CHAPTER 5

MODELING MEASUREMENT ERROR IN EVENT OCCURRENCE FOR SINGLE, NON-RECURRING EVENTS IN DISCRETE-TIME SURVIVAL ANALYSIS

Katherine E. Masyn
University of California at Davis

In the social sciences, research questions around specific behaviors, such as drug use or school drop out, and related events such as juvenile arrest, are often concerned with both the "if" and "when" of the behaviors and events. For example, it may be of interest to understand not only what predicts drug use but also what predicts the timing of first drug use, that is, the age of onset. It may also be of interest to understand how that drug use and its timing affects the likelihood of other negative life course outcomes. Survival analysis, sometimes called event history analysis, refers to a general set of statistical methods developed specifically to model the timing of events. This chapter concerns a subset of those methods dealing with events measured or occurring in discrete-time or grouped-time intervals.

Discrete-time survival analysis is distinguished from its counterpart, continuous-time survival analysis, by the time-scale on which the occurrence of

Advances in Latent Variable Mixture Models, pages 105–145

the event of interest is measured. On a continuous time-scale, an event can occur at any point in time and the exact timing of each event occurring in the sample is recorded such that no two people share the same event time (e.g., measuring event occurrence to the year, month, day, hour and minute). On a discrete time-scale, an event can occur at any point in time[1] but the timing of each event is recorded as occurring in one of a finite number of time intervals or time periods, where the number of intervals is notably fewer than the number of events observed in the sample (e.g., measuring event occurrence only to the year in a 10 year range). This time-scale distinction is necessary because the methods developed for continuous-time survival data do not directly apply to discrete-time survival data, just as regression techniques for continuous dependent variables do not directly apply to categorical outcomes.

As important as determining the time-scale for the event process is defining the nature of the event of interest. For this chapter, we limit ourselves to single, non-recurring events. This means that the event in question can happen once and only once to each individual. This also means that the event is of only one kind or type, meaning that we would consider the event of first juvenile arrest without distinguishing arrests for crimes against property from arrests for crimes against persons. All of what is presented here, however, can be extended to include recurring events and competing risks. We further assume that the duration of the event is short, even instantaneous, relative to the length of the discrete time periods into which time is divided. This means that an event occurrence would not span multiple time intervals but would occur within only one of the possible time periods.

Typically, when discrete-time survival analysis is applied, there is an implicit assumption that the "if" of the observed events is known with complete certainty. In some cases, this assumption might be quite reasonable, especially when dealing with truly definitive outcomes such as death. However, this assumption may be much less tenable when dealing with behavioral or clinical outcomes. Part of what may lessen the plausibility of this assumption is the means by which the event occurrence and timing thereof is assessed. In a retrospective study, each individual may be asked about the timing of an event, for example, "At what age did you have your first full drink of alcohol?" Even in prospective longitudinal studies we often must rely on the self-reports of behavior, such as, "How often did you drink alcohol in the last year?" For clinical outcomes, symptom level data may be collected every year and the occurrence of a clinical diagnosis (e.g., alcohol use disorder) is surmised from the reported symptoms. Techniques for dealing with error and bias in the reporting of event timing have been developed and applied (see, e.g., Gaskell, Wright, & O'Muircheartaigh, 2000). Much less attention has been given to measurement error in the assessment of the event occurrence. This chapter builds upon the limited amount of previous work done on this issue to further our understanding of the impact that measurement error in the determination of event occurrence can have on estimates related to the survival process, and to clarify a reformulation of the survival model as a latent Markov chain model that allows such error to be explicitly incorporated into the survival analysis.

The chapter begins with an overview of discrete-time survival analysis and the main quantities of interest when describing an event history process. The overview is followed by a detailed explanation of how a discrete-time survival analysis can be equivalently reformulated as a restricted Markov chain model. That model is then extended to include event indicators without measurement error and covariate predictors of the event process. The application of the model is illustrated with a real data example. The section after that explores the consequences of having measurement error on the event indicators for the parameter estimates related to the survival process. The chapter concludes with the presentation of a model with multiple event indicators at each time period and a discussion of model limitations and possible modeling extensions.

OVERVIEW OF DISCRETE-TIME SURVIVAL ANALYSIS

Suppose only events occurring in the time range $[a_0, a_J)$ may be observed for a given study with a random sample of n independent individuals[2] i, with $i = 1,2,\ldots,n$. Suppose further that the time range is grouped into J time intervals, $t_j = [a_{j-1}, a_j)$, of equal width[3]. Let J_i be the final complete time period for which individual i is observed, where $J_i \leq J$ for all i. Usually, $J = \max_i(J_i)$. If an individual is lost to follow-up during a certain time period, say time period 5, then the last complete time period of observation for that individual would be $J_i = 4$. Assume for the purposes of this chapter that $a_0 = 0$ so that the beginning of the first time interval is the beginning of the survival process; that is, t_1 is the first time period for which individuals are at-risk for the event and, thus, no one has experienced the event prior to time period 1. Let T_i be the time period in which the event occurs for individual i. If $T_i \leq J_i$, then the event is recorded during the time that the individual is under observation and T_i is known. If $T_i > J_i$, then the event is not recorded during the time that the individual is under observation and the value of T_i is unknown to the researcher; that is, the value of T_i is missing. This kind of missingness is known as *right-censoring*. It is assumed that all individuals are at-risk for the event of interest and that if an individual is not observed to experience the event during the time that he or she is under observation, then he or she will experience the event at some later time, beyond J_i. The key feature of right-censored observations is that there is still some information available about T_i—even though the true value of T_i is unknown, it is known that $T_i > J_i$. This information does contribute something to our understanding of the survival process and can be included in the data analysis. For this chapter, we will limit our treatment of missing event time data to non-informative right-censoring, meaning that the time at which an individual is censored is not related to the event time for that individual ($T_i \perp J_i$), conditional on observed covariates.

There are two main quantities of interest when describing a discrete-time event history process: the survival probability and the hazard probability. The survival probability is likely the most intuitive in understanding the nature of an event history process. The survival probability corresponding to time period j is defined as the probability of an individual "surviving" beyond the interval, j; that is, remaining event-free through time period j. We will denote the survival probability as $P_S(j)$ such that:

$$P_s(j) = \Pr(T > j). \tag{5.1}$$

The hazard probability, although perhaps less intuitive, is the quantity that we deal with most often in survival analysis since, as will be shown later, our models for the event history process are specified in terms of the hazard probabilities. The hazard probability corresponding to time period j is defined as the probability of an individual experiencing the event in time period j given that he or she had not experienced the event prior to time period j. We will denote the hazard probability as $P_h(j)$ such that:

$$P_h(j) = \Pr(T = j \mid T \geq j). \tag{5.2}$$

A specific survival probability can be computed from hazard probabilities using the following relationship:

$$\begin{aligned} P_S(j) &= \Pr(T > j) \\ &= \Pr(T \neq 1 \mid T \geq 1) \cdot \Pr(T \neq 2 \mid T \geq 2) \quad \Pr(T \neq j \mid T \geq j) \\ &= \prod_{v=1}^{j} (1 - \Pr(v)). \end{aligned} \tag{5.3}$$

It is useful to examine both the survival and hazard probabilities for the time periods under study. The hazard probabilities describe how the risk for an event changes over time while the survival probabilities reflect not only the cumulative risk impact on the population, but also quantify the proportion of the original population still susceptible to the risk defined by the hazard probability for each time period.

Given these definitions and the relationship between the survival and hazard probabilities, it is possible to construct the likelihood functions for both uncensored and right-censored observations in terms of the hazard probabilities. We will assume that all individuals in the sample are drawn from the same population, such that $T_i \sim T$ for all $i = 1, \ldots, n$. For an individual i whose event time T_i is observed (i.e., $T_i \leq J_i$), the likelihood is given by:

$$\begin{aligned} l_i &= \Pr(T = t_i) \\ &= \Pr(T = t_i \mid T \geq t_i) \cdot \Pr(T > t_i - 1) \\ &= P_h(t_i) \cdot P_S(t_i - 1) \\ &= P_h(t_i) \cdot \prod_{v=1}^{t_i - 1} (1 - P_h(v)). \end{aligned} \tag{5.4}$$

For an individual i who is right-censored at J_i so that T_i is not observed (i.e., $T_i > J_i$), the likelihood is given by:

$$\begin{aligned} l_i &= \Pr(T > j_i) \\ &= P_S(j_i) \\ &= \prod_{v=1}^{j_i} (1 - P_h(v)). \end{aligned} \tag{5.5}$$

The full observed data likelihood can then be written as:

$$L = \prod_{i=1}^{n} \left[\left(P_h(t_i) \cdot \prod_{v=1}^{t_i - 1} \left(1 - P_h(v)\right) \right)^{I(T_i \leq j_i)} \cdot \left(\prod_{v=1}^{j_i} \left(1 - P_h(v)\right) \right)^{I(T_i > j_i)} \right]. \tag{5.6}$$

Note that the above likelihood may appear less compact than the conventional likelihood expression (see, e.g., Singer & Willet, 1993) where it is usually assumed that if T_i is observed, then $J_i = T_i$; that is, if the event is observed, then the time period in which the event occurs is the final time period of observation so that J_i represents either an event time or a censoring time. We do not make this assumption, which allows for the possibility that an individual may continue to be under observation even *after* the event has occurred. The reason for this becomes clearer in the sections that follow, but is a necessary consequence of probable error in the measurement of event occurrence. The maximum likelihood estimates for the $P_h(j)$ terms under the missing-at-random (MAR) assumption correspond to the same estimates under the assumption of non-informative right-censoring (Little & Rubin, 2002; Masyn, 2003; Muthén & Masyn, 2005).

The various approaches to discrete-time survival analysis differ primarily in the required format of the observed data and how the relationships between the hazard probabilities and covariates are specified. In his seminal 1972 paper, Cox suggested using a logistic regression to relate the discrete-time hazard probabilities to observed covariates. The use of logistic regression for discrete-time survival analysis has been studied further by Singer and Willet (1993, 2003) as well as many others including Prentice and Gloeckler (1978) and Allison (1982, 1984, 1995). Alternate approaches include multilevel ordinal multinomial regression (Hedeker, Siddiqui, & Hu, 2000), mixed Poisson models (Land, Nagin, & McCall, 2001), log-linear models (Laird & Oliver, 1981; Vermunt, 1997), latent class regression models (Masyn, 2003; Muthén & Masyn, 2005; Vermunt, 1997, 2002), multistate models (Lindeboom & Kerkofs, 2000; Steele, Goldstein, & Browne, 2004), and discrete-time Markov chain models (Tuma & Hannan, 1984; Van de Pol & Langeheine, 1990; Vermunt, 1997). This chapter explores in detail the formulation of a discrete-time event history process in a Markov chain framework for the reason that this formulation, of all those listed above, is the one that most readily accommodates the possibility of measurement error in event occurrence. This chapter also details a model specification that allows estimation of the proposed models in more advanced statistical modeling software packages, specifically M*plus* Version 4.2 (Muthén & Muthén, 2006). (For more information on continuous-time survival analysis in a Markov chain framework, see, e.g., Aalen & Johansen, 1978, and Andersen, 1988.)

DISCRETE-TIME SURVIVAL ANALYSIS AS A MARKOV CHAIN MODEL

The phrase *Markov chain* refers to a model for repeated measures of one or more discrete variables where change in a categorical outcome (latent or observed) over time is described through a set of transition matrices. The levels of the categorical outcome are referred to as *states* and the elements of the transition matrices are conditional probabilities for being in a specific state in one time period conditional on the occupied states in the prior time periods. (For more general information on Markov chain models, see, e.g., Langeheine & Van de Pol, 2002; Van de Pol & Langeheine, 1990; Vermunt, Langeheine, & Bockenholt, 1999.) In this section, we reframe the single, non-recurring event history process as a simple first-order Markov chain where the states occupied at each time period are directly observed and measured without error, and where the probability of being in a specific state in one time period is only dependent upon the state occupied in the time period immediately preceding the time period in question.

In order to reformulate an event history process as a Markov chain, we must first define the *states* that may be occupied in any given time period. For a single, non-recurring event, there are three states: pre-event, event, and

post-event. The event state is occupied during the time period in which the event occurs. As in the previous section, we will assume that the first time period begins at $a_0 = 0$ so that all individuals occupy a pre-event state prior to the first time period. From one time period to the next, an individual may either remain in a pre-event state or may transition to the event state. As stated before, we assume that the nature of the event is such that its duration is always contained within a single time period. Thus, once an individual occupies an event state for a given time period, he or she will automatically transition into a post-event state in the next time period where he or she will remain for all subsequent time periods. The post-event state, in Markov chain terms, would be referred to as an *absorbing* state. This means that for a single, non-recurring event, individuals may not occupy an event state more than once.

Let E_j represent the categorical outcome variable for time period j with three categories corresponding to the three potential states. The expressions below formalize the state definitions:

$$E_j = \begin{cases} 0 \text{ (pre-event)} & \text{if } T > j \\ 1 \text{ (event)} & \text{if } T = j \\ 2 \text{ (post-event)} & \text{if } T < j \end{cases} \tag{5.7}$$

where $j = 1,2,\ldots,J$. As explained previously, changes in the categorical outcome, in this case, E, across time periods are described through a set of transition matrices. The elements of each transition matrix are conditional probabilities, denoted τ, where $\tau_{k(j)|m(j-1)}$ is the probability of an individual occupying state k in time period j given that he or she occupied state m in time period $j-1$. That is:

$$\tau_{k(j)|m(j-1)} = \Pr(E_j = k \mid E_{j-1} = m). \tag{5.8}$$

The transition matrices are denoted $\mathbf{T}$, where $\mathbf{T}_{(j)(j-1)}$ is the $M \times K$ transition matrix containing the conditional probabilities for states $(1,\ldots,K)$ in time period j given each state $(1,\ldots,M)$ in time period $j-1$. In our case, there are three states that may be occupied during any given time period,[4] and so a general representation of each transition matrix can be given by:

$$\begin{array}{c} \\ E_{j-1} \\ \mathbf{T}_{(j)(j-1)} = \begin{array}{c} 0 \\ 1 \\ 2 \end{array} \end{array} \begin{array}{c} E_j \\ \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{pmatrix} \tau_{0(j)|0(j-1)} & \tau_{1(j)|0(j-1)} & \tau_{2(j)|0(j-1)} \\ \tau_{0(j)|1(j-1)} & \tau_{1(j)|1(j-1)} & \tau_{2(j)|1(j-1)} \\ \tau_{0(j)|2(j-1)} & \tau_{1(j)|2(j-1)} & \tau_{2(j)|2(j-1)} \end{pmatrix} \end{array}. \tag{5.9}$$

Note that the elements across each row the transition matrix must sum to one. That is:

$$\sum_{k=1}^{K} \tau_{k(j)|m(j-1)} = 1. \quad (5.10)$$

In order for the above transition matrix to properly reflect the nature of the three states as we defined them above, certain restrictions must be placed on the transition probabilities. Once an individual experiences the event, he or she cannot return to a pre-event state. Therefore:

$$\tau_{0(j)|1(j-1)} = \Pr(E_j = 0 \mid E_{j-1} = 1) = \Pr(T > j \mid T = j-1) = 0\,. \quad (5.11)$$

We also assume that occupation in an event state does not extend beyond one time period. Thus:

$$\tau_{1(j)|1(j-1)} = \Pr(E_j = 1 \mid E_{j-1} = 1) = \Pr(T = j \mid T = j-1) = 0\,. \quad (5.12)$$

And once in an event state, the individual automatically transitions to a post-event state where he or she remains for the rest of time. Therefore:

$$\begin{aligned}
\tau_{2(j)|1(j-1)} &= \Pr(E_j = 2 \mid E_{j-1} = 1) = \Pr(T < j \mid T = j-1) = 1;\\
\tau_{0(j)|1(j-1)} &= \Pr(E_j = 0 \mid E_{j-1} = 1) = \Pr(T > j \mid T = j-1) = 0;\\
\tau_{0(j)|2(j-1)} &= \Pr(E_j = 0 \mid E_{j-1} = 2) = \Pr(T > j \mid T < j-1) = 0;\\
\tau_{1(j)|2(j-1)} &= \Pr(E_j = 1 \mid E_{j-1} = 2) = \Pr(T = j \mid T < j-1) = 0;\\
\tau_{2(j)|2(j-1)} &= \Pr(E_j = 2 \mid E_{j-1} = 2) = \Pr(T < j \mid T < j-1) = 1.
\end{aligned} \quad (5.13)$$

Since an individual cannot move to a post-event state without first occupying an event state, we also have:

$$\tau_{2(j)|0(j-1)} = \Pr(E_j = 2 \mid E_{j-1} = 0) = \Pr(T < j \mid T > j-1) = 0\,. \quad (5.14)$$

Inserting the above restrictions into the general transition matrix, we obtain the following:

$$\mathbf{T}_{(j)(j-1)} = \begin{array}{c} \\ E_{j-1} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{ccc} & E_j & \\ 0 & 1 & 2 \\ \tau_{0(j)|0(j-1)} & \tau_{1(j)|0(j-1)} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{array}\,. \quad (5.15)$$

Since each row sums to 1, we then have:

$$\tau_{0(j)|0(j-1)} = 1 - \tau_{1(j)|0(j-1)}. \quad (5.16)$$

Substituting in our definitions for E_j gives:

$$\begin{aligned}
\tau_{1(j)|0(j-1)} &= \Pr(E_j = 1 \mid E_{j-1} = 0) = \Pr(T = j \mid T > j-1) = P_h(j), \text{ and}\\
\tau_{0(j)|0(j-1)} &= 1 - P_h(j).
\end{aligned} \quad (5.17)$$

The transition matrix can now be written in terms of the hazard probabilities as:

$$\mathbf{T}_{(j)(j-1)} = \begin{array}{c} \\ E_{j-1} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{ccc} & E_j & \\ 0 & 1 & 2 \\ 1-P_h(j) & P_h(j) & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{array}\,. \quad (5.18)$$

For each time interval j with $j > 1$, there is an associated transition matrix, $\mathbf{T}_{(j)(j-1)}$ as defined above. For the first time interval, $j = 1$, there is no transition matrix. It is assumed that at the beginning point of the first time interval, everyone is in a pre-event state. Thus, the only two states that an individual may occupy in the first interval is a pre-event state or an event state. Here, the marginal probabilities rather than transition probabilities map onto the hazard probabilities where:

$$\begin{aligned}
\Pr(E_1 = 0) &= \Pr(T > 1) = 1 - P_h(1);\\
\Pr(E_1 = 1) &= \Pr(T = 1) = P_h(1);\\
\Pr(E_1 = 2) &= \Pr(T < 1) = 0.
\end{aligned} \quad (5.19)$$

Instead of a transition matrix, define a 1×2 vector of the initial marginal probabilities, $\Pi_{(1)}$, where:

$$\mathbf{\Pi}_{(1)} = \begin{bmatrix} \pi_0 & \pi_1 \end{bmatrix} = \begin{bmatrix} 1-P_h(1) & P_h(1) \end{bmatrix}. \quad (5.20)$$

The first transition matrix, $\mathbf{T}_{(2)(1)}$, is then a 2×3 matrix rather than a 3×3 matrix given by:

$$\mathbf{T}_{(2)(1)} = \begin{array}{c} \\ E_1 \\ 0 \\ 1 \end{array} \begin{array}{c} E_2 \\ \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{pmatrix} 1-P_h(2) & P_h(2) & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array}. \tag{5.21}$$

In Markov chain models, it is possible to compute the transition probabilities for ending in a particular state in time period j given a starting state in time period $j-p$ by multiplying the transitions matrices between $j-p$ and j. That is:

$$\mathbf{T}_{(j)(j-p)} = \mathbf{T}_{(j-p+1)(j-p)} \cdot \mathbf{T}_{(j-p+2)(j-p+1)} \cdots \mathbf{T}_{(j)(j-1)}. \tag{5.22}$$

Further, the probabilities of ending in a particular state in time period j given an initial starting state are given by:

$$\mathbf{T}_{(j)(1)} = \mathbf{\Pi}_{(1)} \cdot \mathbf{T}_{(2)(1)} \cdot \mathbf{T}_{(3)(2)} \quad \mathbf{T}_{(j)(j-1)}. \tag{5.23}$$

As an example, the transition probabilities in our case from initial state (pre-event) to time period 3 can be expressed as:

$$\begin{aligned} \mathbf{T}_{(3)(1)} &= \mathbf{\Pi}_{(1)} \cdot \mathbf{T}_{(2)(1)} \cdot \mathbf{T}_{(3)(2)} \\ &= \begin{bmatrix} 1-P_h(1) & P_h(1) \end{bmatrix} \cdot \begin{bmatrix} 1-P_h(2) & P_h(2) & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1-P_h(3) & P_h(3) & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} (1-P_h(1)) \cdot (1-P_h(2)) \cdot (1-P_h(3)) \\ (1-P_h(1)) \cdot (1-P_h(2)) \cdot (P_h(3)) \\ (1-P_h(1))(P_h(2)) + P_h(1) \end{bmatrix}^T. \end{aligned} \tag{5.24}$$

Notice that the first element, $T^{(1)}_{(3)(1)}$, of the resultant vector, $\mathbf{T}_{(3)(1)}$, is the survival probability for time period 3. More generally:

$$T^{(1)}_{(j)(1)} = \prod_{v=1}^{j} (1-P_h(v)) = P_S(j), \tag{5.25}$$

$$T^{(2)}_{(j)(1)} = \left(\prod_{v=1}^{j-1} (1-P_h(v)) \right) \cdot P_h(j) = P_S(j-1) \cdot P_h(j).$$

ESTIMATING THE DISCRETE-TIME SURVIVAL ANALYSIS AS A MARKOV CHAIN MODEL

To understand the identification and likelihood construction for the discrete-time survival analysis model for a single, non-recurring event in a Markov chain model framework, it is useful to consider the information available from a complete sample with no missing data during the observation period. This does not mean there are no missing event times, only that all of the right-censoring occurs at the end of the final time period of observation. As an example, take a study with six periods of observation. The possible observed chains for a complete data set are listed in Table 5.1 below.

For six time periods of observation, there are seven potential complete chains. In general, for J time periods, there are $J+1$ possible observed chains for the complete data. With the complete data, all observations could be summarized into a single frequency table. Since the frequencies must add to the total sample size, that is,

$$\sum_{j=1}^{J+1} n_j = n,$$

there are J degrees of freedom available for parameter estimation in a model. Each transition matrix, as defined in the previous section, has exactly one unknown value, $P_h(j)$. There are $J-1$ transition matrices needed to describe the chain across J time periods along with the initial state probabilities vector, also with one unknown value, $P_h(1)$, resulting in J unknown values. Not surprisingly then, the discrete-time survival Markov chain without covariates and without restrictions on the hazard probabilities across time is a just-identified, fully saturated model that reproduces the observed chain frequencies exactly in the complete observations case.

TABLE 5.1 Frequency Table of Observable Chains for Complete Data

E_1	E_2	E_3	E_4	E_5	E_6	f
1	2	2	2	2	2	n_1
0	1	2	2	2	2	n_2
0	0	1	2	2	2	n_3
0	0	0	1	2	2	n_4
0	0	0	0	1	2	n_5
0	0	0	0	0	1	n_6
0	0	0	0	0	0	n_7

The last chain in Table 5.1 represents individuals under observation for all six time periods who do not experience the event during the entire length of observation. These individuals are right-censored in that we do not know their exact event times, only that $T > 6$, but the observations are complete in that we do know their event states for all six periods. Suppose, however, that there are incomplete data in the sense that some individuals are right-censored prior to the end of the study. For example, if there was an individual who was censored during the fourth time period such that the third time period was the last complete time period of observation, his or her observed chain would look like the following:

$$e_i = (0 \quad 0 \quad 0 \quad \bullet \quad \bullet \quad \bullet),$$

where "•" denotes a missing observation. Such an individual contributes some information to estimate the model—we know that he or she could only have a complete chain matching one of the last four in Table 5.1.

It is assumed for the purposes of this chapter that the missing data in the above example are ignorable in the sense that the individual's ultimate event time, though unobserved, is unrelated to the timing of the final (censored) observation on that individual. The conventional assumption of noninformative censoring (i.e., that censoring times are independent of event times conditional on the observed covariates) corresponds to the assumption of ignorable missingness in a general latent variable modeling framework. These models can be estimated using maximum likelihood under the assumption of MAR (Little & Rubin, 2002). The observed data likelihood for uncensored individuals with $T_i = t_i$ where $t_i \leq j_i \leq J$, is given by:

$$\begin{aligned} l_i &= \Pr(T = t_i) \\ &= \Pr(E_1 = 0, E_2 = 0, \ldots, E_{t_i-1} = 0, E_{t_i} = 1, E_{t_i+1} = 2, \ldots, E_{j_i} = 2). \end{aligned} \tag{5.26}$$

We see above that the observed data likelihood can be written in terms of the likelihood of the observed chain which is, in essence, the joint probability of the event indicators at the observed values for each uncensored individual. The joint probability given in Equation 5.26 can be expressed as a product of conditional probabilities given by:

$$\begin{aligned} l_i = {} & \Pr(E_1 = 0)\cdot\Pr(E_2 = 0 \mid E_1 = 0)\cdot\Pr(E_3 = 0 \mid E_2 = 0, E_1 = 0)\cdots \\ & \Pr(E_{t_i-1} = 0 \mid E_{t_i-2} = 0, \ldots, E_1 = 0)\cdot\Pr(E_{t_i} = 1 \mid E_{t_i-1} = 0, \ldots, E_1 = 0)\cdot \\ & \Pr(E_{t_i+1} = 2 \mid E_{t_i} = 1, E_{t_i-1} = 0, \ldots, E_1 = 0)\cdot\Pr(E_{t_i+2} = 2 \mid E_{t_i+1} = 2, \\ & \quad E_{t_i} = 1, E_{t_i-1} = 0, \ldots, E_1 = 0)\cdots \\ & \Pr(E_{j_i} = 2 \mid E_{j_i-1} = 2, \ldots, E_{t_i+1} = 2, E_{t_i} = 1, E_{t_i-1} = 0, \ldots, E_1 = 0). \end{aligned} \tag{5.27}$$

Since the model for the event history is a requisite first-order Markov chain, all the conditional probabilities given in Equation 5.27 reduce to transition probabilities, for example:

$$\Pr(E_{t_i+1} = 2 \mid E_{t_i} = 1, E_{t_i-1} = 0, \ldots, E_1 = 0) = \Pr(E_{t_i+1} = 2 \mid E_{t_i} = 1).$$

So the observed data likelihood further simplifies to:

$$\begin{aligned} l_i = {} & \Pr(E_1 = 0)\cdot\Pr(E_2 = 0 \mid E_1 = 0)\cdots\Pr(E_{t_i-1} = 0 \mid E_{t_i-2} = 0)\cdot \\ & \Pr(E_{t_i} = 1 \mid E_{t_i-1} = 0)\cdot\Pr(E_{t_i+1} = 2 \mid E_{t_i} = 1)\cdot \\ & \Pr(E_{t_i+2} = 2 \mid E_{t_i+1} = 2)\cdots\Pr(E_{j_i} = 2 \mid E_{j_i-1} = 2). \end{aligned} \tag{5.28}$$

Looking at the product pattern in Equation 5.28, the observed data likelihood can be described by a product of a marginal probability in the first time point, transition probabilities from a pre-event to pre-event state in time periods 2 through $t_i - 1$, the transition probability from a pre-event state in time period $t_i - 1$ to an event state in time period t_i, the transition probability from an event state in time period t_i to a post-event state in time period $t_i + 1$, and transition probabilities from a post-event to post-event state in time periods $t_i + 1$ through j_i. Thus, the likelihood of the observed event chain can be expressed in terms of the marginal probability in the first time period and the transition probabilities for all remaining time periods. That is:

$$\begin{aligned} l_i &= \left(\Pr(E_1 = 0)\right)\cdot\left(\prod_{v=2}^{t_i-1}\Pr(E_v = 0 \mid E_{v-1} = 0)\right)\cdot\left(\Pr(E_{t_i} = 1 \mid E_{t_i-1} = 0)\right)\cdot \\ &\quad \left(\Pr(E_{t_i+1} = 2 \mid E_{t_i} = 1)\right)\cdot\left(\prod_{v=t_i+2}^{j_i}\Pr(E_v = 2 \mid E_{v-1} = 2)\right) \\ &= (\pi_0)\cdot\left(\prod_{v=2}^{t_i-1}\tau_{0(v)|0(v-1)}\right)\cdot\left(\tau_{1(t_i)|0(t_i-1)}\right)\cdot\left(\tau_{2(t_i+1)|1(t_i)}\right)\cdot\left(\prod_{v=t_i+2}^{j_i}\tau_{2(v)|2(v-1)}\right) \\ &= (\pi_0)\cdot\left(\prod_{v=2}^{t_i-1}\left(1-\tau_{1(v)|0(v-1)}\right)\right)\cdot\left(\tau_{1(t_i)|0(t_i-1)}\right) \\ &= \left(\prod_{v=1}^{t_i-1}\left(1-P_h(v)\right)\right)\cdot\left(P_h(t_i)\right). \end{aligned} \tag{5.29}$$

The above likelihood holds for all individuals known to be in one of the first six chain patterns in Table 5.1, including those that go missing after the event occurrence but before the conclusion of the study. For right-censored individuals with $T_i > j_i$ and $j_i \leq J$, the observed data likelihood is given by:

$$\begin{aligned} l_i &= \Pr(T > j_i) \qquad (5.30) \\ &= \Pr(E_1 = 0, E_2 = 0, \ldots, E_{j_i} = 0) \\ &= \left(\Pr(E_1 = 0)\right) \cdot \left(\prod_{v=2}^{j_i} \Pr(E_v = 0 \mid E_{v-1} = 0) \right) \\ &= (\pi_0) \cdot \left(\prod_{v=2}^{j_i} \tau_{0(v)|0(v-1)} \right) \\ &= (\pi_0) \cdot \left(\prod_{v=2}^{j_i} (1 - \tau_{1(v)|0(v-1)}) \right) \\ &= \left(\prod_{v=1}^{j_i} (1 - P_h(v)) \right). \end{aligned}$$

The above likelihood holds for individuals in the last chain pattern in Table 5.1 as well as all those in the first six chains who go missing *before* the event occurrence. It follows that the observed data likelihood for the sample is given by:

$$L = \prod_{i=1}^{n} \left[\left(P_h(t_i) \cdot \prod_{v=1}^{t_i - 1} (1 - P_h(v)) \right)^{I(T_i \leq j_i)} \cdot \left(\prod_{v=1}^{j_i} (1 - P_h(v)) \right)^{I(T_i > j_i)} \right], \quad (5.31)$$

which is identical to the likelihood we constructed in an earlier section. Note that we carry over our earlier assumption of $T_i \sim T$ to this likelihood construction. In the Markov chain setting this is referred to as a (population) *homogeneous* Markov chain meaning that the transition probabilities hold for all members of the population from which the sample was drawn.

The discrete-time survival model as a Markov chain model can be represented diagrammatically at shown in Figure 5.1. If we examine the figure in traditional path diagram fashion, we would interpret the paths between each of the adjacent E elements as regression paths. In fact, when we go to specify this model in software such as M*plus* (Muthén

Figure 5.1 Path diagram of a discrete-time survival process as a Markov chain.

& Muthén, 1998–2006), we need to re-express the transition matrices in terms of multinomial logistic regressions.[5] It will then be the regression parameters rather than the actual probabilities that will be estimated in the maximum likelihood estimation procedure, but those parameter estimates are directly translatable into estimates for the transition probabilities of interest.

The general multinomial logistic regression for relating a categorical variable, C, with K categories, to a covariate, X, is given by:

$$\Pr(C = k \mid x) = \frac{\exp(\alpha_k + \beta_k x)}{\sum_{m=1}^{K} \exp(\alpha_m + \beta_m x)}. \quad (5.32)$$

Usually, the last category, K, is selected as the reference class, but in this case we will specify the first category as the reference with $\alpha_1 = 0$ and $\beta_1 = 0$. For the Markov chain without covariates, X is the preceding state variable. In this case, the preceding state variable would have three states (except for the two states of time period 1) that go into the regression model as two dummy variables with the first category ($E_{j-1} = 0$) as the reference group. We will see later how to incorporate observed covariates into the model. Thus, the probabilities for state variable E_j conditional on E_{j-1} in terms of a multinomial logistic regression is given by:

$$\Pr(E_{ji} = k \mid E_{(j-1)i}) = \frac{\exp\left(\alpha_{jk} + \beta_{jk1} \cdot \mathrm{I}(E_{(j-1)i} = 1) + \beta_{jk2} \cdot \mathrm{I}(E_{(j-1)i} = 2)\right)}{\sum_{m=0}^{2} \exp\left(\alpha_{jm} + \beta_{jm1} \cdot \mathrm{I}(E_{(j-1)i} = 1) + \beta_{jm2} \cdot \mathrm{I}(E_{(j-1)i} = 2)\right)}. \quad (5.33)$$

where $k \in \{0,1,2\}$, $\alpha_{j0} = 0$, and $\beta_{j01} = \beta_{j02} = 0$. Substituting the multinomial logistic regressions in place of the transition probabilities in the transition matrix gives (displayed by column because of space restrictions):

$$\mathbf{T}^1_{j j-1} = \begin{bmatrix} \frac{1}{1+\exp(\alpha_{j1})+\exp(\alpha_{j2})} \\ \frac{1}{1+\exp(\alpha_{j1}+\beta_{j11})+\exp(\alpha_{j2}+\beta_{j21})} \\ \frac{1}{1+\exp(\alpha_{j1}+\beta_{j12})+\exp(\alpha_{j2}+\beta_{j22})} \end{bmatrix} = \begin{bmatrix} \tau_{0(j)|0(j-1)} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1-P_h(j) \\ 0 \\ 0 \end{bmatrix}, \quad (5.34)$$

$$\mathbf{T}^2_{j j-1} = \begin{bmatrix} \frac{\exp(\alpha_{j1})}{1+\exp(\alpha_{j1})+\exp(\alpha_{j2})} \\ \frac{\exp(\alpha_{j1}+\beta_{j11})}{1+\exp(\alpha_{j1}+\beta_{j11})+\exp(\alpha_{j2}+\beta_{j21})} \\ \frac{\exp(\alpha_{j1}+\beta_{j12})}{1+\exp(\alpha_{j1}+\beta_{j12})+\exp(\alpha_{j2}+\beta_{j22})} \end{bmatrix} = \begin{bmatrix} \tau_{1(j)|0(j-1)} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} P_h(j) \\ 0 \\ 0 \end{bmatrix},$$

$$\mathbf{T}^3_{j j-1} = \begin{bmatrix} \frac{\exp(\alpha_{j2})}{1+\exp(\alpha_{j1})+\exp(\alpha_{j2})} \\ \frac{\exp(\alpha_{j2}+\beta_{j21})}{1+\exp(\alpha_{j1}+\beta_{j11})+\exp(\alpha_{j2}+\beta_{j21})} \\ \frac{\exp(\alpha_{j2}+\beta_{j22})}{1+\exp(\alpha_{j1}+\beta_{j12})+\exp(\alpha_{j2}+\beta_{j22})} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

The elements of the transition matrix above are a function of six regression parameters: α_{j1}, α_{j2}, β_{j11}, β_{j12}, β_{j21}, and β_{j22} (not including α_{j0}, β_{j01}, and β_{j02} that are already fixed at zero). As noted previously, without observed covariates there is only one degree of freedom available for parameter estimation for each transition matrix. The most direct way of achieving model identification is to fix five of the six regression parameters for each matrix. The key here is to find fixed values for five of the six parameters that will properly correspond to the fixed probabilities in the transition matrix, such as $\tau_{2(j)|2(j-1)} = 1$. Since we are using the logit link function to relate the event state in time period j to the event state in time period $j-1$, we can only approximate probabilities of 0 and 1. Take, for example the regression corresponding to $\tau_{2(j)|0(j-1)}$:

$$\tau_{2(j)|0(j-1)} = \frac{\exp(\alpha_{j2})}{1+\exp(\alpha_{j1})+\exp(\alpha_{j2})} = 0. \quad (5.35)$$

The term $\exp(\alpha_{j2})$ in the numerator can become very small when α_{j2} is very small (that is, a large and negative number) and since $\exp(\alpha_{j1})$ and $\exp(\alpha_{j2})$ will always be greater than zero, the denominator will always be greater than 1. Suppose we choose a values for α_{j2} sufficiently small that, regardless of the value of α_{j1}, the quantity given above will approximate 0, say $\alpha_{j2} = -20$. Then:

$$\tau_{2(j)|0(j-1)} = \frac{\exp(-20)}{1+\exp(\alpha_{j1})+\exp(-20)} = \frac{2.06\cdot 10^{-9}}{1+\exp(\alpha_{j1})+2.06\cdot 10^{-9}} \approx 0.$$

If we fix the value of α_{j2} then we must allow the value of α_{j1} to be freely estimated since the only non-restricted elements of the transition matrix, $\tau_{0(j)|0(j-1)}$ and $\tau_{1(j)|0(j-1)}$, are functions of α_{j1} and α_{j2}. If we leave α_{j1} free to be estimated, then the remaining regression parameters must be fixed. Using similar reasoning as for α_{j2}, we need to choose values for β_{j21} and β_{j22} sufficiently large that $\tau_{2(j)|1(j-1)}$ and $\tau_{2(j)|2(j-1)}$ will approximate 1, say $\beta_{j21} = \beta_{j22} = 40$. And, β_{j11} and β_{j12} must be sufficiently small such that $\tau_{0(j)|1(j-1)}$, $\tau_{0(j)|2(j-1)}$, $\tau_{1(j)|1(j-1)}$, and $\tau_{1(j)|2(j-1)}$ will approximate 0, say $\beta_{j11} = \beta_{j12} = -10$.[6]

Once estimates for the free parameters are obtained, the fixed values for α_{j2}, β_{j11}, β_{j12}, β_{j21}, and β_{j22}, along with the estimate for α_{j1}, can be substituted back into the transition matrix to obtain the estimated transition probabilities, mainly the estimates for $\tau_{1(j)|0(j-1)} = P_h(j)$. For the first time interval, the marginal probabilities can also be expressed through parameters in an unconditional logistic regression given by:

$$\Pr(E_1 = 0) = \frac{1}{1+\exp(\alpha_{11})} = 1 - P_h(1), \quad (5.36)$$

$$\Pr(E_1 = 1) = \frac{\exp(\alpha_{11})}{1+\exp(\alpha_{11})} = P_h(1).$$

Since α_{11} is the only parameter associated with E_1, it can be freely estimated. Notice that there is no α_{12} in the model although there is an implicit α_{10} fixed at zero. Furthermore, the regression for E_2 on E_1 only includes an indicator for $E_1 = 1$, so there are no β_{2k2} coefficients for $k = 0, 1, 2$.

OBSERVED INDICATORS OF EVENT STATES

Up until this point, we have dealt with event history data in the form of a J-element vector of event states, $\mathbf{E}_i = (E_1, E_2, \ldots, E_J)$. Now consider an alternate form of discrete-time event data, especially common in prospective longitudinal studies, where event occurrence is recorded separately for each time period. In other words, for each time period there is an event indicator, U_j, where $U_j = 1$ if the event occurred in time period j and $U_j = 0$ otherwise. It is only by examining the whole string of U_j terms from $j = 1$ to $j = J$ that it is possible to determine whether $U_j = 0$ is indicating a pre-event state or a post-event state. This presupposes that event occurrence does not necessarily prohibit future observation of the individual, that is, that U_j can be observed for $j > T_i$. Let us consider three hypothetical cases. An individual who experienced the event in time period 3 would have the following **U** and **E** vectors:

$$\mathbf{u}_i = (0\ \ 0\ \ 1\ \ 0\ \ 0\ \ 0) \text{ and}$$
$$\mathbf{e}_i = (0\ \ 0\ \ 1\ \ 2\ \ 2\ \ 2).$$

An individual who did not experience the event during the entire observation period would have:

$$\mathbf{u}_i = (0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0) \text{ and}$$
$$\mathbf{e}_i = (0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0).$$

Finally, an individual who was censored during the fourth time period such that the third time period is the last complete time period of observation would have the following **U** and **E** vectors:

$$\mathbf{u}_i = (0\ \ 0\ \ 0\ \ \bullet\ \ \bullet\ \ \bullet) \text{ and}$$
$$\mathbf{e}_i = (0\ \ 0\ \ 0\ \ \bullet\ \ \bullet\ \ \bullet),$$

where "•" denotes a missing observation.

Thinking of the U_j terms as indicators of the E_j elements, we can reframe those E_j as (partially) latent state (or class) variables whose values are determined by the observed event indicators, U_j, and the previous latent event state variable, E_{j-1}. Now the transitions are defined on the structural level of a latent variable model. Instead of a discrete-time Markov chain, we now have a discrete-time *latent* Markov chain. There are advantages to reformulating the model this way that will be discussed later in the chapter. Figure 5.2 displays the path diagram representation of this model.

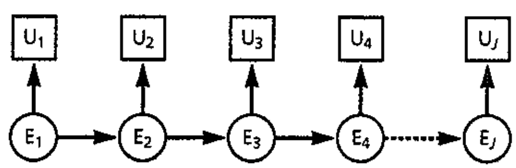

Figure 5.2 Path diagram of a discrete-time survival process as a latent Markov chain.

For the model depicted in Figure 5.2, the event states are still related to one another through a series of multinomial logistic regressions specified exactly as was done in the previous section. The additional part of the model that must be newly specified is the measurement model, relating the U_j terms to the corresponding E_j elements. For now, we will assume that each event indicator, U_j, is a perfect (i.e., without error) indicator of each E_j. That is:

$$\begin{aligned} &\Pr(U_j = 0 \mid E_j = 0) = 0, \\ &\Pr(U_j = 0 \mid E_j = 2) = 0, \text{ and} \\ &\Pr(U_j = 1 \mid E_j = 1) = 1. \end{aligned} \tag{5.37}$$

As with the conditional probabilities in the transition matrices, the probabilities given above must be re-specified by another series of logistic regressions, relating each U_j to the underlying E_j. Here we will again use a regression specification that matches the one used by M*plus* (Muthén & Muthén, 1998-2006) to facilitate implementation of these models in that software. The logistic regression relating U_j to E_j is given by:

$$\Pr(U_j = 1 \mid E_j = k) = \frac{1}{1 + \exp(\omega_{jk})}, \tag{5.38}$$

where $k \in \{0,1,2\}$.[7] In order to impose the restrictions on the regression parameters to match the restrictions on the conditional probabilities above, we must again settle for approximations of 0 and 1. Fixing $\omega_{j0} = \omega_{j2} = 20$ and $\omega_{j1} = -20$ yields

$$\Pr(U_j = 1 \mid E_j = 0) = \frac{1}{1 + \exp(\omega_{j0})} = \frac{1}{1 + \exp(20)} = \frac{1}{1 + 4.85 \cdot 10^{8}} \approx 0,$$

$$\Pr(U_j = 1 \mid E_j = 1) = \frac{1}{1 + \exp(\omega_{j1})} = \frac{1}{1 + \exp(-20)} = \frac{1}{1 + 2.06 \cdot 10^{-9}} \approx 1,$$

$$\Pr(U_j = 1 \mid E_j = 2) = \frac{1}{1 + \exp(\omega_{j2})} = \frac{1}{1 + \exp(20)} = \frac{1}{1 + 4.85 \cdot 10^{8}} \approx 0.$$

Since individuals in time period one can only occupy states 0 or 1, there is no ω_{12} in the model.

The observed data likelihood for an individual i is now given in terms of both U and E elements by:

$$\begin{aligned} l_i &= \Pr(U_1 = u_{1i}, U_2 = u_{2i}, \ldots, U_{j_i} = u_{j_i}) \\ &= \sum_{w=1}^{j_i+1} \left(\Pr\left(U_1 = u_{1i}, U_2 = u_{2i}, \ldots, U_{j_i} = u_{j_i} \mid (E_1, E_2, \ldots, E_{j_i}) = \mathbf{e}_w\right) \cdot \Pr\left((E_1, E_2, \ldots, E_{j_i}) = \mathbf{e}_w\right) \right) \\ &= \sum_{w=1}^{j_i+1} \left[\left(\prod_{v=1}^{j_i} \Pr(U_v = u_{vi} \mid (E_1, E_2, \ldots, E_{j_i}) = \mathbf{e}_w \right) \cdot \Pr\left((E_1, E_2, \ldots, E_{j_i}) = \mathbf{e}_w\right) \right] \\ &= \sum_{w=1}^{j_i+1} \left[\left(\prod_{v=1}^{j_i} \Pr(U_v = u_{vi} \mid E_v = e_{wv}) \right) \cdot \Pr(E_1) \cdot \prod_{v=2}^{j_i} \Pr(E_v = e_{wv} \mid E_{v-1} = e_{w,v-1}) \right], \end{aligned} \tag{5.39}$$

where $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{j_i+1}\}$ represents the set of all possible chains for $(E_1, E_2, \ldots, E_{j_i})$ and e_{wv} represents the value of E_v in chain w. Notice that in this formulation of the likelihood, we assume that the event indicators are independent of each other conditional on the underlying event state variables and that each event indicator only depends directly on the corresponding event state for that time period.

Event History Predictors

Including covariate predictors of event states across time is reasonably straightforward. Since we already have the probability distributions of the E_j event state variables specified as multinomial logistic regressions, we need only add the covariates as additional predictors in the regressions. However, given the restrictions we have imposed on the transition matrices, the only transition probabilities that may be influenced by covariates are $\tau_{1(j)|0(j-1)}$ and, as its complement, $\tau_{0(j)|0(j-1)}$. In other words, a covariate, X, may influence E_j only if $E_{j-1} = 0$. This restriction can be imposed by only allowing X to have an effect on $\Pr(E_j = 1 \mid E_{j-1})$ through an interaction term in the regression equation between the X and the indicator for E_{j-1} as shown below:

$$\Pr(E_{ji} = k \mid E_{(j-1)i}, x_i) = \frac{\exp\left(\alpha_{jk} + \beta_{jk1} \cdot \mathrm{I}(E_{(j-1)i} = 1) + \beta_{jk2} \cdot \mathrm{I}(E_{(j-1)i} = 2) + \gamma_{jk} x_i \cdot \mathrm{I}(E_{(j-1)i} = 0)\right)}{\sum_{m=0}^{2} \exp\left(\alpha_{jm} + \beta_{jm1} \cdot \mathrm{I}(E_{(j-1)i} = 1) + \beta_{jm2} \cdot \mathrm{I}(E_{(j-1)i} = 2) + \gamma_{jk} x_i \cdot \mathrm{I}(E_{(j-1)i} = 0)\right)}. \tag{5.40}$$

In this case, $\gamma_{j0} = \gamma_{j2} = 0$ while γ_{j1} is freely estimated. The parameter γ_{j1} can be interpreted as the log hazard odds ratio for transitioning to an event state in time period j associated with a one unit increase in X among those in a pre-event state in time period $j - 1$.

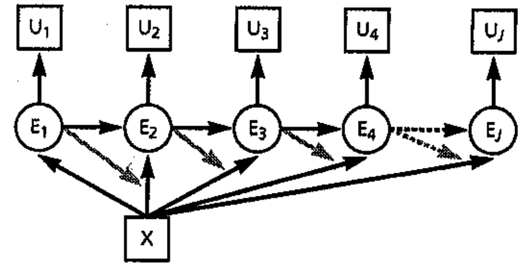

Figure 5.3 Path diagram of a discrete-time survival process as a latent Markov chain with a time-invariant covariate.

For the first event state, E_1, the effect of X is through the logistic regression as given by:

$$\Pr(E_1 = 1 \mid x_i) = \frac{\exp(\alpha_{11} + \gamma_{11} x_i)}{1 + \exp(\alpha_{11} + \gamma_{11} x_i)}. \tag{5.41}$$

Notice that the coefficient on X, γ_{j1}, is indexed by j, which implies that the effect of X on $\Pr(E = 1)$ may differ across the time intervals $j = 1, \ldots, J$. In addition, a covariate X could also be time-varying. Of course, it is possible to have multiple time-varying and time-invariant covariates in a single model. The proportional hazard odds model, equivalent to assuming a time-invariant effect of X, can be obtained by imposing an equality restriction such that $\gamma_{j1} = \gamma_1$ for all j. Figure 5.3 displays the path diagram for the model with a covariate included. In the figure, the arrows pointing from each E_j to the path from X to E_{j+1} are used to represent the interaction terms in the multinomial regressions as given in Equation 5.39. (For more general discrete-time latent Markov chain models with time-invariant and time-varying covariates, see, e.g., Nylund, Muthén, Nishina, Bellmore, & Graham, 2006; Vermunt et al., 1999.)

Recidivism Example

In order to demonstrate how this model may be applied in a real data setting, we utilize a data set from a randomized field experiment originally reported by Rossi, Berk, and Lenihan (1980) that has been used extensively by Allison (1984, 1995) as a pedagogical example in a continuous-time survival analysis framework and also by Muthén and Masyn (2005) as an example in the discrete-time setting. In this study, 432 inmates released from Maryland state prisons were randomly assigned to either an intervention or control condition. The intervention consisted of financial assistance provided to the released inmates for the duration of the study period. Those in the control condition received no aid. The inmates were followed for one year after their release. The event of interest was re-arrest with an emphasis on the influence of a set of explanatory variables (including intervention status) on the likelihood of recidivism. The data available on each inmate is detailed to the week level (i.e., 52 observation intervals). However, for the illustrative purposes of this chapter, the data were recoded into 13 four-week intervals, referred to as "months," identical to the use of the data by Muthén and Masyn (2005). For this chapter, we will estimate the baseline hazard probabilities in the control group as well as the effect of financial aid on the hazard probability of re-arrest under the proportional hazard odds assumption for the first six months following release. Select portions of the M*plus* syntax can be found in the appendix.[8] The observed data consist of six binary event indicators, $U_1,\ldots,U_6$, where $U_j = 1$ if the first re-arrest occurred in month j and is equal to zero otherwise. Results in terms of the regression parameters are given in Table 5.2. Substituting the estimated parameter values back into the regression equations, we can obtain the estimated marginal probabilities for time period 1 and the estimated transition matrices for the control group (financial aid intervention indicator = 0) as

TABLE 5.2 Recidivism Model Results (LL = −233.91)

Parameter	Estimate	Standard Error	*p*-value
$\omega_{10},\ldots,\omega_{60}$	20.00	fixed	—
$\omega_{11},\ldots,\omega_{61}$	−20.00	fixed	—
$\omega_{22},\ldots,\omega_{62}$	20.00	fixed	—
$\beta_{201},\ldots,\beta_{601}$	0.00	fixed	—
$\beta_{302},\ldots,\beta_{602}$	0.00	fixed	—
$\beta_{211},\ldots,\beta_{611}$	−10.00	fixed	—
$\beta_{312},\ldots,\beta_{612}$	−10.00	fixed	—
$\beta_{221},\ldots,\beta_{621}$	40.00	fixed	—
$\beta_{322},\ldots,\beta_{622}$	40.00	fixed	—
$\alpha_{10},\ldots,\alpha_{60}$	0.00	fixed	—
$\alpha_{22},\ldots,\alpha_{62}$	−20.00	fixed	—
α_{11}	−4.59	0.55	<0.001
α_{21}	−3.87	0.38	<0.001
α_{31}	−3.99	0.38	<0.001
α_{41}	−3.84	0.38	<0.001
α_{51}	−3.32	0.31	<0.001
α_{61}	−3.78	0.39	<0.001
γ_1 (Intervention)	−0.18	0.29	0.54

given below. Although the estimated effect of the financial aid intervention is small and statistically non-significant with an estimated hazard odds ratio (hOR) of 0.84, the estimated transition matrices for the intervention group are computed as well for the sake of completeness.

$$\hat{\Pr}(E_1 = 0 \mid \text{control}) = 0.991, \qquad \hat{\Pr}(E_1 = 0 \mid \text{intervention}) = 0.992,$$

$$\Pr(E_1 = 1 \mid \text{control}) = 0.009, \qquad \Pr(E_1 = 1 \mid \text{intervention}) = 0.008,$$

$$\hat{\mathbf{T}}_{21(\text{control})} = \begin{bmatrix} 0.981 & 0.019 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}, \qquad \hat{\mathbf{T}}_{21(\text{intervention})} = \begin{bmatrix} 0.984 & 0.016 & 0.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix},$$

$$\hat{\mathbf{T}}_{32(\text{control})} = \begin{bmatrix} 0.983 & 0.017 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}, \qquad \hat{\mathbf{T}}_{32(\text{intervention})} = \begin{bmatrix} 0.986 & 0.014 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix},$$

$$\hat{\mathbf{T}}_{43(\text{control})} = \begin{bmatrix} 0.981 & 0.019 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}, \qquad \hat{\mathbf{T}}_{43(\text{intervention})} = \begin{bmatrix} 0.984 & 0.016 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix},$$

$$\hat{\mathbf{T}}_{54(\text{control})} = \begin{bmatrix} 0.968 & 0.032 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}, \qquad \hat{\mathbf{T}}_{54(\text{intervention})} = \begin{bmatrix} 0.973 & 0.027 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix},$$

$$\hat{\mathbf{T}}_{65(\text{control})} = \begin{bmatrix} 0.980 & 0.020 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}, \qquad \hat{\mathbf{T}}_{65(\text{intervention})} = \begin{bmatrix} 0.983 & 0.017 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}.$$

These results (log likelihood value, baseline hazard probabilities, and intervention effect) match exactly the results obtained using the latent class regression approach presented in Muthén and Masyn (2005). The results do not match the final analysis presented in previous papers because for this analysis illustration, only six time periods were used and none of the other measured covariates from the original study were included.

Measurement Error on the Event Indicators

The model specification described in the previous sections may seem needlessly complicated for a single, non-recurring event history analysis. However, approaching the discrete-time model in this way allows us to deal with the possibility of measurement error on the event indicators. Consider a simplified scenario where a sample of children, say age 10, none of whom have ever had a drink of alcohol prior to age 10, are asked at 12 yearly follow-ups if they had their first drink in the previous year (yes/no). If we assume they respond with total accuracy, then they will respond in the affirmative for one and only one year and that year will mark the age of onset for alcohol consumption. However, in questions such as the ones in this example, there is likely to be some error (e.g., a child reports having his or her first drink in the previous year at two different follow-ups). To understand the impact that such error in the determination of event status can have on successful estimation of the hazard probabilities and covariate effects, we first need to define the types of error that can be made. Borrowing from the medical and epidemiological literature, we define the *sensitivity* of an event indicator as the likelihood that the indicator will have a value of 1 given that an individual is in a "true" event state. That is:

$$\text{Sensitivity} = \Pr(U_j = 1 \mid E_j = 1). \qquad (5.42)$$

Recall in the previous section, we fixed the sensitivity of the event indicators to 1 by fixing the ω_{j1} parameters to –20. We define the *specificity* of an event indicator as the likelihood that the indicator will have a value of 0 given that an individual is in either a pre-event or post-event state. That is:

$$\text{Specificity} = \Pr(U_j = 0 \mid E_j \neq 1). \qquad (5.43)$$

In the previous section, we fixed the specificity of the event indicators to 1 by fixing the ω_{j0} parameters and ω_{j2} parameters to 20. Note that it would be possible to have unique specificities for each event indicator conditional on the pre-event or post-event state, but here it will be assumed that:

$$\Pr(U_j = 0 \mid E_j = 0) = \Pr(U_j = 0 \mid E_j = 2). \qquad (5.44)$$

(For another example of the use of sensitivity and specificity in a latent class setting, see Rindskopf & Rindskopf, 1986.)

One can intuit the impact that less than perfect sensitivity or specificity might have on the estimated hazard probabilities. With sensitivity less than 1, some individuals in a true event state will not be identified as such and we would expect the hazard probabilities to be under-estimated. With specificity less that 1, some individuals in a true pre-event state will be identified as being in an event state and we would expect the hazard probabilities to be over-estimated. As is the case with any outcome measured with error, we would expect the effects of a covariate on the hazard odds in the presence of less than perfect sensitivity or specificity to be attenuated. For the purposes of this chapter, we assume that the measurement error is random at each time period and is independent of the error at other time periods.[9]

It is possible to directly compute the actual distortion in the hazard and survival probabilities in the case of complete data. Consider an example with only three time periods of measure. Suppose the hazard probability is equal to 0.05 in all three time periods. Suppose also that the sensitivity is equal to 0.80 and the specificity is equal to 0.90. Table 5.3 gives the total probabilities (relative frequencies) of each of the possible chains on the structural level. Based on the relative frequencies in Table 5.3 and the given sensitivity and specificity levels, we can compute the probabilities for each possible event indicator sequence over the three time periods. An example for computing the probability of observing $U = (0, 0, 0)$ is given below:

$$\Pr(U_1=0,U_2=0,U_3=0) = \begin{pmatrix} \Pr(U_1=0,U_2=0,U_3=0 \mid \mathbf{E}=(1,2,2))\cdot\Pr(\mathbf{E}=(1,2,2)) + \\ \Pr(U_1=0,U_2=0,U_3=0 \mid \mathbf{E}=(0,1,2))\cdot\Pr(\mathbf{E}=(0,1,2)) + \\ \Pr(U_1=0,U_2=0,U_3=0 \mid \mathbf{E}=(0,0,1))\cdot\Pr(\mathbf{E}=(0,0,1)) + \\ \Pr(U_1=0,U_2=0,U_3=0 \mid \mathbf{E}=(0,0,0))\cdot\Pr(\mathbf{E}=(0,0,0)) \end{pmatrix}$$

$$= \begin{pmatrix} \Pr(U_1=0 \mid E_1=1)\cdot\Pr(U_2=0 \mid E_2=2)\cdot\Pr(U_3=0 \mid E_3=2)\cdot\Pr(\mathbf{E}=(1,2,2)) + \\ \Pr(U_1=0 \mid E_1=0)\cdot\Pr(U_2=0 \mid E_2=1)\cdot\Pr(U_3=0 \mid E_3=2)\cdot\Pr(\mathbf{E}=(0,1,2)) + \\ \Pr(U_1=0 \mid E_1=0)\cdot\Pr(U_2=0 \mid E_2=0)\cdot\Pr(U_3=0 \mid E_3=1)\cdot\Pr(\mathbf{E}=(0,0,1)) + \\ \Pr(U_1=0 \mid E_1=0)\cdot\Pr(U_2=0 \mid E_2=0)\cdot\Pr(U_3=0 \mid E_3=0)\cdot\Pr(\mathbf{E}=(0,0,0)) \end{pmatrix}$$

$$= \begin{pmatrix} (0.20)\cdot(0.90)\cdot(0.90)\cdot(0.05) + \\ (0.90)\cdot(0.20)\cdot(0.90)\cdot(0.0475) + \\ (0.90)\cdot(0.90)\cdot(0.20)\cdot(0.045) + \\ (0.90)\cdot(0.90)\cdot(0.90)\cdot(0.857) \end{pmatrix}$$

$$= 0.65.$$

Table 5.4 summarizes the relative frequencies for all of the possible event indicator sequences computed in a similar manner to the example above.

If a model were fit that assumed the event indicators functioned without error, then the first four indicator sequences (1–4) in Table 5.4 would all en-

TABLE 5.3 Expected Relative Frequencies of (Latent) Event State Chains for Constant $P_h(j) = 0.05$

Chain #	E_1	E_2	E_3	rf
1	1	2	2	0.050
2	0	1	2	0.048
3	0	0	1	0.045
4	0	0	0	0.857

TABLE 5.4 Expected Relative Frequencies of Event Indicator Sequences for Constant $P_h(j) = 0.05$, Sensitivity = 0.80, and Specificity = 0.90

Sequence #	U_1	U_2	U_3	rf
1	1	1	1	0.002
2	1	1	0	0.015
3	1	0	1	0.015
4	1	0	0	0.100
5	0	1	1	0.014
6	0	1	0	0.100
7	0	0	1	0.100
8	0	0	0	0.650

ter the likelihood as corresponding to the first state chain (1) in Table 5.3, equivalent to entering a true event state in the first time period. So the expected estimated probability for that chain is 0.002 + 0.015 + 0.015 + 0.10 = 0.132—a notable overestimation of the true chain probability (and time period 1 hazard probability) of 0.050. The next two indicator sequences (5–6) in Table 5.4 would enter the likelihood as corresponding to the second state chain (2) in Table 5.3, equivalent to entering a true event state in the second time period, and the expected estimated probability is 0.014 + 0.10 = 0.114—also an overestimation of the true chain probability (and time period 2 hazard probability) of 0.048. The next indicator sequence (7) in Table 5.4 would enter the likelihood as corresponding to the third state chain (3) in Table 5.3, equivalent to entering a true event state in the third time period, and the expected estimated probability is 0.100—another overestimation of the true chain probability (and time period 3 hazard probability) of 0.045. The final indicator sequence (8) in Table 5.4 would enter the likelihood as corresponding to never entering a true event state during the three time periods, i.e., state chain (4) in Table 5.3, and the expected estimated probability is 0.650. This corresponds to the expected estimate for $P_s(3)$ and is less than the true survival probability of 0.857.

Table 5.5 gives a summary of similar calculations for a population with a constant hazard probability of 0.05 across each of six time periods. Notice that with sensitivity and specificity values of 1.0, we obtain the population values as would be expected. Holding the sensitivity level at 1.0, a decrease in the specificity level from 1.0 to 0.9 and 0.8 leads to a remarkable level of distortion in the hazard probabilities—not only is there an upward bias in the hazard probabilities, that bias increases over the six time periods, distorting not only the size of the hazard probabilities but also the change in the hazard probabilities over time. And because of the relationship be-

TABLE 5.5 Expected Estimated Hazard Probabilities and Hazard Odds Ratio by Event Indicator Sensitivity and Specificity

Sensitivity $Pr(U=1 \mid E=1)$	Specificity $Pr(U=0 \mid E=0)$, $Pr(U=0 \mid E=2)$	Expected Estimated Hazard Probability Time Period 1	2	3	4	5	6	Expected Estimated Survival Probability $P_s(6)$	Expected Estimated Hazard Odds Ratio hOR
1.0	1.0	0.05	0.05	0.05	0.05	0.05	0.05	0.74	2.00
	0.9	0.11	0.12	0.13	0.15	0.17	0.21	0.38	1.37
	0.8	0.17	0.18	0.21	0.24	0.30	0.45	0.16	1.23
0.9	1.0	0.045	0.045	0.044	0.045	0.044	0.045	0.76	1.99
	0.9	0.11	0.12	0.13	0.14	0.17	0.20	0.39	1.32
	0.8	0.16	0.18	0.20	0.24	0.30	0.43	0.17	1.21
0.8	1.0	0.04	0.04	0.04	0.04	0.04	0.04	0.78	1.95
	0.9	0.11	0.11	0.13	0.14	0.16	0.19	0.40	1.28
	0.8	0.16	0.17	0.20	0.23	0.29	0.42	0.18	1.15

tween the hazard probabilities and the survival probabilities, the survival probability for the final time period is significantly under-estimated. Holding the specificity level at 1.0, a decrease in the sensitivity level from 1.0 to 0.9 and 0.8 leads to a downward bias in the hazard probability estimates. Of course, the total impact of sensitivity and specificity values less than 1.0 on the hazard probability estimates will depend, in part, on the size of the true hazard probabilities. This table illustrates that the impact on the (baseline) hazard probabilities of measurement error on event indicators can be quite different than the impact of unobserved heterogeneity (or unmeasured covariates) which always results in a downward bias in the baseline hazard.

The last column in Table 5.5 displays the impact of a range of sensitivity and specificity values on the hazard odds ratio assuming a "true" population with a binary covariate ($X = 0/1$) with a population prevalence of 0.5 and an effect of 0.693 on the logit scale (which translates to a hazard odds ratio (hOR) of 2.0). The baseline hazard probabilities ($X = 0$) for the population were a constant value of 0.05. We can see the under-estimation of the covariate effect that occurs. As with the baseline hazard probabilities, the total impact of sensitivity and specificity values less than 1.0 on the covariate effect estimates will depend on the size of the true covariate effect as well as the true baseline hazard probabilities. (For a more general discussion of uncorrelated and correlated measurement error in discrete-time Markov chain models, see, e.g., Bassi, Hagenaars, Croon, & Vermunt, 2000.)

As was discussed in the previous section, the discrete-time survival model without covariates and without restrictions on the hazard probabilities is a just-identified model. With event indicators, there are more degrees of freedom available (theoretically, $2^J - 1$); however, there are likely to be enough empty cells in the joint distribution of the event indicators to result in empirical non-identification when attempting to estimate the ω_j parameters. Consider the case where the U terms function without measurement error. In that situation, there are only J+1 observed patterns for the event indicators and that means there is not enough information to estimate ω_j parameters as well as the unique transition probabilities. Ultimately, this means that even if we believed there to be measurement error on the event indicators, it is unlikely we would be able to estimate any of the ω parameters, even when imposing constraints on model parameters such as measurement invariance (i.e., $\omega_{j0} = \omega_{k0}$, $\omega_{j1} = \omega_{k1}$, and $\omega_{j2} = \omega_{k2}$ for all $j, k = 1, \ldots, J$) and uniformity in the specificity of the event indicators for pre-event and post-event states (i.e, $\omega_{j0} = \omega_{j2}$ for all $j = 1, \ldots, J$). However, if we knew from a previous study or another source the approximate sensitivity and specificity of the event indicators, we could fix the ω parameters to values corresponding to those accuracy rates rather than fixing them at values corresponding to sensitivity and specificity of 1.[10] For example, if we knew the approximate specificity of the event indicator to be 0.80, we could fix the ω_{j0} parameters and ω_{j2} parameters to the value 1.386 (= logit(0.8)) in our estimation model. (For more information about identification of latent Markov models, see, e.g., Van de Pol & Langeheine, 1990.)

It is also possible to improve the chances of having an empirically identified model, and recovering the sensitivity and specificity of the event indicators, by including one or more predictors of event history in the model. To briefly explore this possibility, we conducted a simulation study, drawing 100 samples of size 1,000 from each of three populations. For each population, there was a single observed event indicator at each of six time periods. The sensitivity and specificity for all of the indicators were 0.90 (corresponding to population parameter values $\omega_0 = \omega_2 = 2.197$ and $\omega_1 = -2.197$). For each population, there was a time-invariant binary (0/1) predictor, evenly distributed across each population, with a time-invariant hazard odds ratio of 2.0 (corresponding to population parameter values $\gamma_{11} = \gamma_{21} = \ldots = \gamma_{61} = 0.693$). Each population had a constant baseline hazard probability—the hazard probability when the covariate was equal to zero—across the six time periods. This was the only feature that differed across the three populations. Population 1 had a baseline hazard probability of 0.05 (corresponding to population parameter values $\alpha_{11} = \alpha_{21} = \ldots = \alpha_{61} = -2.944$); Population 2 had a baseline hazard probability of 0.10 (corresponding to population parameter values $\alpha_{11} = \alpha_{21} = \ldots = \alpha_{61} = -2.197$); and Population 3 had a baseline hazard probability of 0.25 (corresponding to population parameter values $\alpha_{11} = \alpha_{21} = \ldots = \alpha_{61} = -1.099$). For each sample of 1,000 observations drawn from each population (100 draws or replications per population), a model was fit that estimated the ω_{jk} parameters under the assumption of measurement invariance and equal specificity for $E_j = 0$ and $E_j = 2$, the γ_{j1} parameters under the assumption of time-invariant effects of the covariate, and the α_{j1} parameters.

Table 5.6 summarizes the model results across the 100 replications for each of the three populations. The rows labeled "Population" give the values of the parameters in the population from which each of the 100 samples of $n = 1{,}000$ was drawn. The rows labeled "Avg. Est." give the average parameter estimate across the 100 estimated models. The rows labeled "Bias" give the difference between the average parameter estimate from the estimated models and the true population value and the "Relative Bias" is the estimated bias divided by the absolute value of the true population parameter value. The rows labeled "MSE" give the mean squared error which decomposes into a sum of the squared bias and variance of the estimates. The last rows, labeled "95% coverage," give the proportion of 95% confidence intervals for the estimate from the replication models that contained the true population parameter value. For both Populations

TABLE 5.6 Simulation Results for 3 Populations Based on 100 Replications of n = 1000

Parameter	ω_0, ω_2	ω_1	α_{11}	α_{21}	α_{31}	α_{41}	α_{51}	α_{61}	γ_1
Population 1: Sensitivity = 0.90; Specificity = 0.90; Baseline Hazard Probability = 0.05; hOR = 2.0									
Population	2.197	−2.197	−2.944	−2.944	−2.944	−2.944	−2.944	−2.944	0.693
Avg. Est.	2.163	−7.153	−8.573	−3.922	−33.253	−3.498	−6.047	−2.304	1.265
Bias	−0.034	−4.956	−5.629	−0.978	−30.309	−0.554	−3.103	0.640	0.572
Relative Bias	−0.015	−2.256	−1.912	−0.332	−10.295	−0.188	−1.054	0.217	0.825
MSE	0.022	78.667	2184.134	21.971	86567.688	27.227	1030.000	30.725	38.420
95% coverage	0.920	0.240	0.850	0.860	0.860	0.900	0.870	0.900	0.890
Population 2: Sensitivity = 0.90; Specificity = 0.90; Baseline Hazard Probability = 0.10; hOR = 2.0									
Population	2.197	−2.197	−2.197	−2.197	−2.197	−2.197	−2.197	−2.197	0.693
Avg. Est.	2.203	−6.560	−2.207	−2.165	−2.179	−2.137	−2.094	−2.047	0.727
Bias	0.006	−4.363	−0.010	0.032	0.018	0.060	0.100	0.150	0.088
Relative Bias	0.003	−1.986	−0.005	0.015	0.008	0.027	0.046	0.068	0.127
MSE	0.005	59.909	0.058	0.063	0.103	0.164	0.187	0.354	0.036
95% coverage	0.970	0.420	0.940	0.940	0.930	0.950	0.950	0.970	0.940
Population 3: Sensitivity = 0.90; Specificity = 0.90; Baseline Hazard Probability = 0.25; hOR = 2.0									
Population	2.197	−2.197	−1.099	−1.099	−1.099	−1.099	−1.099	−1.099	0.693
Avg. Est.	2.190	−2.574	−1.118	−1.092	−1.101	−1.109	−1.092	−1.016	0.688
Bias	−0.007	−0.377	−0.019	0.007	−0.002	−0.010	0.007	0.083	−0.005
Relative Bias	−0.003	−0.172	−0.017	0.006	−0.002	−0.009	0.006	0.076	−0.007
MSE	0.003	3.607	0.016	0.024	0.044	0.093	0.188	0.276	0.016
95% coverage	0.940	0.920	0.960	0.960	0.960	0.950	0.970	0.990	0.960

1 and 2, the parameter ω_1, corresponding to the event indicator sensitivity, is poorly estimated with a large negative relative bias (which translates to an overestimation of the sensitivity), large MSE, and low coverage proportion. However, for Population 1, there is also extremely poor estimation of the baseline hazard probabilities and the covariate effect is overestimated by an average of more than 80%. Note that this stands in contrast to the attenuated covariate effect when measurement error in the event indicators is ignored. The poor performance of the model estimates over the Population 1 replications should not be too surprising when we considered that with a true baseline hazard probability of only 0.05, there is very little information about latent state changes from pre-event to event. With Population 2, the model performance, with the exception of the estimation of event indicator sensitivity, is notably improved over the replications from Population 1, with much smaller relative bias, smaller MSE, and actual coverage near the expect 0.95. For Population 3, with a baseline hazard of 0.25, even the estimation for the parameter corresponding to the item sensitivity is much improved and the relative bias for the covariate effect is less than 1%.

This limited simulation study demonstrates that the actual baseline hazard probability can influence the ability of the model that allows for measurement error on single event indicators with at least one covariate to recover the true population parameters. It appears that for rare events, a less than perfect item sensitivity is difficult to estimate correctly, but that as the baseline hazard probability increases, the likelihood of recovery of hazard probabilities and covariate effects improves. The effect of the baseline hazard probability shown here represents only one of a number of dimensions, including sample size, number of time periods, true covariate effect size, true item specificity and sensitivity, that could influence the quality, in terms of both accuracy and precision, of parameter estimates in these single event indicator models. In general, having the empirical identification of the model dependent upon a covariate is less than ideal, and that will be reflected in model performance. As with any latent variable model, if we want a stable measurement model for the event states, we need multiple event indicators for each time period so that the measurement model at each time period is independently identifiable.

Multiple Event Indicators

If there is more than one measure of event status at each time period, we can extend the model previously presented to include M event indicators at each time point j, U_{mj}. Figure 5.4 displays the path diagram for this extension. The measurement model for E is now defined by the set of event indicators, and the relationship between each of the U_{mj} terms and E_j is

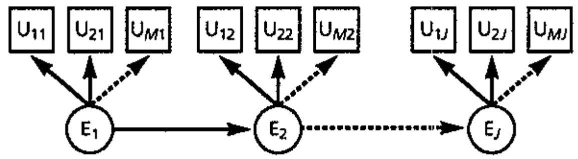

Figure 5.4 Path diagram of a discrete-time survival process as a latent Markov chain with multiple event indicators for each time period.

specified through a series of logistic regressions. The logistic regression relating U_{mj} to E_j is given by:

$$\Pr(U_{mj}=1\,|\,E_j=k)=\frac{1}{1+\exp(\omega_{mjk})}. \tag{5.45}$$

where $k \in \{0,1,2\}$. Essentially, the measurement models for the time periods are a set of latent class models (see, e.g., McCutcheon, 1987). As with a traditional latent class analysis, we assume that the event indicators at each time point j, $\mathbf{U}_j = (U_{1j}, U_{2j}, \ldots, U_{Mj})$, are independent of each other conditional on event status, in addition to assuming that the event indicators across time periods are independent of each other conditional on the underlying state variables and that each set of event indicators depend only on the corresponding event state for that time period. Under those assumptions, the observed data likelihood is given by:

$$\begin{aligned} l_i &= \Pr(\mathbf{U}_1=\mathbf{u}_{1i}, \mathbf{U}_2=\mathbf{u}_{2i}, \ldots, \mathbf{U}_{j_i}=\mathbf{u}_{j_i}) \\ &= \sum_{w=1}^{j_i+1}\left(\begin{aligned}&\Pr\left(\mathbf{U}_1=\mathbf{u}_{1i}, \mathbf{U}_2=\mathbf{u}_{2i}, \ldots, \mathbf{U}_{j_i}=\mathbf{u}_{j_i}\,|\,(E_1,E_2,\ldots,E_{j_i})=e_w\right)\cdot \\ &\Pr\left((E_1,E_2,\ldots,E_{j_i})=e_w\right)\end{aligned}\right) \\ &= \sum_{w=1}^{j_i+1}\left[\left(\prod_{v=1}^{j_i}\prod_{m=1}^{M}\Pr(U_{mv}=u_{mi}\,|\,E_v=e_{vw})\right)\cdot\Pr(E_1)\cdot\prod_{v=2}^{j_i}\Pr(E_v=e_{vw}\,|\,E_{v-1}=e_{w,v-1})\right], \end{aligned} \tag{5.46}$$

where, as before, $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{j_i+1}\}$ represents the set of all possible chains for $(E_1, E_2, \ldots, E_{j_i})$ and e_{vw} represents the value of E_v in chain w.

With multiple event indicators at each time, model identification may not necessitate the imposition of measurement invariance and uniformity of indicator sensitivity and specificity for pre-event and post-event states as it does in the case with only one event indicator per time period. For example, with four event indicators per time period, it would be possible to estimate time- and indicator-specific ω parameters while only imposing constraints corresponding to equal specificities for $E_j = 0$ and $E_j = 2$. It becomes essential with these multiple event indicator models to consider not only the constraints that are necessary for identification, but also the constraints that are necessary for interpretation. It may be quite reasonable to allow for item-specific sensitivity and specificity values since the different event indicators may have different levels of precision in their measurement of event status. However, it may not be reasonable to allow time-specific sensitivity and specificity for one or more of the event indicators, that is, to allow measurement non-invariance of event states over time. In the presence of measurement non-invariance, it becomes more difficult to assert that the levels of E have the same *meaning* across time. And without that consistency of meaning (i.e., "true" pre-event, event, and post-event states), the validity of the constraints on the transition probabilities, and the interpretations of the unconstrained transition probabilities as relating to the hazard probabilities of an event history process, diminishes.

Multiple event indicators can come in various forms. In some cases, the indicators may represent responses to the same question asked at different points in time. Returning to our earlier scenario, suppose that rather than asking the sample of children, who never drank prior to age 10, whether or not they had had their first drink in the previous year (yes/no) at each yearly follow-up, the children are asked at each yearly follow-up how old they were when they had their first drink. In this situation, a child may report two or more different ages of first drink. Instead of averaging the reported ages or favoring one response over another, all the responses can be taken as indicators of the "true" event status at each age. Supposing that all the children are at the age of 10 when the study begins and none have had their first drink prior to age 9,[11] at intake they could report an age of 10 as the age of first drink or report that they had not yet had their first drink; at the first follow-up at the age 11, the children could report an age of 10 or 11 as the age of first drink or report that they had not yet had their first drink; at the age 12 follow-up, children could report an age of 10, 11, or 12 as the age of first drink or report that they had not yet had their first drink; and so on. Supposing a 6-year (or 6-wave) study with yearly assessments starting at age 10, let E_j be the event state at age j where $j = 10, \ldots, J$ and $J = 15$. Let U_{mj} be the indicator for age of first drinking equal to j based on the response given during the wave m where $m = 1, \ldots, M_j$ and $M_j = J - j + 1$. In this case, there are actually a different number of event indicators for each time period. For age $j = 10$, there are a possible $J - j + 1 = 15 - 10 + 1 = 6$ event indicators; for age $j = 15$, there is only one event indicator from the final wave of assessment. Consider the hypothetical response sequence for waves 1-6 given in the response column of Table 5.7. This response sequence provides information on $U_{1,10}, \ldots, U_{6,10}$; $U_{2,11}, \ldots, U_{6,11}$; $U_{3,12}, \ldots, U_{6,12}$; $U_{4,13}, \ldots, U_{6,13}$; $U_{5,14}$, $U_{6,14}$; and $U_{6,15}$ as given in Table 5.7.

It would be important in this setting, as in any other, to carefully consider the constraints placed on the measurement models for the event states

TABLE 5.7 Hypothetical Question Response Sequence and Corresponding Event Indicator Values

Wave	Age	Response	Event Age 10	11	12	13	14	15
1	10	Never had a drink	$U_{1,10} = 0$					
2	11	Never had a drink	$U_{2,10} = 0$	$U_{2,11} = 0$				
3	12	First drink at 12	$U_{3,10} = 0$	$U_{3,11} = 0$	$U_{3,12} = 1$			
4	13	First drink at 11	$U_{4,10} = 0$	$U_{4,11} = 1$	$U_{4,12} = 0$	$U_{4,13} = 0$		
5	14	First drink at 12	$U_{5,10} = 0$	$U_{5,11} = 0$	$U_{5,12} = 1$	$U_{5,13} = 0$	$U_{5,14} = 0$	
6	15	First drink at 13	$U_{6,10} = 0$	$U_{6,11} = 0$	$U_{6,12} = 0$	$U_{6,13} = 1$	$U_{6,14} = 0$	$U_{6,15} = 0$

across time. In this case, it might make sense to assume the same sensitivity and specificity for indicators with the same number of years between the age the response was given and the age to which the response corresponds. For example, $U_{3,10}$, the indicator based on the response at wave 3 (age 12) about event status at age 10, may have the same sensitivity and specificity as $U_{5,12}$, the indicator based on the response at wave 5 (age 14) about event status at age 12.

In others cases, multiple indicators may correspond to a set of measures administered during each time period that are intended to collectively ascertain event status. For example, consider the proposed latent class model for the diagnosis of myocardial infarction (MI) presented by Rindskopf and Rindskopf (1986). Their latent class variable had two levels—MI and no MI—and was measured by a set of four indicators: presence of a positive Q-wave, classical clinical history, flipped LDH, and high CPK-MB. This latent class model could be integrated into an event history analysis of age of first myocardial infarction. In this case, only those hospitalized during a given time period would have measures on these indicators. So one might add a fifth event indicator so that those not hospitalized in a given year were known with a specificity of 1.0 not to have experienced the event in that time period. Here it would probably be quite reasonable to impose measurement invariance on the event indicators across the time periods of observation.

SUMMARY

This chapter has introduced the principles of discrete-time survival analysis and demonstrated, in detail, how an event history model can be reformulated as a restricted discrete-time Markov chain model. The problems that can be caused by measurement error in event occurrence were discussed, and it was shown how a discrete-time latent Markov chain model could be specified to account for measurement error. The presentation here was limited to single, non-recurring event history processes with complete and right-censored data (with non-informative censoring) measured in discrete time periods of equal width. It was assumed that no individual could occupy an event state for more than one time period and that event occurrence could be represented by a single state. It was also assumed that there was neither unobserved heterogeneity nor unmeasured covariates. However, the modeling framework presented in this chapter can accommodate recurring events, competing risks, multiple spells, frailties or unobserved heterogeneity (including the long-term survivor model which could be specified as restricted mover-stayer Markov chain model), informative censoring, and clustered data. Furthermore, there is nothing in the framework that restricts event indicators to binary measures—event indicators could have multiple categories or even be continuous. Some of the same identification issues discussed earlier might still apply depending on the nature and number of the event indicators for each time period. The single necessity for the measurement model for event status is consistency across time so that the event states retain their meaning throughout the observation period. All of the models presented and the extensions mentioned above can be specified and estimated using full-information maximum likelihood enabled by some of the more advanced statistical modeling software packages such as M*plus* (Muthén & Muthén, 1998–2006).

As noted in the beginning of this chapter, there are many approaches to discrete-time survival analysis available to applied researchers. The great advantage of working in a Markov chain modeling framework is gained in situations where the determination of event occurrence is likely to be imperfect and subject to measurement error. These situations are common, even in prospective longitudinal studies, when researchers must rely on self-reports of event occurrence or when the event of interest itself is not directly observable. This approach offers a promising alternative to mitigate the significant distortions in estimation and inference that may occur when not accounting for or presuming an absence of measurement error in the event history process under study.

NOTES

1. Discrete-time survival analysis can also be applied to events that are themselves occurring on a discrete-time scale, such as grade retention.
2. Although beyond the scope of this chapter, it is possible to extend these models to a multilevel setting with individuals clustered in higher-level units. See, for example, Steele (2003).
3. Although we assume here that the time intervals are all of equal length, this assumption is not necessary for any of the models presented.
4. In the first time period, there are actually only two possible states—pre-event and event—as will be explained later in this section.
5. Log-linear models could also be used, but would only allow for the inclusion of categorical predictors of E (see, e.g., Vermunt, 1997).
6. Keep in mind that the values that we have chosen are somewhat arbitrary in that we are selecting values that will allow us to approximate the necessary fixed transition probabilities and that other values of similar magnitude and direction would work equally well.
7. Note that we follow the M*plus* specification for binary and ordered categorical variables where the ω parameters are referred to as *thresholds* and are equivalent to the additive inverses of the intercepts in a traditional binary logistic regression. For more information, see Muthén and Muthén (1998–2006).
8. Full input and output files are available from the author upon request.

9. This situation is quite different from the situation with unmeasured heterogeneity in the event history process when there would be additional correlation in the event states across time due to one or more unmeasured covariates, often referred to collectively as *frailty*. We will not deal with *frailties* in this chapter, but it is possible to extend the models presented here to account for possible unobserved heterogeneity in the form of a higher-order random effect or latent class variable. For more information on modeling unobserved heterogeneity, see, for example, Heckman and Singer (1984); Land et al. (2001); Masyn (2003); Muthén and Masyn (2005); Trussel and Richard (1985); Vaupel, Manton, and Stanton (1979); and Vermunt (1997, 2002).
10. This is analogous to fixing the reliability for a single factor indicator.
11. It is not a necessary assumption that all the subjects be in a pre-event state upon entry into the study. The timeline could extend back before the beginning of the study so that subjects' reports of behavior prior to the study could also be included in the model. For example, some children at age 10 may report having their first drink at age 8 or 9.

REFERENCES

Aalen, O., & Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics*, 5, 141–150.

Allison, P. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology, 13*, 61–98.

Allison, P. (1984). *Event history analysis. Quantitative Applications in the Social Sciences, 46.* Thousand Oaks, CA: Sage.

Allison, P. (1995). *Survival analysis using the SAS system: A practical guide.* Cary, NC: SAS.

Andersen, P. (1988). Multistate models in survival analysis: A study of nephropathy and mortality in diabetes. *Statistics in Medicine, 7*, 661–670.

Bassi, F., Hagenaars, J., Croon, M., & Vermunt, J. (2000). Estimating true changes when categorical panel data are affected by uncorrelated and correlated classification errors: An application to unemployment data. *Sociological Methods and Research, 29*, 230–268.

Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, 34*, 187–220.

Gaskell, G., Wright, D., & O'Muircheartaigh, C. (2000). Telescoping of landmark events: Implications for survey research. *Public Opinion Quarterly, 64*, 77–89.

Heckman, J., & Singer, B. (1984). Economic duration analysis. *Journal of Econometrics, 24*, 63–132.

Hedeker, D., Siddiqui, O., & Hu, F. (2000). Random-effects regression analysis of correlated grouped-time survival data. *Statistical Methods in Medical Research, 9*, 161–179.

Laird, N., & Oliver, D. (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association, 76*, 231–240.

Land, K., Nagin, D., & McCall, P. (2001). Discrete-time hazard regression with hidden heterogeneity: The semiparametric mixed Poisson regression approach. *Sociological Methods and Research, 29*, 342–373.

Langeheine, R., & Van de Pol, F. (2002). Latent Markov models. In J. Hagenaars & McCutcheon (Eds.), *Applied latent class analysis* (pp. 304–344). Cambridge: Cambridge University Press.

Lindeboom, M., & Kerkhofs, M. (2000). Multistate models for clustered duration data: An application to workplace effects on individual sickness absenteeism. *The Review of Economics and Statistics, 82*, 668–684.

Little, R., & Rubin, D. (2002). *Statistical analysis with missing data* (2nd ed.). New York: Wiley.

Masyn, K. (2003). *Discrete-time survival mixture analysis for single and recurrent events using latent variables.* Unpublished doctoral dissertation, University of California, Los Angeles.

McCutcheon, A. (1987). *Latent class analysis.* Newbury Park, CA: Sage.

Muthén, B., & Masyn, K. (2005). Discrete-time survival mixture analysis. *Journal of Educational and Behavioral Statistics, 30*, 27–58.

Muthén, L., & Muthén, B. (1998-2006). *Mplus user's guide* (4th ed.). Los Angeles, CA: Muthén & Muthén.

Nylund, K. L., Muthén, B., Nishina, A., Bellmore, A., & Graham, S. (2006). *Stability and instability of peer victimization during middle school: Using latent transition analysis with covariates, distal outcomes, and modeling extensions.* Manuscript submitted for publication.

Prentice, R., & Gloeckler, L. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics, 34*, 57–67.

Rindskopf, D., & Rindskopf, W. (1986). The value of latent class analysis in medical diagnosis. *Statistics in Medicine, 5*, 21–27.

Rossi, P., Berk, R., & Lenihan, K. (1980). *Money, work, and crime: Some experimental results.* New York: Academic Press.

Singer, J., & Willett, J. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics, 18*, 155–195.

Singer, J., & Willett, J. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York: Oxford University Press.

Steele, F. (2003). A multilevel mixture model for event history data with long-term survivors: An application to an analysis for contraceptive sterilisation in Bangladesh. *Lifetime Data Analysis, 9*, 155–174.

Steele, F., Goldstein, H., & Browne, W. (2004). A general multilevel multistate competing risks model for event history data, with an application to a study of contraceptive use dynamics. *Statistical Modelling, 4*, 145–159.

Trussel, J., & Richards, T. (1985). Correcting for unmeasured heterogeneity in hazard modeling using the Heckman-Singer procedure. *Sociological Methodology*, 242–276.

Tuma, N., & Hannan, M. (1984). *Social dynamics: Models and methods.* New York: Academic Press.

Van de Pol, F., & Langeheine, R. (1990). Mixed Markov latent class models. *Sociological Methodology*, 213–247.

Vaupel, J., Manton, K., & Stallard, E. (1979). The impact of hetergeneity in individual frailty on the dynamics of mortality. *Demography, 16*, 439–454.

Vermunt, J. (1997). *Log-linear models for event histories.* Thousand Oaks, CA: Sage Publications, Inc.

Vermunt, J. (2002). A general latent class approach to unobserved heterogeneity in the analysis of event history data. In J. Hagenaars & A. McCutcheon (Eds.), *Applied latent class analysis* (pp. 383–407). Cambridge: Cambridge University Press.

Vermunt, J., Langeheine, R., & Bockenholt, U. (1999). Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics, 24*, 179–207.

APPENDIX

This appendix contains select M*plus* syntax for the recidivism example with the financial aid covariate under the proportional hazard odds assumption. Note that M*plus* treats the last class of a latent class variable as the reference class both when it is the outcome in a multinomial regression and when it is the predictor, reformulated as dummy variables. Classes are also labeled beginning with "1". Thus, in this syntax, to match the model specification given in the text with the pre-event state as the reference category, the measurement model for e1 is reversed so that e1 = 1 is the event state and e1 = 2 is the pre-event state. For e2–e6, ej = 1 is the post-event state, ej = 2 is the event state, and ej = 3 is the pre-event state.

```
VARIABLE:

Usevariables are u1-u6 finaid;
Categorical are u1-u6;
Classes are e1(2) e2(3) e3(3)
            e4(3) e5(3) e6(3);

MODEL:

%overall%
e2#1 on e1#1@40;
e2#2 on e1#1@-10;
e3#1 on e2#1@40 e2#2@40;
e3#2 on e2#1@-10 e2#2@-10;
e4#1 on e3#1@40 e3#2@40;
e4#2 on e3#1@-10 e3#2@-10;
e5#1 on e4#1@40 e4#2@40;
e5#2 on e4#1@-10 e4#2@-10;
e6#1 on e5#1@40 e5#2@40;
e6#2 on e5#1@-10 e5#2@-10;
```

```
[e2#1@-20 e2#2];
[e3#1@-20 e3#2];
[e4#1@-20 e4#2];
[e5#1@-20 e5#2];
[e6#1@-20 e6#2];
e1#1 on finaid (1);

Model e1:
%e1#1% !event
[u1$1@-20];
e2#2 on finaid@0;
%e1#2% !pre-event
[u1$1@20];
e2#2 on finaid (1);

Model e2:
%e2#1% !post-event
[u2$1@20];
e3#2 on finaid@0;
%e2#2% !event
[u2$1@-20];
e3#2 on finaid@0;
%e2#3% !pre-event
[u2$1@20];
e3#2 on finaid (1);

Model e3:
%e3#1% !post-event
[u3$1@20];
e4#2 on finaid@0;
%e3#2% !event
[u3$1@-20];
e4#2 on finaid@0;
%e3#3% !pre-event
[u3$1@20];
e4#2 on finaid (1);

Model e4:
%e4#1% !post-event
[u4$1@20];
e5#2 on finaid@0;
%e4#2% !event
[u4$1@-20];
e5#2 on finaid@0;
%e4#3% !pre-event
[u4$1@20];
e5#2 on finaid (1);
```



```
Model e5:
%e5#1% !post-event
[u5$1@20];
e6#2 on finaid@0;
%e5#2% !event
[u5$1@-20];
e6#2 on finaid@0;
%e5#3% !pre-event
[u5$1@20];
e6#2 on finaid (1);

Model e6:
%e6#1% !post-event
[u6$1@20];
%e6#2% !event
[u6$1@-20];
%e6#3% !pre-event
[u6$1@20];
```

Advances in Latent Variable Mixture Models

Edited by
Gregory R. Hancock
and
Karen M. Samuelsen

INFORMATION AGE PUBLISHING, INC.
Charlotte, NC • www.infoagepub.com

Library of Congress Cataloging-in-Publication Data

Advances in latent variable mixture models / edited by Gregory R. Hancock and Karen M. Samuelsen.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-1-59311-847-1 (pbk.) ISBN-13: 978-1-59311-848-8 (hardcover)
1. Latent structure analysis. 2. Latent variables. 3. Mixture distributions (Probability theory) I. Hancock, Gregory R. II. Samuelsen, Karen M.
QA278.6.A38 2007
519.5'35–dc22 2007037392

Printed in the United States of America

CONTENTS