Discrete-Time Survival Mixture Analysis

Bengt Muthén and Katherine Masyn

JOURNAL OF EDUCATIONAL AND BEHAVIORAL STATISTICS 2005 30: 27
DOI: 10.3102/10769986030001027

Journal of Educational and Behavioral Statistics
Spring 2005, Vol. 30, No. 1, pp. 27–58

Discrete-Time Survival Mixture Analysis

Bengt Muthén
University of California, Los Angeles

Katherine Masyn
Johns Hopkins University; University of California, Los Angeles

This article proposes a general latent variable approach to discrete-time survival analysis of nonrepeatable events such as onset of drug use. It is shown how the survival analysis can be formulated as a generalized latent class analysis of event history indicators. The latent class analysis can use covariates and can be combined with the joint modeling of other outcomes such as repeated measures for a related process. It is shown that conventional discrete-time survival analysis corresponds to a single-class latent class analysis. Multiple-class extensions are proposed, including the special cases of a class of long-term survivors and classes defined by outcomes related to survival. The estimation uses a general latent variable framework, including both categorical and continuous latent variables and incorporated in the Mplus program. Estimation is carried out using maximum likelihood via the EM algorithm. Two examples serve as illustrations. The first example concerns recidivism after incarceration in a randomized field experiment. The second example concerns school removal related to the development of aggressive behavior in the classroom.

Keywords: *event history, growth mixture modeling, latent classes, maximum likelihood*

1. Introduction

This article considers discrete-time survival analysis to study the probability, or hazard, of experiencing a nonrepeatable event, such as onset of drug use. Unlike logistic regression, which examines the overall probability of an event without regard to the timing of that event, discrete-time survival analysis allows for examination of the longitudinal progression of the probability that an event occurs. Alternative names for this type of analysis are event history analysis and time-to-event analysis. For overviews, see, for example, Allison (1984), Singer and Willett (1993), and Vermunt (1997).

The research was supported for the first author under Grant K02 AA 00230-01 from NIAAA. The research was supported for the second author by both NIMH and NIDA under Grant MH40859, and by NIMH under Grants MH01259, MH38725, MH42968, and T32-MH018834. The work has benefitted from discussions in the Hendricks Brown Prevention Science and Methodology Group and Muthén's Research Apprenticeship Course. We thank Klaus Larsen for helpful comments on an earlier version.

Although continuous-time survival analysis (see, e.g., Hougaard, 2000) is frequently used in many settings, discrete-time survival analysis is often more natural in social and behavioral science applications where time is likely to be measured discretely, for example, in school years. Discrete-time models have the strength that they can easily accommodate time-varying covariates. They also do not require a hazard-related proportionality assumption that is commonly used in continuous-time survival analysis, for example, the Cox proportional hazards model. In addition, these models easily allow for unstructured as well as structured estimation of the hazard function at each discrete time point.

The aim of this article is to show that it is useful to view the discrete-time survival analysis as a latent class model that can be incorporated into a general latent variable modeling framework. This general framework enables interesting model extensions. First, unobserved heterogeneity among the subjects in the study can be captured using multiple latent classes of individuals with different survival functions. Second, the survival analysis can be combined with analysis of other related outcomes, such as a growth mixture model for repeated measures.

The article is organized as follows. In section 2, two data sets are introduced and used to illustrate the general analysis goals of discrete-time survival analysis. Section 3 presents key statistical concepts. Section 4 places the modeling in a general latent variable framework. Using the general framework, section 5 develops modeling extensions for situations with mixtures of unobserved subgroups of individuals differing in their survival functions. Section 6 shows illustrations of the methods returning to the two data sets introduced in section 2. Section 7 concludes.

2. Discrete-Time Survival Analysis Goals

Two data sets are used to illustrate the analysis goals: data on recidivism after incarceration and data on school removal among grade school children. Here, survival concerns time to re-arrest and time to first school removal, respectively. Survival analyses of these data are presented in section 6.

2.1. Recidivism Data

This dataset is from a randomized field experiment originally reported by Rossi, Berk, and Lenihan (1980) and has been used extensively by Allison (1984, 1995) as a pedagogical example in a continuous-time survival analysis framework. 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 1 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 are detailed to the week level, that is, 52 observation intervals. However, for the illustrative purposes of this article, the data are recoded into 13 4-week intervals, referred to as "months." Further justification for a discrete-time treatment of these data is given in section 6.1.

The survival analysis of this dataset will investigate whether the intervention has a significant effect on the event probabilities across the observation periods after controlling for the effects of the other measured covariates: age at time of release, race, prior work experience, marital status at time of release, parole status, number of prior arrests, years of schooling, and employment status.

The first section of Table 1 displays the sample means and standard deviations for the three continuous covariates to be considered in the analysis. The second section of Table 1 displays the sample proportions for the binary covariates. All covariates, with the exception of employment status, are time-invariant. Employment status is a time-varying binary covariate that indicates 1 or more weeks of employment during a given month. The last section of Table 1 displays the sample information about the outcome of interest, defined as the month of re-arrest. This part of the table is further explained in section 3.

2.2. School Removal Data

The second data set is from a school-based preventive intervention study carried out by the Baltimore Prevention Research Center under a partnership among The Johns Hopkins University, the Baltimore City Public Schools, and Morgan State University. In this intervention trial, children were followed from first to seventh grade with respect to the course of aggressive behavior (Kellam, Rebok, Ialongo, & Mayer, 1994). Teacher ratings of a child's aggressive behavior were made during fall and spring for the first two grades and every spring in Grades 3–7. The ratings were made using the Teacher's Observation of Classroom Adaptation-Revised (TOCA-R) instrument (Werthamer-Larsson, Kellam, & Wheeler, 1991), using an average of 10 items, each rated on a 6-point scale from "almost never" to "almost always." The Good Behavior Game intervention was delivered at the classroom level using control group classrooms in the same school (*internal* controls) as well as in other schools matched on school characteristics (*external* controls). A total of 11 elementary schools participated in the study. For this article only the control groups' data are used. At the first-grade fall measurement there were 6 internal and 10 external control classrooms, with a total of 404 children.

The survival analysis of these data will investigate the effects of the measured covariates on trends in both aggression and school removal survival. Here, survival concerns not being removed from school. The analyses explore the relationship between the development of aggressive behavior in Grade 1 and Grade 2 and relate that to first school removal in Grades 3–7 using the discrete-time mixture framework.

Table 2 shows the variables to be used in the survival analyses. The first section of the table gives the sample means and standard deviations for the measures of aggression in first and second grade as well as the two continuous covariates: (1) the percentage of students in each subject's first-grade fall class on free or reduced school lunch and (2) the classroom average aggression for each subject's first-grade fall class. The second section of Table 2 displays the sample proportions for the binary covariates. The last section of Table 2 displays the sample information

TABLE 1
Variable Definitions and Sample Means for Recidivism Data (n = 432)

Variable Name	Description	*M*	*SD*
Age	Age (in Years) at Release	24.60	6.11
Priors	Number of Prior Arrests	2.98	2.89
Educ	Years of Schooling	3.48	0.83
		M	
Emp_1	1st-month employment indicator	0.40	
Emp_2	2nd-month employment indicator	0.52	
Emp_3	3rd-month employment indicator	0.53	
Emp_4	4th-month employment indicator	0.54	
Emp_5	5th-month employment indicator	0.55	
Emp_6	6th-month employment indicator	0.55	
Emp_7	7th-month employment indicator	0.57	
Emp_8	8th-month employment indicator	0.56	
Emp_9	9th-month employment indicator	0.55	
Emp_{10}	10th-month employment indicator	0.55	
Emp_{11}	11th-month employment indicator	0.57	
Emp_{12}	12th-month employment indicator	0.56	
Emp_{13}	13th-month employment indicator	0.55	
Finaid	Financial assistance indicator	0.50	
Black	Black racial indicator	0.88	
Workexp	Prior work experience indicator	0.57	
Married	Married at release indicator	0.12	
Paroled	Parole status indicator	0.62	
		Hazard	
u_1	1st-month re-arrest indicator	$^{4}/_{432} = 0.01$	
u_2	2nd-month re-arrest indicator	$^{8}/_{428} = 0.02$	
u_3	3rd-month re-arrest indicator	$^{7}/_{420} = 0.02$	
u_4	4th-month re-arrest indicator	$^{8}/_{413} = 0.02$	
u_5	5th-month re-arrest indicator	$^{13}/_{405} = 0.03$	
u_6	6th-month re-arrest indicator	$^{8}/_{392} = 0.02$	
u_7	7th-month re-arrest indicator	$^{10}/_{384} = 0.03$	
u_8	8th-month re-arrest indicator	$^{5}/_{374} = 0.01$	
u_9	9th-month re-arrest indicator	$^{11}/_{369} = 0.03$	
u_{10}	10th-month re-arrest indicator	$^{11}/_{358} = 0.03$	
u_{11}	11th-month re-arrest indicator	$^{8}/_{347} = 0.02$	
u_{12}	12th-month re-arrest indicator	$^{9}/_{339} = 0.03$	
u_{13}	13th-month re-arrest indicator	$^{12}/_{330} = 0.03$	

TABLE 2
Variable Definitions and Sample Means for School Removal Data (n = 404)

Variable Name	Description	*M*	*SD*
y_{1F}	First-grade fall TOCA-R measure	1.92	0.94
y_{1S}	First-grade spring TOCA-R measure	2.01	0.91
y_{2F}	Second-grade fall TOCA-R measure	1.81	0.88
y_{2S}	Second-grade spring TOCA-R measure	2.03	0.99
Cavlunch	First-grade fall classroom average lunch	0.45	0.36
Cavtoca1f	First-grade fall classroom average TOCA-R	1.92	0.40
		M	
External	External control group indicator	0.63	
Male	Male gender indicator	0.50	
White	White racial indicator	0.32	
Lunch	Subsidized school lunch indicator	0.46	
		Hazard	
u_3	Third-grade school removal indicator	$^{8}/_{394} = 0.02$	
u_4	Fourth-grade school removal indicator	$^{9}/_{386} = 0.02$	
u_5	Fifth-grade school removal indicator	$^{15}/_{377} = 0.04$	
u_6	Sixth-grade school removal indicator	$^{23}/_{362} = 0.06$	
u_7	Seventh-grade school removal indicator	$^{59}/_{339} = 0.17$	

about the outcome of interest, defined as the grade of school removal. This part of the table is further explained in the next section.

3. Discrete-Time Survival Analysis Methodology

This section introduces the basic statistical components of discrete-time survival analysis. The hazard and survival functions as well as the probability of observing the sample (the likelihood) are presented. Estimations and plots of the marginal hazard and survival probabilities are given for both data examples.

3.1. Event Time and the Likelihood

The most common representation of the event time distribution is the hazard function. Define T as a discrete random variable that indicates the time period when the event occurs. In discrete time, the hazard function is defined as

$$h_j = P(T = j \mid T \geq j). \tag{1}$$

Essentially, h_j is the probability of experiencing the event in time period j given that it was not experienced before j.

Another distributional representation of event time is the survival function. The survival probability at time period j is defined as the probability of not experiencing

the event (i.e., the probability of "surviving," through time period j). In discrete time, the survival function is defined as

$$S_j = P(T > j). \tag{2}$$

The survival probability can be expressed in terms of the hazard by

$$\begin{aligned} S_j = P(T > j) &= P(T \neq j \mid T \geq j)P(T \neq j-1 \mid T \geq j-1)\cdots \\ &\quad P(T_i \neq 2 \mid T_i \geq 2)P(T_i \neq 1 \mid T_i \geq 1) \\ &= \prod_{k=1}^{j}(1-h_k). \end{aligned} \tag{3}$$

Suppose the duration of the study is made up of J time periods. A single nonrepeatable event is considered so that data collection (and the observation of risk) is discontinued for individual i in time period j_i for one of three reasons: (1) The individual experiences the event in j_i; (2) the individual drops out of the study in j_i; or (3) the study concludes. In the first case, $T_i = j_i$. In the second case, it is only known that $T_i > (j_i - 1)$ because the individual dropped out *during* the period j_i, it is not known that $T_i > j_i$. And in the third cases, it is only known that $T_i > J$. Individuals with $T_i > (j_i - 1)$ and $T_i > J$ are *right-censored:* it is unknown whether they experience the event after their observation period. For uncensored individuals with $T_i = j_i$, the likelihood may be expressed in terms of the hazard as

$$\begin{aligned} P(T_i = j_i) = P(T = j_i \mid T \geq j_i)&P(T_i \neq j_i - 1 \mid T_i \geq j_i - 1)\cdots \\ &P(T_i \neq 2 \mid T_i \geq 2)P(T_i \neq 1 \mid T_i \geq 1) \\ &= h_{ij_i}\prod_{k=1}^{j_i-1}(1-h_{ik}). \end{aligned} \tag{4}$$

For individuals with $T_i > (j_i - 1)$, the likelihood may be expressed as

$$P(T_i > j_i - 1) = \prod_{k=1}^{j_i-1}(1-h_{ik}). \tag{5}$$

For individuals with $T_i > J$, setting $j_i = J + 1$ allows the likelihood to be expressed as

$$P(T_i > J) = P(T_i > j_i - 1) = \prod_{k=1}^{j_i-1}(1-h_{ik}), \tag{6}$$

the same as the likelihood for individuals censored before the conclusion of the study.

It follows that the likelihood for the full sample is $L = \Pi^n_{i=1} l_i$, where

$$l_i = (h_{ij_i})^{\delta_i} \prod_{j=1}^{j_i-1} (1 - h_{ij}) \tag{7}$$

where δ_i is a 0/1 indicator with $\delta_i = 1$ if $T_i = j_i$, that is, if individual i experiences the event while under observation.

3.2. Estimating Hazard Probabilities

The sample estimates of the hazard probabilities, also known as the marginal hazard probability estimates, are easily calculated from the event history data. The sample-estimated hazard probability for time period j is simply the number of events that are observed to occur in time period j divided by the total number of subjects at risk in time period j. "Subjects at risk" in this context refers to all those subjects who have not yet experienced the event before period j and are still under observation *in* period j; that is, are not censored during period j. Take, for example, the recidivism example. The hazard probability for a given month would be the probability of being arrested in that month among those inmates who have not been re-arrested before that month. For example, in the 1st month, all 432 released inmates were at risk for re-arrest and four experienced re-arrest. The estimated hazard probability for the 1st month is 4/432 = 0.01. For the 2nd month, only 432 – 4 = 428 inmates were at risk for re-arrest and eight were arrested. The estimated hazard probability for the 2nd month is then 8/428 = 0.02. The 13 sample hazard probabilities for the months of observation are given in the last portion of Table 1. The sample hazard probabilities can be plotted by month as shown in Figure 1. This representation of the sample hazard function suggests that the marginal hazard function may be constant or slightly increasing with mainly random sampling accounting for the fluctuation in the range 0.01–0.03. The proportions of

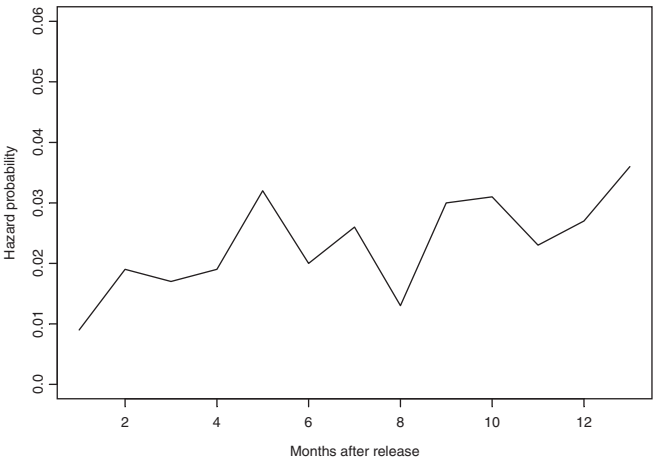

FIGURE 1. *Sample-estimated hazard probabilities of re-arrest.*

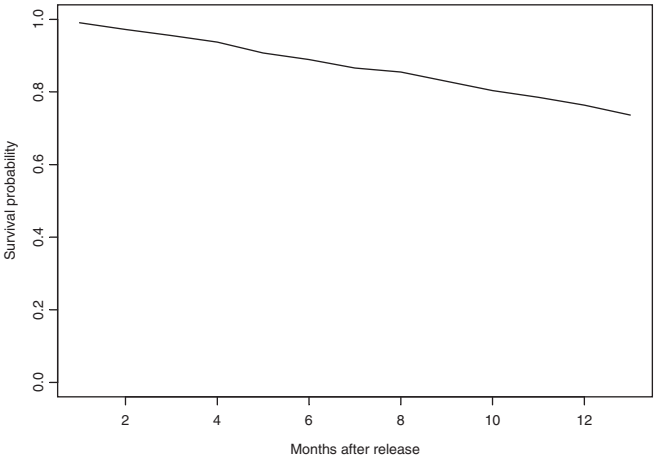

FIGURE 2. *Sample-estimated survival probabilities of re-arrest.*

the initial population of inmates surviving through each month (i.e., the survival probabilities) can be estimated directly from the estimated hazard probabilities using the relationship defined in Equation 3. Figure 2 displays the plot of the estimated survival probabilities by month. There is an increase in the proportion of the total inmates re-arrested over time with almost 30% re-arrested by the end of the 13th month.

Figures 3 and 4 display the same sample-based estimates of the hazard and survival probabilities, stratified by intervention status. There is no clear differ-

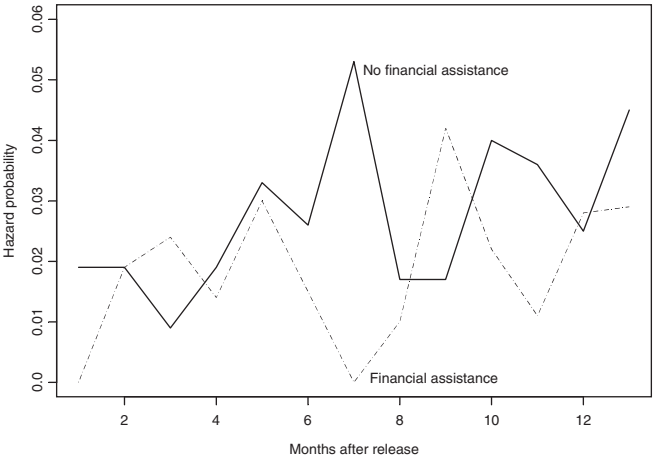

FIGURE 3. *Sample-estimated hazard probabilities of re-arrest by intervention status.*

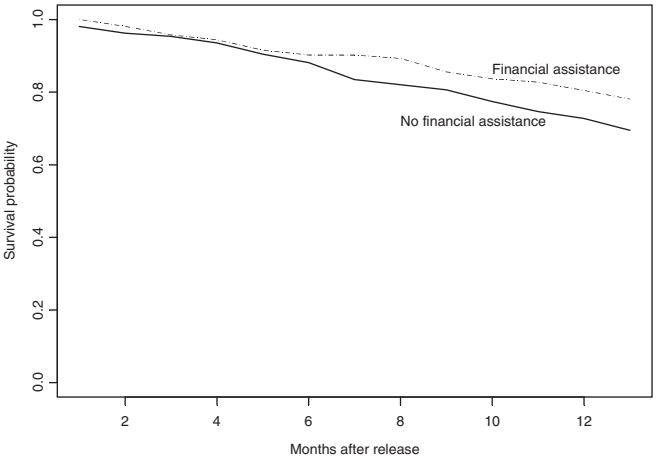

FIGURE 4. *Sample-estimated survival probabilities of re-arrest by intervention status.*

ence in the hazard functions for the two groups, but the group of inmates not receiving the financial aid intervention does have a slightly lower mean survival curve, with almost 10% more of the initial control group re-arrested by the end of the 13th month.

The last portion of Table 2 gives the estimated marginal hazard probabilities of school removal in Grades 3–7 for the second data example. Figure 5 shows sample means for aggression in Grades 1 and 2 and the sample survival curve. The survival curve indicates that by end of Grade 7, about 75% of the children have not experienced school removal. Figure 6 shows the corresponding picture when dividing the sample into high and low-to-average aggression groups based on the upper quartile of the aggression distribution in the fall of first grade. The figure clearly indicates a relationship between aggressive behavior and school removal. The children with higher aggression scores are seen to have a considerably lower mean survival curve, with almost half the children having experienced school removal by the end of Grade 7.

Beyond the marginal hazard estimates, it may also be of interest to investigate the relationship between the hazard probabilities and a set of observed covariates. In line with Singer and Willet (1993), a logistic hazard function is considered, although the use of other link functions, such as the complementary log–log, can also be found in the discrete-time survival literature (e.g., Hedeker, Siddiqui, & Hu, 2000). Let $\mathbf{z}_{ij}$ be a $p \times 1$ vector of values for the set of covariates, $(z_1, \ldots, z_p)$, in time period j for individual i. $\mathbf{z}_j$ represents the set of time-varying (i.e., time-dependent) covariates. Let $\mathbf{x}_i$ be a $q \times 1$ vector of values for the set of

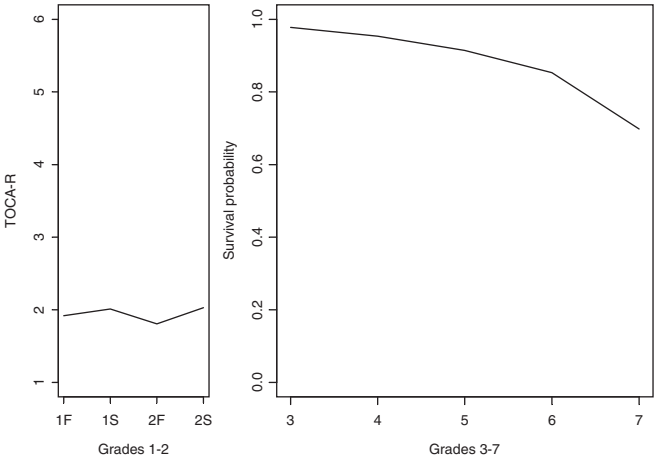

FIGURE 5. *Sample-estimated mean aggression trajectory and survival of school removal.*

time-invariant covariates for individual *i*. Notice that **x** is not indexed by *j* because the values for the *x* covariates are independent of time. Also note that both continuous and categorical covariates can be accommodated which is not the case working in the traditional log-linear framework that only allows categorical variables. The hazard can be related to the covariates using the logistic function as shown below.

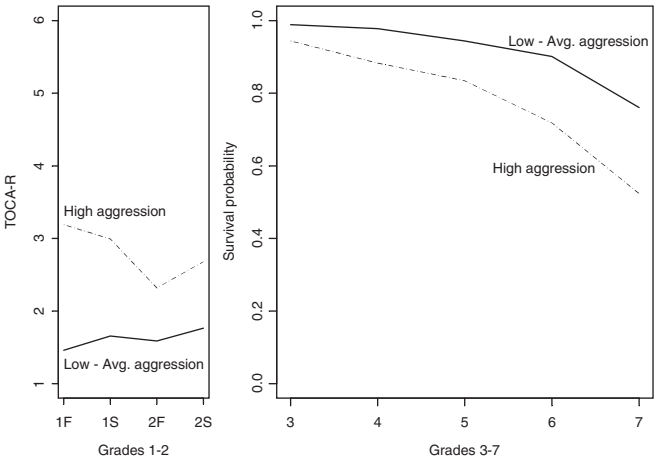

FIGURE 6. *Sample-estimated mean aggression trajectories and survival of school removal by first grade fall baseline aggression measure.*

$$h_{ij} = \frac{1}{1 + e^{-(\text{logit}_{ij})}} \tag{8}$$

where the logit_{ij} is expressed as

$$\text{logit}_{ij} = \beta_j + \kappa'_{zj}\,\mathbf{z}_{ij} + \kappa'_{xj}\,\mathbf{x}_i. \tag{9}$$

By dropping the j subscript from κ_{xj} and κ_{zj} in Equation 9, the effects of the covariates are constrained to be equal across all time periods. This is referred to as the proportional hazard odds model because the hazard odds ratio for the event corresponding to each covariate is constant across the time periods. The inverse logit of β_j is the hazard probability in time period j when $\mathbf{z}_j = \mathbf{0}$ and $\mathbf{x} = \mathbf{0}$. This is known as the *baseline hazard.* A constant baseline hazard function can be imposed by dropping the j subscript from the intercept, β_j.

4. Discrete-Time Survival in a General Latent Variable Framework

Muthén and Shedden (1999) and Muthén and Muthén (2001, Appendix 8) consider a general latent variable modeling framework involving both categorical and continuous latent variables. Estimation is carried out using the EM algorithm to obtain maximum-likelihood estimates. The procedure is incorporated in the Mplus program (Muthén & Muthén, 1998–2004). The model is given in the appendix and relevant parts of it are summarized here, followed by an explanation of how the discrete-time survival model fits into the framework.

The general model can be characterized as a finite mixture model. Mixture modeling allows for unobserved heterogeneity in the sample, where different individuals can belong to different subpopulations without the subpopulation membership being observed but instead inferred from the data. Another way to conceptualize mixture modeling is as a nonparametric approach to estimating an underlying continuous *frailty* distribution; that is, a random effect for the outcome that represents differences in individuals with respect to the survival process. Mixture modeling captures this heterogeneity by a latent categorical variable. Modeling with a continuous frailty in this framework is typically approached by assuming a Normal distribution for the random effect and then using numerical quadrature to integrate over the unobserved frailty where the points and weights for the numerical integration are fixed according to a Normal distribution. Using mixture modeling, in contrast, allows estimation of the points and weights (corresponding to class factor means and class proportions), resulting in a nonparametric estimation of the frailty, free of the normality assumption. Mixture modeling has a wide variety of applications. Overviews with latent class and growth mixture applications are given in Muthén (2001a, 2001b) and Muthén and Muthén (2001). Applications to randomized trials are given in Muthén et al. (2002), Jo (2002), and Jo and Muthén (2000, 2001).

Let c denote a latent categorical variable with K classes, $c_i \in \{1, \ldots, K\}$, where $c_i = k$ if individual i belongs to class k. The model relates c to a $p \times 1$ covariate vector $\mathbf{x}$

by multinomial logistic regression using the $(K-1)$-dimensional parameter vector of logit intercepts, α_c, and the $(K-1)\times p$ parameter matrix of logit slopes, Γ_c, where for $k = 1, 2, \ldots, K$,

$$P(c_i = k \mid \mathrm{x}_i) = \frac{e^{\alpha_{ck} + \gamma'_{ck}\mathbf{x}_i}}{\sum_{s=1}^{K} e^{\alpha_{cs} + \gamma'_{cs}\mathbf{x}_i}}, \tag{10}$$

where the last class is a reference class with coefficients standardized to zero, $\alpha_{c_K} = 0$, $\gamma_{c_K} = \mathbf{0}$.

Define an $r \times 1$ vector $\mathbf{u}$ of binary 0/1 variables with conditional independence given c_i and $\mathbf{x}_i$. That is,

$$P(u_{i1}, u_{i2}, \ldots, u_{ir} \mid c_i, \mathbf{x}_i) = P(u_{i1} \mid c_i, \mathbf{x}_i)\, P(u_{i2} \mid c_i, \mathbf{x}_i) \ldots P(u_{ir} \mid c_i, \mathbf{x}_i). \tag{11}$$

Define $\mathbf{u}_i^* = (u_{i1}^*, u_{i2}^*, \ldots, u_{ir}^*)'$ as continuous latent response propensities underlying $\mathbf{u}$. Here, u_j^* is related to u_j through a threshold parameter τ_j,

$$P(u_{ij} = 1 \mid c_i, \mathrm{x}_i) = \frac{1}{1 + e^{-(-\tau_j + u_{ij}^*)}}, \tag{12}$$

For example, the higher the τ, the higher u^* needs to be to exceed it, and the lower the probability of $u = 1$ (the use of a threshold parameter instead of an intercept parameter is needed when ordered polytomous u's are considered in this framework).

It is convenient to introduce a continuous latent variable vector $\eta_{ui} = (\eta_{u_{1i}}, \eta_{u_{2i}}, \ldots, \eta_{u_{fi}})'$. Conditional on class k,

$$\mathbf{u}_i^* = \Lambda_{u_k}\, \eta_{ui} + \mathbf{K}_{u_k}\, \mathbf{x}_i \tag{13}$$

$$\eta_{ui} = \alpha_{u_k} + \Gamma_{u_k}\, \mathbf{x}_i, \tag{14}$$

where Λ_{u_k} is an $r \times f$ logit parameter matrix varying across the K classes, $\mathbf{K}_{u_k}$ is an $r \times q$ logit parameter matrices varying across the K classes, α_{u_k} is an $f \times 1$ logit parameter vector varying across the K classes, and Γ_{u_k} is an $f \times q$ logit parameter matrix varying across the K classes. The model structure in Equations 13 and 14 is useful when the $\mathbf{u}$ vector represents repeated measures, and the latent classes correspond to different trajectory classes.

4.1. Fitting Discrete-Time Survival Into the General Framework

Section 3.1 described and expressed the likelihood in terms of a final time period of observation for each individual, denoted j_i, and an indicator, δ_i, for whether an event occurred for individual i during the observation. To carry out

discrete-time survival analysis in this general framework, the data on each individual are put in terms of a set of binary 0/1 event history indicators u_j, $j = 1, 2, \ldots, J$, where $u_{ij} = 0$ if individual i is observed to be at risk for the event of interest for the whole of time period j but does not experience the event and $u_{ij} = 1$ if individual i experiences an event in time period j, where J is the last time period of data collection for the study. Consider again the three reasons given for observation of an individual to terminate: (1) individual i experiences the event in time period $j_i < J$; (2) individual i is lost to follow-up during time period $j_i < J$; or (3) individual i does not experience the event of interest, and the study concludes. In cases (1) and (2), individual i has missing values for u_{ij}, $j > j_i$, and u_{ij}, $j \geq j_i$, respectively. These three cases correspond to three patterns of u observations: (1) $u = 0$ for all time periods before the event occurs, $u = 1$ for the period of the event, and u missing for all subsequent periods; (2) $u = 0$ for all time periods before loss to follow-up and u missing for all subsequent periods; and (3) $u = 0$ for all time periods of observation.

As an example consider five time periods. An individual who is censored *after* time period five ($j_i = J + 1 = 6$) has the event history

$$(0 \quad 0 \quad 0 \quad 0 \quad 0),$$

an individual who experiences the event in period four ($j_i = 4$) has the event history

$$(0 \quad 0 \quad 0 \quad 1 \quad 999),$$

and an individual who drops out *after* period three, that is, is censored sometime *during* period four ($j_i = 4$), has the event history

$$(0 \quad 0 \quad 0 \quad 999 \quad 999),$$

where unobserved u information is represented as $u = 999$ to denote missing data. Thus, the complete event history information on individual i may be entered into a $J \times 1$ data vector $\mathbf{u}_i$.

It is assumed that the missing data in the last example is ignorable in the sense that the reason for the individual dropping out after period three is unrelated to individual's event status following drop out. The conventional assumption of noninformative censoring, that is, that censoring times are independent of event times conditional on the observed covariates, corresponds to the assumption of ignorable missingness in the general latent variable model. Under the assumption of MAR (Little & Rubin, 2002), the observed data likelihood for uncensored individuals, in terms of the event indicators, is

$$P(T_i = j_i) = P(u_{ij_i} = 1) \prod_{k=1}^{j_i - 1} P(u_{ik} = 0). \tag{15}$$

For censored individuals, the observed data likelihood may be expressed as

$$P(T_i > j_i - 1) = \prod_{k=1}^{j_i-1} P(u_{ik} = 0). \tag{16}$$

It follows that the observed data likelihood for the full sample is $L = \Pi_{i=1}^{n} l_i$, where

$$l_i = \prod_{k:\, u_{ik} \neq 999} P(u_{ik}) = P(u_{ij_i} = 1)^{\delta_i} \prod_{k=1}^{j_i-1} [1 - P(u_{ik} = 1)], \tag{17}$$

where δ_i is defined as before, and number of classes is specified as singular, that is, $K = 1$.

It is immediately evident that the maximum likelihood estimates for the event indicator probabilities under the assumption of MAR based on the observed data likelihood in Equation 17 are the same as the maximum likelihood estimates for the hazard probabilities based on the likelihood function given in Equation 7. That is,

$$\hat{h}_{ij} = \hat{P}(u_{ij}). \tag{18}$$

Note that the sample means based on the observed data for the event indicators are equal to the estimated hazard probabilities as given in Tables 1 and 2.

Furthermore, expressing the hazard probabilities as a function of the observed covariates using the logit link function is equivalent to the logistic regression of $\mathbf{u}_i$ on the observed covariates. Equation 9 can be rewritten as a special case of the more general latent variable model given in Equations 12–14 with $K = 1$ and $-\tau_j = \beta_j$.

$$P(u_{ij} = 1 \mid \mathbf{x}_i, \mathbf{z}_{ij}) = \frac{1}{1 + e^{-(\text{logit}_{ij})}}, \tag{19}$$

where

$$\text{logit}_{ij} = \beta_j + \kappa'_{zj}\, \mathbf{z}_{ij} + \kappa'_{xj}\, \mathbf{x}_i + \lambda'_{uj}\, \eta_{ui}, \tag{20}$$

$$\eta_{ui} = \alpha_u + \gamma'_u\, \mathbf{x}_i, \tag{21}$$

where β_j is the time-specific logit intercept parameter that may also be interpreted as the logit baseline hazard, κ_{zj} is a $p \times 1$ logit parameter vector that may vary across the J time periods, κ_{xj} is a $q \times 1$ logit parameter vector that may vary across the J time periods, λ_u is an $f \times 1$ logit parameter vector that may vary across the J time periods, α_u is an $f \times 1$ time-invariant logit parameter logit, and γ_u is a $q \times 1$ time-invariant logit parameter vector.

The notational distinction between time-invariant covariates (x) and time-varying covariates (z) is used to make explicit that η may only be a function of covariates that are invariant across the J time periods. Notice that the intercept notation rather than the threshold notation is used here, but it is simple to move between the two specification recalling that $\beta_j = -\tau_j$. Using Λ, functional forms for the logit of the baseline hazard probabilities can be specified; for example, linear trends such as $\Lambda_u = (0, 1, \ldots, J)'$, where J is the number of time periods, and $\beta_j = \beta$ for all $j = 1, \ldots, J$, so that β is the intercept of the linear trend in the logit baseline hazard probabilities, and α_{u1} is the mean slope. Λ_u could also be used to specify a piecewise baseline hazard function; for example, $\Lambda_u = (0, 0, 0, 1, \ldots, 1)'$ with $\beta_j = \beta$ for all $j = 1, \ldots, J$, so that β is the logit of the baseline hazard for time periods $j \leq 3$, and $\beta + \alpha_{u1}$ is the logit of the baseline hazard for time periods $j > 2$. As stated before, a constant baseline hazard probability model can be obtained by setting $\beta_j = \beta$ for all $j = 1, \ldots, J$, and $\alpha = 0$. Also, the proportional hazard odds model can be conveniently obtained by simply dropping the j subscript from κ_{xj} and κ_{zj}. Note that all effects of the x covariates on the hazard probabilities that go through η_u are automatically time-invariant because η_u does not vary across the J time periods.

5. Mixture Analysis

It is often important to take into account unobserved heterogeneity in survival among the subjects studied. Unobserved heterogeneity in the form of unobserved covariates or even random error can result in biased estimates of main effect parameters as well as standard errors if not explicitly taken into account (Vaupel, Manton, & Stallard, 1979). In continuous-time survival modeling it is common to model unobserved heterogeneity using frailties, that is, representing heterogeneity by random effects (continuous latent variables); see, for example, Hougaard (2000). This article takes heterogeneity into account using latent classes of individuals. A general discrete-time survival mixture model is introduced, where different latent classes have different hazard and survival functions. Three different types of survival mixture models will be considered, a generic multiple-class model, a "long-term survival" model with two classes, and a multiple-class model combining the survival model with a growth mixture model. Other examples of this approach for special cases can be found in the discrete-time survival literature; for example, log-linear latent class models used by Vermunt (1997), long-term survivor models applied by Steele (2000), and semiparametric mixed Poisson regression models by Land, Nagin, and McCall (2001).

Consider the multiple-class modification of Equations 19–21 for class k ($k = 1, 2, \ldots, K$),

$$h_{ijk} = P(u_{ij} = 1 \mid c_i = k, \mathbf{x}_i, \mathbf{z}_{ij}) = \frac{1}{1 + e^{-(\text{logit}_{ijk})}}, \tag{22}$$

where

$$\text{logit}_{ijk} = \beta_{jk} + \kappa'_{zjk}\,\mathbf{z}_{ij} + \kappa'_{xjk}\,\mathbf{x}_{i} + \lambda'_{ujk}\,\eta_{ui}, \tag{23}$$

$$\eta_{ui} = \alpha_{uk} + \gamma'_{uk}\,\mathbf{x}_{i}, \tag{24}$$

With multiple classes, the model adds the prediction of class membership by covariates **x** as in (10),

$$P(c_i = k \mid \mathbf{x}_i) = \frac{e^{\alpha_{ck} + \gamma'_{ck}\mathbf{x}_i}}{\sum_{s=1}^{K} e^{\alpha_{cs} + \gamma'_{cs}\mathbf{x}_i}}. \tag{25}$$

The inclusion of multiple classes modifies the likelihood expression in Equation 7, $L = \Pi_{i=1}^{n}\, l_i$, as

$$l_i = \sum_{s=1}^{K}\left[\pi_{ik}\,(h_{ijk})^{\delta_i}\prod_{j=1}^{j_i-1}(1 - h_{ijk})\right], \tag{26}$$

where $\pi_{ik} = P(c_i = k \mid \mathbf{x}_i)$. For multiple-class models, identification of model parameters needs to be carefully considered. The multiple-class model is a special case of latent class analysis with covariates. A recent treatment of identification issues for latent class modeling with covariates is given in Huang and Bandeen-Roche (in press).

A caution should be issued here also about the susceptibility of these multiple-class models to convergence at locally rather than globally optimal solutions, as is true for mixture models in general. Multiple sets of starting values should be used, and the convergence pattern for the likelihood through the iterations of the EM algorithm should be carefully monitored.

As shown earlier, from a latent class point of view, the discrete-time survival model presented in Section 4 can be viewed as a single-class model. When covariates are not present, the discrete-time survival model with unstructured hazard probabilities has the special feature of perfectly fitting the data on the *u*s. A Pearson or likelihood-ratio chi-square statistic has zero value irrespective of the data. As a note, the degrees of freedom computed assuming a unrestricted multinomial model as the alternative is not correct for these models because it does not take into account what might be referred to a "structural zeros" in a categorical data analysis setting (i.e., not all response patterns for the binary *u*s are allowed in the discrete-time setting; e.g., a zero cannot follow a one).

Adding covariate information or restricting the form of the hazard function makes it possible to fit a multiple-class model. This is discussed further in the next section in the context of the special two-class model including a class referred to as long-term survivors. The need to use covariate information to identify unobserved het-

erogeneity is analogous to the need for covariates to identify frailties in continuous-time survival analysis (see, e.g. Nielsen, Gill, Andersen, & Soerensen, 1992)

5.1. Long-Term Survivors

As reported by McLachlan and Peel (2000), the notion of long-term survivors has been used in continuous-time survival modeling at least since Boag (1949); for an overview, see Maller and Zhou (1996). A typical application concerns women treated for breast cancer, ultimately dying of causes other than cancer. For a recent application in the context of discrete-time survival modeling of contraceptive sterilisation, see Steele (2000). Long-term survival means that there is a latent class of individuals whose risk is essentially zero across all time period, that is, individuals who have a zero hazard probability throughout the observation period. Using the u notation of section 4.1, an individual who experiences the event ($u = 1$ observation at any time period) is known to *not* be a member of the long-term survivor class, while individuals who are censored may or may not be members of the long-term survivor class. In this way, the latent class variable c of section 4 is observed in part of the sample. In the general modeling framework, this is handled using the training data feature presented in the appendix. Individuals who experience the event are only allowed to be in the class of non-long-term survivors, while censored individuals have unknown class membership and are classified in the analysis. The model also incorporates a prediction of class membership by covariates.

Because the survival probability is one for the long-term survival class, the survival function for the mixture model may be written as

$$S_{ij} = \pi\, S_{ij}^{NLTS} + (1 - \pi); \tag{27}$$

where S^{NLTS} is the survival function for the non-long-term survivors, and $1 - \pi$ is the probability of being a member of the long-term survivor class. The long-term survivor model fits into the general framework by noting that the zero hazards for the longterm survival class are obtained by setting $\beta_{j,LTS} = -\infty$ (or, equivalently, $\tau_j = \infty$), $\kappa_{xj,LTS} = \mathbf{0}$, $\kappa_{zj,LTS} = \mathbf{0}$, and $\lambda_{u,LTS} = \mathbf{0}$ for all js in Equation 24. The model is completed by the logistic regression for class membership,

$$\log[\pi/(1 - \pi)] = \gamma_c' \mathbf{x}_i, \tag{28}$$

which is a special case of Equation 10.

It may be noted that the long-term discrete-time survival model is not identified unless covariates are present. This is in line with the earlier observation that the single-class discrete-time survival model fits the data on u perfectly, so that more than one class cannot be extracted. Intuitively, there is no information from which to distinguish long-term survivors from other individuals who are censored. With covariate information, however, a distinction between long-term survivors versus those who are at risk for ultimately experiencing the event can be made based on the difference versus similarity in covariate values relative to those who experienced

the event. The covariates may influence the latent class membership probability $\pi_{ik} = P(c_i = k \mid \mathbf{x}_i)$. The covariates may also influence the event history indicator probabilities, h_{ijk}, either directly or via the factor η_{ui}.

It is important to recognize a potential weakness of the long-term discrete-time survival model. Because this model needs covariates to be identified, different sets of covariates may produce nontrivial differences in the latent class formation. In contrast to this situation, the next section presents a model where the latent classes are defined by information that is separate from the event history.

5.2. Combined Discrete-Time Survival and Growth Mixture Modeling

Discrete-time survival analysis can be combined with a growth mixture model. For continuous-time survival analysis, related developments for single-class models include Henderson, Diggle, and Dobson (2000). In the model studied here, the latent classes are defined by the growth mixture model in terms of different developmental trajectory classes and serve as latent categorical predictors in the survival part. Drawing on the general modeling framework of the appendix, this means that the survival model for **u** is analyzed jointly with the growth mixture model for **y**. Maximum-likelihood estimation is also used in this case.

Consider as an example repeated measures on continuous outcomes y_{it} ($i = 1, 2, \ldots, n$) that can be described by only two random effects (growth factors) η_{0i} and η_{1i} and a time-specific residual ϵ,

$$y_{it} = \eta_{0i} + \eta_{1i}\, a_{it} + \epsilon_{it}. \tag{29}$$

Different trajectory classes are allowed for by letting the means, variances, and covariance of η_0 and η_1 vary across the classes. The variances of ϵ_t may also vary across classes. The covariates of **x** may influence class membership as in Equation 10. They may also have class-varying influence on the growth factors ($k = 1, 2, \ldots, K$),

$$\eta_{0i} = \alpha_{0k} + \gamma'_{0k}\, \mathbf{x}_i + \zeta_{0i}, \tag{30}$$

$$\eta_{1i} = \alpha_{1k} + \gamma'_{1k}\, \mathbf{x}_i + \zeta_{1i}, \tag{31}$$

The latent class variable is related to covariates **x** as in the general framework of section 4,

$$P(c_i = k \mid \mathrm{x}_i) = \frac{e^{\alpha_{ck} + \gamma'_{ck}\mathbf{x}_i}}{\sum_{s=1}^{K} e^{\alpha_{cs} + \gamma'_{cs}\mathbf{x}_i}}, \tag{32}$$

The model given in Equations 29–32 is referred to as growth mixture modeling and was introduced in Muthén and Shedden (1999); for overviews, see, e.g., Muthén (2001a, 2001b).

The new feature of the modeling is that the latent class variable for the growth mixture part of the model can be specified to influence the survival part of the model. For example, if the latent growth class variable had a proportional effect on the hazard odds, then the logit of h_{ijk} would be given as

$$\text{logit}_{ijk} = \beta_j + \kappa'_{zjk}\,\mathbf{z}_{ij} + \kappa'_{xjk}\,\mathbf{x}_i + \lambda'_{ujk}\,\eta_{ui}, \tag{33}$$

$$\eta_{ui} = \alpha_{uk} + \gamma'_{uk}\,\mathbf{x}_i, \tag{34}$$

where the the intercept parameters, β_js, are held invariant across the latent classes and $\alpha_{uK} = \mathbf{0}$ (for a reference class), so that the latent class membership for the growth model influences the hazard function through the class-varying α and the class-varying γ influence from $\mathbf{x}$.

More complex models may also be fitted in the general modeling framework. Without covariates, it may be noted that the added *y* information makes it possible to identify more than one class for the *u* variables. That is, even when the distribution of the *u*s does not require more than one class, more than one class can be specified for the *u*s. When covariates are present, the growth mixture model may be combined with a multiple-class discrete-time survival model. This means that two different latent class variables are needed. Modeling with several latent class variables using the general framework was described in Muthén (2001b).

6. Examples

This section illustrates the methodology using the recidivism and school removal examples presented in section 2. The recidivism example is used to examine a single-class survival model with time-invariant and time-varying covariates. The school removal example is used to illustrate the combined analysis of a growth mixture model and a survival model. All analyses are carried out using the Mplus program. Input for the analyses are found at www.statmodel.com.

6.1. Recidivism Analyses

The recidivism data were described in section 2.1. The primary interest for this analysis is to assess accurately the effects of the financial assistance intervention while accounting for the other covariates related to re-arrest. In his series of continuous-time analyses of this data, Allison (1984, 1995) found consistently significant effects for age at release and number of prior arrests, with nonsignificant or borderline significant intervention effects. For example, when applying the exponential regression model, the estimated hazard for re-arrest of those in the financial assistance group was approximately 72% of that for individuals in the control group who received no aid, with a two-tailed *t* test *p* value of approximately .09 (Allison, 1984). In the discrete-time analyses to be presented, the effects of the aid intervention on the hazard for re-arrest are also examined. Instead of the 52 week-long intervals treated as continuous-time observations, the outcomes in this article have been grouped into 13 4-week intervals to be modeled as discrete-time

observations. In the original study, inmates were assessed on a monthly basis, so this treatment of the data may have greater reliability with regard to time-varying covariate effects. In addition, the discrete-time framework allows testing of the modeling assumptions such as constancy of the baseline hazard function and proportionality of covariate effects that could only be informally evaluated in the continuous-time setting using sensitivity analysis. It is also possible to expand the evaluation of the intervention to allow for its influence on latent survival class membership.

A first analysis step separately evaluates the proportionality assumption for each of the covariates. The fit of the model using the hazard logit defined in Equation 21, which allows for time-specific covariate effects, is compared to the model that constrains the covariate effects to be equal across time using the factor η_u. The second model is the proportional hazard odds model. The models with and without the proportionality assumption are shown in diagrammatic form in Figures 7 and 8, respectively. Considering intervention status as a time-invariant covariate, the chi-square difference for these two models is 12.2 with 12 degrees of freedom, suggesting that there is little evidence in the data to reject the proportional hazard odds assumption. Looking at each covariate in turn, no evidence was found to reject the proportionality assumption for any of the covariates, including the time-varying employment status.

As the next step, a model with all the covariates is constructed, allowing for relaxation of the proportionality assumption when called for by the first step in the analysis. This model may then be used to evaluate the functional form of the baseline hazard. In the preceding analysis step, the hazard is completely unstructured. A specific structure may now be imposed on the logit baseline hazard, such as constancy or linear trend, and model fit compared to the unstructured case. A model with constancy of the hazard may be defined as in Equation 21, removing the subscript *j* from the intercept β and setting $\alpha_u = \mathbf{0}$. Considering the constant hazard model, the chi-square difference, compared to the unstructured hazard model, is 8.8 with 12 degrees of freedom, suggesting that there is little evidence in the data to reject the constant baseline hazard assumption. Table 3 shows the results from the model with the proportionality and constant hazard assumptions applied. These results are consistent

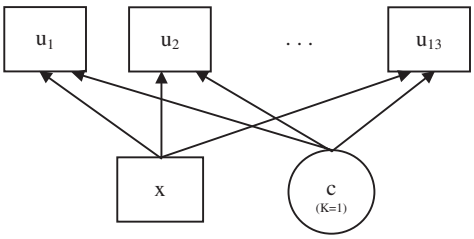

FIGURE 7. *Recidivism path diagram: Survival model with time-varying covariate effects.*

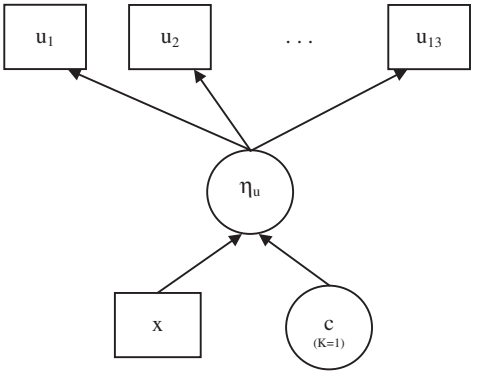

FIGURE 8. *Recidivism path diagram: Survival model with proportional hazard odds assumption applied to time-invariant covariates.*

with the previous continuous-time analyses (Allison, 1984). Figure 9 shows the model-estimated mean survival plots for both the one-class model for the two intervention groups at the overall sample mean values for the other covariates.

Figure 10 displays the diagram for a multiple-class model with covariates. However, for these particular data with the set of covariates used here and the relatively low base rate of re-arrest, there is not sufficient information to make a convincing inference about multiple latent classes of survival, and results of such an analysis are not presented.

TABLE 3
1-Class Survival Model with Constant Hazard and Proportional Odds Assumptions

Thresholds (τ)	*Est.*	*SE*	*t*
u_1–u_{13}	1.80	0.82	–2.20
Latent Class Growth Factor (η_u)	*Est.*	*SE*	*t*
Finaid	–0.33	0.19	–1.72
Black	0.37	0.29	1.27
Workexp	0.01	0.21	0.05
Married	–0.29	0.39	–0.75
Paroled	–0.07	0.20	–0.36
Age	**−0.05**	0.02	–2.07
Priors	**0.07**	0.03	2.55
Educ	–0.21	0.13	–1.67
Event Indicator Regression	*Est.*	*SE*	*t*
Emp_1–Emp_{13}	**−1.04**	0.21	–4.90

Log likelihood = –514.18, BIC = 1089.04, 10 free parameters. Boldface entries indicate significance at the 5% level.

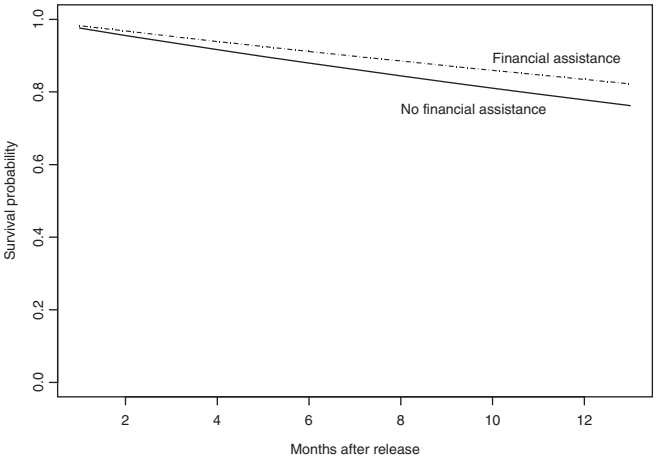

FIGURE 9. *Model-estimated survival probabilities of re-arrest by intervention status for one-class model.*

6.2. School Removal Analyses

School removal data were described in section 2. It was seen that aggressive behavior in the classroom in the fall of Grade 1 was associated with a higher risk for school removal in later grades. The measure of aggressive behavior may, however, contain considerable time-to-time variation as well as measurement error. It may not represent a more sustained level of aggressive behavior and does not capture the trend of behavioral development. In the current analyses, information will,

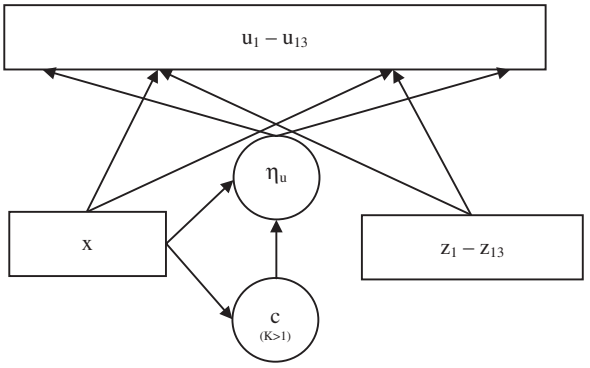

FIGURE 10. *Recidivism path diagram: Multiple-class survival model with time-varying and time-invariant covariates.*

therefore, be incorporated from repeated measures of the child's aggressive behavior. Behavioral development during the four time points of fall and spring in Grades 1 and 2 is used to predict survival in terms of school removal during Grades 3–7. This is achieved using the combination of growth mixture modeling and survival analysis discussed in section 5.2. In this way, a latent trajectory class variable serves to capture the growth shape of aggressive behavior development and is used as a latent class predictor added to the set of observed covariates for the survival process.

In their growth mixture analysis of the aggressive behavior data, Muthén et al. (2002) found evidence of at least three trajectory classes for the development during Grades 1–7: a class with initially high but decreasing aggression trajectory; a class with medium but increasing aggression; and a class with a low stable aggression level. Therefore, a three-class model will also be used here. Muthén et al. (2002) used a linear model for development in Grades 1 and 2 (Model 3).

The covariates to be used are those given in Table 2. The school removal data are obtained as students within classrooms, where some covariates are observed on the individual level and some on the classroom level. For these data there are 16 different classrooms. Such multilevel data need special procedures to obtain correct standard errors and drawing on Muthén et al. (2002), a "sandwich estimator" is used here.

A first analysis step investigates the three-class growth mixture analysis of the four aggressive behavior measures in Grades 1 and 2. The model is given in Equations 29–32. In this model the means of the growth factors are allowed to vary across classes, whereas the slopes in the regressions of the growth factors on the covariates are taken to be class-invariant for simplicity. In line with Muthén et al. (2002), the low class is allowed to have its own variances for the intercept growth factor and for the time-specific residual variances, while the other two classes have the same variances and the same covariance between the growth factors. The high, medium, and low classes were found to contain 8%, 48%, and 44% of the children, respectively.

As a second step, the survival part for Grades 3–7 was added to the model. The model with a single latent class variable is shown in diagrammatic form in Figure 11. In this model, the latent trajectory classes influence the survival part of the model by letting the α_u parameter in Equation 24 vary across classes. For simplicity, the γ_u parameters in Equation 24 are held invariant across classes.

The addition of the survival part did not alter the class percentages to a large degree; the new percentages were 10%, 48%, and 43%, respectively, for the three classes. The estimated mean growth curves in each class also did not change much. The stability of the results may indicate that the growth mixture model is rather well defined. In principle, however, the survival information does contribute to the definition of the latent classes. The fact that the addition of the survival information did not alter the classes much could mean that the survival information is either weaker than the growth information or that it concurs with the growth information. The estimated coefficients for the growth mixture growth factors, class membership, and sur-

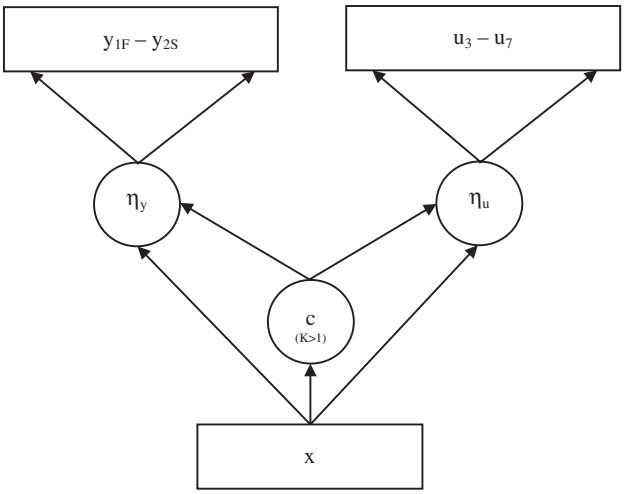

FIGURE 11. *School removal path diagram: Growth + Survival model with time-invariant covariates and single latent class variable.*

vival regressed on the covariates are shown below in Table 4. For simplicity, entries have been left empty when estimates are held equal to a class to the left in the table.

In the three-class model, the latent class regression part of the model finds that the log odds of being in the high class relative to the low class is significantly increased by being in the external control group relative to the internal control group, being male relative to being female, and having a high class average aggression score. The regression coefficients for the intercept and slope factors show influence of covariates within each class. The intercept factor is significantly increased by an individual not being in the external control group, not being white, and being in a class with a low class average lunch value (a poverty indicator). The slope factor is significantly increased by a high class average lunch value and a low class average aggression value in fall of first grade. For the survival part of the model, the latent class growth factor coefficients show an increase in the hazard for school removal by being male, having a high class average lunch value, and having a low class average aggression value. Here, the class-varying intercept values indicate the influence of latent class on hazards. Using the low class as comparison group, membership in the high class gives a significantly increased hazard, as does membership in the medium class. Figure 12 shows the model-estimated mean aggression trajectories and survival of school removal for the three-class model.

Figure 13 shows a diagram for the model with two latent class variables, one for the trajectory classes and one for the survival classes. The model is that of Equations 29–32 combined with Equations 22–24. The parameters of the growth mixture part of the model only vary across the three trajectory classes, while the

TABLE 4
Three-Class Growth + Survival Mixture Model Parameter Estimates

Parameter	High Class		Medium Class		Low Class	
Intercept Factor (η_0)	*Est.*	*SE*	*Est.*	*SE*	*Est.*	*SE*
Intercept	3.89	0.38	2.13	0.22	1.34	0.13
External	**−0.28**	0.05				
Male	0.06	0.05				
White	**−0.13**	0.06				
Lunch	–0.01	0.04				
Cavlunch	**−0.22**	0.07				
Cavtoca1f	0.13	0.07				
Slope Factor (η_1)	*Est.*	*SE*	*Est.*	*SE*	*Est.*	*SE*
Intercept	0.17	0.27	0.35	0.02	0.28	0.19
External	–0.01	0.07				
Male	0.03	0.04				
White	0.03	0.07				
Lunch	–0.02	0.03				
Cavlunch	**0.19**	0.08				
Cavtoca1f	**−0.19**	0.10				
Thresholds (τ)	*Est.*	*SE*	*Est.*	*SE*	*Est.*	*SE*
u_3	3.86	0.64				
u_4	3.68	0.63				
u_5	3.02	0.69				
u_6	2.44	0.64				
u_7	1.10	0.54				
Latent Class Growth Factor (η_u)	*Est.*	*SE*	*Est.*	*SE*	*Est.*	*SE*
Intercept	**2.41**	0.43	**0.79**	0.28	0.00	fixed
External	0.05	0.26				
Male	**0.68**	0.23				
White	–0.48	0.33				
Lunch	–0.28	0.26				
Cavlunch	**1.37**	0.39				
Cavtoca1f	**−0.94**	0.26				
Latent Class Regression	*Est.*	*SE*	*Est.*	*SE*	*Est.*	*SE*
Intercept	–10.97	1.16	–3.99	0.77		
External	**1.69**	0.42	**0.68**	0.33		
Male	**1.71**	0.61	0.53	0.41		
White	–0.14	0.44	–0.31	0.35		
Lunch	0.78	0.79	**0.77**	0.26		
Cavlunch	0.55	0.93	–0.88	0.65		
Cavtoca1f	**3.41**	0.40	**1.90**	0.45		
Class Proportions	0.10		0.48		0.43	

Log likelihood = −1432.49, BIC = 3218.92, 59 free parameters. Boldface entries indicate significance at the 5% level.

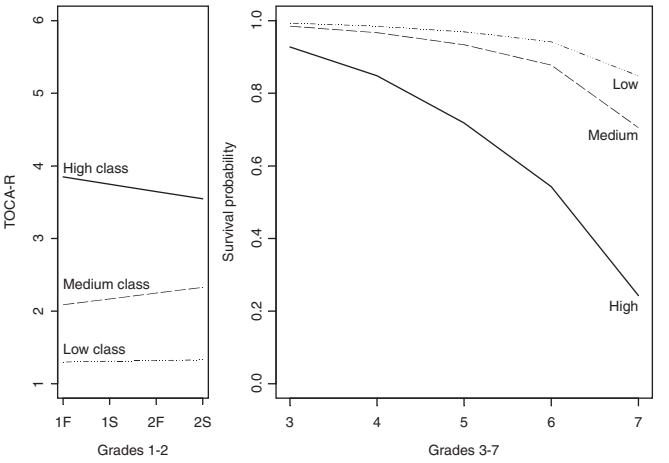

FIGURE 12. *Model-estimated mean aggression trajectory and survival of school removal for 3-class growth + survival mixture model.*

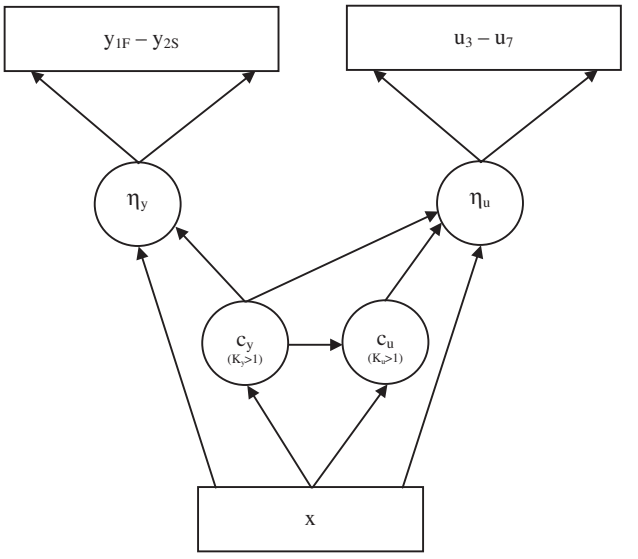

FIGURE 13. *School removal path diagram: Growth + Survival model with time-invariant covariates and both trajectory and survival latent class variables.*

intercept parameters α_u in Equation 24 of the survival part of the model vary across the trajectory classes as well as the survival classes. This model will not explored here, but future analysis efforts may focus on attempts to predict latent classes of survival using aggression trajectory classes in addition to observed covariates.

7. Conclusions

This article has introduced an approach to discrete-time survival analysis using a general latent variable framework. Conventional discrete-time survival analysis is a special case within this framework where a single-class latent class analysis of event history indicators is performed. The single-class model can be identically estimated in a traditional logistic regression model (see, e.g., Singer & Willett, 1993). The great advantage of this more general framework is that it allows for powerful and, at the same time, straightforward, modeling extensions, some of which were proposed and exemplified here. First, unobserved heterogeneity among subjects can be captured using multiple latent classes, where each class was allowed to have its own survival function. A special case of this is long-term survivor model in which a subgroup of individuals have zero hazard for experiencing the event. The mixture extension can also be used for nonparametric estimation of an unspecified continuous frailty distribution. Second, the general modeling framework makes it possible to place the survival analysis in a larger analytic as well as conceptual model in order to study the relationship of survival to other outcomes. As an example, survival analysis was combined with growth mixture modeling of repeated measures. These extensions show the usefulness of integrating the survival analysis in the broader framework.

Many further extensions are of interest in this framework. Some analyses, including those described in this article, have been previously available in that the estimation of the general framework as presented had been implemented in an earlier version of the computer program Mplus (Muthén & Muthén, 1998–2004), which was utilized here. As noted throughout the article, special cases of the more general model have been explored by other authors using alternative model specifications. Examples of other extensions that have been investigated within this framework include survival modeling of recurrent event and multiple spell processes (Masyn, 2003) and discrete-time competing risk models. In addition, recent advances in model specification and estimation in this framework, found in the latest Version 3 of Mplus, permit more sophisticated event history analysis using continuous frailties, multilevel data, and measurement models for covariates as well as outcomes. This also presents even more opportunities for simultaneous modeling of survival processes with other parallel and adjacent longitudinal processes than that which has formerly been explored. It is hoped that this article will stimulate further innovative applications beyond the analysis possibilities presented here.

Appendix

Consider the observed variables **x**, **y**, and **u**, where **x** denotes a $q \times 1$ vector of covariates, **y** denotes a $p \times 1$ vector of continuous outcome variables, and **u** denotes an $r \times 1$ vector of binary and ordered polytomous categorical outcome variables.

Consider latent variables η denoting an $m \times 1$ vector of continuous variables and **c** denoting a latent categorical variable with K classes, $c_i \in \{1, \ldots, K\}$, where $c_i = k$ if individual i belongs to class k. The model has three parts: c related to **x**; **u** related to c and **x**; and **y** related to c and **x**. The following summary draws on Muthén and Muthén (2001, Appendix 8).

The model relates c to **x** by multinomial logistic regression using the K-1-dimensional parameter vector of logit intercepts α_c and the $(K-1) \times q$ parameter matrix of logit slopes Γ_c, where for $k = 1, 2, \ldots, K$

$$P(c_i = k \mid \mathbf{x}_i) = \frac{e^{\alpha_{ck} + \gamma'_{ck}\mathbf{x}_i}}{\sum_{s=1}^{K} e^{\alpha_{cs} + \gamma'_{cs}\mathbf{x}_i}}, \tag{35}$$

where the last class is a reference class with coefficients standardized to zero, $\alpha_{c_K} = 0$, $\gamma_{c_k} = \mathbf{0}$. The latent classes of c influence both **u** and **y**. Consider first the **u** part of the model.

For **u**, conditional independence is assumed given c_i and $\mathbf{x}_i$,

$$P(u_{i1}, u_{i2}, \ldots, u_{ir} \mid c_i, \mathbf{x}_i) = P(u_{i1} \mid c_i, \mathbf{x}_i)\, P(u_{i2} \mid c_i, \mathbf{x}_i) \ldots P(u_{ir} \mid c_i, \mathbf{x}_i) \tag{36}$$

The categorical variable u_{ij} ($j = 1, 2, \ldots, r$) with S_j ordered categories follows an ordered polytomous logistic regression (proportional odds model), where for categories $s = 0, 1, 2, \ldots, S_j - 1$ and $\tau_{j,k,0} = -\infty$, $\tau_{j,k,S_j} = \infty$,

$$u_{ij} = s, \text{ if } \tau_{j,k,s} < u^*_{ij} \leq \tau_{j,k,s+1}, \tag{37}$$

$$P(u_{ij} = s \mid c_i, \mathbf{x}_i) = F_{s+1}(u^*_{ij}) - F_s(u^*_{ij}), \tag{38}$$

$$F_s(u^*) = \frac{1}{1 + e^{-(\tau_s - u^*)}}, \tag{39}$$

where for $\mathbf{u}^*_i = (u^*_{i1}, u^*_{i2}, \ldots, u^*_{ir})'$, $\eta_{ui} = (\eta_{u_1 i}, \eta_{u_2 i}, \ldots, \eta_{u_f i})'$, and conditional on class k,

$$\mathbf{u}^*_i = \Lambda_{u_k} \eta_{ui} + \mathbf{K}_{u_k} \mathbf{x}_i, \tag{40}$$

$$\eta_{ui} = \alpha_{u_k} + \Gamma_{u_k} \mathbf{x}_i, \tag{41}$$

where Λ_{u_k} is an $r \times f$ logit parameter matrix varying across the K classes, $\mathbf{K}_{u_k}$ is an $r \times q$ logit parameter matrix varying across the K classes, α_{u_k} is an $f \times 1$ vector logit parameter vector varying across the K classes, and Γ_{u_k} is an $f \times q$ logit parameter matrix varying across the K classes. The thresholds may be stacked in the

$$\sum_{j=1}^{r} (S_j - 1) \times 1$$

vectors τ_k varying across the K classes.

It should be noted that Equation 40 does not include intercept terms given the presence of τ parameters. Furthermore, τ parameters have opposite signs than u^* in Equation 40 because of their interpretation as thresholds or cutpoints that a latent continuous response variable u^* exceeds or falls below (see also Agresti, 1990, pp. 322–324). For example, with a binary u_j scored 0/1, (38) leads to

$$P(u_{ij} = 1 \mid c_i, \mathrm{x}_i) = 1 - \frac{1}{1 + e^{-(\tau - u^*)}}, \tag{42}$$

$$= \frac{1}{1 + e^{-\mathrm{logit}}}, \tag{43}$$

where logit $= -\tau + u^*$. For example, the higher the τ the higher u^* needs to be to exceed it, and the lower the probability of $u = 1$.

The model structure in Equations 40 and 41 is useful when the **u** vector represents repeated measures, and the latent classes correspond to different trajectory classes. In this case, the elements of η_u correspond to growth factors in random effects growth modeling, except that η_u has zero variance conditional on **x**.

Consider next the **y** part of the model. Multivariate normality is assumed for **y** conditional on **x** and class *k,*

$$\mathbf{y}_i = \boldsymbol{\upsilon}_k + \Lambda_k \eta_i + \mathbf{K}_k \mathbf{x}_i + \boldsymbol{\epsilon}_i, \tag{44}$$

$$\eta_i = \alpha_k + \mathrm{B}_k \eta_i + \Gamma_k \mathbf{x}_i + \zeta_i, \tag{45}$$

where the residual vector $\boldsymbol{\epsilon}_i$ is $N(0, \Theta_k)$ and the residual vector ζ_i is $N(0, \Psi_k)$, both assumed to be uncorrelated with other variables. This part of the mixture model builds on a general structural equation model generalized to the K classes of the mixture.

The Mplus mixture model is estimated by maximum-likelihood using the EM algorithm. Missing data on **u** and **y** are handled using the MAR assumption (Little & Rubin, 2002). The analysis makes it possible to incorporate knowledge about class membership for certain individuals. Individuals with known class membership are referred to as training data (see also Hosmer, 1973; McLachlan & Basford, 1988). The training data typically consists of 0 and 1 class membership values for all individuals, where 1 denotes which classes an individual may belong to. Known class membership for an individual corresponds to having training data value of 1 for the known class and 0 for all other classes. Unknown class membership for an individual is specified by the value 1 for all classes. With class membership training data, the class probabilities are renormed for each individual to add to one over the admissible set of classes.

For comparison of fit of models that have the same number of classes and are nested, the usual likelihood-ratio chi-square difference test can be used. Comparison of models with different numbers of classes, however, is accomplished by a Bayesian information criterion (BIC; Kass & Raftery, 1993; Schwartz, 1978),

$$\text{BIC} = -2\log L + d \ln n, \tag{46}$$

where d is the number of free parameters in the model. The lower the BIC value, the better the model.

When the model contains only **u,** Pearson and likelihood ratio chi-square tests against the unrestricted multinomial alternative can be computed,

$$\chi^2_P = \sum_{\text{cells}} \frac{(o_i - e_i)^2}{e_i}, \tag{47}$$

$$\chi^2_L = 2\sum_{\text{cells}} o_i \log o_i/e_i, \tag{48}$$

where o_i is the observed frequency in cell i of the multivariate frequency table for **u** and e_i is the corresponding frequency estimated under the model. With missing data on **u**, the EM algorithm described in Little and Rubin (2002; chapter 9.3, pp. 181–185) is used to compute the estimated frequencies in the unrestricted multinomial model.

References

Allison, P. D. (1984). *Event history analysis. Regression for longitudinal event data. Quantitative applications in the social sciences, No. 46.* Thousand Oaks, CA: Sage.

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

Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society* B, *11,* 15–53.

Hedeker, D., Siddiqui, O., & Hu, F. B. (2000). Random-effects regression analysis of correlated group-time survival data. *Statistical Methods in Medical Research, 9*(2), 161–179.

Henderson, R., Diggle, P., & Dobson, A. (2000). Joint modeling of longitudinal measurements and event time data. *Biostatistics, 1,* 465–480.

Hosmer, D. W. (1973). A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three different types of sample. *Biometrics, 29,* 761–770.

Hougaard, P. (2000). *Analysis of multivariate survival data.* New York: Springer.

Huang, G. H., & Bandeen-Roche, K. (2004). Building an identifiable latent variable model with covariate effects on underlying and measured variables. *Psychometrika, 69*, 5–32.

Jo, B. (2002). Estimation of intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics, 27,* 385–409.

Jo, B., & Muthén, B. (2000). Intervention studies with noncompliance: Complier average causal effect estimation in growth mixture modeling. In N. Duan & S. Reise (Eds.), *Multilevel modeling: Methodological advances, issues, and applications* (pp. 112–139), Multivariate Applications Book Series. Hillsdale, NJ: Erlbaum.

Jo, B., & Muthén, B. (2001). Modeling of intervention effects with noncompliance: A latent variable approach for randomized trials. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 57–87). Hillsdale, NJ: Erlbaum.

Kass R. E., & Raftery, A. E. (1993). Bayes factors. *Journal of the American Statistical Association, 90,* 773–795.

Kellam, S. G., Rebok, G. W., Ialongo, N., & Mayer, L. S. (1994). The course and malleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiologically based preventive trial. *Journal of Child Psychology and Psychiatry, 35,* 359–382.

Land, K. C., Nagin, D. S., & McCall, P. L. (2001). Discrete-time hazard regression models with hidden heterogeneity: The semiparametric mixed Poisson regression approach. *Sociological Methods and Research, 29*(3), 342–373.

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

Maller, R. A., & Zhou, X. (1996). *Survival analysis with long-term survivors.* 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.

McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and applications to clustering.* New York: Marcel Dekker.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models.* New York: Wiley.

Muthén, B. (2001a). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In L. M. Collins, & A. Sayer, (Eds.), *New methods for the analysis of change* (pp. 291–322). Washington, DC: APA.

Muthén, B. (2001b). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–33). Hillsdale, NJ: Erlbaum.

Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S. T., Yang, C. C., Wang, C. P., Kellam, S., Carlin, J., & Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics, 3*(4), 459–475.

Muthén, B., & Muthén, L. (1998–2004). *Mplus* [Computer software]. Los Angeles: Muthén & Muthén.

Muthén, B., & Muthén, L. (2001). *Mplus User's Guide.* Los Angeles: Muthén & Muthén.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55,* 463–469.

Nielsen, G. G., Gill, R. D., Andersen, P. K., & Soerensen, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics, 19,* 25–43.

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

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461–464.

Singer, J. D., & Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics, 18*(2), 155–195.

Steele, F. (2000). A multilevel mixture model for even history data with long-term survivors: An application to an analysis of contraceptive sterilization in Bangladesh. Paper presented at the Population Association of America Annual Meetings, March 2000, Los Angeles.

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics to mortality. *Demography, 16*(3), 439–454.

Vermunt, J. K. (1997). *Log-linear models for event histories. Advanced quantitative techniques in the social sciences, vol. 8.* Thousand Oaks, CA: Sage.

Werthamer-Larsson, L., Kellam, S. G., & Wheeler, L. (1991). Effect of first-grade classroom environment on child shy behavior, aggressive behavior, and concentration problems. *American Journal of Community Psychology, 19,* 585–602.

Authors

BENGT MUTHÉN is Professor at the Graduate School of Education & Information Studies, UCLA, Los Angeles, CA 90095: bmuthen@ucla.edu. His research interests focus on the development of applied statistical methodology in education and public health. Dr. Muthén's areas specialization include latent variable modeling, analysis of individual differences in longitudinal data, analysis of categorical data, and multilevel modeling. He is one of the developers of the Mplus computer program, which implements many of his statistical procedures, and currently has an independent Scientist Award from the National Institutes of Health for methodology development in the alcohol field.

KATHERINE MASYN is Assistant Professor, Department of Human and Community Development, University of California, Davis, CA 95616; kmasyn@ucdavis.edu. Her areas of specialization are discrete time survival analysis, latent variable growth modeling, and finite mixture modeling.

Manuscript received October 3, 2003
Revision received August 5, 2004
Accepted August 6, 2004