer without being occupationally exposed to asbestos powder GIVEN that s/he was initially smoking. Treating this same intervening variable as a prior factor considers the probability that a specific individual will develop lung cancer GIVEN that s/he was initially smoking and will not experience asbestos powder exposure. These approaches differ in the way they handle the intervening variable constraints. The posterior conditioning analysis treats the intervening variable as a condition which may or may not occur during the period of observation whereas the prior conditioning controls for it before observation begins. The resulting incidence rates may differ significantly and, as will be seen, require distinct formulae.

Researchers may analyze intervening constraints in a direct epidemiological analysis by using characteristic and intervening variables to partition the data into distinct cohorts; they then compute cohort specific rates. In our example, treating asbestos powder exposure as a POSTERIOR factor is exemplified by the comparison of workers in an industrial setting which may, but need not, involve asbestos powder exposure (e.g., shipbuilding). At the termination of the study, one computes lung cancer rates within each cohort (smoker, non-smoker) separately for those individuals who do (do not) experience asbestos powder exposure.

Considering asbestos powder exposure as a PRIOR factor requires that the subcohorts of smokers, for example, be demarcated at the initiation of the study. One subcohort consists of those smokers who will be exposed to asbestos powder during the period of observation and the other of those who will not. When the study terminates, one computes the appropriate lung cancer rate within each subcohort; a comparison of those rates serves to assess the effect of exposure to asbestos powder on the incidence of lung cancer within a cohort of smokers. By defining similar subcohorts of non-smokers we may complete the analysis of smoking and lung cancer while controlling, in a PRIOR sense, for the effect of asbestos powder exposure.

From this example it is clear that both of these traditional approaches necessitate a large amount of data in order to obtain accurate estimates of the rates under consideration. The data demand becomes a major concern when the condition and/or intervening variable is rare or when a number of intervening variables are considered simultaneously; the latter situation requires the formation of many subcohorts, each of which must be "reasonably" large.

Efficiency criteria, or simply the problem of obtaining sufficient information, indicate the advantage of using analytic models rather than following up multiple cohorts; the time-dependent, probabilistic nature of many empirical processes implies that such analytic models be stochastic processes. Investigators may then use these models to estimate the probability distributions of (i) the time until the process arrives at a predetermined state (e.g., a disease condition), (ii) the length of stay of the process in certain states, and (iii) the number of visits to key states. The scientist may condition all of these distributions by intervening variables as previously stipulated, thereby utilizing the model to account for various confounding or concomitant variables.

In Section 2 we assume that the underlying process is a stationary (time-homogeneous) MC, present formal analytic definitions of both types (prior and posterior) of intervening variable analyses, and derive a relationship between them. We then show how a researcher may perform both analyses, as well as the ordinary (non-intervening variable) analysis, using functions of the MC transition matrix. Thus all analyses demand only sufficient data to estimate the transition matrix. To illustrate these concepts, we will present a numerical example based on an MC model for a study of induced abortion. We employ a 79 state, time-homogeneous model described in Shachtman and Hogue (2).

1. CONDITIONING THE RATES

A. TRANSITION FUNCTIONS WITHOUT INTERVENING VARIABLES

Assuming that the phenomena under investigation are representable by a stationary MC, let $p_{jk} = P(X_n = k \mid X_{n-1} = j)$ be the probability that the underlying process is in state k at time n given that it was in state j at time $n-1$; note that because of stationarity these probabilities do not depend on n. The chain's transition matrix is then $P = ((p_{jk}))$. We may obtain the n-step transition probabilities, $p_{jk}(n) = P(X_n = k \mid X_0 = j)$, by the expression $P^n = ((p_{jk}(n)))$; that is, $p_{jk}(n)$ is the (j,k)-th element of the n-th power of the transition matrix. Finally, we express the first passage time probabilities, $f_{jk}(n) = P(X_n = k; X_m \ /= k, 1 <= m <= n-1 \mid X_0 = j)$, by

$$p_{jk}(n) = \text{SUM}/m(1,n) \ f_{jk}(m) \ p_{kk}(n-m) \ , \qquad n >= 1$$

and hence may obtain them iteratively by the formula

$$f_{jk}(1) = p_{jk}$$

$$f_{jk}(n) = p_{jk}(n) - \text{SUM}/m(1,n-1)\ f_{jk}(m)p_{kk}(n-m)\ , \qquad n >= 2.$$

(The symbol $/=$ means unequal and $\text{SUM}/m(1,n)$ represents the sum from $m=1$ to $m=n$.) Using these probabilities, a researcher may employ known techniques to compute such quantities as mean time necessary to "travel" between any two states, mean number of visits to any state, and mean length of stay in a given state.

To be definite let us now suppose the researcher is interested in the distribution of the time required to reach state k from state j. As expressed in Shachtman, Schoenfelder, and Hogue (3) this distribution is given by $((F_{jk}(n): n = 0,1,...))$ where $F_{jk}(n) = \text{SUM}/m(1,n)\ f_{jk}(m)$ is the probability that the process visits state k by time n given that it was in state j at time 0. The unconditioned probability of visiting state k by time n is then expressible as $G_k(n) = \text{SUM}/j\ a_j(0) F_{jk}(n)$, where the summation is over all states of the chain and $((a_j(0) = P(X_0 = j):$ all states $j))$ constitutes the initial probability distribution. Thus, for example, if n indexes months and states k,j represent diagnosis of lung cancer and performing a "pulmonary function" task respectively, then $F_{jk}(n)$ will be the probability that a person is diagnosed to have lung cancer by month n given that he was performing a "pulmonary function" task at time 0. The quantity $G_k(n)$ probability probability that a person develops lung cancer by month n irrespective of his status at time 0.

These distributions are of particular interest when state k is absorbing. In this situation $((F_{jk}(n): n = 0,1,...))$ is the distribution for time to absorption from state j and $((G_k(n): n = 0,1,...))$ is the time to absorption distribution. Moreover, $F_{jk}(n) = p_{jk}(n)$ for k absorbing. Throughout the remainder of this paper we will, for ease of expression, treat k as if it were an absorbing state; we will, however, make no explicit use of the assumption. Hence it could be dropped, in which case we would be computing "time to visit" rather than "time to absorption" distributions.

B. TRANSITION FUNCTIONS WITH INTERVENING VARIABLES AS POSTERIOR FACTORS

As previously mentioned, interest will often extend beyond an analysis of initial and absorbing states; that is, a researcher will want to know the probability of absorption without visiting one or more prespecified states. Letting h represent such a taboo state, an informative conditional probability, called a post-conditioned taboo probability (POSTAB) is given by

$$_h p_{jk}(n) = P(X_n = k; X_m \neq h, 1 \leq m \leq n-1 \mid X_0 = j),$$

the probability of being in state k at time n without having entered state h given that the process started in state j. The corresponding post-conditioned first passage probability (FOSTAB) is

$$_h f_{jk}(n) = P(X_n = k; X_m \neq h,k , 1 \leq m \leq n-1 \mid X_0 = j).$$

Note that for $h = k$, $_h f_{jk}(n) = f_{jk}(n)$, the ordinary, unconditioned first passage time probability.

Analogous to the non-taboo situation, the probability distribution $(({}_hF_{jk}(n): n = 0,1,...))$, where ${}_hF_{jk}(n) = \text{SUM}/m(1,n)\ {}_hf_{jk}(m)$, is the post-conditioned taboo time to absorption from state j distribution. Likewise $(({}_hG_k(n): n = 0,1,...))$, where ${}_hG_k(n) = \text{SUM}/j\ a_j(0)\ {}_hF_{jk}(n)$, is the post-conditioned taboo time to absorption distribution.

LEMMA: The POSTABs and FOSTABs follow the same functional relationships as unconditioned probabilities; in particular, assuming $h \ne k$, we have

(i) $\quad {}_hp_{jk}(1) = {}_hf_{jk}(1) = f_{jk} = p_{jk}$

$\quad {}_hp_{jk}(n) = \text{SUM}/m(1,n)\ {}_hf_{jk}(m)\ {}_hp_{kk}(n-m)\ ,\quad n \ge 2.$

(ii) $\quad {}_hp_{jk}(n) = \text{SUM}/r(r \ne h)\ {}_hp_{jr}(n-1)\ p_{rk}\ ,\quad n \ge 2.$

PROOF: Obvious by inspection.

□

By virtue of part (ii) of this lemma one may easily derive POSTABs from the unconditioned transition probabilities whenever $h \ne k$. (Recall that for $h = k$, the POSTABs are the usual first-passage time probabilities.) Define P^* to be the matrix formed by replacing the h-th row of P, the row corresponding to the taboo state, by zeroes.

LEMMA: The sequence of matrices defined by $R^1 = P$, $R^n = R^{n-1} P^*$, $n \geq 2$, is such that $R^n = ((_h p_{jk}(n)))$.

PROOF: Obvious by inspection.

□

Note that R^n is this the matrix of n-step POSTABS.

C. TRANSITION FUNCTIONS WITH INTERVENING VARIABLES AS PRIOR FACTORS

We have previously suggested another way of incorporating intervening variables into an analysis. Rather than being interested in time to absorption for a "subcohort" which does not visit a taboo state during the time of observation, the researcher may want to know the time to absorption distribution for a cohort which is explicity restricted from visiting a taboo state during the period of observation. In this case the taboo restriction is placed behind rather than in front of the conditioning bar.

These probabilities are expressed as

$$_h r_{jk}(n) = P(X_n = k \mid X_m /= h, \ 1 <= m <= n-1 \ ; \ X_0 = j),$$

the pre-conditioned taboo probability (PRETAB), and

$$_h s_{jk}(n) = P(X_n = k \mid X_m /= h,k, \ 1 <= m <= n-1; \ X_0 = j),$$

the pre-conditioned first passage taboo probability (FRETAB). We define

$$_h r_{jk}(1) = \ _h s_{jk}(1) = p_{jk} .$$

D. RELATING THE PRE- AND POST-CONDITIONED RATES

To this point we have introduced two methods of defining n-step transition functions for intervening variable analysis. As discussed in the introduction each has a distinct interpretation involving either prior or posterior conditioning. The following theorem, relating pre-conditioned and post-conditioned first passage probabilities, provides a direct comparison of these types of taboo conditioning.

Furthermore, by expressing the FRETABs as a function of the FOSTABs, the latter being iteratively obtainable from the unconditioned transition probabilities, this theorem also provides a convenient method of computing the former. Although not explicitly stated, a similar theorem relates PRETABs to POSTABs; since the latter are obtainable from the unconditioned transition probabilities, so are the former. Hence all taboo probabilities and distributions discussed are computable as functions of the initial transition matrix governing the MC.

THEOREM: (FRETABs from FOSTABs)   Assuming that $h \neq k$, we have

(i) $_h s_{jk}(n) = {}_h f_{jk}(n) / (1 - {}_h F_{jk}(n-1) - {}_k F_{jh}(n-1))$

where $_r F_{ji}(n) = \text{SUM}/m(1,n) \; {}_r f_{ji}(m),$ $i,r = h,k$ & $n \geq 2.$

(ii)   If for all $n \geq 1$ we have $\text{SUM}/m(1,n) ({}_h f_{jk}(m) + {}_k f_{jh}(m)) < 1,$

it follows that $_h s_{jk}(n) \geq {}_h f_{jk}(n).$

PROOF:

(i) Define the following sets: $A = ((X_n = k))$

$B = ((X_m \neq h, k \quad 1 \leq m \leq n-1))$

$C = ((X_0 = j))$

Note that by definition we have

$f_{h \, jk}(m) = P(X_m = k; X_u \neq h, k, \quad 1 \leq u \leq m-1 \mid X_0 = j)$

$= P(X_m = k$ for the first time and no h has occurred $\mid X_0 = j)$.

Similarly,

$f_{k \, jh}(m) = P(X_m = h$ for the first time and no k has occurred $\mid X_0 = j)$.

Now, define d as the union of states h and k. Then, since $h \neq k$, it follows that $X_m = k \longrightarrow X_m \neq h$ and $X_m = h \longrightarrow X_m \neq k$, we see

$f_{h \, jk}(m) + f_{k \, jh}(m) = P(X_m = d$ for the first time; $X_m = k \mid X_0 = j)$

$+ P(X_m = d$ for the first time; $X_m = h \mid X_0 = j)$

$= P(X_m = d$ for the first time $\mid X_0 = j)$.

Hence,

$F_{h \, jk}(n) + F_{k \, jh}(n) = \text{SUM}/m(1, n) \; (f_{h \, jk}(m) + f_{k \, jh}(m))$

$= \text{SUM}/m(1, n) \; P(X_m = d$ for the first time $\mid X_0 = j)$

$= P(X_m = d$ for some $m, \; 1 \leq m \leq n \mid X_0 = j)$.

and

$$1 - F_{h\,jk}(n) - F_{k\,jh}(n) = 1 - P(X_m = d \text{ for some } m, 1 \le m \le n \mid X_0 = j)$$

$$= P(X_m \,/\!= d, 1 \le m \le n \mid X_0 = j)$$

$$= P(X_m \,/\!= h,k, 1 \le m \le n \mid X_0 = j).$$

Now,

$$P(B\mid C) = P(X_m \,/\!= h,k, 1 \le m \le n-1 \mid X_0 = j)$$

$$= 1 - F_{h\,jk}(n-1) - F_{k\,jh}(n-1).$$

Then

$$s_{h\,jk}(n) = P(A\mid BC) = P(ABC)/P(BC) = P(AB\mid C)P(C)/P(B\mid C)P(C) = P(AB\mid C)/P(B\mid C)$$

$$= f_{h\,jk}(n) \,/(\,1 - F_{k\,jh}(n-1) - F_{h\,jk}(n-1)\,).$$

(ii) Obvious as $0 < 1 - F_{k\,jh}(n-1) - F_{h\,jk}(n-1) \le 1$.

◻

The condition necessary to have $s_{h\,jk}(n) \ge f_{h\,jk}(n)$ is that neither

state h nor k be absorbing, or, if they are, that there be an infinite "waiting time" before absorption. One may extend this theorem, and the corresponding analyses, to a collection H of taboo states rather than the single state h.

COROLLARY: If $h = k$, then (i) reduces to

$$s_{k\,jk}(n) = f_{jk}(n) \,/\, (1 - F_{jk}(n)\,).$$

PROOF: Obvious by inspection.

◻

E. TRANSITION FUNCTIONS WITH INTERVENING VARIABLES AS DUAL PRIOR AND

POSTERIOR FACTORS

In many epidemiological investigations a researcher is interested

in particular intervention strategies which may be time dependent.

For instance, one may ask what is the probability of absorption in state k

without visiting state h (or k) for the first t time units? To answer

this question we consider a combination of the taboo conditioned

probabilities described above. These probabilities, merging FRE- and

FOSTABs, are called dual-conditioned taboo probabilities and, for the case

of the first passage times, are written as

$$\,_{hk}fs_{jk}(n|t) = P(X_n = k; \; X_u \,/= \, h,k \, , \quad t+1 <= u <= n-1 \, |$$

$$X_v \,/= \, h,k \, , \; 1 <= v <= t \, ; \, X_0 = j). \quad n >= t+1.$$

Using the following corollary to the previous theorem, we relate this

dual-conditioned probability to the FOSTABs, thereby obtaining an iterative

method for computing the former.

$$\text{COROLLARY:} \quad \,_{hk}fs_{jk}(n|t) = \,_{h}f_{jk}(n) \, / \, (\, 1 - \,_{k}F_{jh}(t) - \,_{h}F_{jk}(t) \,), \quad n >= t+1.$$

D

Thus the distribution formed by summing the $\,_{hk}fs_{jk}(n|t)$ probabilities

over n >= t+1 gives the desired dual-conditioned taboo time to absorption

from state j distributions, $\,_{hk}FS_{jk}(N|t) = \text{SUM}/n(t+1,N) \; \,_{hk}fs_{jk}(n|t).$

By computing this distribution for all initial states and averaging over

the initial distribution, we arrive at the general (irrespective of

initial state) dual-conditioned taboo time to absorption distribution,

$$GS_{hk}(N!t) = SUM/_J \; a_{hk}(O) \; FS_{jk}(N!t).$$ By varying the value of t,

the researcher may assess the efficacy of a proposed intervention strategy of explicitly preventing a visit to the taboo state as a function of time.

By altering the conditioning statements, a researcher may obtain interesting variations of the above dual-conditioned taboo probabilities, e.g., one may consider one of the following expressions:

(i) $P(X_n = k; X_v /= k, 1 <= v <= n-1 \mid X_u /= k, 1 <= u <= t; X_0 = j),$
$$n >= t$$

(ii) $P(X_n = k; X_v /= h,k, 1 <= v <= n-1 \mid X_u /= k, 1 <= u <= t; X_0 = j),$
$$n >= t$$

(iii) $P(X_n = k; X_v /= h,k, 1 <= v <= n-1 \mid X_u /= h, 1 <= u <= t; X_0 = j),$
$$n >= t$$

One may express each of these taboo probabilities in terms of the initial transition probabilities and post-conditioned taboo probabilities by deriving an appropriate corollary to the previous theorem. To obtain (i) one replaces the numerator of the theorem by $f_{jk}(n)$ and the denominator by $1-F_{jk}(t)$; to obtain (ii) and (iii) one replaces the denominator of the theorem by $1-F_{jk}(t)$ and $1-F_{jh}(t)$, respectively.

2. AN APPLICATION TO POPULATION EPIDEMIOLOGY.

To demonstrate these techniques we consider an MC model originally designed for an inferential study of the effect of induced abortion on subsequent pregnancy outcome (2). Data used in this example come from a historical prospective study of Macedonian women residing in Skopje,

Yugoslavia who aborted their first pregnancy during 1968-69, Hogue (1).

Interviews conducted in 1972 produced, for each woman, a chronological

record of her reproductive behavior between the beginning of menstruation and

the interview date. Quality checks verified that the women interviewed

were indeed representative of the target population of Yugoslavian women (1).

The interval between the initial abortion and the first delivery is an

important consideration to both population epidemiologists and demographers.

Let a represent an induced abortion, d a delivery, and assume that the

total mass of the initial distribution is concentrated on a. Then the

basic time to delivery distribution is ((F_{ad} (n))) where

F_{ad} (n) = SUM/m(1,n) f_{ad} (m) and n indexes months since the initial

abortion. This distribution, which appears in column 2 of Table 1,

necessarily reflects diverse contraceptive use patterns; consequently,

further investigation is warranted. The state space of this MC partitions

all types of contraception into one of three states: i = ineffective method

(e.g., rhythm or coitus interruptus), m = moderately effective method (e.g.,

condom), and e = effective method (e.g., birth control pills). By treating

each of these states as a taboo state, the corresponding post-conditioned

taboo time to delivery distribution assesses the probability of a delivery

by month n without having resorted to the taboo type of contraception.

These distributions are listed in columns 3-5 of Table 1.

The dual-conditioning technique provides an assessment of a t-month

limited intervention. For example, suppose a researcher wishes to evaluate

a program which would explicitly prevent the use of ineffective methods of

contraception for t months. Consequently, women who desire to employ a

method of contraception must use either an effective or moderately effective

method. (NOTE: This is not to imply that the women who avoid ineffective

contraceptive methods will continually, or even temporarily, employ other,

more effective methods.) By treating i as a taboo state and evaluating

the dual-conditioned taboo time to delivery distribution FS_{id} $(N|t)$, we

may assess the effect of such a policy on those women who do not deliver

during the first t months. We list these distributions in Table 2.

Tables 1,2

This application addresses epidemiologic questions about the effect of
different patterns of contraceptive use on the distribution of time to
delivery. By computing the post-conditioned time to delivery distributions
(Table 1) we are able to obtain insight into the population dynamics of
cohorts of women electing to use different levels of contraception. Note
the large magnitude of the relative decrease in the unconditioned time to
delivery distribution (18% = (100%) x (.5344 - .4366)/.5344 by month 60)
which results when ineffective contraceptive methods, i, are regarded as
taboo. Thus the women who avoid i are less likely to deliver during
a set period of time than the group as a whole; possibly this is because
the members of this subgroup who elect to use contraceptive methods
employ a method which is at least moderately effective. Even more
surprising is the similarity of the distributions which result when
moderately effective and effective contraceptive practices are treated
as taboo. As expected, however, both of the associated subgroups, the
avoiders of m and the avoiders of e, are more likely to deliver than
the subgroup which avoid i.

Finally, by computing the dual-conditioned taboo time to delivery
distribution using i as the taboo state (Table 2) we are able to assess
the efficacy of preventing ineffective contraceptive practices. Note that
the taboo time to delivery distribution following the termination of this

implementation policy appears to be invariant across the duration of that policy. For example, the probability of a delivery less than 24 months following termination of policy is approximately 0.29 for all three values of t studied. Also note that these t-month limited distributions in columns 2,3,4 are not the unconditioned (column 2) distribution with a t-month delay.

Clearly, states other than i may be of interest as the taboo state in the dual-conditioning investigation. Of particular concern might be the state of susceptibility to pregnancy; similar distribution calculations would give insight into the efficacy of an implementation program encouraging all women to employ some type of contraceptive practice. These distributions would provide benchmarks of total acceptance against which family planning administrators could compare actual practice. Furthermore, by modifying dual-condition representations, one may calculate similar distributions for more complicated patterns of contraceptive switching within cohorts of users.

3. SUMMARY

We have made use of a typical epidemiologic problem to illustrate the use of conditional rate information. Using a traditional cohort formulation such an investigation would have required large amounts of data -- an expensive proposition, even if data were available. Alternatively, we are able to employ the concepts of pre- and post-conditioned probabilities and the theorem relating them.

In the example presented above, the MC was submitted to formal goodness-of-fit tests in order to verify the Markov property (3). Depending on the application, the researcher may prefer to (i) employ structural assumption validation, i.e., verify the Markov property by comparing probabilities for all pairs of states visited over appropriate time lengths, (ii) accept

the model as a reasonable approximation to the empirical process under study, or (iii) loosely validate the property by examining functions of the transition matrix (e.g., those leading to the time to absorption in a state) and seeing if these functions replicate equivalent empirical distribution functions derived directly from the underlying data. We caution that the latter is not a strong validation since, as seen in (3), distinctly different populations may produce similar functions from their respective transition matrices.

Making use of an MC offers the additional advantage of employing partial path information which is unavailable in cohort studies. That is, we may follow an individual up to a certain point, lose that person to observation (or follow-up) and not have use of that "path" for a cohort study. However, all transitions made up to the point of loss are valid for computing maximum likelihood estimators of the chain's transition probabilities. Hence the behavior, for example in visiting various combinations of contraceptive states, of a "lost" individual may be reflected in the transition probability estimators. When applied to the dual-conditioned taboo probability estimators, this information is especially advantageous in maximizing the use of available data.

In addition, by using either a parametric analysis varying particular transition probabilities or simulation, one may investigate the effects of alternative assumptions about the magnitude of certain probabilities on final cumulative distributions. In our example, for instance, one may wish to raise or lower estimates of particular contraceptive uses, and/or probabilities of jumping to "susceptible" states, and test the effect on functions of the transition matrix, while simultaneously taking into account one or more intervening variables.

In summary, we have shown how a stochastic model allows researchers to assess the effect of implementing intervention in situations where

financial and/or ethical considerations prohibit cohort studies. Furthermore,
we have established new analytic versions of conditional rates for use in
biological and epidemiological studies and have generated various way to
produce such rates in the face of intervening variables. Finally, we have
presented relationships among them, thereby further illustrating the use of
a Markov chain as a tool of (statistical) inference.

REFERENCES

1. C. J. Hogue, "Low Birth Weight Subsequent to Induced Abortion: A Historical Prospective Study of 948 Women in Skopje, Yugoslavia," AM. J. OBSTET. GYNECOL. 123, 675-81 (1975). .

2. R. H. Shachtman and C. J. Hogue, "Markov Chain Model for Events Following Induced Abortion," OPNS. RES. 24, 916-32 (1976).

3. R. H. Shachtman, J. R. Schoenfelder and C. J. Hogue, "Using a Stochastic Model to Investigate Time to Absorption Distributions," Submitted to OPNS. RES. (1979).

TABLE 1

Time to Delivery Distributions*
(initial abortion to delivery)

Month	Unconditioned	Post-Conditioned		
n	$F_{ad}(n)$	$F_{i\,ad}(n)$	$F_{m\,id}(n)$	$F_{e\,id}(n)$
12	0.0392	0.0356	0.0384	0.0390
24	0.2047	0.1797	0.1991	0.2025
36	0.3382	0.2892	0.3278	0.3334
48	0.4465	0.3727	0.4314	0.4386
60	0.5344	0.4366	0.5149	0.5230

* States:
    a = induced abortion
    d = delivery
    i = ineffective contraceptive methods
    m = moderately effective contraceptive methods
    e = effective contraceptive methods

Table 2

Split-Conditioned Time to Delivery Distributions following an
induced abortion.   Taboo state is ineffective contraceptive methods*

Month	Unconditioned	Split-Conditioned		
N	$F_{ad}(N)$	$FS_{id\ ad}(N!t)$		
		t=12	t=24	t=36
12	.0392	.....	.....	.....
24	.2047	.1651	.....	.....
36	.3382	.3027	.1594	.....
48	.4465	.4067	.2936	.1541
60	.5344	.4878	.3962	.2840

* States:
  a = induced abortion
  d = delivery
  i = ineffective contraceptive methods