DISCRETE-TIME MULTILEVEL HAZARD ANALYSIS

Jennifer S. Barber*
Susan Murphy**
William G. Axinn***
Jerry Maples****

Combining innovations in hazard modeling with those in multilevel modeling, we develop a method to estimate discrete-time multilevel hazard models. We derive the likelihood of and formulate assumptions for a discrete-time multilevel hazard model with time-varying covariates at two levels. We pay special attention to assumptions justifying the estimation method. Next, we demonstrate file construction and estimation of the models using two common software packages, HLM and MLN. We also illustrate the use of both packages by estimating a model of the hazard of contraceptive use in rural Nepal using time-varying covariates at both individual and neighborhood levels.

The first and second authors contributed equally to this paper. The paper benefited substantially by comments and suggestions from the editors and reviewers. This research was supported by NICHD grant HD32912, by NSF grants SBR 9811983 and DMS 9802885, by grant P50 DA 10075 from NIDA to the Pennsylvania State University s Methodology Center, and by a P30 center grant from NICHD to the University of Michigan s Population Studies Center. Direct correspondence to the first author at the Institute for Social Research, University of Michigan, 426 Thompson St., Ann Arbor, Michigan, 48106-1248 or via e-mail at jebarber@umich.edu.

* Institute for Social Research, University of Michigan
** Department of Statistics and Institute for Social Research, University of Michigan
*** Department of Sociology and Institute for Social Research, University of Michigan
**** The Methodology Center and Department of Statistics, Pennsylvania State University

1. INTRODUCTION

Over the past two decades, one of the central themes in Sociology has been the study of individual life courses: understanding the timing and sequencing of life events such as cohabitation, marriage, labor force entry and exit, and educational attainment (Elder 1977, 1983; Rindfuss, Morgan, and Swicegood 1988; Thornton, Axinn and Teachman 1995). As a result, sociological models of individual behavior have become increasingly dynamic, even incorporating measures of individual characteristics that change over time. A wide range of advances in the estimation of hazard models with time-varying covariates has fueled this explosion in dynamic modeling (Allison 1984; Petersen 1991; Yamaguchi 1991).

Another central theme in both classic and modern Sociology has been the relationship between macro level social changes and micro level behavior (Alexander 1988; Coleman 1990; Durkheim [1933] 1984; Smith 1989; Weber 1922). Recent advances in multilevel modeling have dramatically improved efforts to include macro characteristics, sometimes called ecological, neighborhood, or contextual characteristics, in micro level models of behavior (Bryk and Raudenbush 1992; DiPrete and Forristal 1994; Goldstein 1995; Ringdal 1992). These advances include not only multilevel linear models, but also multilevel generalized linear models such as logistic regression and loglinear regression (see Bryk, Raudenbush, and Congdon 1996; Goldstein 1995; Wong and Mason 1985).

In this paper we combine the dynamic approach to modeling provided by hazard models with multilevel models. We have three goals: 1) To develop a discrete-time multilevel hazard model; 2) To illustrate the use of well known software to estimate models of macro level effects on micro level behavior where there is change over time at both the micro and the macro levels of analysis; and 3) To provide details regarding the assumptions that allow the regression coefficients of both the micro level covariates and the macro level covariates to be estimated in a multilevel hazard analysis framework.

Estimation procedures for multilevel models of social behavior must accommodate multilevel data structures. Classical statistical procedures, such as hazard analysis, assume that subjects (or individuals) behave independently, yet it is likely that individuals in the same macro context behave more similarly than individuals from different contexts. As a consequence, statistical procedures that ignore the multilevel data structure underestimate standard errors leading to hypothesis tests with elevated Type 1 error rates (rejecting the null hypothesis in error). Kreft (1994, p. 151) gives a thorough discussion of this issue in the linear regression setting and Muthén (1997, p. 455-458) gives a similar discussion in the context of linear latent variable models.

The use of a single level hazard model for multilevel data creates additional complications due to the fact that hazard models are both nonlinear and dynamic. First, if one wants to compare individuals with different characteristics within the same or similar macro contexts, then regression coefficients in the single level, nonlinear model applied to data on individuals grouped within contexts are not the desired regression coefficients (Diggle, Liang, and Zeger 1994). Rather, regression coefficients from the single level, nonlinear model reflect marginal comparisons of individuals from the variety of contexts. In other words, these models do not statistically control for macro contextual characteristics. Thus, for instance, if the goal is to compare the contraceptive behavior of educated women to the contraceptive behavior of uneducated women, within neighborhoods with similar access to schooling opportunities, then the coefficients from the single level nonlinear model are inappropriate.

Second, hazard models are dynamic in that the event under study, such as initiation of permanent contraception, unfolds over time. If a single level hazard model is applied to multilevel data, duration bias results (Trussell and Richards 1985; Vaupel and Yashin 1985; Yamaguchi 1991). Suppose that the event is the initiation of permanent contraception and suppose that the macro context is the neighborhood and neighborhoods are heterogeneous, e.g. some neighborhoods are special in that their characteristics lead to delayed permanent contraceptive use. As time proceeds most of the women who have yet to experience the event will be from the neighborhoods that delay the event time. The estimator of the baseline hazard will not reflect the hazard for any one type of individual, rather it reflects an average hazard, averaged over the variety of macro level contexts. At later times this average will be primarily over individuals from the special contexts. The unobserved heterogeneity of the contexts results in an underestimation of the baseline hazard. One way to reduce the bias is to include a random intercept in the regression model for the hazard (Vaupel and Yashin 1985).

Suppose that the association of an individual level covariate, such as woman s educational level, with the event, contraceptive timing, varies across contexts. Then the unobserved heterogeneity of the contexts results in a second form of duration bias, a biased regression coefficient. At earlier times the regression coefficient of woman s education level reflects a comparison of groups of women wherein each group is composed of women from the full variety of neighborhoods. But at later times, the regression coefficient for women s education reflects a comparison of groups of women wherein the groups are primarily composed of women from the special neighborhoods. To prevent this bias, we propose a multilevel model including random coefficients. The crux is that multilevel models of social behavior demand not only multilevel data but also multilevel statistical procedures.

A first step in avoiding the above problems is to use statistical procedures that acknowledge the multilevel data structure. However multilevel *hazard models* are uncommon, particularly models involving both individual and macro level time-varying covariates**.** (For examples of multilevel hazard models, see Brewster 1994; Guo 1993; Guo and Rodríguez 1992; Hedeker, Siddiqui, and Hu 1998; Ma and Willms 1999; Massey and Espinosa 1996; Sastry 1996, 1997; Vaupel 1988). Few use a fully dynamic multilevel model with interactions between macro and individual level covariates and dynamic time-varying measures at both macro and individual levels in models of individual behavior. One reason has been a dearth of data providing measures of change over time in macro level characteristics. However recent advances in data collection methods have led to the development of techniques for collecting a continuous record of change over time in macro level (e.g. neighborhood) characteristics (Axinn, Barber, and Ghimire 1997). Another reason has been lack of widely available estimation procedures. Widely available software programs developed for multilevel data (e.g. HLM, MLN, Proc Mixed in SAS) have not been explicitly extended to discrete-time hazard analysis with time-varying covariates and most software programs developed for hazard models (e.g.S-PLUS, STATA) have not been extended to fit multilevel data. The ideal multilevel hazard analysis program would allow both time-varying macro and individual level covariates.

2. AN EMPIRICAL EXAMPLE

Our example comes from the Chitwan Valley Family Study. The purpose of the study was to collect detailed information about historical social changes in the neighborhoods in the valley, and to analyze how those social changes relate to individual level behavioral change in the propensity to use contraceptives, to delay marriage, and to limit childbearing. Data were collected from 171 neighborhoods in the Chitwan Valley, located in central Nepal. Every individual in each of the 171 neighborhoods was interviewed. The study collected retrospective histories of change in each neighborhood using the Neighborhood History Calendar method (Axinn et al. 1997), and retrospective histories of each individual s behavior using a Life History Calendar adapted specifically to the setting (Axinn, Pearce, and Ghimire 1999). Our analysis examines women age 49 and younger who have had at least one birth.

We will test three hypotheses concerning the timing of permanent contraceptive use, chosen to highlight the model s flexibility for estimating the effects of different types of covariates. The hypotheses are:

H1 (individual level): Educated women have a higher hazard of using a permanent contraceptive method[1] compared to uneducated women.

H2 (neighborhood level): Increased access to nearby schooling opportunities is associated

[1] We focus on *permanent* contraceptive use because the vast majority of contraception in Nepal is used for stopping childbearing rather than spacing births. We consider a woman s own sterilization, her spouse s sterilization, IUD, Norplant, and depo-provera to be permanent methods in this setting (Axinn and Barber 1999). Because permanent contraception among women who have no children is extremely rare in this setting, we only consider women to be at risk of contracepting after they have given birth to their first child.

with a higher hazard of using a permanent contraceptive method.

H3 (cross level): The association between education and permanent contraceptive use is stronger in neighborhoods with a school nearby.

With H1, we compare the timing of permanent contraceptive use between women of different levels of education but within the same or similar neighborhoods.[2] With H2, we test whether women who live in neighborhoods with access to schools are more likely to limit their childbearing via contraceptive use. When schooling opportunities are convenient and nearby, a woman is more likely to expect that her children will attend school, which increases the hazard of permanent contraceptive use. H3, our cross-level hypothesis, is a macro-micro interaction: living in a neighborhood with a school nearby will strengthen the individual level relationship between a woman s own education and her propensity to use a permanent contraceptive method. Women with formal education may be more likely to limit their family size precisely because they want to send their children to school (Axinn 1993), thus the relationship between a woman s own education and her family-limiting behavior will be stronger if she believes that she will actually have the opportunity to send her children to school.

To construct a model to test these hypotheses, we use the following variables with subscripts denoting the t^{th} calendar year, and the j^{th} woman in the k^{th} neighborhood.

Y_{tjk} = a dichotomous indicator of whether woman j in neighborhood k initiates permanent contraceptive use during year t. **This is the dependent variable.**

p_{tjk} = the hazard of initiating permanent contraception by woman j in neighborhood k

---

[2]Much has been written about the effects of education on contraceptive use, and we refer readers to this vast literature rather than elaborating on this hypothesis here (see Axinn 1993; Axinn and Barber 1999).

during year t (given no prior contraceptive use). This is the mean of Y_{tjk} given no prior contraceptive use and all prior covariate measurements.

$Educ_j$ = a dichotomous indicator of whether woman j attended school (before the birth of her first child). This is a **time-invariant individual level covariate**.

$Chldrn_{tj}$ = the total number of children woman j has had by year t, minus one.[3] This is a **time-varying individual level covariate**.[4]

$School_{tk}$ = a dichotomous indicator of whether there is a school within a five-minute walk from neighborhood k during year t. This is the **time-varying group level covariate**.

Dis_k = distance from the neighborhood to nearest town. Distance in miles to the nearest town was computed using global positioning systems technology. This is a **time-invariant group level covariate**.

Finally, the following two variables are used to indicate the baseline hazard in the models. They are counter variables where the first person-year for each woman is coded 0, and each subsequent year is incremented accordingly. These two measures form the baseline hazard of permanent contraceptive use.

$Time_{tj}$ = number of years since woman j's first birth.

---

[3]Total number of children is coded as the actual number the woman gave birth to minus one because we analyze only women who have given birth to at least one child. If we did not transform the variable in this way, the baseline hazard would be estimated for a woman with zero children, which is outside the valid range in this analysis (Kreft, DeLeeuw, and Aiken 1995).

[4]Our time-varying measures of individual and neighborhood characteristics are measured in the year *prior* to the current year of permanent contraceptive risk. For example, we use the total number of children in the prior year to predict the hazard of permanent contraceptive use in the current year. In other words, all time-varying covariates are lagged by one year.

$Time^2_{tj}$ = number of years since woman j's first birth, squared.[5]

We estimate a multilevel hazard model that allows the effects of education and total number of children to vary by neighborhood. We model the hazard by the logit link; thus, the parameters represent additive effects on the log-odds of contraceptive use. The following represents the multilevel model, which we call the conceptual model, or CM. Using multilevel terminology, the individual level model, or hazard model, for woman j in neighborhood k is

$$\text{Logit}\,(p_{tjk}) = {}_{0k} + {}_{1k}\,Educ_j + {}_{2k}\,Chldrn_{tj} + {}_{3}\,Time_{tj} + {}_{4}\,Time^2_{tj}. \qquad \text{(1a) CM}$$

Note that ${}_0$, ${}_1$, and ${}_2$ are indexed by k. This is to allow these effects to vary by neighborhood. We allow ${}_0$ (the intercept in the individual model) to vary by neighborhood so that the overall level of contraceptive use is a function of the neighborhood in which the respondent lives. One way in which the intercept may vary is according to whether a school is nearby. Another way it may vary is by how far from the nearest town the neighborhood lies. We allow ${}_1$ to vary by neighborhood so that the effect of a woman s education on her initiation of permanent contraceptive use may vary by neighborhood because of the different schooling opportunities in each neighborhood. And, we allow ${}_2$ to vary by neighborhood because the effect of having a large number of children may be different in different neighborhoods.

Thus, the neighborhood level model is

$$_{0k} = {}_{00} + {}_{01}Dis_k + {}_{02}School_{tk} + {}_{0k} \qquad \text{(1b) CM}$$

$$_{1k} = {}_{10} + {}_{11}\,School_{tk} + {}_{1k}$$

$$_{2k} = {}_{20} + {}_{2k}$$

---

[5]We also estimated models with dichotomous indicators of each time period. The estimates of the other coefficients in these models did not change. Thus, we chose to parsimoniously represent time in our models with time and $time^2$.

$_3 = \ _{30}$

$_4 = \ _{40}$

The _ (= $_{0k}$, $_{1k}$, $_{2k}$) are unobserved error terms or random effects that model the correlation between the timing of contraceptive use by women in the same neighborhood; women in the same neighborhood share the same error terms. We assume that _ is multivariate normal with mean zero and unknown variance-covariance matrix.

In this model, $_{0k}$ represents the overall level of contraceptive use in neighborhood k, which varies by the neighborhood's distance to the nearest town and the presence of a school in the neighborhood. $_{1k}$ represents the effect of women's education for neighborhood k, which varies by the presence of a school in the neighborhood. $_{2k}$ represents the effects of the woman's total number of children for neighborhood k. Finally, $_3$, and $_4$ represent the effects of time since first birth and time^2 since first birth, respectively.

Note that this model, although expressed in models at each level, is actually only one model; to see this, substitute the level two equations for the s in the level 1 equation,

$$\text{Logit}(p_{tjk}) = (\ _{00} + \ _{01}\text{Dis}_k + \ _{02}\text{School}_{tk} + \ _{0k}) + (\ _{10} + \ _{11}\text{School}_{tk} + \ _{1k})\text{Educ}_j + (\ _{20} + \ _{2k})\text{Chldrn}_{tj} + \ _{30}\text{Time}_{tj} + \ _{40}\text{Time}^2_{tj}. \quad \text{(1c) CM}$$

This presentation of the model is likely to be more intuitive to those most familiar with single-level hazard analysis.

This conceptual model CM is used to make the likelihood derivations in the next section concrete. A multilevel model may be of more levels and may include more covariates. Furthermore, the link function need not be the logistic link; other link functions can be used and do not alter the likelihood derivations. For example, the log(-log) function, or the probit function could be used.

3. THE LIKELIHOOD

In this section we demonstrate that conditionally on the _, the likelihood for a group (neighborhood in our example) is given by the same formula as in Allison (1982, equations 18-20; also see Laird and Olivier 1981). This means that any program for maximum likelihood estimation in a multilevel regression analysis of a dichotomous dependent variable can be used to estimate the regression coefficients. This is analogous to Allison s (1982, pg. 74) result that single level discrete time hazard models can be estimated using programs for the analysis of dichotomous dependent variables such as Proc Logistic in SAS.

We formulate the model and the likelihood for one group; so in this section we omit the subscript k. We first derive the likelihood in the complete data setting, acting as if the response is not censored. Next, we allow for censoring and include all groups in the likelihood. To highlight our assumptions, we give this derivation for only two levels (individual and group); the derivation for three or more levels is similar. Suppose M is the total number of time periods possible for the entire study. In our example, time (t) is calender time, a period is one year, and M is 60 years. The j^{th} subject s response is the time to the event, T_j (initiation of contraception in our example). Alternately we denote the j^{th} response by Y_{tj}, t=1, . . ., M, where Y_{tj} is a time-varying dichotomous random variable with the value 1 if $T_j = t$ and 0 otherwise. Thus $Y_{tj} = 0$ for t=1, . . ., M except in the year of the event, when $Y_{tj} = 1$. Or, $Y_{tj} = 0$ for all t=1, . . ., M if $T_j > M$ (i.e. the event does not occur within the M periods). The j^{th} subject s at-risk variable, R_{tj}, is 1 if the j^{th} subject is at risk of the event at time t and is 0 otherwise. In our example, a woman is not at risk of contraceptive initiation before the birth of her first child or after age 49. Thus R_{tj} = 0 until the period after the j^{th} woman s first childbirth. The year after her first child s birth, R_{tj} = 1; it remains 1 until the year after she initiates contraception or until she reaches age 50, whichever occurs first. Sometimes when discussing multiple subjects from the same group simultaneously, we use the subscript i (in addition to the subscript j) to denote the "other" subject.

The collection of error terms in the group level model is denoted by _ (these are the errors, ${}_0$, ${}_1$, ${}_2$, in our neighborhood level model). The group level covariates at time t are GZ_t. IZ_{tj} denotes the j^{th} subject s individual level covariates at time t. After a subject experiences the event (initiation of contraception), we do not use the individual level covariates. Thus if t > T_j, IZ_{tj} is left undefined. All of the time-invariant individual level covariates are contained in IZ_{1j} and GZ_1 contains all of the time-invariant group level covariates. The variables for a group of size n are summarized in Table 1.

(Table 1, about here)

Our primary assumption is a *modeling assumption* for the conditional hazard probability that the j^{th} subject will experience the event at time t. That is, conditional on T_j t, on the error terms (_), on the covariates {GZ_s = gz_s, IZ_{sj} = iz_{si}, s   t, i=1, . . ., n}, and on the past responses by the other group members {Y_{si}, s< t, i   j}, the probability that T_j = t (or equivalently Y_{tj} = 1), is

_

That is, the hazard probability (p_{tj}) is assumed to be a function of only the j^{th} subject s past covariates and not the covariates of other subjects. Other subjects covariates thought to be predictive of the j^{th} subject s event time should be included in the j^{th} subject s covariates. For example if we believe that other subjects views on contraceptive use may influence the contraceptive use of the j^{th} subject, then we should include those views in the covariates for the j^{th} subject. Thus, in our example we make the modeling assumption that (1) given the woman does not initiate contraceptive use prior to year t; and (2) given her total number of children ($Childrn_{tj}$), education level ($Educ_j$), distance to the nearest town (Dis_k), presence of a nearby school ($School_{tk}$), and _, the chance that the j^{th} woman initiates contraceptive use in year t does not depend on other characteristics of her neighbors. Additionally we assume that the covariates are related to the hazard via the logistic function as given in the conceptual model CM.

The second assumption is *conditional independence*. In the multilevel nonlinear model for binary responses, the residual correlation between responses within a group are modeled by the error terms (e.g. see the conditional likelihoods in Rodriguez and Goldman (1995, eq. 7) or Hedeker and Gibbons (1994, eq. 2)). That is, conditional on the error terms (_) and the covariates, the responses within the group are modeled as independent. We make a similar conditional independence assumption for the multilevel hazard model. The *conditional independence* assumption is that the responses (Y_{tj}, j=1, . . ., n) are independent conditional on the error terms (_), on the covariates ({GZ_s, s t}, {IZ_{sj}, s t, j=1, . . ., n}), and on the past responses by all group members {Y_{sj}, s < t, j=1, . . ., n}. In our example, we are assuming that the j^{th} woman s decision to initiate contraceptive use at time t is independent of her neighbors contraceptive decisions at time t, given that she has yet to initiate contraceptive use, all other available observations on her neighbors at time t and conditional on the error terms in the neighborhood model. This assumption implies that, at least approximately, all of one woman s influence on her neighbor in terms of contraceptive use is via unmeasured neighborhood level influences (the _) and via observed shared neighborhood and individual characteristics.

Denote the conditional density of the group level covariates at time t given the error terms, past covariates, and past responses by $g_t(gz_t \mid \_, \{gz_s, iz_{sj}, y_{sj}, s < t\}, j=1, \ldots, n)$ and denote the conditional density of the n group members individual level covariates at time t given the error terms, past covariates, past responses, and time t group level covariates by $f_t(iz_{tj}, j=1, \ldots, n \mid \_, \{gz_s, s \quad t\}, \{iz_{sj}, y_{sj}, s < t\}, j=1, \ldots, n)$. The *modeling assumption* and the *conditional independence* assumption imply that the likelihood for Y_{tj}, GZ_t, IZ_{tj}, t=1, . . ., M; j=1, . . ., n given _ is

where we abbreviated p_{tj}, g_t, and f_t by omitting the conditioning sets.

To form the likelihood for the group s observed data we include the distribution of the error term and integrate out the error term, resulting in the rather complicated formula for the observed likelihood for a group,

_ _

where is the multivariate normal distribution with mean zero and unknown variance-covariance matrix. Without further assumptions, this likelihood is of little use. This is because the integral is quite complex, due to the fact that the covariate distributions (g_t and f_t) may depend on the error terms (_). Thus, our third assumption is that the covariates are *noninformative* of the error terms. That is, the covariates at time t (GZ_t, IZ_{tj}, j=1, . . ., n) are independent of the error term (_) given past responses and covariates (Y_{sj}, GZ_s, IZ_{sj}, $s< t$, $j=1, \ldots, n$). This assumption means that both g_t and f_t are not functions of _ and only p_{tj} remains a function of _; thus we may rewrite the above likelihood as,

_ _ (2)

This is one group s contribution to the likelihood.

The *noninformative* covariates assumption is not innocuous. If we conceptualize the error terms as unmeasured group level covariates, then we are assuming that conditional on past observations, the unmeasured group level covariates are independent of the covariates measured at time t. Thus at first it appears that the covariates should simply be independent of the error terms. However this independence property must hold conditional on the past responses. To illustrate the stringent nature of this assumption suppose that there is an unmeasured common cause of both the response at time 2 (Y_{2j}) and the individual level covariate at the future time 3 (IZ_{3j}). Denote this unmeasured common cause by U_{1j}. Our model assumes that _ is a predictor of Y_{2j}. Since Y_{2j} is predicted by both U_{1j} and _, we may expect that conditional on Y_{2j}, U_{1j} is correlated with _. Combining this (conditional on Y_{2j}, U_{1j} is correlated with _) with the fact that U_{1j} is a cause of IZ_{3j} implies that conditionally on Y_{2j}, _ will be a predictor of IZ_{3j}. Thus, although _ is marginally independent of the covariate IZ_{3j}, the presence of an unmeasured common cause of both IZ_{3j} and the response (Y_{2j}) will, in general, lead to a violation of the *noninformative* covariates assumption. Of course if the covariate is exogenous, for example a randomized treatment, then there are no unmeasured common causes of both the covariate and the response and the *noninformative* covariates assumption is satisfied. Alternatively we may believe that we have included all important common predictors of future responses and future covariate values in our collection of past covariates and that all measured covariates are uncorrelated with the error terms. In our extremely simplified model, we are assuming that the distance to the nearest town (Dis) and whether the woman ever went to school (Educ), combined with total number of children (Childrn) and the presence of a school (School) *in the past*, contain all common predictors of current contraceptive initiation, current number of children (Children), and current presence of a school (School). Furthermore, we are assuming that the total number of children (Chldrn), presence of a school (School), distance to the nearest town (Dis), and whether the woman ever went to school (Educ) are independent of _.

Suppose that some members of the group are observed for a shorter, nonrandom, period than the entire M intervals. For example, suppose that the jth subject is followed from time 1 until time c_j where c_j is a fixed constant. In this case the data for a group is Y_{sj}, IZ_{sj}, R_{sj}, for s c_j j=1, . . ., n and GZ_s, for s M. The likelihood from (2) is then,

– – (3)

where f_t is the conditional density of the individual level covariates (IZ_{tj}) for which t c_j, given the error terms, past covariates, past responses, and time t group level covariates.

Even if the covariate distributions in (3) contain information about the hazard regression coefficients, we can use only the last term in (3) in the estimation of the hazard regression coefficients (Gill 1992). This is because the last term in (3) is a partial likelihood as defined by Cox (1975) and Wong (1986). A partial likelihood is the product of conditional densities under the restriction that the conditioning sets are nested. The results of Gill (1992) imply that the last term in (3) is equal to

–   –

Because the set {GZ_s=gz_s, IZ_{sj}=iz_{sj}, s   min(u, c_j); Y_{sj}=y_{sj}, s < u, s c_j, j=1,...,n} contains the set {GZ_s=gz_s, IZ_{sj}=iz_{sj}, s   min(t, c_j),; Y_{sj}=y_{sj}, s < t, s c_j, j=1,...,n} for u less than t, this product is a partial likelihood.

We can write the (partial) likelihood for a sample of N groups by adding an additional subscript to denote the k^{th} group and multiplying across groups,

–   –   (4)

Wong (1986) showed that estimators found by maximizing a partial likelihood behave like maximum likelihood estimators. That is, the distribution of the estimators can be approximated by a normal distribution in large samples and the information matrix can be used to form standard errors; therefore we can use maximum likelihood software for inference on the parameters in the hazard probabilities, p_{tjk}. Note that the products of p_{tjk} in the integral from (4) form the same formulae as given by Allison (1982, equations 18-20) and Singer and Willet (1993, equations 7-10) for the partial likelihood when there are no unobserved error terms (_ is a constant).

The above arguments are valid when the c_{jk} s are fixed constants. However in most prospective longitudinal studies, subjects may attrit from the study and their response is censored at the time of attrition. In these cases, c_j is the time at which the subject leaves the study; thus CN_{jk} = c_{jk} where the use of CN indicates that the censoring time may be chosen by the subject. It is possible and plausible that CN_{jk} will be predictive of future (but unobserved) responses; in this case maximizing (4) to form estimators of the regression coefficients may not result in large-sample-unbiased estimators. Estimators based on (4) will be unbiased under further assumptions. We make a simpler, stronger version of Heitjan and Rubin s (1991) assumption, *coarsening at random*.[6] We assume that

and that the former probability is neither a function of the unknown regression coefficients nor of the unknown variance covariance matrix for _. In words, we assume that conditional on responses and covariates up to and including time c_{jk}, future values of the responses and covariates are not predictive of the chance that $CN_{jk} = c_{jk}$. When *coarsening at random* holds, Hietjan and Rubin (1991) show that we may treat the observed c_{jk} s as fixed constants in likelihood inference. Thus under this assumption we may use (4) as our (partial) likelihood.

Because censoring in longitudinal studies is common and may be related to the response, the *coarsening at random* assumption is often difficult to make. However, in our example, the retrospective nature of the data collection implies that this assumption is satisfied. Censoring

---

[6]See Heitjan and Rubin (1991) or Gill, Van Der Laan, and Robins (1997) for the weaker version of *coarsening at random*.

occurs at the time of the interview and is thus independent not only of the timing of contraception, but also of the covariates. In general, one attempts to measure all common predictors of future censoring and future responses/covariates and then includes these covariates in the regression model for time t.

Recall that the at-risk indicator (R_{tjk}) is 1 if the j^{th} subject in the k^{th} group is at risk of the event at time t, and is 0 otherwise. Thus when we have censoring, we set R_{tjk} to 0 for $t > CN_{jk}$. We may then rewrite (4) as,

(5)

As Allison notes (1982, p. 74),

is the same as the k^{th} group s contribution to the likelihood for the regression analysis of dichotomous dependent variables and thus, when there are no unobserved error terms, discrete-time hazard models can be estimated using programs for the analysis of dichotomous data (for example Proc Logistic in SAS or logistic in STATA). Similarly, conditional on the unobserved error terms, the above product is the same as the k^{th} group s contribution to the likelihood for a multilevel regression analysis of dichotomous dependent variables. Thus the partial likelihood in (5) is the same as the likelihood we would use in a multilevel regression analysis of a dichotomous dependent variable. Thus under the *modeling, conditional independence*, *noninformative covariates*, and *coarsening at random* assumptions, we may use any program for maximum likelihood estimation in a multilevel regression analysis of a dichotomous dependent variable. Goldstein (1995), without stating assumptions, notes this when all covariates are time-invariant. The extension described here incorporates time-varying covariates at both the individual and group levels.

4. ESTIMATION WITH COMMON SOFTWARE

We use two software packages, HLM and MLN, to illustrate the estimation of discrete-time multilevel hazard models (Bryk, Raudenbush and Congdon 1996; Goldstein, Rasbash, Plewis, Draper, Browne, Yang, Woodhouse, and Healy 1998). These software packages are widely used by sociologists. In order to estimate nonlinear multilevel models, both HLM and MLN use penalized quasi-likelihood to estimate the regression coefficients and variance-covariance matrix of the error terms (Breslow and Clayton 1993). This method is useful when the integral, as in equation (5), cannot be explicitly calculated; these methods approximate the integral and thus the resulting estimation method can be viewed as an approximation to maximum likelihood estimation for the regression coefficients. When the intraclass correlation[7] is large the approximation may not perform well; see for example Rodríguez and Goldman (1995). Better approximations are in progress (Raudenbush and Yang 1998). Additionally MLN provides a second order approximation which, in simulations, appears to work well even when the intraclass correlation is large (Goldstein and Rasbash 1996).

Our conceptual model CM in equation 1 must be modified for implementation in both

---

[7] The intraclass correlation is the correlation between two responses from the same group (Bryk and Raudenbush 1992:18).

HLM and MLN. Table 2 provides an overview of how CM is implemented differently in HLM and MLN. Nonetheless, both packages can be used to provide appropriate estimates of the parameters in a multilevel hazard model that includes time-varying covariates at the group as well as the individual level. Below, we illustrate how to implement our conceptual model using these software packages. Note also that SAS version 7 includes an experimental procedure, proc NLMIXED, that can be used with multilevel data structures and nonlinear dependent variables. Because proc NLMIXED uses only one input data file, CM in equation 1 must be modified similarly to the MLN modification.

4.1. *Using HLM*

Recall that HLM requires the input of two data sets (or three data sets for a three level model): an individual level data set and a group level data set. Because we are estimating a discrete-time hazard model, the individual level data set is a person-year data set. The person-year data set has multiple lines for each individual with the number of lines corresponding to the number of years the person is at risk. (See appendix Table 1.) The appropriate value for the time-varying individual level covariate in a particular year can be entered on the line corresponding to that year. The group level data set is composed of only one line per group. (See appendix Table 2.) In other words, there cannot be multiple observations (at multiple time points) of a particular variable for each neighborhood. Thus, when using HLM we must include the time-varying group level characteristics ($School_{tk}$ in our example) in the individual (level 1) equation. Thus, we can no longer include the time-varying group level characteristic as a predictor of HLM s level one intercept in the HLM implementation of CM. However, this HLM implementation is algebraically equivalent to CM, as we demonstrate below.

As before, we substitute the level two CM equations for the s in the level 1 CM equation to express CM in one equation:

$$\text{Logit}\ (p_{tjk}) = ({}_{00} + {}_{01}\text{Dis}_k + {}_{02}\text{School}_{tk} + {}_{0k}) + ({}_{10} + {}_{11}\text{School}_{tk} + {}_{1k})\text{Educ}_j + ({}_{20} + {}_{2k})\text{Chldrn}_{tj} + {}_{30}\text{Time}_{tj} + {}_{40}\text{Time}^2{}_{tj}$$

By rearranging terms and collecting the time-varying presence of a school (School_{tk}) terms in the last line we get

$$\text{Logit}\ (p_{tjk}) = ({}_{00} + {}_{01}\text{Dis}_k + {}_{0k}) + ({}_{10} + {}_{1k})\text{Educ}_j + ({}_{20} + {}_{2k})\text{Chldrn}_{tj} + {}_{30}\text{Time}_{tj} + {}_{40}\text{Time}^2{}_{tj} + {}_{02}\text{School}_{tk} + {}_{11}\text{School}_{tk}\text{Educ}_j$$

Now we rewrite this as an HLM two level model. We get the HLM <u>level 1 implementation</u>:

$$\text{Logit}\ (p_{tjk}) = {}'_{0k} + {}'_{1k}\text{Educ}_j + {}_{2k}\text{Chldrn}_{tj} + {}_3\text{Time}_{tj} + {}_4\text{Time}^2{}_{tj} + {}'_5\text{School}_{tk} + {}'_6\text{School}_{tk}\text{Educ}_{jk}, \qquad \text{(6a) HLM}$$

where we have placed primes on the s that do not coincide with CM, as the following HLM <u>level 2 implementation</u> indicates:

$$'_{0k} = {}_{00} + {}_{01}\text{Dis}_k + {}_{0k} \qquad \text{(6b) HLM}$$

$$'_{1k} = {}_{10} + {}_{1k}$$

$${}_{2k} = {}_{20} + {}_{2k}$$

$${}_3 = {}_{30}$$

$${}_4 = {}_{40}$$

$$'_5 = {}_{02}$$

$$'_6 = {}_{11}$$

Essentially, we are treating the time-varying neighborhood level covariate, distance to the nearest school, as a time-varying individual level covariate. The main difference between this and a real individual level time-varying covariate is that we do not allow $_5$ and $_6$ to vary by neighborhood. This two level re-expression of our model can be estimated by the HLM software because the new level 2 equations vary by neighborhood but not over time.

Table 2 explicitly compares CM and its implementation in HLM. Note that if a term in the HLM equation coincides with the in CM, we use the same symbol. However, we use a prime symbol (') to indicate that a does not correspond to CM. For example, because we moved $School_{tk}$ from the level 2 equation to the level 1 equation, the level 2 equations for $_{0k}$ and $_{1k}$ change. We indicate this by using $'_{0k}$ and $'_{1k}$ in the HLM implementation. However, because the level 2 equations for $_{2k}$, $_3$, and $_4$ *do not change*, we continue to use $_{2k}$, $_3$ and $_4$. And, because $'_5$ and $'_6$ are new terms, we use the prime symbol.

(Table 2, about here)

It must be emphasized that the s and -primes in equation 6 are used only as a conceptual way to group the terms. All implementations produce estimates of the terms (and their standard errors) in our conceptual model (CM). HLM forces us to group the terms differently because it does not allow time-varying covariates in the level 2 equation. Only the conceptual model corresponds to a grouping of the terms that matches the usual understanding of a multilevel model (see description of s on page 10). Note that $'_0$ *can not* be interpreted as the average level of contraceptive use in neighborhood k according to our model, CM. $'_0$ is used in HLM simply because the software program requires the input of two separate equations: one level 1 equation and a separate level 2 equation for each of the coefficients in the level 1 equation. The predicted level of contraceptive use for each neighborhood can be constructed from the HLM parameters according to the CM level 2 model ($_{0k} = \ _{00} + \ _{01}Dis_k + \ _{02}School_{tk}$). Table 2 illustrates the distinct groupings of terms within s that are required when implementing CM in HLM or MLN.

4.2. *Using MLN*

In MLN, *all* of the neighborhood characteristics must be included in the individual level equation. This is because MLN only uses one input data file. (See Appendix Table 3). Thus, to use MLN we include all individual and neighborhood characteristics in the level one equation. Table 2 illustrates the differences between the new model (MLN) and the previous models CM and HLM. In the MLN implementation, distance to the nearest town (Dis_k) is moved from the level 2 equation to the level 1 equation, resulting in a new parameter, $''_7$. This results in a further change to $'_{0k}$, indicated by $''_{0k}$. This model is still equivalent to CM, and its derivation is similar to the derivation for the HLM implementation, so we do not show it in detail. The resulting <u>level 1 implementation</u> in MLN is as follows:

$$\text{Logit}(p_{tjk}) = ''_{0k} + '_{1k}\text{Educ}_j + {}_{2k}\text{Chldrn}_{tj} + {}_{3}\text{Time}_{tj} + {}_{4}\text{Time}^2_{tj} + '_5\text{School}_{tk} + '_6\text{School}_{tk}\text{Educ}_j + ''_7\text{Dis}_k \quad \text{(7a) MLN}$$

The MLN <u>level 2 implementation</u> is

$$''_{0k} = {}_{00} + {}_{0k} \quad \text{(7b) MLN}$$

$$'_{1k} = {}_{10} + {}_{1k}$$

$${}_{2k} = {}_{20} + {}_{2k}$$

$${}_{3} = {}_{30}$$

$${}_{4} = {}_{40}$$

$$'_{5} = {}_{02}$$

$$'_{6} = {}_{11}$$

$''_{7} = \ {}_{01}$

Essentially, we have included all of the neighborhood characteristics in the level 1 implementation and use the level 2 implementation only to specify which effects are modeled as random. This model is the statistical equivalent of our more intuitive conceptual model, CM, presented in equation (1) where neighborhood characteristics are included in the level 2 model.

5. RESULTS OF APPLYING THIS TECHNIQUE TO OUR EMPIRICAL EXAMPLE

Now we return to our empirical example to demonstrate estimation of the discrete-time multilevel hazard model in HLM and MLN. We relate our results below to the parameters in the conceptual model CM presented in equation 1. Recall that Table 2 illustrates how to reconcile CM with its implementation in both HLM and MLN, illustrated in equations (6) and (7). Descriptive statistics for each of the measures used in analyses are presented in Table 3.

(Table 3, about here)

Model 1 in Table 4 includes the cross level effect term that allows us to test H3. The estimate of the cross level effect (${}_{11}$) is not statistically significant and therefore indicates that the influence of the woman s education level does not vary with the availability of educational opportunities within the neighborhood in which she lives.

(Table 4, about here)

Model 2 includes neighborhood characteristics as predictors of the mean hazard of contraceptive use in a particular neighborhood (${}_{0k}$), but not as predictors of the effect of a woman s education on her propensity to use contraception (${}_{1k}$). In other words, the cross level effect has been deleted from Model 2. The estimates in model 2 correspond to the following conceptual model, which is very similar to CM, except that $School_{tk}$ is not a predictor of $_{1k}$ in equation (8).

<u>Individual level model</u>

$$\text{Logit}(p_{tjk}) = {}_{0k} + {}_{1k}Educ_j + {}_{2k}Chldrn_{tj} + {}_{3}Time_{tj} + {}_{4}Time^2_{tj} \quad (8a)$$

<u>Neighborhood level model</u>

$$ {}_{0k} = {}_{00} + {}_{01}Dis_k + {}_{02}School_{tk} + {}_{0k} \quad (8b)$$

$$ {}_{1k} = {}_{10} + {}_{1k}$$

$$ {}_{2k} = {}_{20} + {}_{2k}$$

$$ {}_{3} = {}_{30}$$

$$ {}_{4} = {}_{40}$$

This model allows us to evaluate H1, which asks whether formal education is associated with a higher hazard of permanent contraceptive use. In this model $_{10}$ indicates that formal education *is* associated with a higher hazard of contraceptive use, holding constant other factors in the model. The magnitude of $_{10}$ indicates that having attended school is associated with a .58 higher log-odds of permanent contraceptive use. Note that the estimates of the terms obtained using HLM and MLN are similar, but not exactly the same. This is because HLM and MLN use slightly different methods to approximate the likelihood and because with MLN we used the second-order approximation option.

Model 2 also allows us to evaluate H2, whether nearby schooling opportunities are associated with a higher hazard of permanent contraceptive use. This hypothesis can be evaluated by examining $_{02}$. The coefficient, .29 (.30 in HLM), indicates that having a school within a 5-

minute walk is associated with a .29 higher log-odds of permanent contraceptive use. We do not describe the other parameters in the model, but their interpretations are straightforward.

Again, note that the s are not directly estimated in either statistical package. However, we can compute predicted s using the estimates of the terms provided by either HLM or MLN. For instance, according to our modified conceptual model in equation (8), the predicted value of $_{0k} = {}_{00} + {}_{01}\text{Dis}_k + {}_{02}\text{School}_{tk}$. Thus, according to model 2 in Table 4, for a neighborhood with average distance to the nearest town (8.24) and a school within a 5 minute walk, the predicted value of $_{0k}$ = -4.48 + -.04(8.24) + .29(1) = -4.5196.

Finally, note the estimates of the variances of the random effects at the bottom of Table 4. Var($_{0k}$), which is estimated to be .01 by HLM and .00 by MLN, is the variance in the intercept that is not explained by the neighborhood level variables in the model. For the intercept term, this is quite small. In other words, there is little variation between neighborhoods in the intercept that is not explained by the presence of a school (School) and the distance to the nearest town (Dis). Similarly, Var($_{1k}$) is the variance in the coefficient for woman s education (Educ) that is not explained by the presence of a school. The estimate of Var($_{1k}$) computed by HLM in both model 1 and model 2 is quite large relative to the size of the effect; in other words, this indicates substantial variance across neighborhoods in the impact of woman's education on the hazard of contraceptive use net of the presence of a school. Note, however, that the estimate of this variance component computed by MLN is zero. Var($_{2k}$) is the variance in the estimated effect of total number of children (Childrn) net of the mean. The estimates produced by both HLM and MLN are substantial relative to the size of the effect of total number of children. This also indicates variability in the effect of total number of children across neighborhoods. Overall, however, note that these variances are not estimated precisely. The different approximations to the likelihood used by HLM and MLN produce different estimates, and the random error inherent in sampling procedures adds to the lack of precision. Thus, the variances of the random effects should be interpreted with caution.

6. CONCLUSION

In this paper we described conditions under which any software package using maximum likelihood estimation for multilevel logistic regression models may be used to perform a multilevel discrete-time hazard analysis with time varying covariates at both individual and group levels. In particular we have demonstrated the use of HLM and MLN software to estimate this discrete-time multilevel hazard model. Both of these software packages are widely used by sociologists, and either can be used to estimate this type of model. The keys to their use lie in creation of the input data sets and interpretation of the output coefficients. The results generated by these estimation procedures are quite similar, though minor differences result from slight variations in the approximations used in the two packages. Both packages are easily available and a wide range of sociologists will find them useful for estimating discrete-time multilevel hazard models.

In order to use this method we made *modeling, conditional independence*, *noninformative covariates* and *coarsening at random* assumptions. The last two assumptions imply that we must measure all common predictors of the event time, covariates, and censoring in order to use multilevel hazard analysis. This is rarely successful in sociological studies. Further research is needed on methods for relaxing these assumptions. To test sociological models of macro-micro linkage, it is particularly important to devise methods that provide unbiased estimates of group level effects (neighborhoods, schools, businesses) even when more proximate individual level predictors are omitted. Such methods are required to establish unbiased estimates of the total impact of macro characteristics on micro behavior and outcomes. The research reported here is an initial step toward that goal.

APPENDIX. CONSTRUCTING THE DATA FILES

1. *Data File for Use with HLM*

To fit a two level model with HLM, two data files must be created. The first data file will hold the information on the groups. The second data file will have the information on the individuals. Note that the current version of HLM (version 4) does not allow missing data when estimating nonlinear models.

In our example, the neighborhood file has one data line per neighborhood, for a total of 171 lines of data. The variables on each line will be the neighborhood ID and the value of the time-invariant neighborhood covariate, distance to the nearest town. Appendix Table 1 shows our group level data file for use with HLM.

The second data file contains information about the 1,395 women. Because we are conducting a discrete-time hazard analysis, one line of data represents a person-year. For example, if a woman is at risk of permanent contraceptive use for five years (from the year after her first marital birth until contraceptive use or censoring), that woman will be in the data set five times. The first field in the file is the group ID, which in our example is the ID of the neighborhood in which the woman lives. Note that the HLM software requires that all of the individual level observations are grouped together by their respective group level ID. The next columns contain the individual level covariates, both time-invariant and time-varying. The next column contains the time-varying neighborhood level covariate, the presence of a school within a five-minute walk, which we are treating as an individual level covariate for use in HLM. The subsequent column contains the cross-product of the neighborhood level time-varying covariate and individual level variable. In our case, there is one cross-product; the product of the presence of a school and the woman s education. Finally, the last column contains the response variable, contraceptive use. In this data, the response will be 0 if the woman did not use a permanent contraceptive method during that year, and 1 if the woman did use a permanent contraceptive method during that year. Note that most records will have contraceptive use equal to zero. Only the last year for each woman can have a response equal to 1. On the woman s last year she was either censored (did not use a permanent contraceptive method before the end of the study), where contraceptive use is coded 0, or used a permanent contraceptive method, where contraceptive use is coded 1, and is subsequently no longer at risk (and thus subsequent observation-years are not in the data set for this woman). Appendix Table 2 shows the individual level data file for use with HLM.

2. *Data File for Use with Mln*

The data set for use with MLN is simpler because there is only one, rather than two, data files used to estimate the models. Thus, time-varying and time-invariant group and individual level covariates are included in the same file. This data set is similar to the individual level file used with HLM; however, it also includes the group level variables. Thus, in this data set, one line again represents one person-year. Note that MLN requires three additional variables, CONS, BCONS, and DENOM that are equal to 1 in all observations. (See Goldstein et al. 1998: 97-101.) Appendix Table 3 illustrates the MLN data set.

References

Alexander, Jeffrey C. 1988. Action and Its Environments: Toward a New Synthesis. New York: Columbia University Press.

Allison, Paul D. 1982. Discrete-time Methods for the Analysis of Event Histories. In Sociological Methodology 1982, edited by Samuel Leinhardt, 61 - 98. San Francisco: Jossey-Bass.

-----. 1984. Event History Analysis: Regression for Longitudinal Event Data. Beverly Hills: Sage Publications.

Axinn, William G. 1993. The Effects of Children s Schooling on Fertility Limitation. Population Studies 47:481-493.

Axinn, William G. and Jennifer S. Barber. 1999. "The Spread of Mass Education and Fertility Limitation. Paper presented at the annual meetings of the Population Association of America, March 27-29, New York, NY.

Axinn, William G., Jennifer S. Barber, and Dirgha J. Ghimire. 1997. The Neighborhood History Calendar: A Data Collection Method Designed for Dynamic Multilevel Modeling. In Sociological Methodology 1997, edited by Adrian Raftery, 355-392. Boston: Blackwell Publishers.

Axinn, William G., Lisa D. Pearce, and Dirgha J. Ghimire. 1999. Innovations in Life History Calendar Applications. Social Science Research 28:243-264.

Breslow, N. and D. G. Clayton. 1993. Approximate Inference in Generalized Linear Mixed Models. Journal of the American Statistical Association 88: 9-25.

Brewster, Karin L. 1994. Race Differences in Sexual Activity Among Adolescent Women: The Role of Neighborhood Characteristics. American Sociological Review 59: 408-424.

Bryk, Anthony S. and Stephen W. Raudenbush. 1992. Hierarchical Linear Models: Applications and Data Analysis Methods. Newbury Park, CA: Sage.

Bryk, Anthony, Stephen Raudenbush, and Richard Congdon. 1996. HLM: Hierarchical Linear and Nonlinear Modeling with the HLM/2l and HLM/3l Programs. Chicago: Scientific Software International.

Coleman, James S. 1990. Foundations of Social Theory. Cambridge: Harvard University Press.

Cox, D. R. 1975. Partial Likelihood. Biometrika 62(2): 269-276.

DiPrete, Thomas A. and Jerry D. Forristal. 1994. Multilevel Models: Methods and Substance. Annual Review of Sociology 20:331-357.

Diggle, Peter J., Kung-Yee Liang, and Scott L. Zeger. 1994. Analysis of Longitudinal Data. Oxford : Clarendon Press.

Durkheim, Emile. 1984 [1933]. The Division of Labor in Society. New York: Free Press.

Elder, Glen H., Jr. 1983. The Life Course Perspective. In The American Family in Social-Historical Perspective, 3rd Edition, edited by Michael Gordon, 54-60. New York: St. Martin s Press.

Elder, Glen H., Jr. 1977. Family History and the Life Course. Journal of Family History 2: 279-304.

Gill, Richard D. 1992. Marginal Partial Likelihood. Scandinavian Journal of Statistics. Theory and Applications 19: 133-137.

Gill, Richard D., M. J. Van Der Laan, and J. M. Robins. 1997. Coarsening at Random: Characterizations, Conjectures, Counter-Examples. In The Proceedings of the First Seattle Symposium on Biostatistics: Survival Analysis, Vol. 123 of Springer Lecture Notes in Statistics, edited by D. Y. Lin and T. R. Fleming, 255-294. New York: Springer Verlag. .

Goldstein, H. 1995. Multilevel Statistical Models, 2nd Ed. New York: Halsted Press.

Goldstein, H. and J. Rasbash. 1996. Improved Approximations for Multilevel Models with Binary Responses. Journal of the Royal Statistical Society, Series A 159: 505-513.

Goldstein, H., J. Rasbash, I. Plewis, D. Draper, W. Browne, M. Yang, G. Woodhouse and M. Healy. 1998. A User s Guide to MLwiN. London: Multilevel Models Project.

Guo, Guang. 1993. Use of Sibling Data to Estimate Family Mortality Effects in Guatemala. Demography 30: 15-32.

Guo, Guang and Germán Rodríguez. 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: 960-976.

Hedeker, Donald and Robert D. Gibbons. 1994. A Random-Effects Ordinal Regression Model for Multilevel Analysis. Biometrics 50: 993-944.

Hedeker, Donald, Ohidul Siddiqui, and Frank B. Hu. 1998. Random-effects Regression Analysis of Correlated Grouped-time Survival Data. Unpublished Manuscript. University of Illinois at

Chicago.

Heitjan, Daniel F., and Donald B. Rubin. 1991. Ignorability and Coarse Data. The Annals of Statistics 19: 2244-2253.

Kreft, Ita G. G. 1994. Multilevel Models for Hierarchically Nested Data: Potential Applications in Substance Abuse Prevention Research. In Advances in Data Analysis for Prevention Intervention Research, edited by Linda M. Collins and Larry A. Seitz, 140-183. NIDA Research Monograph No. 142. Washington, D.C.: U.S. Department of Health and Human Services.

Kreft, Ita G. G., Jan DeLeeuw, and Leona S. Aiken. 1995. The Effect of Different Forms of Centering in Hierarchical Linear Models. Multivariate Behavioral Research 30(1):1-21.

Laird, Nan and Donald Olivier. 1981. Covariance Analysis of Censored Survival Data Using Log-linear Analysis Techniques. Journal of the American Statistical Association 76: 231-240.

Ma, Xin and J. Douglas Willms. 1999. "Dropping Out of Advanced Mathematics: How Much Do Studnets and Schools Contribute to the Problem?" Educational Evaluation and Policy Analysis 21:365-383.

Massey, Douglas S. and Kristin E. Espinosa. 1996. What s Driving Mexico-U.S. Migration? A Theoretical, Empirical, and Policy Analysis. American Journal of Sociology 102:939-999.

Muthén, Bengt. 1997. Latent Variable Modeling of Longitudinal and Multilevel Data. In Sociological Methodology 1997, edited by Adrian E. Raftery, 453-480. Boston: Blackwell Publishers.

Petersen, Trond. 1991. The Statistical Analysis of Event Histories. Sociological Methods and Research 19:270-323.

Raudenbush, Stephen W. and M. Yang. 1998. Numerical Integration via High-Order, Multivariate LaPlace Approximation with Application to Multilevel Models. In Multilevel Modeling Newsletter, vol. 10(2), pp. 11-14. London: Multilevel Models Project.

Rindfuss, Ronald R., S. Philip Morgan, and Gray Swicegood. 1988. First Births in America: Changes in the Timing of Parenthood. Berkeley: University of California Press.

Ringdal, Kristen. 1992. Recent Developments in Methods for Multilevel Analysis. Acta Sociologica 35:235-243.

Rodríguez, Germán and Noreen Goldman. 1995. An Assessment of Estimation Procedures for Multilevel Models with Binary Responses. Journal of the Royal Statistical Society, Series A, 158: 73-89.

Sastry, Narayan. 1996. Community Characteristics, Individual and Household Attributes, and Child Survival in Brazil. Demography 33: 211-229.

Sastry, Narayan. 1997. Family-level Clustering of Childhood Mortality Risk in Northeast Brazil. Population Studies 51: 245-261.

Singer, J. D. and J. B. Willet. 1993. It s about Time: Using Discrete-time Survival Analysis to Study Duration and the Timing of Events. Journal of Education Statistics 18: 155-195.

Smith, Herbert L. 1989. Integrating Theory and Research on the Institutional Determinants of Fertility. Demography 26:171-184.

Thornton, Arland, William G. Axinn, and Jay Teachman. 1995. The Influence of Educational Experiences on Cohabitation and Marriage in Early Adulthood. American Sociological Review 60(5):762-764.

Trussell, James and Toni Richards. 1985. Correction for Unmeasured Heterogeneity in Hazard Models Using the Heckman-Singer Procedure. In Sociological Methodology 1985, edited by Nancy Brandon Tuma, 242-276. San Francisco: Josey-Bass.

Vaupel, James W. 1988. Inherited Frailty and Longevity. Demography, 25: 277-287.

Vaupel, James W. and Analtoli I. Yashin. 1985. Heterogeneity s Ruses: Some Surprising Effects of Selection on Population Dynamics. The American Statistician 39: 176-185.

Weber, Max. 1922. The Sociology of Religion. London: Methuer and Co.

Wong, George Y. and William M. Mason. 1985. The Hierarchical Logistic Regression Model for Multilevel Analysis. Journal of the American Statistical Association 80:513-524.

Wong, Wing Hung. 1986. Theory of Partial Likelihood . The Annals of Statistics 14:88-123.

Yamaguchi, Kazuo. 1991. Event History Analysis. Newbury Park, CA: Sage.