# Discrete-Time Survival Mixture Analysis

Bengt Muthén
Katherine Masyn
University of California, Los Angeles
*

Version 3, November 12, 2001

*The research was supported for the first author under grant K02 AA 00230-01 from NIAAA, by both NIMH and NIDA under grant MH40859, and by NIMH under grants No. MH01259, No. MH38725, and No. MH42968. 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. The email address for the first author is bmuthen@ucla.edu.

## Abstract

This paper proposes a general latent variable approach to discrete-time survival analysis of non-repeatable 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 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.

Key words: event history, latent classes, long-term survivors, growth mixture modeling, maximum likelihood.

Complete address of first author: Graduate School of Education & Information Studies, Moore Hall, Box 951521, Los Angeles CA 90095-1521.

# 1 Introduction

This paper considers discrete-time survival analysis to study the probability, or hazard, of experiencing a non-repeatable 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).

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

The aim of this paper 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 new types of analyses. 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 paper 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 will be 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 will be presented in Section 6.

## 2.1 Recidivism Data

This data set 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 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 paper, the data is recoded into 13 four-week intervals, referred to as "months". Further justification for a discrete-time treatment of this data is given in Section 6.1.

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 one 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. For the thirteen months, there are thirteen corresponding binary indicators, labeled $u_1 - u_{13}$. A subject has a value of 1 if he was arrested in a given month and a value of 0 if he was not. An inmate is only at risk for re-arrest if he has not already been arrested after his release. If a subject is no longer at risk, i.e., no longer part of the risk set, then he has a missing value for all subsequent indicators. Thus, the sample average for each of the event indicators is the estimated probability of being arrested in given month among those inmates who have not been re-arrested prior to that month. For example, in the first month, all 432 released inmates were at risk for re-arrest and 4 experienced re-arrest. The estimated indicator mean for the first month is $4/432 = 0.01$. For the second month, only $432 - 4 = 428$ inmates were at risk for re-arrest and 8 were arrested. The estimated indicator mean for the second month is then $8/428 = 0.02$. The sample event indicator mean is also known as the marginal hazard probability. The sample hazard probabilities can be plotted by month as shown in Figure 1. This representation of the sample hazard function suggests that the actual hazard function may be constant with random sampling accounting for the fluctuation in the range 0.01 to 0.03. The proportions of the initial population of inmates surviving through each month, termed survival probabilities, can be estimated directly from the estimated hazard probabilities. The relationship between the survival and hazard functions is described in greater detail in Section 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 thirteenth month.

INSERT TABLE 1 HERE

INSERT FIGURES 1-2 HERE

Figures 3 and 4 display the same sample-based estimates of the hazard and survival probabilities, stratified by intervention status. There is no clear difference in the hazard functions for the two groups but the group of inmates not receiving the financial aid intervention does have a slightly lower survival curve, with almost 10% more of the initial control group re-arrested by the end of the thirteenth month. The survival analysis of this data set will investigate whether the intervention has a significant effect on the hazard probabilities during the study period 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. Extending to discrete-time mixture analysis, the effect of intervention on the long-term probability of re-arrest as well as its effect on the pattern of recidivism during the duration of the study period will be statistically evaluated.

INSERT FIGURES 3-4 HERE

## 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 & Mayers, 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 six-point scale from almost never to almost always. A 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 paper only the control groups' data will be used. At the first grade fall measurement there were 6 internal and 10 external control classrooms, with a total of 404 children.

Table 2 shows the variables to be used in the survival analyses. Here, survival concerns not being removed from school. The analyses will focus on 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. 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 3/4 of the children have not experienced school removal. Figure 6 shows the corresponding picture when dividing the sample into a high and a normal/low aggression group 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 a higher aggression score are seen to have a considerably lower survival curve with almost half the children having experienced school removal by the end of grade 7. The survival analysis of these data will investigate the effects of the measured covariates on trends in both aggression and school removal survival. In the discrete-time mixture framework, the influence of latent trajectory classes of aggression in first and second grade on classes of survival trends in grades 3-7 will also be explored.

INSERT TABLE 2

INSERT FIGURES 5-6 HERE

# 3 Discrete-Time Survival Analysis Methodology

This section introduces the basic statistical components of discrete-time survival analysis. The probability of observing the sample (the likelihood), the hazard function, and the survival function are presented.

## 3.1 Event History Indicators

Discrete-time survival analysis considers a set of binary 0/1 event history indicators $u_j, j = 1, 2, \ldots, j_i$, where $u_{ij} = 1$ if individual $i$ experiences an event in time period $j$ and $j_i$ is the last time period of data collection for individual $i$. A single non-repeatable event is considered so that data collection ends for individual $i$ when $u = 1$ has been observed. This means that there are only two types of patterns of $u$ observations. One pattern has $u = 0$ for every time period that was observed during data collection. Here, the data collection ends before an event has been observed. These individuals are referred to as censored (right-censoring) because it is unknown if and when they experienced the event after data collection ended. The other pattern has $u = 0$ for every time period except the last for which $u = 1$. These individuals are referred to as uncensored because their survival status is known. Note that for both types of patterns, the last time period of data collection may differ across individuals.

### 3.2 The Likelihood

One approach to constructing the likelihood for a given data set is to begin with the actual survival time. Define $T$ as a discrete random variable that indicates the time period $j$ when the event occurs.

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 $j$,

$$S_j = P(T > j). \tag{1}$$

The hazard probability is defined as the probability of experiencing the event in time period $j$ given that it was not experienced prior to $j$,

$$h_j = P(T = j | T \geq j). \tag{2}$$

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

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

Suppose the duration of the study is made up of $J$ time periods. Then each individual $i$ in the sample is observed until some period $j_i$, where $j_i \leq J$. Observation of the individual may discontinue for three reasons: 1) The individual experiences the event;

2) The individual drops out of the study; or 3) The study concludes. In the first case, $T_i = j_i$. In the second and the third case, it is only known that $T_i > j_i$. Individuals with $T_i > j_i$ are *right-censored*: it is unknown whether they experience the event after their observation period. For individuals with $T_i = j_i$, the likelihood may be expressed in terms of the hazard as

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

For individuals with $T_i > j_i$, the likelihood may be expressed as

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

As shown in detail in Singer and Willet (1993), it follows that the likelihood for the full sample is $L = \prod_{i=1}^{n} l_i$, where

$$l_i = \prod_{j=1}^{j_i} h_{ij}^{u_{ij}}\ (1 - h_{ij})^{1 - u_{ij}} \tag{6}$$

and

$$h_{ij} = P(u_{ij} = 1) \tag{7}$$

using the $u_{ij}$ event history indicators defined in the preceding section.

### 3.3 The Hazard Function

In line with Singer and Willet (1993) a logistic hazard function is considered. Adding $q$ covariates $\mathbf{x}$ the hazard can be written as the logistic function

$$h_{ij} = \frac{1}{1 + e^{-(logit_{ij})}}. \tag{8}$$

where the $logit_{ij}$ is expressed as,

$$logit_{ij} = -\tau_j + \boldsymbol{\kappa}_j' \, \mathbf{x}_i, \tag{9}$$

where $\mathbf{x}_i$ refers to both time-invariant and time-varying covariates. The conventional proportionality assumption may be applied, which is conveniently represented via a factor $\eta_{ui}$,

$$logit_{ij} = -\tau_j + \eta_{ui}, \tag{10}$$

$$\eta_{ui} = \boldsymbol{\gamma}_u' \, \mathbf{x}_i, \tag{11}$$

so that a given $x$ variable has the same influence on the hazard during all time periods $j$. The specification in (11) is typically used with time-invariant covariates. Equality of effects of the time-varying covariates may be introduced by dropping the $j$ subscript from $\boldsymbol{\kappa}$ in (9). It may also be of interest to specify functional forms for the logits in (9), for example linear trends.

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

Muthén and Shedden (1999), Muthén, Shedden, and Spisic (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, 1998-2001). 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. Mixture modeling captures this heterogeneity by a latent categorical variable. Mixture modeling has a wide variety of applications. Overviews with latent class and growth mixture applications are given in Muthén (2001a, b) and Muthén and Muthén (2000). Applications to randomized trials are given in Muthén, Brown, Masyn, Jo, Khoo, Yang, Wang, Kellam, Carlin, and Liao (2000), Jo (1999), and Jo and Muthén (2000, 2001).

Let $\mathbf{c}$ denote a latent categorical variable with $K$ classes, $\mathbf{c}_i = (c_{i1}, c_{i2}, \ldots, c_{iK})'$, where $c_{ik} = 1$ if individual $i$ belongs to class $k$ and zero otherwise. The model relates

$\mathbf{c}$ to a $q \times 1$ covariate vector $\mathbf{x}$ by multinomial logistic regression using the $(K-1)$-dimensional parameter vector of logit intercepts $\boldsymbol{\alpha}_c$ and the $(K-1) \times q$ parameter matrix of logit slopes $\boldsymbol{\Gamma}_c$, where for $k = 1, 2, \ldots, K$

$$P(c_{ik} = 1|\mathbf{x}_i) = \frac{e^{\alpha_{c_k} + \boldsymbol{\gamma}'_{c_k} \mathbf{x}_i}}{\sum_{k=1}^{K} e^{\alpha_{c_k} + \boldsymbol{\gamma}'_{c_k} \mathbf{x}_i}}, \tag{12}$$

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

Define an $r \times 1$ vector $\mathbf{u}$ of binary (0/1) variables with conditional independence given $\mathbf{c}_i$ and $\mathbf{x}_i$,

$$P(u_{i1}, u_{i2}, \ldots, u_{ir}|\mathbf{c}_i, \mathbf{x}_i) = P(u_{i1}|\mathbf{c}_i, \mathbf{x}_i)\ P(u_{i2}|\mathbf{c}_i, \mathbf{x}_i) \ldots P(u_{ir}|\mathbf{c}_i, \mathbf{x}_i). \tag{13}$$

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 $\tau_j$,

$$P(u_{ij} = 1|\mathbf{c}_i, \mathbf{x}_i) = \frac{1}{1 + e^{-(-\tau + u^*)}}, \tag{14}$$

For example, the higher the $\tau$, 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).

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

$$\mathbf{u}_i^* = \boldsymbol{\Lambda}_{u_k}\ \boldsymbol{\eta}_{ui} + \mathbf{K}_{u_k}\ \mathbf{x}_i, \tag{15}$$

$$\boldsymbol{\eta}_{ui} = \boldsymbol{\alpha}_{u_k} + \boldsymbol{\Gamma}'_{u_k}\ \mathbf{x}_i, \tag{16}$$

where $\mathbf{\Lambda}_{u_k}$ 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, $\boldsymbol{\alpha}_{u_k}$ is an $f \times 1$ logit parameter vector varying across the $K$ classes, and $\mathbf{\Gamma}_{u_k}$ is an $f \times q$ logit parameter matrix varying across the $K$ classes. The model structure in (15) and (16) is useful when the $\mathbf{u}$ vector represents repeated measures and the latent classes correspond to different trajectory classes. In this case, the elements of $\boldsymbol{\eta}_u$ correspond to growth factors in random effects growth modeling, except that $\boldsymbol{\eta}_u$ has zero variance conditional on $\mathbf{x}$. In the present survival context, $\boldsymbol{\eta}_u$ is used both to conveniently specify a proportional-odds assumption and to impose trends.

## 4.1 Fitting Discrete-Time Survival into the General Framework

The likelihood in (6) indicates how discrete-time survival analysis can be carried out in the general latent variable model. Consider first how to specify the event history indicators $\mathbf{u}$.

As an example consider five time periods. An individual who is censored after time period five ($j_i = 5$) has the event history

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

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

$$\begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix},$$

while an individual who drops out after period four before the study ends ($j_i = 4$) has the event history

$$\begin{pmatrix} 0 & 0 & 0 & 0 \end{pmatrix}.$$

The event history information may be entered into an $r \times 1$ data vector $\mathbf{u}_i'$ where $r$ denotes the maximum value of $j_i$ over all individuals and unobserved $u$ information is represented as $u = 999$ to denote missing data. Using the examples above, the first data vector $\mathbf{u}_i'$ is

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

the second data vector $\mathbf{u}_i'$ is

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 999 \end{pmatrix},$$

and the third data vector $\mathbf{u}_i'$ is

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 999 \end{pmatrix}.$$

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 four is unrelated to the event being studied. The conventional assumption of noninformative censoring, i.e., that censoring times are independent of event times, corresponds to the assumption of ignorable missingness in the general latent variable model. The likelihood expression in (6) is obtained when applying maximum-likelihood estimation under MAR (Little & Rubin, 1987) to the general modeling framework with missing data on $\mathbf{u}$ as indicated above.

The likelihood in (6) suggests how to specify discrete-time models in the general framework. The likelihood is the same as for $\mathbf{u}$ related to $\mathbf{c}$ and $\mathbf{x}$ in a single-class model ($K = 1$), due to the fact that the $u$'s are independent conditional on $\mathbf{x}$. The $\boldsymbol{\kappa}'_j$ parameter vector of (9) is the $j^{th}$ row of $\mathbf{K}_u$ in (15), while (11) is a special case of (16). Using $\boldsymbol{\Lambda}_{u_k}$ in (15), functional forms for the logits in (9) can be specified, for example linear trends such as

$$\boldsymbol{\Lambda}_{u_k} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & T-1 \end{pmatrix},$$

where $T$ is the number of time periods.

# 5 Mixture Analysis

It is often important to take into account unobserved heterogeneity in survival among the subjects studied. In continuous-time survival modeling it is common to take unobserved heterogeneity into account using "frailties", that is representing heterogeneity by random effects (continuous latent variables); see e.g. Hougaard (2000). This paper 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.

Consider the multiple-class modification of (8) and (9) for class $k$ ($k = 1, 2, \ldots, K$),

$$h_{ijk} = \frac{1}{1 + e^{-(logit_{ijk})}}, \tag{17}$$

$$logit_{ijk} = -\tau_{jk} + \boldsymbol{\kappa}'_{jk}\,\mathbf{x}_i. \tag{18}$$

With multiple classes, the model adds the prediction of class membership by covariates $\mathbf{x}$ as in (12),

$$P(c_{ik} = 1|\mathbf{x}_i) = \frac{e^{\alpha_{c_k} + \boldsymbol{\gamma}'_{c_k}\mathbf{x}_i}}{\sum_{k=1}^{K} e^{\alpha_{c_k} + \boldsymbol{\gamma}'_{c_k}\mathbf{x}_i}}. \tag{19}$$

The inclusion of multiple classes modifies the likelihood expression in (6), $L = \prod_{i=1}^{n} l_i$, as

$$l_i = \sum_{k=1}^{K} \pi_{ik} \prod_{j=1}^{j_i} h_{ijk}^{u_{ij}} (1 - h_{ijk})^{1-u_{ij}}, \tag{20}$$

where $\pi_{ik} = P(c_{ik} = 1|\mathbf{x}_i)$ in (12). 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 (2001).

As mentioned earlier, from a latent class point of view, the discrete-time survival model presented in Section 3 can be viewed as a single-class model. When covariates are not present, the discrete-time survival model has the special feature of perfectly fitting the data on the $u$'s. A Pearson or likelihood-ratio chi-square test of fit of the model against the unrestricted multinomial distribution ($2^r - 1 - r$ degrees of freedom) has zero value irrespective of the data. This is because the $u$ variables are by construction independent, i.e. $P(u_j, u_k) = P(u_j)\ P(u_k)$ for all pairs of $j, k$. This follows from the definition of $u_j$ as representing events conditional on previous events. Consider for example the $(u_1\ u_2)$ outcomes (0 0) and (0 1),

$$P(u_1 = 0,\ u_2 = 0) = P(T > 2)$$

$$= P(T_i \neq 1|T_i \geq 1)\ P(T_i \neq 2|T_i \geq 2) = P(u_1 = 0)\ P(u_2 = 0), \tag{21}$$

$$P(u_1 = 0,\ u_2 = 1) = P(T = 2)$$

$$= P(T_i \neq 1 | T_i \geq 1)\ P(T_i = 2 | T_i \geq 2) = P(u_1 = 0)\ P(u_2 = 1). \quad (22)$$

The independence feature implies that there is no information in the joint distribution of the $u$ variables from which multiple classes can be defined. Adding covariate information, however, 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 heterogeneity 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 who are not in the risk set, but who have zero hazard during all time periods. Using the $u$ notation of Section 3.1, 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 3 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

$$S_{ij} = \pi \, S_{ij}^{NLTS} + (1 - \pi); \tag{23}$$

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 long-term survival class are obtained by setting $\tau_j = \infty$ and $\boldsymbol{\kappa}_j = \mathbf{0}$ for all $j$'s in (9); see also (15). The model is completed by the logistic regression for class membership,

$$log[\pi/(1 - \pi)] = \boldsymbol{\gamma}_c' \, \mathbf{x}_i, \tag{24}$$

which is a special case of (12).

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 $\pi_{ik} = P(c_{ik} = 1|\mathbf{x}_i)$. The covariates may also influence the event history indicator probabilities (the hazards) $h_{ijk}$, either directly or via the factor $\eta_{ui}$ (using the proportional-odds specification),

$$h_{ij} = \frac{1}{1 + e^{-(logit_{ij})}}, \tag{25}$$

$$logit_{ij} = -\tau_j + \eta_{ui}, \tag{26}$$

$$\eta_{ui} = \boldsymbol{\gamma}'_u \, \mathbf{x}_i. \tag{27}$$

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 to be 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 $\mathbf{u}$ is analyzed jointly with the growth mixture model for $\mathbf{y}$. Maximum-likelihood estimation is used also in this case.

Consider as an example repeated measures on continuous outcomes $y_{it}$ ($i = 1, 2, \ldots, n$; $t = t_1, t_2, \ldots, t_T$) that can be described by only two random effects (growth factors) $\eta_{0i}$ and $\eta_{1i}$ and a time-specific residual $\epsilon$,

$$y_{it} = \eta_{0i} + \eta_{1i}\, a_{it} + \epsilon_{it}, \tag{28}$$

where it will be assumed that the time scores are common to all individuals, $a_{it} = a_t$ (deviations from this can be handled via missing data techniques). Different trajectory classes are allowed for by letting the means, variances and covariance of $\eta_0$ and $\eta_1$ vary across the classes. The variances of $\epsilon_t$ may also vary across classes. The covariates of $\mathbf{x}$ may influence class membership as in (12). They may also have class-varying influence on the growth factors ($k = 1, 2, \ldots, K$),

$$\eta_{0i} = \alpha_{0k} + \boldsymbol{\gamma}'_{0k}\, \mathbf{x}_i + \zeta_{0i}, \tag{29}$$

$$\eta_{1i} = \alpha_{1k} + \boldsymbol{\gamma}'_{1k}\, \mathbf{x}_i + \zeta_{1i}, \tag{30}$$

The latent class variable is related to covariates $\mathbf{x}$ as in the general framework of Section 4,

$$P(c_{ik} = 1|\mathbf{x}_i) = \frac{e^{\alpha_{c_k} + \boldsymbol{\gamma}'_{c_k} \mathbf{x}_i}}{\sum_{k=1}^{K} e^{\alpha_{c_k} + \boldsymbol{\gamma}'_{c_k} \mathbf{x}_i}}, \tag{31}$$

The model given in (28) - (31) is referred to as growth mixture modeling and was introduced in Muthén and Shedden (1999); for overviews, see e.g. Muthén (2001a, b).

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, using a proportional-odds formulation,

$$h_{ijk} = \frac{1}{1 + e^{-(logit_{ijk})}}, \tag{32}$$

$$logit_{ijk} = -\tau_j + \eta_{uik}, \tag{33}$$

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

It is seen in (33) that the threshold parameters are held invariant across the latent classes,

$$\tau_{j1} = \tau_{j2} = \ldots = \tau_{jK} = \tau_j; j = 1, 2, \ldots, r, \tag{35}$$

so that the latent class membership for the growth model influences the hazard function through the class-varying $\alpha$ and the class-varying $\boldsymbol{\gamma}$ influence from $\mathbf{x}$, as seen in (34). Here, $\alpha$ is standardized to zero for a reference class, while estimated in the remaining classes.

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 $y$'s does not require more than one class, more than one class can be specified for the $u$'s. For example, a long-term survivor class can be specified for the $u$'s because the $y$ information makes it possible to identify the long-term survivors among the censored individuals. When covariates are present, the growth mixture model may be combined with a multiple-class discrete-time survival model, e.g. in the form of a long-term survivor 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 covariates, testing the proportional-odds assumption. It is also used to illustrate modeling with a long-term survivor class. 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 accurately assess 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 non-significant 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 0.09 (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 paper have been grouped into 13 four-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 regards to time-varying covariate effects. In addition, the discrete time framework allows testing of the modeling assumptions such as constancy of the 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 (9), which allows for time-specific covariate effects, is compared to the model with the hazard logit defined in (10) - (11), which constrains the covariate effects to be equal across time using the factor $\eta_u$. 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.

INSERT FIGURES 7-8 HERE

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 hazard. In the preceding analysis step, the hazard is completely unstructured. A specific structure may now be imposed on the logit 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 (9) or (10), removing the subscript $j$ from the threshold $\tau$. 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 hazard assumption. Table 3 shows the results from the model with the proportionality and constant hazard assumptions applied. These results are consistent with the previous continuous-time analyses (Allison, 1984).

INSERT TABLE 3 HERE

Next the possibility of a latent class of long-term survivors is considered. In the current application, this is a class of inmates who have been fully "reformed" and, once released, are no longer at risk for arrest. The set of time-invariant explanatory variables is allowed to influence the hazard both directly, through influence on the factor $\eta_u$ in (34), and indirectly, through influence on the probability of latent class membership in (31). The diagram for the model with covariates and a long-term survivor class is shown in Figure 9. In this two-class situation, the hazard for the non-long-term survivor class was kept unstructured. The results for the model with long-term survivors are given in Table 4. A caution should be issued here for the fitting of a long-term survivor model. This special case of latent class analysis, like the more general forms of LCA, can be susceptible to convergence at locally rather than globally optimal solutions. Because of this, 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.

INSERT FIGURE 9 HERE

INSERT TABLE 4 HERE

The two-class model has a log likelihood value of $-496.23$ with a BIC (Bayesian Information Criterion; see Appendix) value of 1180.58 (31 parameters). These values may be compared to the previous one-class model with unstructured hazard that does not include long-term survivors; the one-class model has a log likelihood value of $-505.36$ with a BIC value of 1144.23 (22 parameters). Based on the BIC values, there is no real evidence for the need to include a long-term survivor class. However, the two-class model provides interesting insight into the intervention effect that is worth examining more closely even if the model itself is not statistically superior to the one-class model. For the two-class model it is estimated that 21% of the inmates are long-term survivors, not at risk for re-arrest. This implies that nearly 30% of those inmates not re-arrested during the study period may have no possibility of future incarceration. The direct effects of the covariates on the hazard change in both magnitude and significance when compared to the one-class model. And the set of significant covariates directly associated with the hazard is different from the set of significant predictors for latent class membership. The intervention effect on the hazard is estimated as a much larger and significant effect in the non-long-term survivor class compared to the estimated overall effect in the one-class model. In the one-class model, the hazard odds ratio for the intervention group compared to the control group is estimated as 0.72 and this effect is not significant. In the two-class model, the hazard odds ratio for the intervention group compared to the control group in the non-long-term survivor class is 0.43 and this effect is statistically significant. This could indicate that the financial assistance intervention is effective at extending the time to re-arrest for those in the recidivism risk set. In addition, age remains a significant influence on the hazard odds for re-arrest but the number of prior convictions, found to be significant in the one-class model, does not have a significant effect among non-long-term survivors in the two-class model. The latent class regression part of the model finds that the log odds of being in the long-term survivor class relative to the non-long-term survivor class is significantly associated with prior work experience, age, and years of education. It should also be noted that there is not a significant effect of the intervention on the probability of long-term survivor class membership. One might infer from this that the benefits of the intervention extend only as long as the financial aid is provided and that the long-term likelihood of recidivism is not altered. Figures 10 and 11 show the model-estimated mean survival plots for both the one- and two-class model, respectively, for the two intervention groups at the overall sample mean values for the other covariates. Figure 11, when compared to Figure 10, shows a worse course of recidivism for non-long-term survivors but a larger effect of financial aid.

INSERT FIGURES 10-11 HERE

A problem with the presented solution should, however, be noted and illustrates the potential fragility of the long-term survivor model alluded to earlier. In this example, a local optimum was found that was only a few log likelihood points lower than the solution presented in Table 4. This competing solution had markedly different estimates for the latent class regression parameters and for the percentage of long-term survivors. For these particular data with the set of covariates used here there is therefore not sufficient information to make convincing inference about a latent class of long-term survivors.

## 6.2 School Removal Analyses

The school removal data were described in Section 2. It was seen that aggressive behavior in the classroom in 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 therefore be incorporated from repeated measures of the child's aggressive behavior. The behavioral development during the four time points of Fall and Spring in grade 1 and grade 2 will be 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.

In their growth mixture analysis of the aggressive behavior data, Muthén, Brown, Masyn, Jo, Khoo, Yang, Wang, Kellam, Carlin, and Liao (2000) found evidence of at least three trajectory classes for the development during grade 1 - grade 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. A three-class model will therefore be used also here. Muthén et al (2000) used a linear model for development in grades 1 and 2 (Model 3).

For the survival part of the model one can argue on substantive grounds for a subgroup of children who are never at risk for school removal. Two model versions, with and without a long-term survival class, will therefore be examined. 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 (2000) 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 (28) - (31). 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 (2000), 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 without a long-term survivor class is shown in diagrammatic form in Figure 12. In this model, the latent trajectory classes influence the survival part of the model by letting the $\alpha_u$ parameter in (34) vary across classes. For simplicity, the $\boldsymbol{\gamma}_u$ parameters in (34) are held invariant across classes.

INSERT FIGURE 12 HERE

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. To investigate this further, a class of long-term survivors for the survival part is explored next.

Figure 13 shows a diagram for the model with three trajectory classes and two survival classes. In line with Muthén (2001b), having these two latent class variables in the model is handled by a single latent class variable with six classes. The model is that of (28) - (35). For the long-term survivor classes (with high, medium, or low trajectories), $\tau_j$ in (33) is fixed at 10 (to represent $\infty$) and $\alpha_u$ and $\boldsymbol{\gamma}_u$ in (34) are fixed at zero. The parameters of the growth mixture part of the model only vary across the three trajectory classes, while the intercept parameters $\alpha_u$ in (34) of the survival part of the model vary across all classes, except the last two classes with low trajectory where it is fixed at zero as a standardization.

INSERT FIGURE 13 HERE

The analysis of the six-class model showed a very low class count for long-term survivors in the high trajectory class, giving rise to problems in estimating regression coefficients for class membership related to covariates. This class was therefore removed. The resulting 5-class solution had a log likelihood value of $-1,419.01$, while the BIC (Bayesian Information Criterion; see Appendix) value was $3,275.94$ (73 parameters). These values may be compared to those of the previous 3-class model that does not include long-term survivors, $-1,432.49$ and $3,218.92$ (59 parameters), respectively. Because the BIC value is higher (worse) for the 5-class model than for the 3-class model, there is no evidence based on BIC for the need to include a long-term survivor class. Considering the estimated class counts for those with medium and low trajectories in the 5-class model, however, shows a majority being long-term survivor individuals. Of the estimated 44 individuals with a low trajectory, 26 are in the long-term survivor class, while among the estimated 49 individuals with a medium trajectory, 28 are in the long-term survivor class. In total, it is estimated that 54% of the children are long-term survivors, not at risk for school removal.

The estimated coefficients for the growth mixture growth factors, class membership, and survival regressed on the covariates are shown below in Table 5 and Table 6 for the 3-class and 5-class models, respectively. For simplicity, entries have been left empty when estimates are held equal to a class to the left in the table. The results are quite different for the two models, reflecting the fact that the 5-class model considers only a subset of the children to be in the risk set for school removal.

In the 3-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 14 shows the model-estimated mean aggression trajectories and survival of school removal for the 3-class model.

INSERT TABLE 5 HERE

INSERT FIGURE 14 HERE

Comparing the 5-class model in Table 6 to the 3-class model in Table 5, the major differences lie in the covariate and latent class membership influence on the hazards, shown in the latent class growth factor part of the tables. Unlike the 3-class model, the 5-class model of Table 6 shows that none of the covariates nor latent class membership (see the intercept entries) has a significant influence on the hazards. It should be kept in mind here that the non-significant covariate influence concerns the subgroup of non-long-term survivors, estimated as 46% of the sample, and is perhaps due to this group being more homogeneous. Loss of power due to considering a smaller subgroup may also play a role. The non-significant influence of class membership refers to influence of aggression trajectory class among non-long-term survivors. Again, this group is presumably more homogeneous. The non-significant influence of trajectory class membership in the 5-class model suggests that adding the long-term survivor distinction makes aggressive behavior development a less powerful predictor of school removal. Future analysis efforts may focus on attempts to predict long-term survivorship. Figure 15 shows the model-estimated mean aggression trajectories and survival of school removal for the 5-class model. Compared to Figure 14, the inclusion of the long-term survivor class in Figure 15 makes the survival curve for children in the high aggressive trajectory class less different from the survival curves of the other two trajectory classes.

INSERT TABLE 6 HERE

INSERT FIGURE 15 HERE

# 7 Conclusions

This paper 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 more general framework presented here allows for powerful new modeling extensions, two of which were proposed and exemplified. First, unobserved heterogeneity among subjects was captured using multiple latent classes, where each class was allowed to have its own survival function. A particularly interesting example of this is the situation where a subgroup of individuals are never at risk for experiencing the event. Second, the general modeling framework made 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. The extensions show the usefulness of integrating the survival analysis in the broader framework. The analyses are readily available in that the estimation of the general framework may be carried out in the existing computer program Mplus (Muthén & Muthén, 1998-2001). Many further extensions are of interest in this framework, including modeling of repeatable events and competing risks. It is hoped that this paper will stimulate further innovative applications beyond the new analysis possibilities presented here.

# APPENDIX

Consider the observed variables $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{u}$, where $\mathbf{x}$ denotes a $q \times 1$ vector of covariates, $\mathbf{y}$ denotes a $p \times 1$ vector of continuous outcome variables, and $\mathbf{u}$ denotes an $r \times 1$ vector of binary and ordered polytomous categorical outcome variables. Consider latent variables $\boldsymbol{\eta}$ denoting an $m \times 1$ vector of continuous variables and $\mathbf{c}$ denoting a latent categorical variable with $K$ classes, $\mathbf{c}_i = (c_{i1}, c_{i2}, \ldots, c_{iK})'$, where $c_{ik} = 1$ if individual $i$ belongs to class $k$ and zero otherwise. The model has three parts: $\mathbf{c}$ related to $\mathbf{x}$; $\mathbf{u}$ related to $\mathbf{c}$ and $\mathbf{x}$; and $\mathbf{y}$ related to $\mathbf{c}$ and $\mathbf{x}$. The following summary draws on Muthén and Muthén (2001, Appendix 8),

The model relates $\mathbf{c}$ to $\mathbf{x}$ by multinomial logistic regression using the $K-1$-dimensional parameter vector of logit intercepts $\boldsymbol{\alpha}_c$ and the $(K-1) \times q$ parameter matrix of logit slopes $\boldsymbol{\Gamma}_c$, where for $k = 1, 2, \ldots, K$

$$P(c_{ik} = 1|\mathbf{x}_i) = \frac{e^{\alpha_{c_k} + \boldsymbol{\gamma}'_{c_k}\mathbf{x}_i}}{\sum_{k=1}^{K} e^{\alpha_{c_k} + \boldsymbol{\gamma}'_{c_k}\mathbf{x}_i}}, \tag{36}$$

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

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

$$P(u_{i1}, u_{i2}, \ldots, u_{ir}|\mathbf{c}_i, \mathbf{x}_i) = P(u_{i1}|\mathbf{c}_i, \mathbf{x}_i)\; P(u_{i2}|\mathbf{c}_i, \mathbf{x}_i) \ldots P(u_{ir}|\mathbf{c}_i, \mathbf{x}_i). \tag{37}$$

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 $\tau_{j,k,0} = -\infty$, $\tau_{j,k,S_j} = \infty$,

$$u_{ij} = s, \; if \quad \tau_{j,k,s} < u^*_{ij} \leq \tau_{j,k,s+1}, \tag{38}$$

$$P(u_{ij} = s | \mathbf{c}_i, \mathbf{x}_i) = F_{s+1}(u^*_{ij}) - F_s(u^*_{ij}), \tag{39}$$

$$F_s(u^*) = \frac{1}{1 + e^{-(\tau_s - u^*)}}, \tag{40}$$

where for $\mathbf{u}^*_i = (u^*_{i1}, u^*_{i2}, \ldots, u^*_{ir})'$, $\boldsymbol{\eta}_{ui} = (\eta_{u_{1i}}, \eta_{u_{2i}}, \ldots, \eta_{u_{fi}})'$, and conditional on class $k$,

$$\mathbf{u}^*_i = \boldsymbol{\Lambda}_{u_k} \, \boldsymbol{\eta}_{ui} + \mathbf{K}_{u_k} \, \mathbf{x}_i, \tag{41}$$

$$\boldsymbol{\eta}_{ui} = \boldsymbol{\alpha}_{u_k} + \boldsymbol{\Gamma}_{u_k} \, \mathbf{x}_i, \tag{42}$$

where $\boldsymbol{\Lambda}_{u_k}$ 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, $\boldsymbol{\alpha}_{u_k}$ is an $f \times 1$ vector logit parameter vector varying across the $K$ classes, and $\boldsymbol{\Gamma}_{u_k}$ 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 $\boldsymbol{\tau}_k$ varying across the $K$ classes.

It should be noted that (41) does not include intercept terms given the presence of $\tau$ parameters. Furthermore, $\tau$ parameters have opposite signs than $u^*$ in (41) 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 (39) leads to

$$P(u_{ij} = 1 | \mathbf{c}_i, \mathbf{x}_i) = 1 - \frac{1}{1 + e^{-(\tau - u^*)}}, \tag{43}$$

$$= \frac{1}{1 + e^{-\,logit}}, \tag{44}$$

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

The model structure in (41) and (42) is useful when the $\mathbf{u}$ vector represents repeated measures and the latent classes correspond to different trajectory classes. In this case, the elements of $\boldsymbol{\eta}_u$ correspond to growth factors in random effects growth modeling, except that $\boldsymbol{\eta}_u$ has zero variance conditional on $\mathbf{x}$. The parameterization of this type of growth model is shown in the section Latent Class Growth Analysis below.

Consider next the $\mathbf{y}$ part of the model. Multivariate normality is assumed for $\mathbf{y}$ conditional on $\mathbf{x}$ and class $k$,

$$\mathbf{y}_i = \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_k\, \boldsymbol{\eta}_i + \mathbf{K}_k\, \mathbf{x}_i + \boldsymbol{\epsilon}_i, \tag{45}$$

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha}_k + \mathbf{B}_k\, \boldsymbol{\eta}_i + \boldsymbol{\Gamma}_k\, \mathbf{x}_i + \boldsymbol{\zeta}_i, \tag{46}$$

where the residual vector $\boldsymbol{\epsilon}_i$ is $N(\mathbf{0}, \boldsymbol{\Theta}_k)$ and the residual vector $\boldsymbol{\zeta}_i$ is $N(\mathbf{0}, \boldsymbol{\Psi}_k)$, 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 $\mathbf{u}$ and $\mathbf{y}$ are handled using the MAR assumption (Little & Rubin, 1987). 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 McLachlan & Basford, 1988; Hosmer, 1973). 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; Schwartz, 1978; Kass & Raftery, 1993),

$$BIC = -2\ logL + r\ ln\ n, \tag{47}$$

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

When the model contains only $\mathbf{u}$, Pearson and likelihood ratio chi-square tests against the unrestricted multinomial alternative can be computed,

$$\chi^2_P = \sum_{cells} \frac{(o_i - e_i)^2}{e_i}, \tag{48}$$

$$\chi^2_L = 2 \sum_{cells} o_i\ log\ o_i/e_i, \tag{49}$$

where $o_i$ is the observed frequency in cell $i$ of the multivariate frequency table for $\mathbf{u}$ and $e_i$ is the corresponding frequency estimated under the model. With missing data on $\mathbf{u}$, the EM algorithm described in Little and Rubin (1987; 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: Sage Publications.

Allison, P.D. (1995), Survival Analysis Using the SAS System: A Practical Guide, Cary, NC: SAS Publications.

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.

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

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

Huang, G.H. & Karen, B.R (2001). Latent variable regression with covariate effects on underlying and measured variables: An approach of analyzing multiple polytomous surrogates. Manuscript under review.

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.

Maller, R.A. & Zhou, X. (1996). Survival Analysis with Long-Term Survivors. New York: Wiley.

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 & Sons.

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 Collins, L.M. & Sayer, A. (Eds.), New Methods for the Analysis of Change (pp. 291-322). Washington, D.C.: 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). Lawrence Erlbaum Associates.

Muthén, B., Brown, C.H., Masyn, K., Jo, B., Khoo, S.T., Yang, C.C., Wang, C.P., Kellam, S., Carlin, J., & Liao, J. (2000). General growth mixture modeling for randomized preventive interventions. Under review.

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

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

Muthén, B., Shedden, K., & Spisic. D. (2000). General latent variable mixture modeling.

Technical report.

Nagin, D.S. (1999). Analyzing developmental trajectories: a semi-parametric, group-based approach. Psychological Methods, 4, 139-157.

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.

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 sterilisation in Bangladesh. Paper presented at the Population Association of America Annual Meetings, March 2000, Los Angeles.

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

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.

Table 1

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

| Variable name | Description | Mean | 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 |
| | | Mean | |
| $emp_1$ | First month employment indicator | 0.40 | |
| $emp_2$ | Second month employment indicator | 0.52 | |
| $emp_3$ | Third month employment indicator | 0.53 | |
| $emp_4$ | Fourth month employment indicator | 0.54 | |
| $emp_5$ | Fifth month employment indicator | 0.55 | |
| $emp_6$ | Sixth month employment indicator | 0.55 | |
| $emp_7$ | Seventh month employment indicator | 0.57 | |
| $emp_8$ | Eighth month employment indicator | 0.56 | |
| $emp_9$ | Ninth month employment indicator | 0.55 | |
| $emp_{10}$ | Tenth month employment indicator | 0.55 | |
| $emp_{11}$ | Eleventh month employment indicator | 0.57 | |
| $emp_{12}$ | Twelfth month employment indicator | 0.56 | |
| $emp_{13}$ | Thirteenth 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 | |
| $u_1$ | First month re-arrest indicator | $^4/_{432}$ = 0.01 | |
| $u_2$ | Second month re-arrest indicator | $^8/_{428}$ = 0.02 | |
| $u_3$ | Third month re-arrest indicator | $^7/_{420}$ = 0.02 | |
| $u_4$ | Fourth month re-arrest indicator | $^8/_{413}$ = 0.02 | |
| $u_5$ | Fifth month re-arrest indicator | $^{13}/_{405}$ = 0.03 | |
| $u_6$ | Sixth month re-arrest indicator | $^8/_{392}$ = 0.02 | |
| $u_7$ | Seventh month re-arrest indicator | $^{10}/_{384}$ = 0.03 | |
| $u_8$ | Eighth month re-arrest indicator | $^5/_{374}$ = 0.01 | |
| $u_9$ | Ninth month re-arrest indicator | $^{11}/_{369}$ = 0.03 | |
| $u_{10}$ | Tenth month re-arrest indicator | $^{11}/_{358}$ = 0.03 | |
| $u_{11}$ | Eleventh month re-arrest indicator | $^8/_{347}$ = 0.02 | |
| $u_{12}$ | Twelfth month re-arrest indicator | $^9/_{339}$ = 0.03 | |
| $u_{13}$ | Thirteenth month re-arrest indicator | $^{12}/_{330}$ = 0.03 | |

Table 2

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

| Variable name | Description | Mean | 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 grave fall classroom average TOCA-R | 1.92 | 0.40 |
| | | Mean | |
| 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 | |

Table 3

1-class Survival Model with Constant Hazard and Proportional Odds Assumptions

| Thresholds ($\tau$) | Est. | SE | t |
|---|---|---|---|
| $u_1 - u_{13}$ | 1.80 | 0.82 | -2.20 |
| Latent Class Growth Factor ($\eta_u$) | 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. Bold-face entries indicate significance at the 5% level.

Table 4

2-class Survival Model with Non-long-term Survivor Class Parameter Estimates

| Thresholds ($\tau$) | Est. | SE | t |
|---|---|---|---|
| $u_1$ | 3.21 | 1.11 | 2.90 |
| $u_2$ | 2.33 | 1.07 | 2.19 |
| $u_3$ | 2.41 | 1.07 | 2.26 |
| $u_4$ | 2.22 | 1.10 | 2.01 |
| $u_5$ | 1.66 | 1.09 | 1.52 |
| $u_6$ | 2.09 | 1.12 | 1.87 |
| $u_7$ | 1.82 | 1.12 | 1.63 |
| $u_8$ | 2.47 | 1.11 | 2.23 |
| $u_9$ | 1.61 | 1.14 | 1.42 |
| $u_{10}$ | 1.54 | 1.13 | 1.37 |
| $u_{11}$ | 1.76 | 1.17 | 1.50 |
| $u_{12}$ | 1.60 | 1.16 | 1.39 |
| $u_{13}$ | 1.21 | 1.16 | 1.05 |
| Latent Class Growth Factor ($\eta_u$) | Est. | SE | t |
| finaid | **-0.85** | 0.27 | -3.21 |
| black | -0.05 | 0.35 | -0.14 |
| workexp | -0.56 | 0.33 | -1.70 |
| married | -0.73 | 0.40 | -1.83 |
| paroled | 0.35 | 0.30 | 1.16 |
| age | **-0.07** | 0.03 | -2.67 |
| priors | 0.04 | 0.04 | 1.20 |
| educ | 0.34 | 0.22 | 1.52 |
| Event Indicator Regression | Est. | SE | t |
| $emp_1$ - $emp_{13}$ | **-1.15** | 0.23 | -4.96 |
| Latent Class Regression* | Est. | SE | t |
| finaid | -2.15 | 1.36 | -1.58 |
| black | -1.62 | 1.14 | -1.42 |
| workexp | **-1.83** | 0.90 | -2.04 |
| married | -2.88 | 1.50 | -1.92 |
| paroled | 1.89 | 1.29 | 1.47 |
| age | **-0.15** | 0.07 | -2.20 |
| priors | -0.24 | 0.13 | -1.81 |
| educ | **1.79** | 0.65 | 2.74 |

*Class 1 is long-term survivors, 21%, Class 2 is non-long-term survivors, 79%

Log likelihood = -496.23, BIC = 1180.58, 31 free parameters. Bold-face entries indicate significance at the 5% level.

Table 5

3-class Growth + Survival Mixture Model (with no LTS class) Parameter Estimates

| | High Class | | Medium Class | | Low Class | |
|---|---|---|---|---|---|---|
| Intercept Factor ($\eta_0$) | 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 ($\eta_1$) | 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 ($\tau$) | 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 ($\eta_u$) | 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. Bold-face entries indicate significance at the 5% level.

Table 6

5-class Growth + Survival Mixture Model (with LTS class) Parameter Estimates

| | High Non-LTS | | Medium Non-LTS | | Medium LTS | | Low Non-LTS | | Low LTS | |
|---|---|---|---|---|---|---|---|---|---|---|
| Intercept Factor ($\eta_0$) | Est. | SE | Est. | SE | Est. | SE | Est. | SE | Est. | SE |
| intercept | 3.86 | 0.53 | 2.20 | 0.28 | | | 1.31 | 0.13 | | |
| external | **-0.27** | 0.05 | | | | | | | | |
| male | 0.05 | 0.05 | | | | | | | | |
| white | -0.11 | 0.06 | | | | | | | | |
| lunch | -0.01 | 0.05 | | | | | | | | |
| cavlunch | **-0.22** | 0.06 | | | | | | | | |
| cavtoca1f | 0.15 | 0.08 | | | | | | | | |
| Slope Factor ($\eta_1$) | Est. | SE | Est. | SE | Est. | SE | Est. | SE | Est. | SE |
| intercept | 0.26 | 0.30 | 0.32 | 0.30 | | | 0.28 | 0.23 | | |
| external | -0.01 | 0.08 | | | | | | | | |
| male | 0.02 | 0.04 | | | | | | | | |
| white | 0.03 | 0.08 | | | | | | | | |
| lunch | -0.03 | 0.03 | | | | | | | | |
| cavlunch | **0.19** | 0.08 | | | | | | | | |
| cavtoca1f | -0.19 | 0.13 | | | | | | | | |
| Thresholds ($\tau$) | Est. | SE | Est. | SE | Est. | SE | Est. | SE | Est. | SE |
| $u_3$ | 5.57 | 0.96 | | | 10.0 | fixed | 5.57 | 0.96 | 10.0 | fixed |
| $u_4$ | 5.30 | 0.86 | | | 10.0 | fixed | 5.30 | 0.86 | 10.0 | fixed |
| $u_5$ | 4.49 | 0.76 | | | 10.0 | fixed | 4.49 | 0.76 | 10.0 | fixed |
| $u_6$ | 3.62 | 0.72 | | | 10.0 | fixed | 3.62 | 0.72 | 10.0 | fixed |
| $u_7$ | 1.54 | 0.71 | | | 10.0 | fixed | 1.54 | 0.71 | 10.0 | fixed |
| Latent Class Growth Factor ($\eta_u$) | | | | | | | | | | |
| intercept | 1.63 | 0.87 | 0.78 | 0.95 | 0.0 | fixed | 0.0 | fixed | 0.0 | fixed |
| external | 0.20 | 0.30 | | | 0.0 | fixed | 0.20 | 0.30 | 0.0 | fixed |
| male | 1.13 | 0.98 | | | 0.0 | fixed | 1.13 | 0.98 | 0.0 | fixed |
| white | 1.18 | 0.62 | | | 0.0 | fixed | 1.18 | 0.62 | 0.0 | fixed |
| lunch | -0.93 | 0.62 | | | 0.0 | fixed | -0.93 | 0.62 | 0.0 | fixed |
| cavlunch | 0.56 | 0.87 | | | 0.0 | fixed | 0.56 | 0.87 | 0.0 | fixed |
| cavtoca1f | 0.41 | 0.56 | | | 0.0 | fixed | 0.41 | 0.56 | 0.0 | fixed |
| Latent Class Regression | | | | | | | | | | |
| intercept | -10.2 | 2.16 | -0.61 | 1.97 | -5.76 | 3.57 | 1.32 | 2.50 | | |
| external | **1.78** | 0.62 | 0.70 | 0.54 | **1.33** | 0.62 | 0.69 | 0.73 | | |
| male | 2.35 | 2.46 | 0.51 | 1.09 | 0.66 | 0.69 | -0.16 | 1.13 | | |
| white | -0.60 | 1.94 | **-1.24** | 0.60 | -0.20 | 0.94 | -1.27 | 0.99 | | |
| lunch | 0.84 | 1.37 | 1.14 | 0.66 | 0.51 | 0.73 | 0.04 | 0.95 | | |
| cavlunch | 3.28 | 3.57 | 1.19 | 1.88 | 0.31 | 1.89 | 3.87 | 2.06 | | |
| cavtoca1f | **2.24** | 1.07 | -0.39 | 0.94 | 2.28 | 1.66 | -1.98 | 1.18 | | |
| Class Proportions | 0.07 | | 0.21 | | 0.28 | | 0.18 | | 0.26 | |

Log likelihood = -1419.01, BIC = 3275.94, 73 free parameters. Bold-face entries indicate significance at the 5% level.

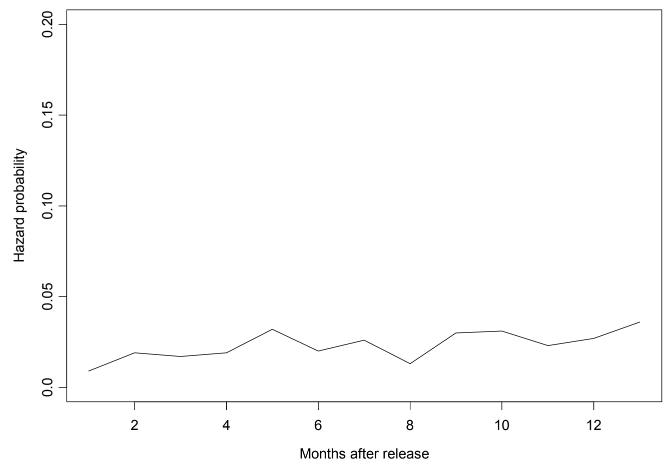


Figure 1. Sample-estimated hazard probabilities of re-arrest

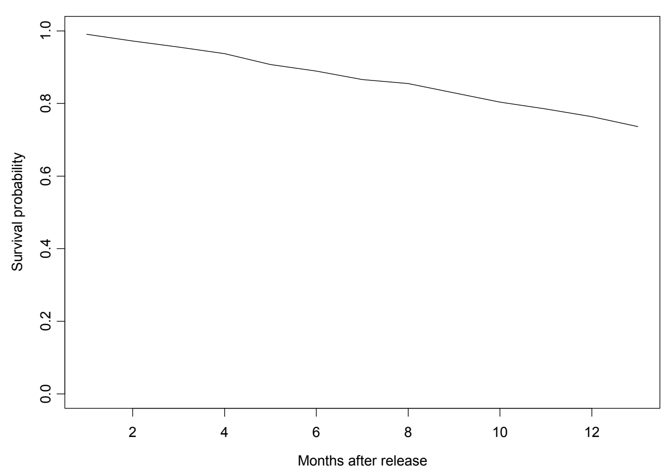


Figure 2. Sample-estimated survival probabilities of re-arrest

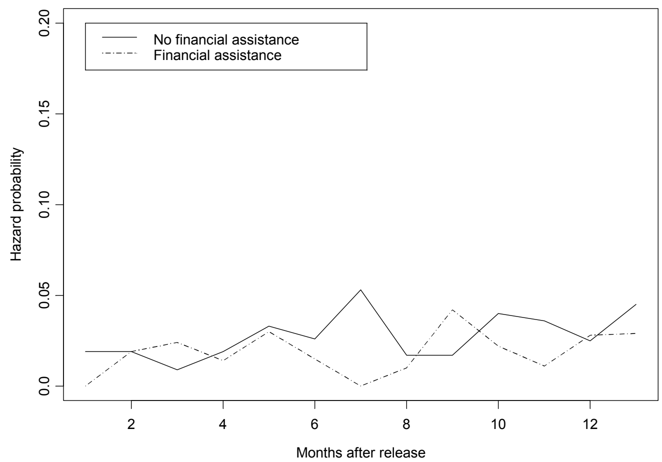


Figure 3. Sample-estimated hazard probabilities of re-arrest by intervention status

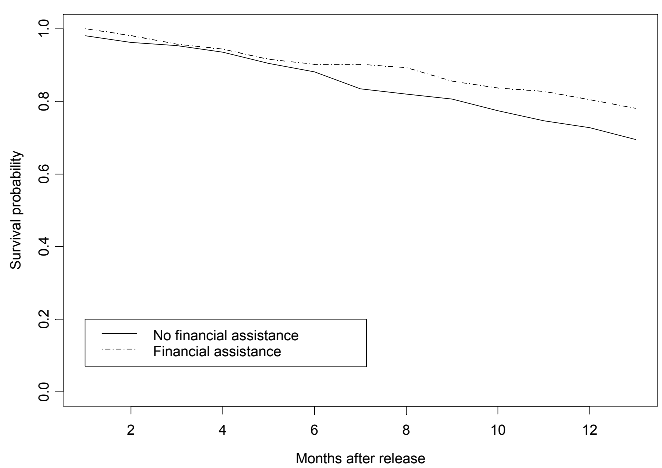


Figure 4. Sample-estimated survival probabilities of re-arrest by intervention status

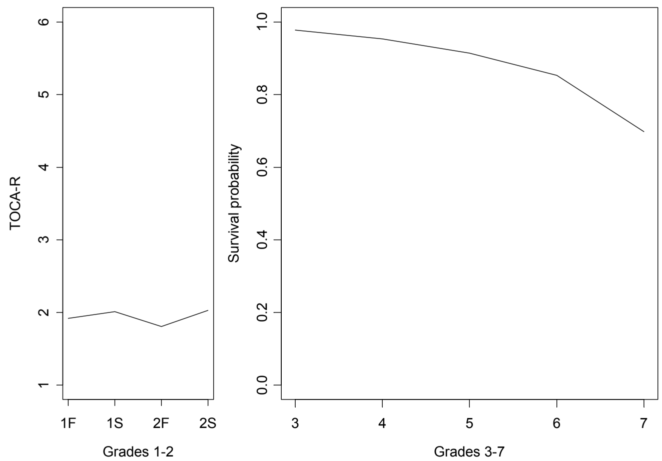


Figure 5. Sample-estimated mean aggression trajectory and survival of school removal

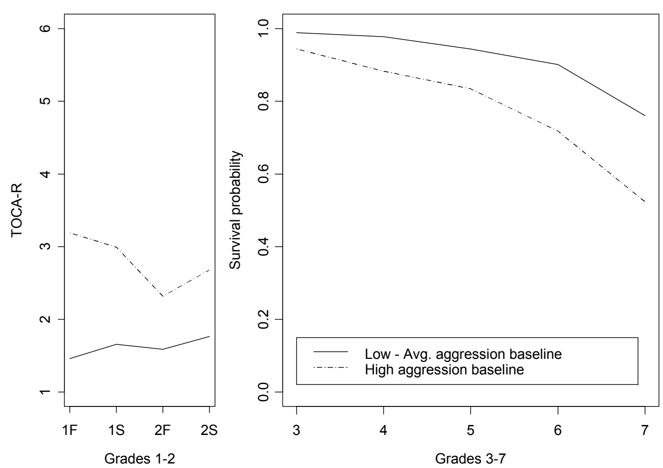


Figure 6. Sample-estimated mean aggression trajectories and survival of school removal by first grade fall baseline aggression measure

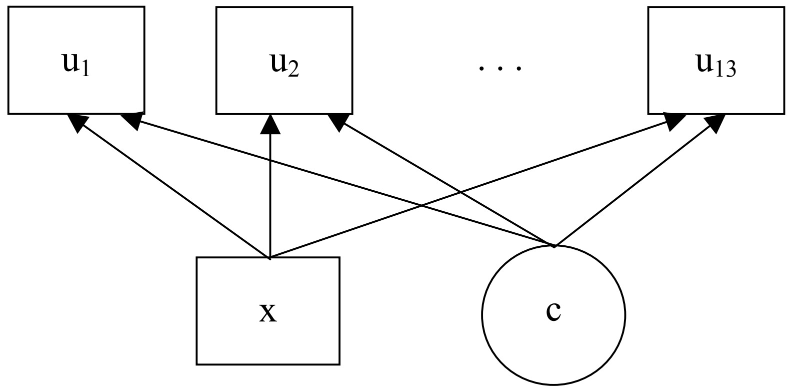


Figure 7. Recidivism path diagram: Survival model with time-varying covariate effects

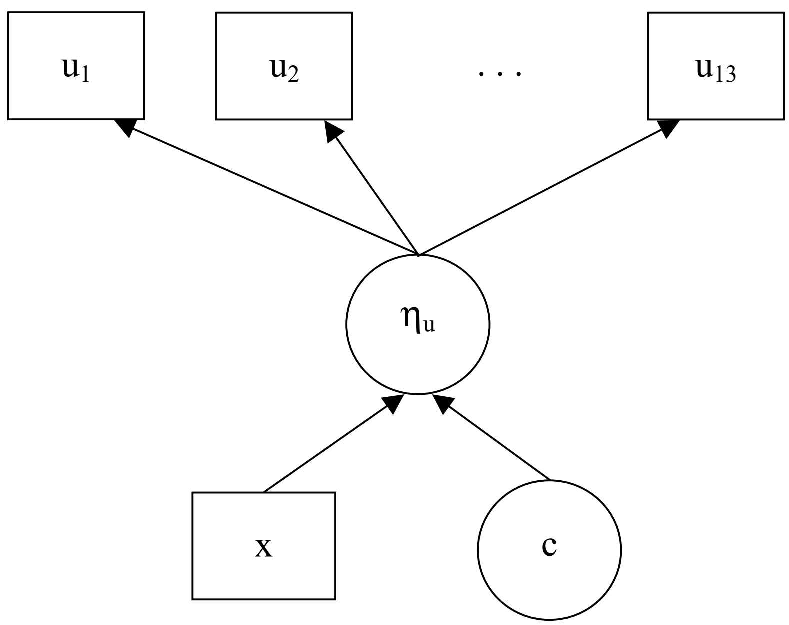


Figure 8. Recidivism path diagram: Survival model with proportional hazard odds assumption applied to time-invariant covariates

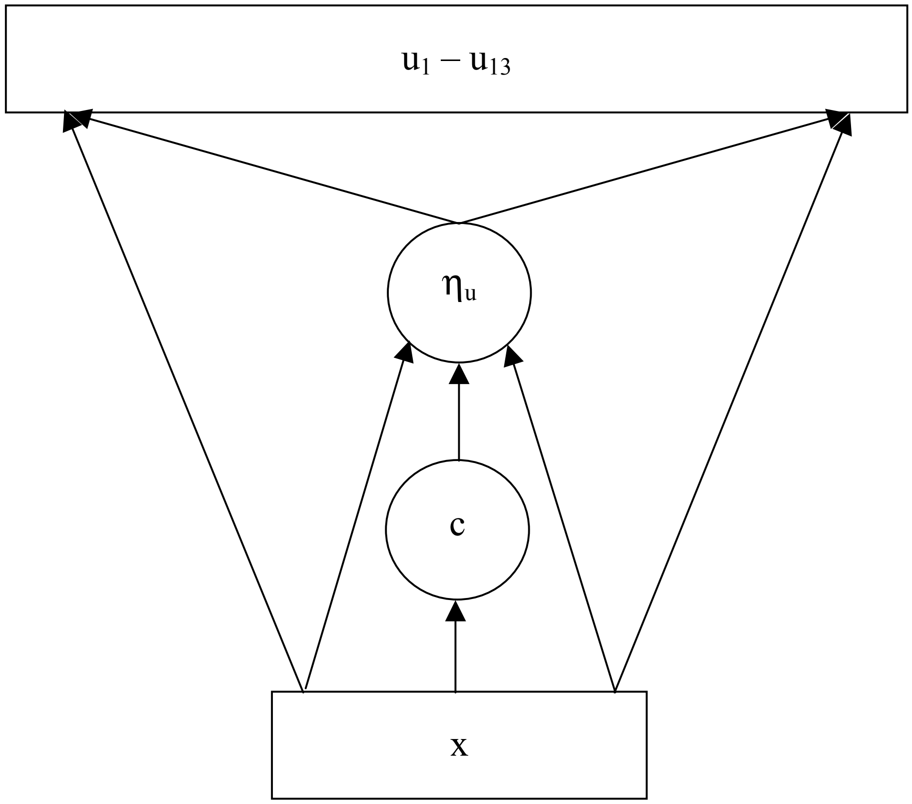


Figure 9. Recidivism path diagram: Survival model with time-varying and time-invariant covariates and long-term survivor class

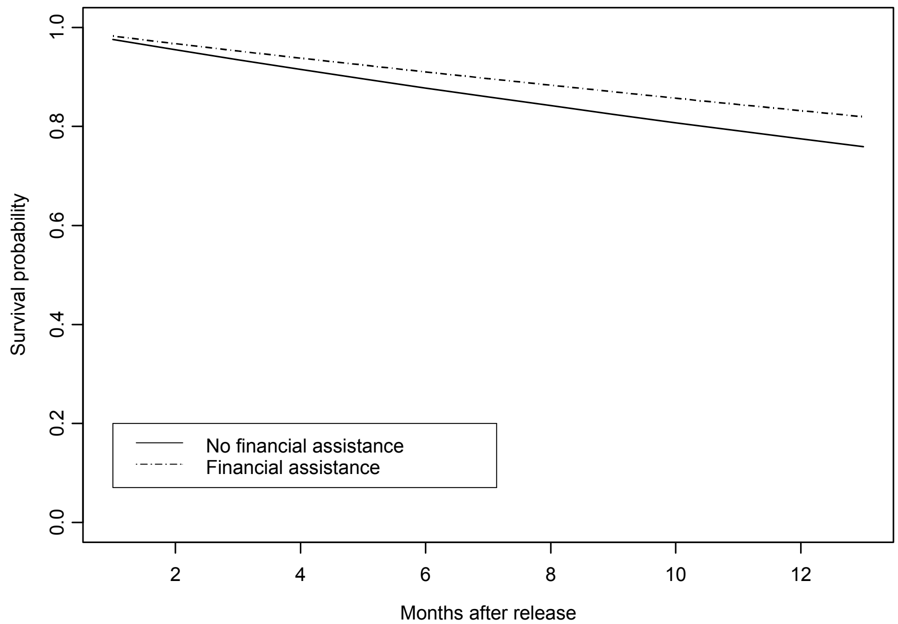


Figure 10. Model-estimated survival probabilities of re-arrest by intervention status for one-class model

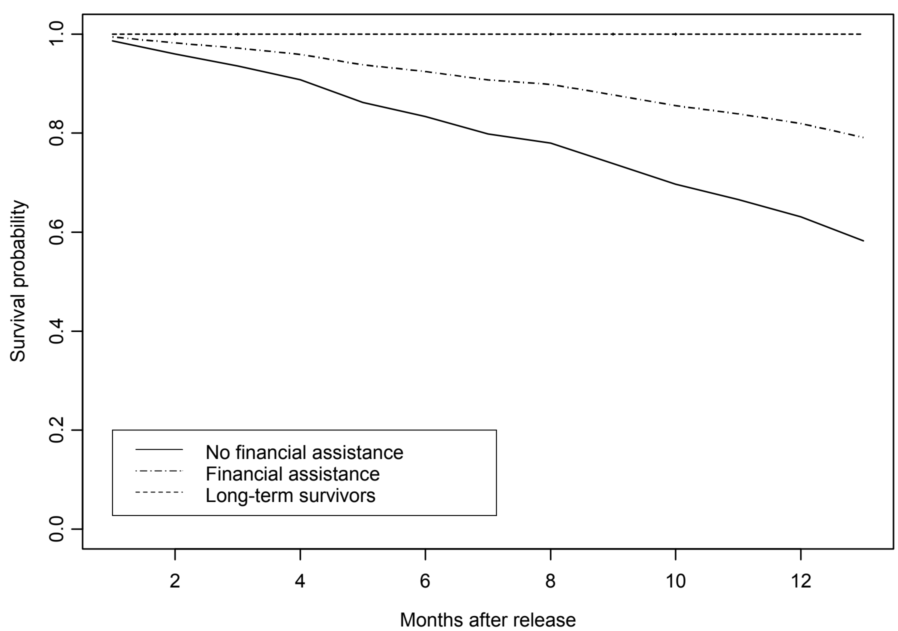


Figure 11. Model-estimated survival probabilities of re-arrest by intervention status for two-class model with long-term survivors

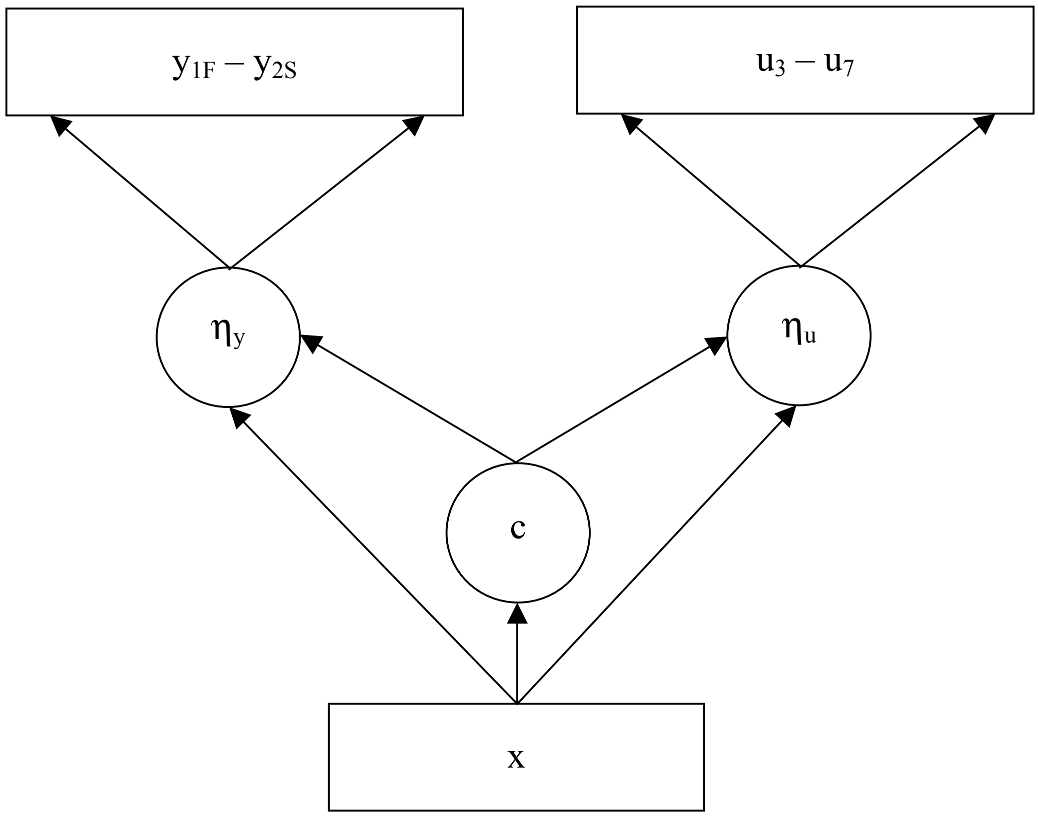


Figure 12. School removal path diagram: Growth + Survival model with time-invariant covariates and single latent class variable

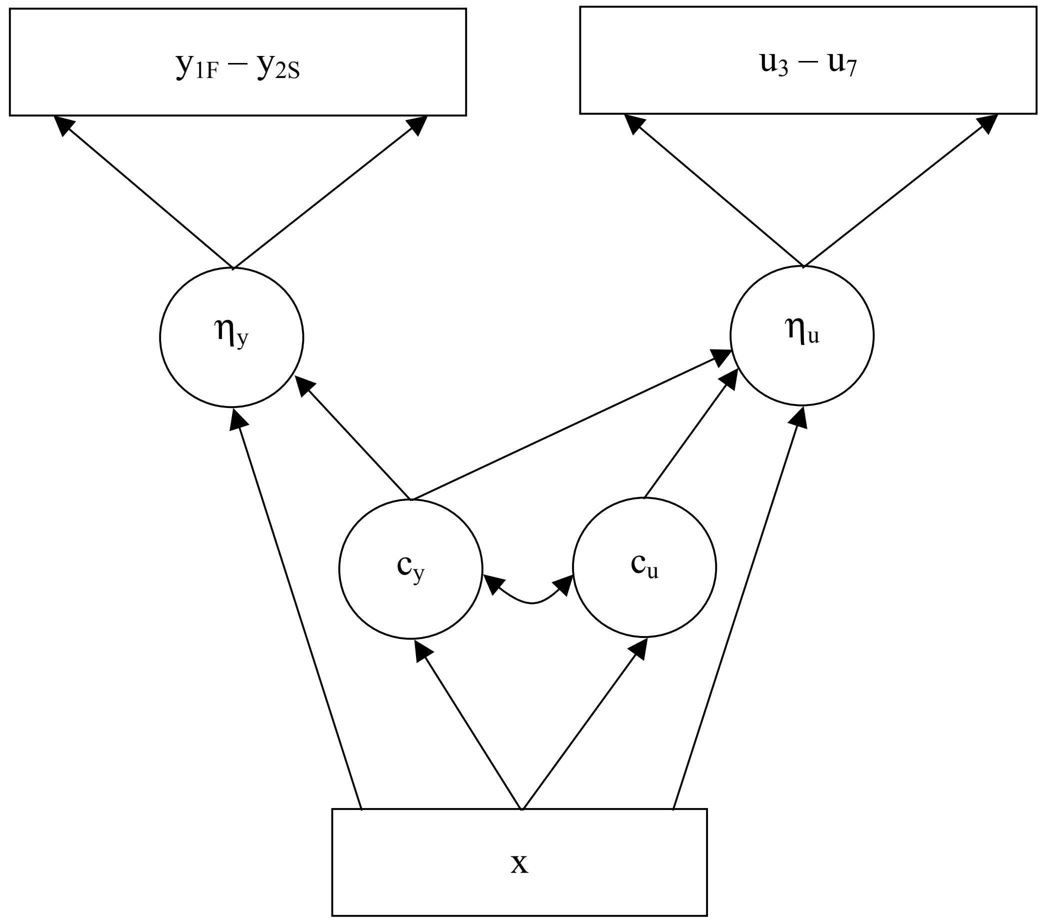


Figure 13. School removal path diagram: Growth + Survival model with time-invariant covariates and both trajectory and survival latent class variables

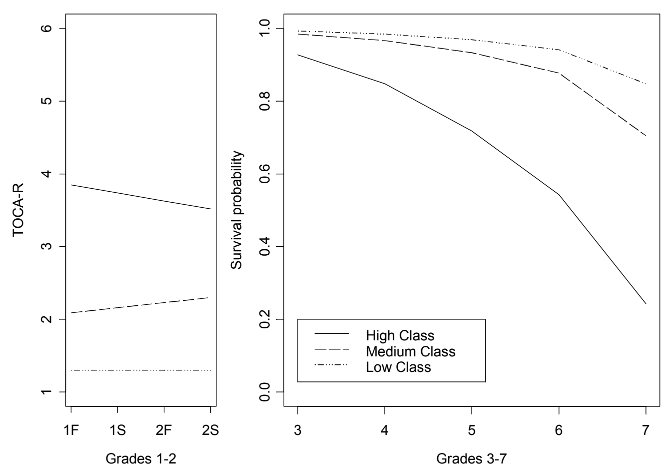


Figure 14. Model-estimated mean aggression trajectory and survival of school removal for 3-class growth + survival mixture model

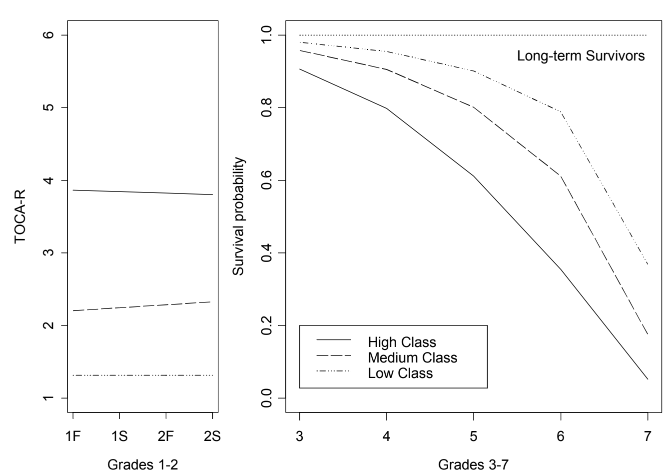


Figure 15. Model-estimated mean aggression trajectory and survival of school removal for 5-class growth + survival mixture model with long-term survivor class