Multilevel Analysis of Ordinal Outcomes Related to Survival Data

Donald Hedeker and Robin J. Mermelstein

University of Illinois at Chicago

Correspondence concerning this article should be addressed to Donald Hedeker, Division of Epidemiology and Biostatistics (M/C 923), School of Public Health, University of Illinois at Chicago, 1603 West Taylor Street, Room 955, Chicago, IL, 60612-4336, USA.
Phone: (312) 996-4896, fax: (312) 996-0064, email: hedeker@uic.edu.

Multilevel Analysis of Ordinal Outcomes Related to Survival Data

Summary

Multilevel modeling is described for analysis of correlated grouped-time survival data. Two analysis approaches are considered. The first treats survival time as an ordinal outcome, which is either right-censored or not. The second approach treats survival time as a set of dichotomous indicators of whether the event occurred for time periods up to the period of the event or censor. For either approach both proportional hazards and proportional odds versions of the multilevel model are presented, as well as partial proportional hazards and odds generalizations. Estimation using standard software (*i.e.*, `SAS PROC NLMIXED`) can be accomplished using a full-information maximum (marginal) likelihood solution. Several examples will be provided, along with software syntax, to illustrate features of multilevel grouped-time survival analysis.

Short Running Title: Multilevel survival model

Keywords: censored observations, proportional hazards model; discrete survival data; frailty; heterogeneity; clustering

1 Introduction

Models for grouped-time survival data are useful for analysis of failure-time data when subjects are measured repeatedly at fixed intervals in terms of the occurrence of some event, or when determination of the exact time of the event is only known within grouped intervals of time. For example, in many school-based studies of substance use, students are typically measured annually regarding their smoking, alcohol, and other substance use during the past year. An important question is then to determine the degree to which covariates are related to substance use initiation. In these studies it is often of interest to model the student outcomes while controlling for the nesting of students in classrooms and/or schools. In analysis of such grouped-time initiation (or survival) data, use of grouped-time regression models that assume independence of observations [Thompson, 1977; Prentice & Gloeckler, 1978; Allison, 1982] is therefore problematic because of this clustering of students. More generally, this same issue arises for other types of clustered datasets in which subjects are observed nested within various types of clusters (*e.g.*, hospitals, firms, clinics, counties), and thus cannot be assumed to be independent. To account for the data clustering, multilevel models (also called hierarchical linear or mixed models) provide a useful approach for simultaneously estimating the parameters of the regression model and the variance components that account for the data clustering [Goldstein, 1995; Raudenbush & Bryk, 2002].

For continuous-time survival data that are clustered, several authors [Vaupel et al., 1979; Lancaster, 1979; Clayton & Cuzick, 1985; Self & Prentice, 1986; G. Guo & Rodriquez, 1992; Paik et al., 1994; Shih & Louis, 1995] have developed mixed-effects survival models. These models are often termed frailty models or survival models including heterogeneity, and review articles describe many of these models [Pickles & Crouchley, 1995; Hougaard, 1995]. An alternative approach for dealing with correlated data uses the generalized estimating equations (GEE) method to estimate model parameters. In this regard, Lee et al. [1992] and Wei et al. [1989] have developed continuous-time survival models.

Application of these continuous-time models to grouped or discrete-time survival data is generally not recommended because of the large number of ties that result. Instead, models

specifically developed for grouped or discrete-time survival data have been proposed. Both Han & Hausman [1990] and Scheike & Jensen [1997] have described proportional hazards models incorporating a log-gamma distribution specification of heterogeneity. Also, Ten Have [1996] developed a discrete-time proportional hazards survival model incorporating a log-gamma random effects distribution, additionally allowing for ordinal survival and failure categories. Ten Have & Uttal [1994] used Gibbs sampling to fit continuation ratio logit models with multiple normally distributed random effects. In terms of a GEE approach, S. W. Guo & Lin [1994] have developed a multivariate model for grouped-time survival data.

Several authors have noted the relationship between ordinal regression models (using complementary log-log and logit link functions) and survival analysis models for grouped and discrete time [Han & Hausman, 1990; McCullagh, 1980; Teachman et al., 1994]. Similarly, others [Allison, 1982; D'Agostino et al., 1990; Singer & Willett, 1993] have described how dichotomous regression models can be used to model grouped and discrete time survival data. The ordinal approach simply treats survival time as an ordered outcome that is either right-censored or not. Alternatively, in the dichotomous approach each survival time is represented as a set of indicators of whether or not an individual failed in each time unit (until a person either experiences the event or is censored). As a result, the dichotomous approach is more useful for inclusion of time-dependent covariates and relaxing of the proportional hazards assumption.

Several authors have generalized these fixed-effects regression models for categorical responses to the multilevel setting [Ten Have & Uttal, 1994; Ten Have, 1996; Scheike & Jensen, 1997; Barber et al., 2000; Hedeker et al., 2000; Rabe-Hesketh et al., 2001; Reardon et al., 2002; Muthén & Masyn, 2005; Grilli, 2005]. The resulting models are generally based on dichotomous and ordinal mixed-effects regression models, albeit with the extension of the ordinal model to allow for right-censoring of the response. Typically, these models allow multiple random effects and a general form for model covariates. In many cases, proportional or partial proportional hazards or odds models are considered. In this chapter, this class of two-level models will be described where the random effects are assumed to be normally distributed. Because we assume the normal distribution for the random effects, standard

software (*e.g.*, SAS PROC NLMIXED) can be used to estimate these models, and therefore broaden the potential application of this approach. Syntax examples will be provided to facilitate this.

2 Multilevel Grouped-Time Survival Analysis Model

Let i denote the level-2 units ($i = 1, \ldots, N$) and let j denote the level-1 units ($j = 1, \ldots, n_i$). Note that this use of i for level-2 units and j for level-1 units is consistent with usage often found in statistics [Verbeke & Molenberghs, 2000] and biostatistics [Diggle et al., 2002], but is opposite from the typical usage in the multilevel literature [Goldstein, 1995; de Leeuw & Meijer, 2008]. If subjects are nested within clusters, the subjects and clusters represent the level-1 and level-2 units, respectively. Alternatively, if there are multiple failure times per subject, then the level-2 units are the subjects and the level-1 units are the repeated failure times. Suppose that there is a continuous random variable for the uncensored time of event occurrence (which may not be observed), however assume that time (of assessment) can take on only discrete positive values $t = 1, 2, \ldots, m$. For each level-1 unit, observation continues until time t_{ij} at which point either an event occurs or the observation is censored, where censoring indicates being observed at t_{ij} but not at $t_{ij} + 1$. Define P_{ijt} to be the probability of failure, up to and including time interval t, that is,

$$P_{ijt} = \Pr\ [t_{ij} \leq t] \tag{1}$$

and so the probability of survival beyond time interval t is simply $1 - P_{ijt}$.

Because $1-P_{ijt}$ represents the survivor function, McCullagh [1980] proposed the following grouped-time version of the continuous-time proportional hazards model:

$$\log[-\log(1 - P_{ijt})] = \alpha_{0t} + \boldsymbol{x}'_{ij}\boldsymbol{\beta}\ . \tag{2}$$

This is the so-called complementary log-log function, which can be re-expressed in terms of the cumulative failure probability, $P_{ijt} = 1 - \exp(-\exp(\alpha_{0t} + \boldsymbol{x}'_{ij}\boldsymbol{\beta}))$. In this model, $\boldsymbol{x}_{ij}$ is a $p \times 1$ vector including covariates that vary either at level 1 or 2, however they do not vary with time (*i.e.*, they do not vary across the ordered response categories). They may,

however, represent the average of a variable across time or the value of the covariate at the time of the event.

Since the integrated hazard function equals $-\log(1 - P_{ijt})$, this model represents the covariate effects ($\boldsymbol{\beta}$) on the log of the integrated hazard function. The covariate effects are identical to those in the grouped-time version of the proportional hazards model described by Prentice & Gloeckler [1978]. As such, the $\boldsymbol{\beta}$ coefficients are also identical to the coefficients in the underlying continuous-time proportional hazards model. Furthermore, as noted by Allison [1982], the regression coefficients of the model are invariant to interval length. Augmenting the coefficients $\boldsymbol{\beta}$, the intercept terms α_{0t} are a set of m constants that represent the logarithm of the integrated baseline hazard (*i.e.*, when $\boldsymbol{x} = \mathbf{0}$). As such, these terms represent cutpoints on the integrated baseline hazard function; these parameters are often referred to as threshold parameters in descriptions of ordinal regression models. While the above model is the same as that described in McCullagh [1980], it is written so that the covariate effects are of the same sign as the Cox proportional hazards model. A positive coefficient for a regressor then reflects increasing hazard with greater values of the regressor.

Adding random effects to this model, we get

$$\log[-\log(1 - P_{ijt})] = \alpha_{0t} + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{w}'_{ij}\boldsymbol{v}_i \ , \tag{3}$$

or

$$P_{ijt} = 1 - \exp(-\exp(\alpha_{0t} + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{w}'_{ij}\boldsymbol{v}_i)) = 1 - \exp(-\exp z_{ijt}) \ , \tag{4}$$

where $\boldsymbol{v}_i$ is the $r \times 1$ vector of unknown random effects for the level-2 unit i, and $\boldsymbol{w}_{ij}$ is the design vector for the r random effects. The distribution of the r random effects $\boldsymbol{v}_i$ is assumed to be multivariate with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_v$. An important special case is when the distribution is assumed to be multivariate normal. For convenience, the random effects are often expressed in standardized form. Specifically, let $\boldsymbol{v} = \boldsymbol{S\theta}$, where $\boldsymbol{SS}' = \boldsymbol{\Sigma}_v$ is the Cholesky decomposition of $\boldsymbol{\Sigma}_v$. The model for z_{ijt} then is written as:

$$z_{ijt} = \alpha_{0t} + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{w}'_{ij}\boldsymbol{S\theta}_i \ . \tag{5}$$

As a result of the transformation, the Cholesky factor $\boldsymbol{S}$ is estimated instead of the covariance matrix $\boldsymbol{\Sigma}_v$. As the Cholesky factor is essentially the square-root of the covariance matrix, this allows more stable estimation of near-zero variance terms.

2.1 Proportional Odds Model

As applied to continuous-time (cross-sectional) survival data, the proportional odds model is described by Bennett [1983]. For grouped-time, the multilevel proportional odds model is written in terms of the logit link function as

$$\log[P_{ijt}/(1 - P_{ijt})] = z_{ijt} \tag{6}$$

or alternatively as $P_{ijt} = 1/[1 + \exp(-z_{ijt})]$. The choice of which link function to use is not always clear-cut. Bennett [1983] noted that the proportional odds model is useful for survival data when the hazards of groups of subjects are thought to converge with time. This contrasts to the proportional hazards model where the hazard rates for separate groups of subjects are assumed proportional at all timepoints. However, this type of non-proportional hazards effect can often be accommodated in the complementary log-log link model by including interactions of covariates with the baseline hazard cutpoints [Collett, 1994]. Also as Doksum & Gasko [1990] note, large amounts of high quality data are often necessary for link function selection to be relevant. Since these two link functions often provide similar fits, Ten Have [1996] suggests that the choice of which to use depends upon whether inference should be in terms of odds ratios or discrete hazard ratios. Similarly, McCullagh [1980] notes that link function choice should be based primarily on ease of interpretation.

2.2 Pooling of Repeated Observations and Non-proportional Hazards

Thus far, survival time has been represented as an ordered outcome t_{ij} that is designated as censored or not. An alternative approach for grouped-time survival data, described by several authors [Allison, 1982; D'Agostino et al., 1990; Singer & Willett, 1993] and others,

treats each individual's survival time as a set of dichotomous observations indicating whether or not an individual failed in each time unit until a person either experiences the event or is censored. Thus, each survival time is represented as a $t_{ij} \times 1$ vector of zeros for censored individuals, while for individuals experiencing the event the last element of this $t_{ij} \times 1$ vector of zeros is changed to a one. These multiple person-time indicators are then treated as distinct observations in a dichotomous regression model. In the case of clustered data, a multilevel dichotomous regression model is used. This method has been called the pooling of repeated observations method by Cupples et al. [1985] and is extensively described in Singer & Willett [2003]. For multilevel models, Reardon et al. [2002] provide a useful illustration of this approach. The dichotomous treatment is particularly useful for handling time-dependent covariates and fitting non-proportional hazards models because the covariate values can change across each individual's t_{ij} timepoints.

For the dichotomous approach, define p_{ijt} to be the probability of failure in time interval t, conditional on survival prior to t:

$$p_{ijt} = \Pr\left[t_{ij} = t \mid t_{ij} \geq t\right] \tag{7}$$

Similarly, $1 - p_{ijt}$ is the probability of survival beyond time interval t, conditional on survival prior to t. The proportional hazards model is then written as

$$\log[-\log(1 - p_{ijt})] = \alpha_{0t} + \boldsymbol{x}'_{ijt}\boldsymbol{\beta} + \boldsymbol{w}'_{ij}\boldsymbol{S}\boldsymbol{\theta}_i \ , \tag{8}$$

and the corresponding proportional odds model is

$$\log[p_{ijt}/(1 - p_{ijt})] = \alpha_{0t} + \boldsymbol{x}'_{ijt}\boldsymbol{\beta} + \boldsymbol{w}'_{ij}\boldsymbol{S}\boldsymbol{\theta}_i \ , \tag{9}$$

where now the covariates $\boldsymbol{x}$ can vary across time and so are denoted as $\boldsymbol{x}_{ijt}$. Augmenting the model intercept α_{01}, the remaining intercept terms α_{0t} $(t = 2, \ldots m)$ are obtained by including as regressors $m - 1$ dummy codes representing deviations from the first timepoint. Because the covariate vector $\boldsymbol{x}$ now varies with t, this approach automatically allows for time-dependent covariates, and relaxing the proportional hazards assumption only involves including interactions of covariates with the $m - 1$ timepoint dummy codes.

Under the complementary log-log link function, the two approaches characterized by (3) and (8) yield identical results for the parameters that do not depend on t, namely the regression coefficients of time-independent covariates and the Cholesky factor [Engel, 1993; Läärä & Matthews, 1985]. For the logit link, similar, but not identical, results are obtained for these parameters. Comparing these two approaches, notice that for the ordinal approach each observation consists of only two pieces of data: the (ordinal) time of the event and whether it was censored or not. Alternatively, in the dichotomous approach each survival time is represented as a vector of dichotomous indicators, where the size of the vector depends upon the timing of the event or censoring. Thus, the ordinal approach can be easier to implement and offers savings in terms of the dataset size, especially as the number of timepoints gets large, while the dichotomous approach is superior in its treatment of time-dependent covariates and relaxing of the proportional hazards or odds assumption.

2.3 Proportional Hazards/Odds Assumption

Relaxing the proportional hazards or odds assumption in the ordinal model is possible; Hedeker & Mermelstein [1998] have described a multilevel partial proportional odds model. For this, the model can be rewritten as:

$$z_{ijt} = \alpha_{0t} + (\boldsymbol{u}^*_{ij})'\boldsymbol{\alpha}^*_t + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{w}'_{ij}\boldsymbol{S}\boldsymbol{\theta}_i \quad , \tag{10}$$

or absorbing α_{0t} and $\boldsymbol{\alpha}^*_t$ into $\boldsymbol{\alpha}_t$,

$$z_{ijt} = \boldsymbol{u}'_{ij}\boldsymbol{\alpha}_t + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{w}'_{ij}\boldsymbol{S}\boldsymbol{\theta}_i \quad , \tag{11}$$

where, $\boldsymbol{u}_{ij}$ is a $(l+1)\times 1$ vector containing (in addition to a 1 for α_{0t}) the values of observation ij on the set of l covariates for which interactions with the cutpoints of the integrated baseline hazard are desired. Here, $\boldsymbol{\alpha}_t$ is a $(l+1)\times 1$ vector of regression coefficients associated with the l variables (and the intercept) in $\boldsymbol{u}_{ij}$. Under the complementary log-log link this provides a partial proportional hazards model, while under the logistic link it would be a partial proportional odds model. Of course, all covariates could be in $\boldsymbol{u}$, and none in $\boldsymbol{x}$, to provide purely non-proportional hazards/odds models.

Tests of the proportional hazards or odds assumption for a set of covariates can then be performed by running and comparing models: (a) requiring proportional odds/hazards (*i.e.*, covariates are in $\boldsymbol{x}$); (b) relaxing proportional odds/hazards assumption (*i.e.*, covariates are in $\boldsymbol{u}$). Comparing the model deviances from these two using a likelihood ratio test then provides a test of the proportional odds/hazards assumption for the set of covariates under consideration.

Note that because the dichotomous and ordinal approaches only yield identical results under the proportional hazards model (*i.e.*, the complementary log-log link and covariates with effects that do not vary across time), differences in interpretation emerge for covariates allowed to have varying effects across time under these two approaches. For covariates of this type, the dichotomous approach is generally preferred because it models the covariate's influence in terms of the conditional probability of failure given prior survival (*i.e.*, the hazard function), rather than the cumulative probability of failure (*i.e.*, the integrated or cumulative hazard function) as in the ordinal representation of the model.

3 Maximum Likelihood Estimation

For the dichotomous approach, standard methods and software for multilevel analysis of dichotomous outcomes can be used. This is well-described in Barber et al. [2000]. For the ordinal treatment of survival times, the solution must be extended to accommodate right-censoring of the ordinal outcome. For this, let $\delta_{ij} = 0$ if level-1 unit ij is a censored observation and equal to 1 if the event occurs (fails). Thus, t_{ij} denotes the value of time ($t = 1, \ldots, m$) when either the ijth unit failed or was censored. It is assumed that the censoring and failure mechanisms are independent. In the multilevel model the probability of a failure at time t for a given level-2 unit i, conditional on $\boldsymbol{\theta}$ (and given $\boldsymbol{\alpha}_t, \boldsymbol{\beta}$, and $\boldsymbol{S}$) is:

$$\Pr(t_{ij} = t \cap \delta_j = 1 \mid \boldsymbol{\theta}; \boldsymbol{\alpha}_t, \boldsymbol{\beta}, \boldsymbol{S}) = P_{ijt} - P_{ij,t-1} \tag{12}$$

where $P_{ij0} = 0$ and $P_{ij,m+1} = 1$. The corresponding probability of being right censored at time t equals the cumulative probability of not failing at that time, $1 - P_{ijt}$.

Let $\boldsymbol{t}_i$ denote the vector pattern of failure times from level-2 unit i for the n_i level-1 units nested within. Similarly, let $\boldsymbol{\delta}_i$ denote the vector pattern of event indicators. The joint probability of patterns $\boldsymbol{t}_i$ and $\boldsymbol{\delta}_i$, given $\boldsymbol{\theta}$, assuming independent censoring is equal to the product of the probabilities of the level-1 responses:

$$\ell(\boldsymbol{t}_i, \boldsymbol{\delta}_i \mid \boldsymbol{\theta}; \boldsymbol{\alpha}_t, \boldsymbol{\beta}, \boldsymbol{S}) = \prod_{j=1}^{n_i} \prod_{t=1}^{m} \left[(P_{ijt} - P_{ij,t-1})^{\delta_{ij}} (1 - P_{ijt})^{1-\delta_{ij}} \right]^{d_{ijt}} \tag{13}$$

where $d_{ijt} = 1$ if $t_{ij} = t$ (and $= 0$ if $t_{ij} \neq t$).

The marginal density of $\boldsymbol{t}_i$ and $\boldsymbol{\delta}_i$ in the population is expressed as the following integral of the conditional likelihood, $\ell(\cdot)$, weighted by the prior density $g(\cdot)$:

$$h(\boldsymbol{t}_i, \boldsymbol{\delta}_i) = \int_{\boldsymbol{\theta}} \ell(\boldsymbol{t}_i, \boldsymbol{\delta}_i \mid \boldsymbol{\theta}; \boldsymbol{\alpha}_t, \boldsymbol{\beta}, \boldsymbol{S})\, g(\boldsymbol{\theta})\, d\boldsymbol{\theta} \tag{14}$$

where $g(\boldsymbol{\theta})$ represents the multivariate distribution of the standardized random effects vector $\boldsymbol{\theta}$ in the population. The marginal log-likelihood for the patterns from the N level-2 units is then written as $\log L = \sum_i^N \log h(\boldsymbol{t}_i, \boldsymbol{\delta}_i)$. Maximizing this log-likelihood yields maximum likelihood (ML) estimates, which are sometimes referred to as maximum marginal likelihood estimates [Bock, 1989] because integrating the joint likelihood of random effects and responses over the distribution of random effects translates to marginalization of the data distribution. Specific details of the solution, utilizing numerical quadrature to integrate over the random effects distribution, are provided in Hedeker et al. [2000]. As mentioned, SAS PROC NLMIXED can be used to obtain the ML estimates; several syntax examples are provided in the Appendix.

4 Examples

Three examples will be presented to illustrate the ordinal representation of the grouped-time survival analysis multilevel model. The first two examples are from school-based studies and treat students nested within schools. In the first example, there is right-censoring only at the last timepoint; these censored observations then form an additional category of the ordered time to event outcome. Thus, the model is akin to an ordinary multilevel ordinal regression using a complementary log-log link function to yield a proportional hazards model. The

second example has intermittant right-censoring (*i.e.*, observations can be right-censored at any timepoint) and so the likelihood function must be adapted for the censored observations, as described in the section on Estimation. This is perhaps the more usual situation in survival or time-to-event data. The final example is a joint longitudinal and survival model that allows the two processes to be correlated. In this example, time until study dropout is the survival outcome which is related to the longitudinal outcomes via the random effects of the latter.

4.1 Example 1: EMA Study

The data for this example are drawn from a natural history or Ecological Momentary Assessment (EMA, [Stone & Shiffman, 1994; Smyth & Stone, 2003]) study of adolescent smoking. Participants included in this study were in either 9th or 10th grade at baseline, and reported on a screening questionnaire 6-8 weeks prior to baseline that they had smoked at least one cigarette in their lifetimes. The majority (57.6%) had smoked at least one cigarette in the past month at baseline. A total of 461 students completed the baseline measurement wave. Baseline meaurement was coordinated in the schools of these students. In all, there were 16 schools in this study.

The study utilized a multi-method approach to assess adolescents including a week-long time/event sampling method via hand-held computers (EMA). Adolescents carried the hand-held computers with them at all times during a data collection period of seven consecutive days and were trained both to respond to random prompts from the computers and to event record (initiate a data collection interview) in conjunction with smoking episodes. Random prompts and the self-initiated smoking records were mutually exclusive; no smoking occured during random prompts. Questions concerned place, activity, companionship, mood, and other subjective variables. The hand-held computers date and time-stamped each entry.

In a previous paper based on an earlier EMA dataset [Hedeker et al., 2006], our group observed that adolescent smoking reports were most commonly observed on Fridays and Saturdays (*i.e.*, weekend smoking), but that mid-week smoking was more informative in determining the level of an adolescent's smoking behavior. Here, our interest is in modeling

time until the first smoking report following the weekend.[1] The idea being that earlier post-weekend smoking is a potentially greater indicator of dependency among adolescents. Thus, for each subject, we recorded the first day in which a smoking report was made, ordering the days as Sunday, Monday, ... , Friday, Saturday (this variable will be denoted as `Smk`). Table 1 lists the frequencies of responses across these seven days, and also the number of students who did not provide a smoking report during the week. `Smk` is thus an ordinal outcome with eight response categories.

Insert Table 1 here

In terms of predictor variables, for simplicity, we only considered a subject's level of social isolation (denoted as `SocIso`), which has previously been shown to be related to smoking in adolescents [Ennett & Bauman, 1993; Johnson & Hoffmann, 2000]. This variable was based on responses from the random prompts and consisted of a subject's average, across all prompts, on several individual mood items, each rated from 1 (not at all) to 10 (very much), that were identified via factor analysis. Specifically, `SocIso` consisted of the average of the following items that reflected a subject's assessment of their social isolation before the prompt signal: I felt lonely, I felt left out, and I felt ignored. Over all prompts, and ignoring the clustering of the data, the marginal mean of `SocIso` was 2.709 (sd=1.329) reflecting a relatively low level of social isolation on average.

A proportional hazards model was fit to these data using `Smk` as the time to event variable and `SocIso` as the independent variable. First, a model was run that ignored the clustering of students in schools. The effect of social isolation was observed to be significant in this analysis ($\hat{\beta} = .1085$, se $= .04650$, $p < .02$). Thus, students with higher average social isolation scores had increased hazards for earlier post-weekend smoking. However, when the clustering of students in schools was taken into account, by including a random school effect in the model, the effect was no longer significant at the .05 level ($\hat{\beta} = .0986$, se $= .04813$, $p < .059$).

[1] Here, we consider Friday and Saturday to be the weekend days, and Sunday to be the first post-weekend day. This, of course, is technically incorrect, but gets at the notion that for adolescents most social events involving peers occur on Friday and Saturday.

While not a dramatic change between these two, nonetheless, the conclusion based on the multilevel model would be that the effect of social isolation was only marginally significant.

The random effect variation, expressed as a standard deviation ($\hat{\sigma}_v$), was estimated to be .2926 (se = .09882). This estimated school variability can be expressed as an intraclass correlation, $\hat{\sigma}_v^2/(\hat{\sigma}_v^2 + \sigma^2)$, where σ^2 represents the variance of the latent continuous event time variable. For the complementary log-log link, the standard variance $\sigma^2 = \pi^2/6$, while for the logit link, $\sigma^2 = \pi^2/3$ [Agresti, 2002]. Applying this formula, the estimated intraclass correlation equals .04947 under the proportional hazards model (*i.e.*, complementary log-log link). This value is certainly in the range reported by Siddiqui et al. [1996], who examined school-based ICCs for a number of similar outcomes, and suggests a fair degree of similarity in terms of time to smoking within schools.

To test the proportional hazards assumption, a model was also fit allowing the effect of `SocIso` to vary across the ordinal `Smk` outcome. With eight categories of `Smk`, seven coefficients for `SocIso` were estimated, one for each cumulative comparison of these categories. The deviance for this extended model was 1450.8, while the deviance for the above proportional hazards model was 1453.2. Based on these values, the likelihood ratio chi-square statistic is 2.4, which on six degrees of freedom is not significant. Thus, the proportional hazards assumption is reasonable. Appendix 1 provides the `SAS PROC NLMIXED` syntax for both of these analyses.

4.2 Example 2: TVSFP Study

This example is taken from the Hedeker et al. [2000] article which used MIXOR [Hedeker & Gibbons, 1996] to estimate model parameters. Here, we replicate the results using `SAS NLMIXED` code. Relative to the previous example, this one will have right-censored observations at all timepoints, so the program code must reflect this, as noted just below equation (12). Namely, the probability of being right censored at time t equals the cumulative probability of not failing at that time, $1 - P_{ijt}$.

The Television School and Family Smoking Prevention and Cessation Project (TVSFP) study [Flay et al., 1988] was designed to test independent and combined effects of a school-

based social-resistance curriculum and a television-based program in terms of tobacco use prevention and cessation. The sample consisted of seventh-grade students who were assessed at pretest (Wave A), immediate post-intervention (Wave B), one-year follow-up (Wave C), and two-year follow-up (Wave D). A cluster randomization design was used to assign schools to the design conditions, while the primary outcome variables were at the student level. Schools were randomized to one of four study conditions: (a) a social-resistance classroom curriculum (CC); (b) a media (television) intervention (TV), (c) a combination of CC and TV conditions; and (d) a no-treatment control group.

An outcome of interest from the study is the onset of cigarette experimentation. At each of the four timepoints, students answered the question: "have you ever smoked a cigarette?" In analyzing the data below, because the intervention was implemented following the pretest, we focused on the three post-intervention timepoints and included only those students who had not answered yes to this question at pretest. In all, there were 1556 students included in the analysis of smoking initiation. Of these students, approximately forty percent ($n = 634$) answered yes to the smoking question at one of the three post-intervention timepoints, while the other sixty percent ($n = 922$) either answered no at the last timepoint or were censored prior to the last timepoint. The breakdown of cigarette onset for gender and condition subgroups is provided in Table 2.

Insert Table 2 here

In terms of the clustering, these 1556 students were from 28 schools with between 13 to 151 students per school ($\bar{n} = 56$, $sd = 38$).

A proportional hazards model was fit to these data to examine the effects of gender and intervention group on time to smoking initiation. Specifically, gender was included as a dummy variable expressing the male versus female difference. For the condition terms, because the CC by TV interaction was observed to be non-significant, a main effects model was considered. The maximum likelihood estimates (standard errors) were: Male = 0.05736 (0.07983), CC = 0.04461 (0.08418), and TV = 0.02141 (0.08311). Thus, none of the regressors are close to being significant, though the direction of the effects are increased hazards of

smoking initiation for males and those exposed to the CC and TV interventions. The random effect variation, expressed as a standard deviation ($\hat{\sigma}_v$), was estimated to be 0.05119 (0.1242). So the clustering effect attributable to schools is not large, and is much smaller than the previous example. Expressed as in intraclass correlation, it equals .00159, which reflects a rather low degree of similarity in smoking initiation times within schools. The `SAS PROC NLMIXED` syntax for this example is listed in Appendix 2.

4.3 Example 3: Joint Longitudinal and Survival Model

The data for this example come from the National Institute of Mental Health Schizophrenia Collaborative Study. In terms of the longitudinal outcome, we will examine Item 79 of the Inpatient Multidimensional Psychiatric Scale (IMPS; Lorr & Klett [1966]). Item 79, "Severity of Illness," was originally scored on a 7-point scale ranging from *normal, not at all ill* (0) to *among the most extremely ill* (7). Here, as in Hedeker & Gibbons [1994], we will analyze the outcome as an ordinal variable, specifically recoding the original seven ordered categories of the IMPS 79 severity score into four: (1) normal or borderline mentally ill, (2) mildly or moderately ill, (3) markedly ill, and (4) severely or among the most extremely ill.

In this study, patients were randomly assigned to receive one of four medications: placebo, chlorpromazine, fluphenazine, or thioridazine. Since our previous analyses revealed similar effects for the three anti-psychotic drug groups, they were combined in the present analysis. The experimental design and corresponding sample sizes are presented in Table 3. In this study, the protocol called for subjects to be measured at weeks 0, 1, 3, and 6; however, a few subjects were additionally measured at weeks 2, 4, and 5. There was some intermittent missingness in this study, however, dropout was a much more common pattern of missingness. In all, 102 of 437 subjects did not complete the trial.

Insert Table 3 here

The main question of interest is addressing whether there is differential change across time for the drug groups, here combined, relative to the control group. We have previously

addressed this question by analyzing these data using multilevel ordinal regression using both a probit [Hedeker & Gibbons, 1994] and logistic link [Hedeker & Gibbons, 2006]. A potential issue with these previous analyses is that the assumption of missing at random (MAR), inherent in the full maximum likelihood estimation of model parameters of the ordinal multilevel models, may not be plausible. That is, it could be that there was an association between missingness and the value of the dependent variable (*i.e.*, severity of illness) that would have been measured. Though the observed data cannot confirm or refute this possibility, one can fit missing not at random (MNAR) models as a means of doing a sensitivity analysis.

One class of MNAR models augment the usual multilevel model for longitudinal data with a model of dropout (or missingness), in which dropout depends on the random effects of the multilevel model. These models have been called random-coefficient selection models [Little, 1995], random-effects-dependent models [Hogan & Laird, 1997], and shared parameter models [De Gruttola & Tu, 1994; Wu & Carroll, 1988; Wu & Bailey, 1989; Schluchter, 1992; Ten Have et al., 1998] in the literature. Appealing aspects of this class of models is that they can be used for nonignorable missingness and can be fit using some standard software. X. Guo & Carlin [2004] present an excellent description for longitudinal outcomes and continuous time until event data, and provide a Web page with `PROC NLMIXED` syntax to carry out the analysis. Here, we will modify their approach and syntax for a longitudinal ordinal outcome and a grouped-time survival event (*i.e.*, time until study dropout).

In terms of the ordinal severity of illness outcome (denoted `imps79`), as in Hedeker & Gibbons [2006], we will use a multilevel ordinal logistic regression model including random subject intercepts and time trends. Here, if i represents subjects, j timepoints, and c ordinal response categories, the longitudinal outcome model is

$$\log\left[\frac{P_{ijc}}{1-P_{ijc}}\right] = \gamma_c - [\beta_0 \; + \; \beta_1\sqrt{\mathtt{W}}_j \; + \; \beta_2\mathtt{Tx}_i \; + \; \beta_3(\mathtt{Tx}_i \times \sqrt{\mathtt{W}}_j) \; + \; \upsilon_{0i} \; + \; \upsilon_{1i}\sqrt{\mathtt{W}}_j], \qquad (15)$$

where, `Tx` denotes treatment group (0=placebo and 1=drug) and `W` denotes week (the square root of week is used to linearize the relationship between the cumulative logits and week).

This is a cumulative logit model where $P_{ijc} = \Pr\,(\texttt{imps79}_{ij} \leq c)$, and the threshold parameters γ_c represent the cumulative logit values when the covariates and random subject effects equal zero. As parameterized above, a positive regression coefficient β indicates that as the regressor increases so does the probability of a higher value of `imps79`.

For time to dropout, let the variable $\texttt{D}_i = j$ if subject i drops out after the jth timepoint; namely, $\texttt{imps}_{ij}$ is observed, but $\texttt{imps}_{i,j+1}, \ldots, \texttt{imps}_{i,n}$ are all missing (here n represents the last possible timepoint, week 6). Note that we are ignoring intermittant missingness and focusing on time until a subject is no longer measured. Because there were no subjects who were only measured at week 0 in this study, $\texttt{D}_i$ will take on values of 1 to 5 for subjects dropping out prior to the end of the study, and a value of 6 for subjects completing the study (*i.e.*, was measured at week 6). Table 4 lists the frequencies of time to dropout (`D`) by treatment group.

Insert Table 4 here

For analysis of time to dropout, consider the following proportional hazards survival model (*i.e.*, ordinal regression model with complementary log-log link):

$$\log(-\log(1-\mathrm{P}(\texttt{D}_i \leq j))) \;=\; \alpha_{0j} + \alpha_1 \texttt{Tx}_i + \alpha_2(1-\texttt{Tx}_i)\upsilon_{0i} + \alpha_3(1-\texttt{Tx}_i)\upsilon_{1i} + \alpha_4 \texttt{Tx}_i \upsilon_{0i} + \alpha_5 \texttt{Tx}_i \upsilon_{1i}\;. \tag{16}$$

This model specifies that the time of dropout is influenced by a person's treatment group (α_1) and their intercept and trend in `imps79`. For the latter, α_2 and α_3 represent the effect of these on dropout among the placebo group (*i.e.*, when $\texttt{Tx}_i = 0$) , whereas α_4 and α_5 are the analogous effects for the drug group (*i.e.*, when $\texttt{Tx}_i = 1$).

Table 5 lists the regression coefficient estimates from separate and shared parameter modeling of these data. The separate parameter model sets $\alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$.

Insert Table 5 here

The separate parameter model yields identical parameter estimates and standard errors as running these two models, one for the longitudinal outcome and one for time to dropout, separately (not shown). Thus, the results for the longitudinal component represent the usual ordinal model with random subject intercepts and time-trends assuming MAR. As can be seen, both the $\sqrt{\texttt{W}}$ and $\texttt{Tx} \times \sqrt{\texttt{W}}$ terms are significant. This indicates, respectively, that the placebo group is improving across time and that the drug group is improving at an even faster rate across time. Additionally, because the `Tx` term is not significant, these groups are not different when week=0 (*i.e.*, at baseline). The dropout component indicates that the drug group has a significantly diminished hazard. Exponentiating the estimate of $-.693$ yields approximately .5, indicating that the hazard for dropout is double in the placebo group, relative to the drug group.

The shared paramter model fits these data significantly better, as evidenced by the likelihood ratio test, $X_4^2 = 18.2, p < .002$. This is not necessarily a rejection of MAR, but it is a rejection of this particular MAR model in favor of this particular MNAR shared parameter model. In terms of the longitudinal component, we see that the conclusions are the same as in the MAR model. If anything, the results are slightly stronger for the drug group in that the drug by time interaction is somewhat larger in the MNAR shared parameter model. In terms of the dropout component, all of the terms are significant. The significant `Tx` effect indicates that the drug group has a significantly diminished hazard of dropping out. For the terms involving the random effects, higher (*i.e.*, more positive) intercepts and slopes are associated with dropout for the placebo group, whereas for the drug group it is lower (*i.e.*, more negative) intercepts and slopes that are associated with dropout. In other words, among placebo subjects, those who start off worse (*i.e.*, higher severity scores) and who are not improving, or improving at a slower rate, are more likely to drop out. Conversely, for the drug subjects, those who start off relatively better and who have more negative slopes (*i.e.*, greater improvement) are more likely to drop out. `SAS PROC NLMIXED` code for the shared parameter model is provided in Appendix 3.

5 Discussion

Multilevel categorical regression models have been described for analysis of clustered grouped-time survival data, using either a proportional or partial proportional hazards or odds assumption. For models without time-dependent covariates, and assuming proportional hazards or odds, the data are analyzed utilizing an ordinal mixed-effects regression model. In this approach, survival times are represented as ordinal outcomes that are right-censored or not. Alternatively, in the dichotomous representation of the model, survival times are represented as sets of binary indicators of survival and analyzed using multilevel methods for dichotomous outcomes. In this chapter we have focused on the ordinal representation of the model; for extensive information on the dichotomous version see Singer & Willett [2003] and Barber et al. [2000].

Three examples were presented to illustrate the flexibility of this approach. The first did not have any intermittant right-censoring, and so the model was akin to a standard ordinal multilevel model using a complementary log-log link. Intermittant right-censoring was present for the second example, and the resulting likelihood function was modified to account for this. The final example illustrated how the ordinal representation of the survival analysis model can readily be used in longitudinal trials where there is interest in jointly modeling the longitudinal process and time to study dropout. For all examples, `SAS PROC NLMIXED` code was provided to allow data analysts to apply these methods.

Acknowledgments

Thanks are due to Siu Chi Wong for assisting with data analysis. This work was supported by National Cancer Institute grant 5PO1 CA98262.

References

Agresti, A. (2002). *Categorical data analysis, 2nd edition.* New York: Wiley.

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. In S. Leinhardt (Ed.), *Sociological methodology 1982* (p. 61-98). San Francisco: Jossey-Bass.

Barber, J. S., Murphy, S. A., Axinn, W. G., & Maples, J. (2000). Discrete-time multilevel hazard analysis. *Sociological Methodology*, *30*, 201-235.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine*, *2*, 273-277.

Bock, R. D. (1989). Measurement of human variation: a two stage model. In R. D. Bock (Ed.), *Multilevel analysis of educational data.* New York: Academic Press.

Clayton, D., & Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *Journal of the Royal Statistical Society, Series A*, *148*, 82-117.

Collett, D. (1994). *Modelling survival data in medical research.* New York: Chapman and Hall.

Cupples, L. A., D'Agostino, R. B., Anderson, K., & Kannel, W. B. (1985). Comparison of baseline and repeated measure covariate techniques in the framingham heart study. *Statistics in Medicine*, *7*, 205-218.

D'Agostino, R. B., Lee, M.-L., Belanger, A. J., Cupples, L. A., Anderson, K., & Kannel, W. B. (1990). Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. *Statistics in Medicine*, *9*, 1501-1515.

De Gruttola, V., & Tu, X. M. (1994). Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics*, *50*, 1003-1014.

de Leeuw, J., & Meijer, E. (Eds.). (2008). *Handbook of multilevel analysis.* New York: Springer.

Diggle, P. J., Heagerty, P., Liang, K.-Y., & Zeger, S. L. (2002). *Analysis of longitudinal data, 2nd edition*. New York: Oxford University Press.

Doksum, K. A., & Gasko, M. (1990). On a correspondence between models in binary regression analysis and in survival analysis. *International Statistical Review*, *58*, 243-252.

Engel, J. (1993). On the analysis of grouped extreme-value data with GLIM. *Applied Statistics*, *42*, 633-640.

Ennett, S. T., & Bauman, K. E. (1993). Peer group structure and adolescent cigarette smoking: a social network analysis. *Journal of Health and Social Behavior*, *34*, 226-236.

Flay, B. R., Brannon, B. R., Johnson, C. A., Hansen, W. B., Ulene, A., Whitney-Saltiel, D. A., et al. (1988). The television, school and family smoking prevention/cessation project: I. theoretical basis and program development. *Preventive Medicine*, *17(5)*, 585-607.

Goldstein, H. (1995). *Multilevel statistical models, 2nd edition*. New York: Halstead Press.

Grilli, L. (2005). The random-effects proportional hazards model with grouped survival data: a comparison between the grouped continuous and continuation ratio versions. *Journal of the Royal Statistical Society: Series A*, *168*, 8394.

Guo, G., & Rodriquez, G. (1992). Estimating a multivariate proportional hazards model for clustered data using the em algorithm. with an application to child survival in guatemala. *Journal of the American Statistical Association*, *87*, 969-976.

Guo, S. W., & Lin, D. Y. (1994). Regression analysis of multivariate grouped survival data. *Biometrics*, *50*, 632-639.

Guo, X., & Carlin, B. P. (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, *58*, 16-24.

Han, A., & Hausman, J. A. (1990). Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics*, *5*, 1-28.

Hedeker, D., & Gibbons, R. D. (1994). A random effects ordinal regression model for multilevel analysis. *Biometrics*, *50*, 933-944.

Hedeker, D., & Gibbons, R. D. (1996). MIXOR: a computer program for mixed-effects ordinal probit and logistic regression analysis. *Computer Methods and Programs in Biomedicine*, *49*, 157-176.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. New York: Wiley.

Hedeker, D., & Mermelstein, R. J. (1998). A multilevel thresholds of change model for analysis of stages of change data. *Multivariate Behavioral Research*, *33*, 427-455.

Hedeker, D., Mermelstein, R. J., & Flay, B. R. (2006). Application of item response theory models for intensive longitudinal data. In T. A. Walls & J. F. Schafer (Eds.), *Models for intensive longitudinal data* (p. 84-108). New York: Oxford.

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

Hogan, J. W., & Laird, N. M. (1997). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine*, *16*, 259-272.

Hougaard, P. (1995). Frailty models for survival data. *Lifetime Data Analysis*, *1*, 255-273.

Johnson, R. A., & Hoffmann, J. P. (2000). Adolescent cigarette smoking in u.s. racial/ethnic subgroups: findings from the national education longitudinal study. *Journal of Health and Social Behavior*, *41*, 392-407.

Läärä, E., & Matthews, J. N. S. (1985). The equivalence of two models for ordinal data. *Biometrika*, *72*, 206-207.

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica*, *47*, 939-956.

Lee, E. W., Wei, L. J., & Amato, D. A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In J. P. Klein & P. K. Goel (Eds.), *Survival analysis: State of the art* (p. 237-247). Dordrecht: Kluwer.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, *90*, 1112-1121.

Lorr, M., & Klett, C. J. (1966). *Inpatient multidimensional psychiatric scale: Manual*. Palo Alto, CA: Consulting Psychologists Press.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society, Series B*, *42*, 109-142.

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

Paik, M. C., Tsai, W.-Y., & Ottman, R. (1994). Multivariate survival analysis using piecewise gamma frailty. *Biometrics*, *50*, 975-988.

Pickles, A., & Crouchley, R. (1995). A comparison of frailty models for multivariate survival data. *Statistics in Medicine*, *14*, 1447-1461.

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

Rabe-Hesketh, S., Yang, S., & Pickles, A. (2001). Multilevel models for censored and latent responses. *Statistical Methods in Medical Research*, *10*, 409-427.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models, 2nd edition*. Thousand Oaks, CA: Sage.

Reardon, S. F., Brennan, R., & Buka, S. L. (2002). Estimating multi-level discrete-time hazard models using cross-sectional data: neighborhood effects on the onset of adolescent cigarette use. *Multivariate Behavioral Research*, *37*, 297-330.

Scheike, T. H., & Jensen, T. K. (1997). A discrete survival model with random effects: an application to time to pregnancy. *Biometrics*, *53*, 318-329.

Schluchter, M. D. (1992). Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine*, *11*, 1861-1870.

Self, S. G., & Prentice, R. L. (1986). Incorporating random effects into multivariate relative risk regression models. In S. H. Moolgavkar & R. L. Prentice (Eds.), *Modern statistical methods in chronic disease epidemiology* (p. 167-178). New York: Wiley.

Shih, J. H., & Louis, T. A. (1995). Assessing gamma frailty models for clustered failure time data. *Lifetime Data Analysis*, *1*, 205-220.

Siddiqui, O., Hedeker, D., Flay, B. R., & Hu, F. B. (1996). Intraclass correlation estimates in a school-based smoking prevention study: outcome and mediating variables, by gender and ethnicity. *American Journal of Epidemiology*, *144*, 425-433.

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 and Behavioral Statistics*, *18*, 155-195.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York: Oxford University Press.

Smyth, J. M., & Stone, A. A. (2003). Ecological momentary assessment research in behavioral medicine. *Journal of Happiness Studies*, *4*, 35-52.

Stone, A., & Shiffman, S. (1994). Ecological Momentary Assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*, *16*, 199-202.

Teachman, J. D., Call, V. R. A., & Carver, K. P. (1994). Marital status and the duration of joblessness among white men. *Journal of Marriage and the Family*, *56*, 415-428.

Ten Have, T. R. (1996). A mixed effects model for multivariate ordinal response data including correlated discrete failure times with ordinal responses. *Biometrics*, *52*, 473-491.

Ten Have, T. R., Kunselman, A. R., Pulkstenis, E. P., & Landis, J. R. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics*, *54*, 367-383.

Ten Have, T. R., & Uttal, D. H. (1994). Subject-specific and population-averaged continuation ratio logit models for multiple discrete time survival profiles. *Applied Statistics*, *43*, 371-384.

Thompson, W. A. (1977). On the treatment of grouped observations in life studies. *Biometrics*, *33*, 463-470.

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

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. New York: Springer.

Wei, L. J., Lin, D. Y., & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, *84*, 1065-1073.

Wu, M. C., & Bailey, K. R. (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics*, *45*, 939-955.

Wu, M. C., & Carroll, R. J. (1988). Estimation and comparison of changes in the presence of right censoring by modeling the censoring process. *Biometrics*, *44*, 175-188.

Appendix 1

In this listing for Example 1, and in subsequent syntax listings, expressions with all uppercase letters are used for SAS-specific syntax, while expressions including lowercase letters are used for user-defined entities. In this example, SocIso is the regressor and bSocIso is its regression coefficient. Note that PROC NLMIXED requires the user to name all model parameters in the syntax. The variable Smk indicates the day of the first smoking event with Sunday=1, Monday=2, ... , Saturday=7, and Never Smoked=8. The final category of Never Smoked represents all right-censored observations (*i.e.*, there is no intermittant right-censoring). Under the complementary log-log link, the cumulative probability of an event occuring up to a particular timepoint is given by (4). Because this is a cumulative probability, the actual probability for a given timepoint is obtained by subtraction of these cumulative probabilities, except for the first and last categories. The last category (*i.e.*, Never Smoked) is obtained as 1 minus the cumulative probability of smoking up to and including Saturday (*i.e.*, the last day). The parameters a1, a2, ... , a7 represent the baseline hazard (*i.e.*, hazard when all covariates equal 0); there are seven of these in this example because the total number of Smk categories is eight. These are akin to the threshold parameters in ordinal regression models and the values of these parameters should be increasing to reflect increased hazard across time. Finally, the variable Schoolid is the cluster (level-2) id, which indicates which students belong to which schools. The random effect variance attributable to schools is estimated as a standard deviation and named sd. For clustered data, where the cluster variance is thought to be small, it is usually better to estimate the standard deviation than the variance because the latter will be much smaller and close to zero. Also, the random effect, named theta, is multiplied by its standard deviation in the model, as in (5), and so it is in standardized form (*i.e.*, the variance of theta equals 1 on the RANDOM statement).

```
PROC NLMIXED;
PARMS a1=-1.9 a2=-1.7 a3=-1.4 a4=-1.1 a5=-.9 a6=-.8 a7=-.6 bSocIso=.1 sd=.2;
   z = bSocIso*SocIso + sd*theta;
```

```
IF (Smk=1) THEN
   p = 1 - EXP( - EXP(a1+z));
ELSE IF (Smk=2) THEN
   p = (1 - EXP( - EXP(a2+z))) - (1 - EXP( - EXP(a1+z)));
ELSE IF (Smk=3) THEN
   p = (1 - EXP( - EXP(a3+z))) - (1 - EXP( - EXP(a2+z)));
ELSE IF (Smk=4) THEN
   p = (1 - EXP( - EXP(a4+z))) - (1 - EXP( - EXP(a3+z)));
ELSE IF (Smk=5) THEN
   p = (1 - EXP( - EXP(a5+z))) - (1 - EXP( - EXP(a4+z)));
ELSE IF (Smk=6) THEN
   p = (1 - EXP( - EXP(a6+z))) - (1 - EXP( - EXP(a5+z)));
ELSE IF (Smk=7) THEN
   p = (1 - EXP( - EXP(a7+z))) - (1 - EXP( - EXP(a6+z)));
ELSE IF (Smk=8) THEN
   p = EXP( - EXP(a7+z));
logl = LOG(p);
MODEL Smk ~ GENERAL(logl);
RANDOM theta ~ NORMAL(0,1) SUBJECT=Schoolid;
```

Users must provide starting values for all parameters on the `PARMS` statement. To do so, it is beneficial to run the model in stages using estimates from a prior stage as starting values and setting the additional parameters to zero or some small value. For example, one can start by estimating a fixed-effects model to provide starting values for the regression coefficients using `SAS PROC PHREG`.

In order to test the proportional hazards assumption, one can compare the above model to one in which the effect of `SocIso` is allowed to vary across the cumulative comparisons of the ordinal outcome (*i.e.*, a non-proportional hazards model). For this, seven response models (named `z1, z2, ... , z7`) with varying effects of `SocIso` (named `bSocIso1, bSocIso2, ... , bSocIso7`) are defined. The appropriate response models are then indicated in the

calculations for the category probabilities.

```
PROC NLMIXED;
PARMS a1=-1.9 a2=-1.7 a3=-1.4 a4=-1.1 a5=-.9 a6=-.8 a7=-.6 sd=.2
      bSocIso1=.1 bSocIso2=.1 bSocIso3=.1 bSocIso4=.1 bSocIso5=.1
      bSocIso6=.1 bSocIso7=.1;
   z1 = bSocIso1*SocIso + sd*theta;
   z2 = bSocIso2*SocIso + sd*theta;
   z3 = bSocIso3*SocIso + sd*theta;
   z4 = bSocIso4*SocIso + sd*theta;
   z5 = bSocIso5*SocIso + sd*theta;
   z6 = bSocIso6*SocIso + sd*theta;
   z7 = bSocIso7*SocIso + sd*theta;
IF (Smk=1) THEN
   p = 1 - EXP( - EXP(a1+z1));
ELSE IF (Smk=2) THEN
   p = (1 - EXP( - EXP(a2+z2))) - (1 - EXP( - EXP(a1+z1)));
ELSE IF (Smk=3) THEN
   p = (1 - EXP( - EXP(a3+z3))) - (1 - EXP( - EXP(a2+z2)));
ELSE IF (Smk=4) THEN
   p = (1 - EXP( - EXP(a4+z4))) - (1 - EXP( - EXP(a3+z3)));
ELSE IF (Smk=5) THEN
   p = (1 - EXP( - EXP(a5+z5))) - (1 - EXP( - EXP(a4+z4)));
ELSE IF (Smk=6) THEN
   p = (1 - EXP( - EXP(a6+z6))) - (1 - EXP( - EXP(a5+z5)));
ELSE IF (Smk=7) THEN
   p = (1 - EXP( - EXP(a7+z7))) - (1 - EXP( - EXP(a6+z6)));
ELSE IF (Smk=8) THEN
   p = EXP( - EXP(a7+z7));
logl = LOG(p);
MODEL Smk ~ GENERAL(logl);
RANDOM theta ~ NORMAL(0,1) SUBJECT=Schoolid;
```

Appendix 2

For Example 2, `Male`, `Cc`, and `Tv` are indicator variables of male, CC intervention, and TV intervention, respectively. The regression coefficients for these variables are named `bMale`, `bCc`, and `bTv`. The variable `Onset` indicates whether the student either answered yes to the smoking question or was censored at immediate post-intervention (`Onset` = 2), the one-year follow-up (`Onset` = 3), or the second year follow-up (`Onset` = 4). The indicator `statuse` distinguishes censored observations (`Statuse` = 0) from smoking observations (`Statuse` = 1). Notice that if the event occurs, then the cumulative probability, under the complementary log-log link, is given by (4), whereas if the event is censored then the probability equals 1 minus this expression. The parameters `a1`, `a2`, and `a3` represent the baseline hazard; there are three in this example because there are three timepoints and there is right-censoring. As in Example 1, the variable `Schoolid` is the cluster (level-2) id, and the random effect variance attributable to schools is estimated as a standard deviation and named `sd`.

```
PROC NLMIXED;
PARMS a1=-3.5 a2=-2.9 a3=-2.4 bMale=.2 bCc=.1 bTv=.1 sd=.2;
   z = bMale*Male + bCc*Cc + bTv*Tv +sd*theta;
IF (Onset=2 AND Statuse=1) THEN
   p = 1 - EXP( - EXP(a1+z));
ELSE IF (Onset=2 AND Statuse=0) THEN
   p = EXP( - EXP(a1+z));
ELSE IF (Onset=3 AND Statuse=1) THEN
   p = (1 - EXP( - EXP(a2+z))) - (1 - EXP( - EXP(a1+z)));
ELSE IF (Onset=3 AND Statuse=0) THEN
   p = EXP( - EXP(a2+z));
ELSE IF (Onset=4 AND Statuse=1) THEN
   p = (1 - EXP( - EXP(a3+z))) - (1 - EXP( - EXP(a2+z)));
ELSE IF (Onset=4 AND Statuse=0) THEN
   p = EXP( - EXP(a3+z));
logl = LOG(p);
MODEL Onset ~ GENERAL(logl);
RANDOM theta ~ NORMAL(0,1) SUBJECT=Schoolid;
```

Appendix 3

To estimate the shared parameter model in Example 3 with NLMIXED, first, one must do a bit of data managing to produce a dataset with both the longitudinal outcomes and the time to dropout. The code for this is listed below.

```
DATA one; INFILE 'c:\schzrepo.dat'; INPUT id imps79 week tx ;

/* The coding for the variables is as follows:
id = subject id number
imps79 = overall severity (ordinal version from 1 to 4)
week = 0,1,2,3,4,5,6 (most of the obs. are at weeks 0,1,3, and 6)
tx 0=placebo 1=drug (chlorpromazine, fluphenazine, or thioridazine)

/* compute the square root of week to linearize relationship */
sweek = SQRT(week);

/* calculate the maximum value of WEEK for each subject */
PROC MEANS NOPRINT; CLASS id; VAR week;
OUTPUT OUT=two MAX(week)=d;

/* adding the max of week (d) to the original dataset */
DATA all; MERGE one two; BY id; IF id NE .;

/* create an indicator of the last observation of a subject (last) */
PROC SORT DATA=all; BY id week;
DATA all; SET all; BY id;
last = LAST.id;
RUN;
```

Each subject contributes n_i records in the input data file, namely, the file `schzrepo.dat` only contains datalines for which a subject has a valid measurement at a given timepoint. Thus, the maximum value of the time variable `week` for a given subject is the last timepoint that a subject is measured on. This variable, named `d`, then serves as the time to dropout variable in the analysis.

Below is the `NLMIXED code` for the shared parameter model, which, as mentioned, is based on the code of the X. Guo & Carlin [2004] article. The first part of the code is for the

time to dropout model, and the latter part is for the ordinal longitudinal outcome.

For time to dropout, `Zsurv` represents the response model and `Psurv` is the probability for a given observation based on the complementary log–log link function. As time to dropout is ordinal, with values of 1 to 6 indicating the final week of measurement for the individual, the cumulative probabilities represent the probability of response in a given category and below. Individual category probabilities, the `Psurv`'s, are therefore obtained by subtraction. As noted, the `i1` to `i5` parameters represent the cumulative baseline hazard, and the parameters `aTx` to `aDslp` are the effects on time to dropout. In particular, the latter four, `aPint` to `aDslp`, indicate the effect of the random subject intercepts and time-trends of the longitudinal model on time to dropout.

In terms of the longitudinal model (*i.e.*, the ordinal imps79 scores), the `b` terms are for the regression coefficients (*i.e.*, the β's), the `u` terms are for the random subject effects, and the `v` terms are for the random-effects variance–covariance parameters. The Cholesky factorization of the variance–covariance matrix of the random effects is used in the code below. The ordinal outcome `imps79` can take on values from 1 to 4 and the cumulative probabilities (*i.e.*, the `Plong`'s) are calculated using the logistic response function. As in the ordinal survival model, individual category probabilities are obtained by subtraction. The `t` parameters represent the threshold parameters. Because an intercept is in the model (`b0`), the number of thresholds equals the number of ordinal categories minus two.

```
PROC NLMIXED DATA=all;

PARMS b0=7.3 bTx=0 bSwk=-.88 bTxSwk=-1.7 t2=4 t3=6.5 s11=1 s12=0 s22=.5
     aTx=0 aPint=0 aPslp=0 aDint=0 aDslp=0 i1=-1 i2=-.7 i3=-.5 i4=0 i5=.2;

/* Compute log likelihood contribution of the survival data part */
/* when the last observation of a subject is reached */
IF (last) THEN DO;
    Zsurv = aTx*Tx + aPint*(1-Tx)*u0 + aPslp*(1-Tx)*u1 + aDint*Tx*u0 + aDslp*Tx*u1;
    IF (d=1) THEN
     Psurv = 1 - EXP( - EXP(i1+Zsurv));
    ELSE IF (d=2) THEN
```

```
     Psurv = (1 - EXP( - EXP(i2+Zsurv))) - (1 - EXP( - EXP(i1+Zsurv)));
    ELSE IF (d=3) THEN
     Psurv = (1 - EXP( - EXP(i3+Zsurv))) - (1 - EXP( - EXP(i2+Zsurv)));
    ELSE IF (d=4) THEN
     Psurv = (1 - EXP( - EXP(i4+Zsurv))) - (1 - EXP( - EXP(i3+Zsurv)));
    ELSE IF (d=5) THEN
     Psurv = (1 - EXP( - EXP(i5+Zsurv))) - (1 - EXP( - EXP(i4+Zsurv)));
    ELSE IF (d=6) THEN
     Psurv = EXP( - EXP(i5+Zsurv));
    IF (Psurv > 1e-8) THEN Lsurv = LOG(Psurv);
    ELSE Lsurv = -1e100;
END; ELSE Lsurv=0;

/* Cholesky parameterization of the random effects var-covar matrix */
/* This ensures that the matrix is non-negative definite */
v11 = s11*s11;
v12 = s11*s12;
v22 = s12*s12 + s22*s22;

/* Compute the contribution of the longitudinal part */
/* Every observation in the data set makes a contribution */
Zlong = b0 + bTx*tx + bSwk*sweek + bTxSwk*tx*sweek + u0 + u1*sweek;
IF (imps79=1) THEN
    Plong = 1 / (1 + EXP(-(-Zlong)));
ELSE IF (imps79=2) THEN
    Plong = (1/(1 + EXP(-(t2-Zlong)))) - (1/(1 + EXP(-(-Zlong))));
ELSE IF (imps79=3) THEN
    Plong = (1/(1 + EXP(-(t3-Zlong)))) - (1/(1 + EXP(-(t2-Zlong))));
ELSE IF (imps79=4) THEN
    Plong = 1 - (1 / (1 + EXP(-(t3-Zlong))));
IF (Plong > 1e-8) THEN Llong = LOG(Plong);
ELSE Llong = -1e100;

/* Any numeric variable can be used as the response in the MODEL statement */
/* It has no bearing on the results */
MODEL last ~ GENERAL(Llong + Lsurv);
```

```
RANDOM u0 u1 ~ NORMAL([0, 0],[v11,v12,v22]) SUBJECT=id;

/* Compute the variances and covariance of the random effects */
ESTIMATE 'Var[u0]' v11;
ESTIMATE 'Cov[u0,u1]' v12;
ESTIMATE 'Var[u1]' v22;
RUN;
```

Table 1

Onset of Smoking Event Across Days

Frequencies (and Percentages), N = 461

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Never
68	56	39	21	18	24	8	227
(14.75)	(12.15)	(8.46)	(4.56)	(3.90)	(5.21)	(1.74)	(49.24)

Table 2

Onset of Cigarette Experimentation Across Three Waves
Frequencies (and Percentages) for Gender and Condition Subgroups

	Wave B			*Wave C*			*Wave D*		
	event	censored	total	event	censored	total	event	censored	total
Males	156 (21.0)	83 (11.2)	742	89 (17.7)	134 (26.6)	503	63 (22.5)	217 (77.5)	280
Females	130 (16.0)	105 (12.9)	814	117 (20.2)	154 (26.6)	579	79 (25.6)	229 (74.4)	308
Control	66 (16.5)	60 (15.0)	401	53 (19.3)	69 (25.1)	275	34 (22.2)	119 (77.8)	153
CC only	75 (19.1)	27 (6.9)	392	53 (18.3)	61 (21.0)	290	49 (27.8)	127 (72.2)	176
TV only	71 (17.3)	54 (13.2)	410	60 (21.1)	79 (27.7)	285	38 (26.0)	108 (74.0)	146
CC and TV	74 (21.0)	47 (13.3)	353	40 (17.2)	79 (34.1)	232	21 (18.6)	92 (81.4)	113

Table 3

Experimental Design and Weekly Sample Sizes Across Time

	Sample Size at Week							
Group	0	1	2	3	4	5	6	Total
Placebo	107	105	5	87	2	2	70	108
Drug	327	321	9	287	9	7	265	329

Drug = Chlorpromazine, fluphenazine, or thioridazine.

Table 4

Crosstabulation of Treatment Group by Time to Dropout — Frequencies (Percentages)

	Time to Dropout						
Treatment Group	1	2	3	4	5	6	Total
Placebo	13	5	16	2	2	70	108
	(.12)	(.05)	(.15)	(.02)	(.02)	(.65)	
Drug	24	5	26	3	6	265	329
	(.07)	(.02)	(.08)	(.01)	(.02)	(.81)	

Table 5
—em Separate and Shared Parameter Models

	Separate			Shared		
Parameter	Estimate	SE	$p <$	Estimate	SE	$p <$
Outcome						
`Tx` β_1	.048	.392	.90	.149	.382	.70
$\sqrt{\texttt{W}}$ β_2	−.887	.218	.0001	−.708	.221	.002
`Tx` $\times$ $\sqrt{\texttt{W}}$ β_3	−1.692	.252	.0001	−1.909	.257	.0001
Dropout						
`Tx` α_1	−.693	.205	.0008	−.719	.281	.02
Placebo random intercept α_2				.242	.094	.02
Placebo random slope α_3				.570	.289	.05
Drug random intercept α_4				−.150	.071	.04
Drug random slope α_5				−.553	.177	.002
Deviance	4056.7			4038.5		