It's about Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events

Judith D. Singer; John B. Willett

It's About Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events

Judith D. Singer and **John B. Willett**
Harvard University
Graduate School of Education
and
National Center for Education Statistics

Key words: *survival analysis, event history analysis, longitudinal data*

Educational researchers frequently ask whether and, if so, when events occur. Until relatively recently, however, sound statistical methods for answering such questions have not been readily available. In this article, by empirical example and mathematical argument, we demonstrate how the methods of discrete-time survival analysis provide educational statisticians with an ideal framework for studying event occurrence. Using longitudinal data on the career paths of 3,941 special educators as a springboard, we derive maximum likelihood estimators for the parameters of a discrete-time hazard model, and we show how the model can be fit using standard logistic regression software. We then distinguish among the several types of main effects and interactions that can be included as predictors in the model, offering data analytic advice for the practitioner. To aid educational statisticians interested in conducting discrete-time survival analysis, we provide illustrative computer code (SAS, 1989) for fitting discrete-time hazard models and for recapturing fitted hazard and survival functions.

How long do teachers stay in teaching? How many years does it take to earn a doctorate? How long do children with disabilities receive special education services? Educational researchers often pose such questions about whether and, if so, when a variety of events occur. But until relatively recently, sound statistical methods for answering these questions about time have not been readily available. In this article, by empirical example and mathematical argument, we demonstrate how the set of statistical methods known as *discrete-time survival analysis* provide educational researchers with an ideal framework not only for answering these descriptive questions but for modeling the relationship between event occurrence and predictors as well.

Since the 1972 publication of Cox's seminal article on statistical models for lifetime data, survival methods—especially those for continuous-time data—have enjoyed increasing popularity in a variety of disciplines ranging from medicine and industrial product testing to sociology and economics. Statistical packages now routinely include programs for fitting at least one type of model, usually the continuous-time proportional hazards model (Goldstein et al., 1989; Harrell & Goldstein, in press). And in addition to the classic technical references (Cox & Oakes, 1984; Kalbfleisch & Prentice, 1980; Miller, 1981), several excellent applied overviews are now in print (Allison, 1982; Blossfeld, Hamerle, & Mayer, 1989; Yamaguchi, 1991).

These facilitators notwithstanding, we believe that researchers contemplating the use of survival analysis to model educational data still face three obstacles. First, most readily available software is designed for fitting models that incorporate only time-invariant predictors (those whose values are constant over time). Yet the values of many predictors of educational processes—such as financial aid, the availability of support and remedial programs, and the nature of the peer support network—fluctuate naturally with time. Second, the most popular model in use today (the continuous-time proportional hazards model) is predicated on the often unrealistic assumption that the effect of a predictor on event occurrence is constant over time. Yet in many educational applications, the effects of predictors—such as teacher salary or peer pressure—will vary over time. Third, continuous-time models (in which researchers assume that they know the precise instant when the event occurs) do not adapt readily to school contexts, where time is so often measured discretely, in quarters, semesters, or years.

For these reasons, our presentation focuses on discrete-time survival analysis, an intuitively appealing, general approach to survival analysis that overcomes all of these problems. As we will show, a variety of models can be fit under the discrete-time rubric. Predictors whose values vary over time can be modeled as can time-invariant ones. So, too, can predictors whose effects vary over time and those whose effects are constant. By postulating that events occur in discrete time periods, we need not worry about the

inaccuracy of statistical adjustments invoked when adapting continuous-time methods to discrete-time data (Pierce, Stewart, & Kopecky, 1979; Prentice & Gloeckler, 1978). And as an added boon to empirical researchers, the models of discrete-time survival analysis can be fit using standard logistic regression analysis software.

We begin our presentation heuristically, by examining longitudinal data from Singer (1993) describing the careers of 3,941 special educators. Although somewhat unorthodox, by using real data to identify the problems inherent in studying event occurrence, we lay the groundwork for our later arguments, substantiating the need for survival methods and providing a context for our work. We then describe the foundations of discrete-time survival analysis, presenting the underlying statistical model and deriving maximum likelihood procedures for estimating its parameters that, surprisingly, can be implemented using logistic regression software. Finally, we provide practical data-analytic advice for fitting discrete-time hazard models, distinguishing among the different types of effects that can be explored within the discrete-time survival analysis framework, and showing how these models can be used to address substantively interesting questions about time.

Why Are Special Statistical Methods
Needed for Studying Event Occurrence?

Statistical obstacles await researchers studying duration and the timing of events. The core problem is that, no matter when data collection begins and no matter how long any subsequent follow-up takes, some people may not experience the target event before data collection ends, some students may not graduate, some teachers may not quit teaching, some students may still be in special education. These observations are *censored*.[1]

Individuals with censored data provide incomplete information about event occurrence. If an individual's event time is censored, the researcher knows only that, if the person ever experiences the event, he or she will do so after data collection ends. The researcher knows neither when the event will occur nor even whether it will happen. All the researcher knows is that by the end of data collection, a time that may be arbitrary with respect to the underlying process, the event has not yet occurred.

Censoring's toll can be seen in Table 1, which presents data describing the career paths of 3,941 special educators hired by the Michigan public schools between 1972 and 1978 (for a description of the sample and a full analysis, see Singer, 1993). The entries indicate whether and, if so, when these teachers stopped teaching in the state between their year of hire and 1985, when data collection ended. The term *year* in the first column refers not to chronological year but to the year of the teaching career—1st year, 2nd year, through the 12th year; Year 1 is 1972 for those hired in 1972, 1973 for those

TABLE 1
What do survival data look like? Number of consecutive years in teaching for 3,941 special educators hired in Michigan between 1972 and 1978

	Number who			Proportion of	
Year	Were teaching at the beginning of the year	Left during the year	Were censored at the end of the year	All teachers still teaching at the end of the year	Teachers at the beginning of the year who left by the end of the year
1	3,941	456	0	0.88	0.12
2	3,485	384	0	0.79	0.11
3	3,101	359	0	0.70	0.12
4	2,742	295	0	0.62	0.11
5	2,447	218	0	0.57	0.09
6	2,229	184	0	0.52	0.08
7	2,045	123	280	*0.49*	0.06
8	1,642	79	307	*0.46*	0.05
9	1,256	53	255	*0.44*	0.04
10	948	35	265	*0.43*	0.04
11	648	16	241	*0.42*	0.02
12	391	5	386	*0.41*	0.01

hired in 1973, and so on. The next three columns tally the number of teachers employed at the beginning of each year, the number who stopped teaching by the end of the year, and the number whose careers were censored at the end of the year (still teaching when data collection ended). By the end of data collection (1985), 2,207 teachers had left teaching; the 1,734 who did not (44% of the sample) are censored.

How can we describe the distribution of these special educators' careers? Empirical researchers faced with similar data structures have used a variety of ad hoc strategies, none completely satisfactory. Some have focused exclusively on those people with noncensored event times (e.g., Abedi & Benkin, 1987; Boli, Katchadourian, & Mahoney, 1988). Implementing this strategy in Table 1, using data from only the 2,207 teachers who left teaching before the end of data collection, yields an estimated mean career duration of 3.7 years. But excluding the 1,734 censored teachers obviously distorts the complete distribution of career duration. The "average" employment duration for all teachers must be longer than 3.7 because this estimate uses data from only those teachers with short careers.

Cognizant of these concerns, some researchers have included censored cases in their analyses by imputing unknown event times. The most popular

approach is to assign censored cases their duration value when data collection ended (Frank & Keith, 1984; Kahn & Kulick, 1975). Using this approach in our example (e.g., assigning career lengths of 7 to the 280 teachers censored at 7 years) yields an estimated mean career duration of 7.5. While suitably longer than our first underestimate, this estimate, too, is surely wrong. Imputing event times for censored cases changes a "nonevent" into an "event," and it further assumes that all these new events occur at the earliest times possible!

Given these errors, others have chosen a more conservative approach. Rather than imputing event times for censored cases, they dichotomize at some particular (and often arbitrary) time and ask whether the event occurs by that time (e.g., Ott, 1988; Stage, 1988). Although better than data elimination or incorrect imputation, dichotomization does not resolve the censoring dilemma; it simply obscures it from view. Dichotomization discards information, eliminates potentially meaningful variation in event times, and prevents a researcher from answering such key questions as "How long does the average teacher teach?"

Clearly a better analytic strategy is needed, and, in the next section, we begin to formulate one. But before doing so, let us continue heuristically, by returning to Table 1 and speculating about meaningful ways of summarizing the distribution of event times, even in the face of censoring. The fifth column of Table 1 presents one useful summary—the proportion of all teachers still teaching at the end of each year. The survival analysis literature uses the term *survival probability* to refer to the proportion of an initial population that survives through each of several successive time periods and the term *survivor function* to refer to the chronological pattern of these probabilities over time. Examining the *sample survivor function* in the fifth column of Table 1, we see that 88% of the teachers teach (survive) more than 1 year, 79% teach more than 2 years, and that 52% teach more than 6 years. Because we do not know what happens to the 1,734 censored teachers, we cannot directly compute the sample survivor function beyond Year 6 (although, as we show below, it can be estimated indirectly).

The sixth column of Table 1 presents another useful summary—the proportion of teachers known to be teaching at the beginning of the year who left by the end of the year. Defining the *risk set* as the group of people known to be eligible to experience the event in a particular time period (those at risk), the survival analysis literature uses the term *hazard probability* to refer to the proportion of the risk set who experiences the event in that time period and the term *hazard function* to refer to the chronological pattern of these probabilities over time. In the top panel of Figure 1, we have plotted the hazard function for our sample of special educators (following Miller's 1981 suggestion, we have interpolated linearly between discrete sample hazard probabilities, rather than plotting the hazard profile as a step func-
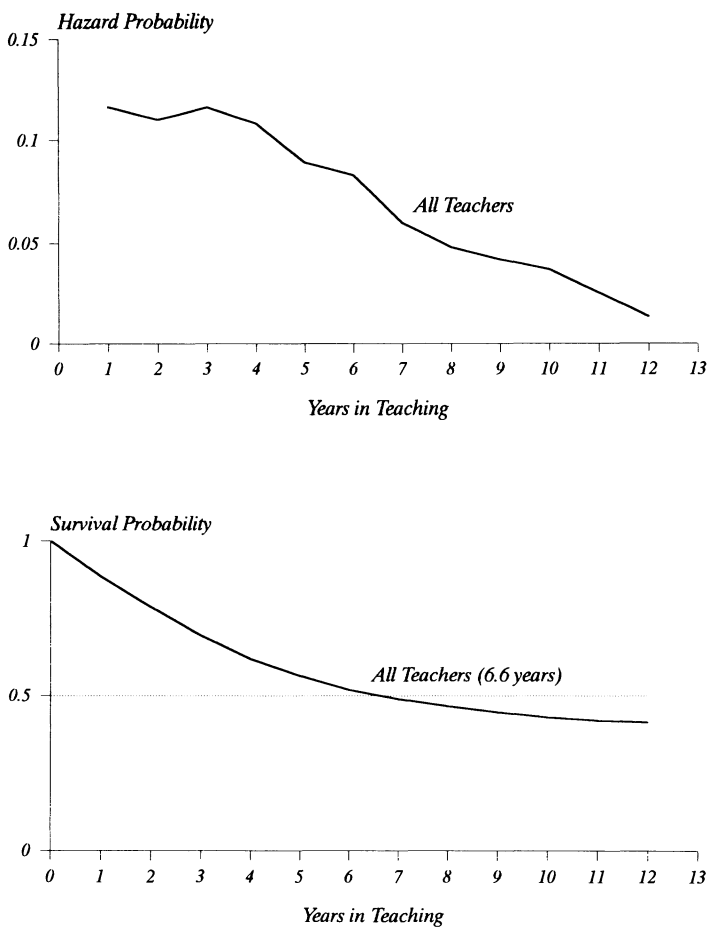
FIGURE 1. *Sample survivor and hazard functions for all special educators hired in Michigan between 1972 and 1978 (with estimated median lifetime in parentheses on the plot of the sample survival function)*

tion). Examining this panel, we see that, among these 3,941 special educa-tors, 12% left by the end of their first year. Of the 3,485 teachers who stayed 1 year, 11% left by the end of their second. Unlike the sample survivor function, we can compute the sample hazard function in every year of data collection regardless of censoring. Of the 2,045 teachers who taught contin-uously for 6 years, for example, 6% left by the end of their seventh; of the 948 teachers who taught continuously for 9 years, 4% left at the end of their 10th.

A key feature of hazard is that it is computed using each year's risk set,

which consists of those people known to be eligible to experience the event in that year. The 12th-year risk set, for example, consists of 391 teachers known to be eligible to leave in Year 12: the 648 teachers in the 11th-year risk set less the 16 who left in Year 11 and the 241 who were censored in Year 11. Basing hazard on the risk set ensures that we can compute hazard in every year even in the face of censoring.

Of course, this definition's credibility requires that censoring is unrelated to event occurrence, which is known as *independent censoring*. Under independent censoring, each year's risk set is representative of all teachers who would be teaching in that year; censored individuals do not differ from those who remain. This allows us to generalize the behavior of people in the risk set back to the entire sample and hence back to the original population. If censoring is not independent, individuals in the risk set differ systematically from censored individuals, and the generalization may be incorrect. For these reasons, we assume, as others generally do, that censoring is independent.[2]

The hazard function has many appealing properties which, taken together, explain why it—and not the survivor function—forms the cornerstone of survival analysis. First and foremost, the hazard function assesses exactly what we want to know—whether and, if so, when events occur. Its magnitude in each year indicates the risk of event occurrence in that year— the higher the hazard, the greater the risk. Examining the sample hazard function in Figure 1, for example, we see that special educators are most likely to leave teaching early in their careers. After the initial hazardous years, only a small proportion of continuing teachers leaves. Second, the hazard function appropriately includes data from both noncensored and censored cases; we do not discard teachers or arbitrarily impute their event times even if we do not know when or whether they will leave teaching. Third, the sample hazard function can be computed in every year when event occurrence is recorded; information on variation in the timing of events is not ignored.

Last but not least, under the assumption of independent censoring, we can use the sample hazard function to estimate the sample survivor function in those years when censoring precludes its direct computation. There is an inextricable link between the survival and hazard probabilities that we will illustrate using data from the sixth and seventh years of the teaching career displayed in Table 1. From row 6, we know that 52% of all teachers survive through the sixth year of their careers. From row 7, where the sample hazard probability is 6%, we know that 94% of the entering seventh-year risk set will not leave teaching. Therefore, we estimate that, at the end of the seventh year, 94% of the entering 52% will remain. An estimate of the survival probability at the end of Year 7 is simply (.52) × (.94), or 49%. In other words, the sample survival probability in any year is simply one minus

the hazard probability for that year multiplied by the sample survival probability from the previous year. Providing censoring is independent of event occurrence, this formula can be used to fill in the sample survival probabilities in Table 1 for Years 7 through 12 (shown in italics). We have plotted the obtained sample survivor function in the lower panel of Figure 1.

Having used the sample hazard function to estimate the remaining values of the sample survivor function, we can compute a third summary statistic—the *median lifetime*. When the sample survivor function equals .50, half of the teachers in the sample have left teaching; half have not. From the lower panel of Figure 1, using linear interpolation between years (Miller, 1981), the estimated median lifetime in this sample is 6.6 years. Unlike the earlier biased estimates of mean career duration, the estimated median lifetime correctly answers the initial descriptive question "How long does the average teacher teach," and it does so in a meaningful metric using data from both noncensored and censored cases.

Formalizing a Discrete-Time Survival Analysis Model

Consider a homogeneous population of individuals, each at risk of experiencing a single target event—leaving teaching, graduating from college, being decertified from special education. Assume that, for each person, the target event is nonrepeatable; once it occurs, it cannot occur again. Once a newly hired teacher quits for the first time, he or she cannot quit for the first time again; once a student graduates from high school, he or she cannot graduate from high school again. Models for repeated events will be described in a subsequent article; interested readers should consult Allison (1984), Tuma and Hannan (1984), or Yamaguchi (1991).

To record event occurrence in discrete intervals, divide continuous time into an infinite sequence of contiguous time periods $(0, t_1], (t_1, t_2], \ldots,$ $(t_{j-1}, t_j], \ldots,$ and so forth. Let the letter j index periods; the jth period begins immediately after time t_{j-1} (hence the initial parenthesis) and ends at, and includes, time t_j (and hence the concluding bracket). If time is measured in years, as in Table 1, an event occurring any time after t_1 (the last day of Year 1) and up to and including t_2 (the last day of Year 2) is classified as happening during the 2nd time interval $(t_1, t_2]$.[3]

Research interest centers on whether and, if so, when (in which time period) the single nonrepeatable event occurs. Notice that, because each individual can experience the target event only once, event occurrence is inherently *conditional*. An individual can experience the event in a time period j only if he or she did not already experience it any of the earlier time periods prior to j. Similarly, once an individual experiences the event, he or she cannot experience it again in any later time period.

Defining the Discrete-Time Hazard Function

Let T represent the discrete random variable that indicates the time period j when the event occurs for a randomly selected individual from the population. The most common ways of characterizing the distribution of a discrete random variable are through its probability density function and cumulative density function. But because event occurrence is intrinsically conditional, we characterize T by its *conditional probability density function*—the distribution of the probability that an event will occur in each time period given that it has not already occurred in a previous time period—known as the discrete-time hazard function. Owing to its genesis in modeling human lifetimes, the hazard function is also known as the conditional death density function, and its realizations at any given time are known as the force of mortality (Gross & Clark, 1975).

Discrete-time hazard, h_j, is defined as the conditional probability that a randomly selected individual will experience the target event in time period j, given that he or she did not experience the event prior to j:

$$h_j = Pr[T = j \mid T \geq j]. \tag{1}$$

These conditional probabilities, the h_j, are the fundamental parameters of the discrete-time survival process. As the central focus of all analyses, we estimate their values and investigate their dependence on selected covariates; as probabilities, their value always lies between 0 and 1.

We refer to the set of discrete-time hazard probability parameters h_j as a function of time period j as the population *discrete-time hazard function*. This function can be visualized as a plot, against time, of the population risk of the event occurring in each time period, conditional on the event having not occurred in any earlier time period. Depending on the distribution of risk over time, the hazard function takes on a different shape. Singer and Willett (1991) provide additional examples of hazard functions and the kinds of information that can be retrieved from them.

Building Observed Heterogeneity Into the Definition of Hazard

Interest usually centers on detecting whether the risk of an event's occurrence differs systematically across people; in other words, researchers want to identify predictors of heterogeneity in hazard. After discovering that the risk of leaving teaching declines over time, for example, researchers want to know what characteristics distinguish low-risk from high-risk teachers (Charters, 1970; Murnane, Singer, & Willett, 1988). Or after learning that the median time to doctorate is 4 years, researchers ask whether the careers of men differ from those of women or whether financial aid makes a difference (Civian, 1990; Zwick & Braun, 1988).

We address questions about systematic heterogeneity in hazard by deter-

mining whether different types of individuals, distinguished by their values on specific predictors, have different hazard functions. We introduce observed heterogeneity into our definition by considering P predictors, Z_p ($p = 1, 2, \ldots, P$), each of which characterizes the members of the population on a specific dimension. In a study of teachers' careers, for example, Z_1 might be sex, Z_2 might be attitude towards the job, and Z_3 might be salary (corrected for inflation). Because the values of some predictors (such as salary) may themselves vary over time, during data collection, we record the values of these predictors in every time period. We denote individual i's values for each of the P predictors in time period j as the vector $z_{ij} = [z_{1ij} z_{2ij}, \ldots, z_{Pij}]$. We assume that the values of the predictors remain constant within time periods.

Although interest remains focused on the population hazard probabilities, we introduce the subscript i into our definition so that h_{ij} defines the probability that individual i—as distinguished by his or her predictor values $z_{1ij}, z_{2ij}, \ldots, z_{Pij}$—experiences the event in time period j, given that he or she survived through all prior periods. Thus, Equation 1 becomes:

$$h_{ij} = Pr\{T_i = j \mid T_i \geq j, Z_{1ij} = z_{1ij}, Z_{2ij} = z_{2ij}, \ldots Z_{Pij} = z_{Pij}\}. \qquad (2)$$

Towards a Statistical Model for Discrete-Time Hazard

Equation 2 indicates that hazard depends on each individual's values on a vector of predictors; it does not specify the functional form of that dependence. In this section, we develop a formal model of a hypothesized relationship between the population hazard probabilities and predictors. We presage the specification by returning to the Michigan special educator data and asking how we might express the relationship between a single simple time-invariant predictor—teacher sex—and hazard. By beginning with an empirical example, we hope to lay the groundwork for the statistical model soon to be postulated and clarify the meaning of the parameters that the model will include.

If a teacher's sex is related to the risk of leaving teaching, the hazard functions for men and women will differ systematically. One way of exploring this relationship is by examining the sample hazard profiles estimated separately for the 615 men and the 3,326 women in the sample, as shown in Figure 2. Notice that the profiles of risk for men and women have similar shapes: The risk of leaving teaching is highest immediately after hire and declines steadily over time. Over and above this similarity in shape, there is a vertical separation between the profiles: In every year except the 2nd and the 12th, the hazard function for women is higher relative to its location for men. So in this sample at least, in nearly every year of their careers, women are at greater risk than men of leaving teaching.

When we make these two statements—one about the general profile of risk over time and another about a difference in level of risk between
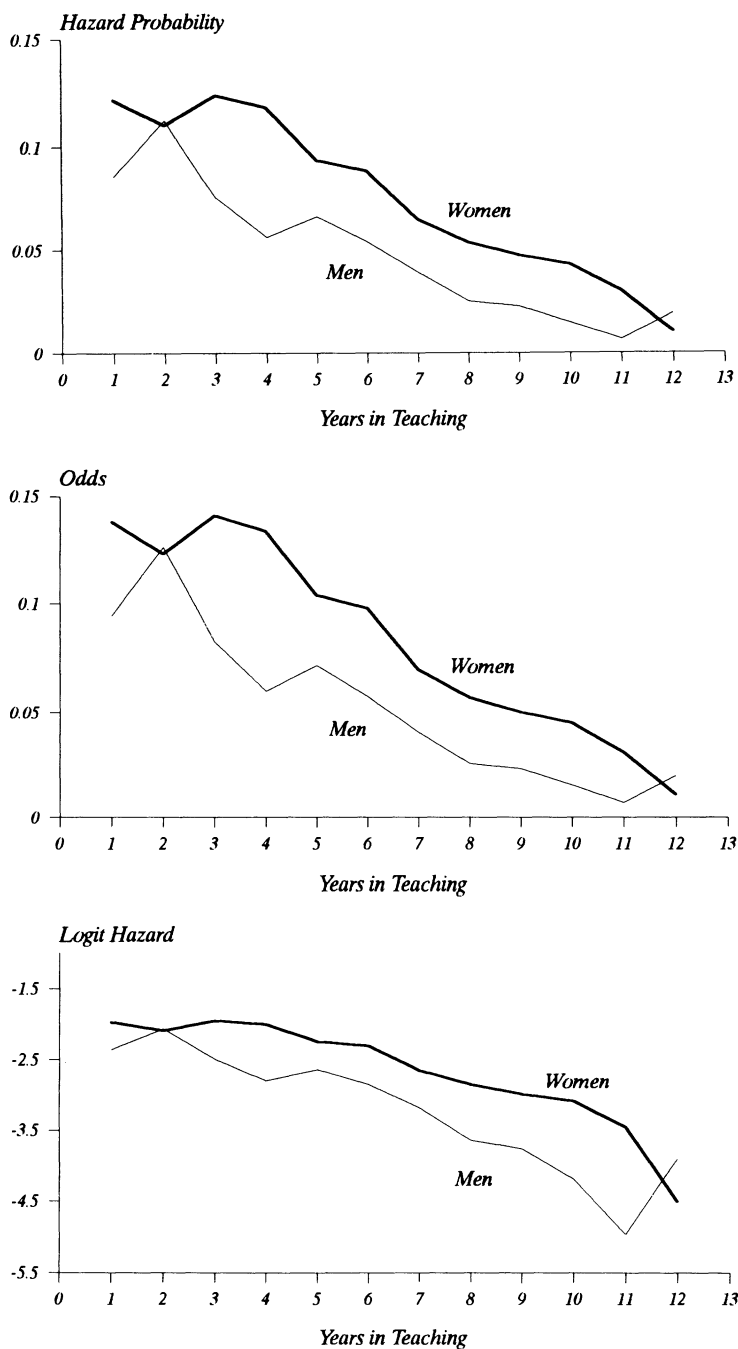
FIGURE 2. *Sample hazard functions—plotted on three scales: raw hazard, odds, and logit—for male and female special educators hired in Michigan between 1972 and 1978*

groups—we are essentially ascribing the variation in hazard to two compo-
nents: a baseline profile of risk over time and a shift in risk associated with
variation in teacher sex. Even in the absence of formal model, we can
systematize this partition further by defining a dichotomous predictor
FEMALE (0 for men; 1 for women) and thinking about how variation in this
predictor is associated with variation in the conceptual outcome—the entire
hazard profile. One approach is to treat the hazard function for men (when
FEMALE = 0) as a baseline pattern of risk—the profile of the risk of event
occurrence over time for a baseline group of people or, in other words, those
sharing the predictor value of 0, which, in this example, happens to be
men.[4] Having done so, think of estimating a parameter that describes the
way in which variation in the value of the predictor FEMALE (as it goes from
0 to 1) shifts the baseline hazard function to generate the profile of risk for
the other group, which, in this case, happens to be women.

Conceptually, then, we are drawn to a statistical model in which variation
in the predictor acts to vertically displace an entire baseline hazard function,
yielding another entire function. While admittedly coarse, this conceptual-
ization captures the two essential attributes of a discrete-time hazard model:
(a) a baseline profile of risk and (b) a shift parameter that captures the effect
of the predictor on the baseline profile.

Cox (1972) proposed that, because the h_{ij} are probabilities, they can be
reparameterized so that they have a logistic dependence on the predictors
and the time periods (also Allison, 1982; Brown, 1975; Efron, 1988; Laird
& Oliver, 1981; Thompson, 1977).[5] Such a model represents the log-odds
of event occurrence as a function of predictors and also has the attributes
of a baseline profile and a shift which are implied above. Our proposed
population discrete-time hazard model is therefore:

$$h_{ij} = \frac{1}{1 + e^{-[(\alpha_1 D_{1ij} + \alpha_2 D_{2ij} + \cdots + \alpha_J D_{Jij}) + (\beta_1 Z_{1ij} + \beta_2 Z_{2ij} + \cdots + \beta_P Z_{Pij})]}}, \tag{3}$$

where $[D_{1ij} D_{2ij}, \ldots, D_{Jij}]$ are a sequence of dummy variables, with values
$[d_{1ij} d_{2ij}, \ldots, d_{Jij}]$ indexing time periods. J refers to the last time period
observed for anyone in the sample. If j_i represents the last time period when
individual i was observed (and at which time he or she was either censored
or experienced the target event), then $J = \sup\{j_i\}$. The time-period dum-
mies are defined identically for everyone: $d_{1ij} = 1$ when $j = 1$, and $d_{1ij} = 0$
when j takes on any other value (2 through J); $d_{2ij} = 1$ when $j = 2$, and
$d_{2ij} = 0$ otherwise; and so on. As we will see, the intercept parameters
$[\alpha_1 \alpha_2 \ldots \alpha_J]$ capture the baseline level of hazard in each time period, and
the slope parameters $[\beta_1 \beta_2, \ldots, \beta_p]$ describe the effects of the predictors on
the baseline hazard function, albeit on a logistic scale.

Taking logistic transformations of both sides of the equation, we have:

$$\log_e\left(\frac{h_{ij}}{1 - h_{ij}}\right) = (\alpha_1 D_{1ij} + \alpha_2 D_{2ij} + \cdots + \alpha_J D_{Jij})$$

$$+ (\beta_1 Z_{1ij} + \beta_2 Z_{2ij} + \cdots + \beta_P Z_{Pij}). \qquad (4)$$

Written in this form, it is easy to see that we are postulating that the *conditional log-odds* that the event will occur in each time period j (given that it did not occur before) is a linear function of a constant term, α_j, specific to period j, and of the values of the predictors period j multiplied by the appropriate slope parameters. This reexpression clarifies the model's interpretation both algebraically and graphically. Specifically, we are assuming that the predictors are linearly associated with the logistic transformation of hazard (*logit-hazard*), not with the hazard probabilities themselves, nor with the natural logarithm of the hazard probabilities (as in a continuous-time hazard model). We are not positing that the hazard function for women is a vertically shifted version of the baseline hazard function for men, as was initially suggested by the sample plots in Figure 2; rather we are positing that the shifts occur in the scale of logit-hazard. Were we to antilog these logit-hazard profiles to yield profiles describing the conditional odds of event occurrence at different values of the predictors, the odds profiles would be magnifications or diminutions of one another—they would be proportional. We return to this proportionality property later when we discuss the assumptions of the discrete-time hazard model.

Notice also that the discrete-time hazard model contains no single stand-alone intercept. Instead the alpha parameters $[\alpha_1 \alpha_2, \ldots, \alpha_J]$ act as multiple intercepts, one per time period. When the values of all the covariates, $Z_1 - Z_P$, are set to zero, the population discrete-time hazard depends only on $[\alpha_1 \alpha_2, \ldots, \alpha_J]$ and the values of the time dummies. The $[\alpha_1 \alpha_2, \ldots, \alpha_J]$ represent the population *baseline logit-hazard function* because it captures the time-period by time-period conditional log-odds that individuals whose covariate values are all zero (the baseline group) will experience the event in each time period, given that they have not already done so. Let us return to the Michigan special educator data, for instance. Were we to model the effect of the single time-invariant predictor FEMALE on the risk of leaving teaching, the parameter α_1 would represent the log-odds that a man would leave in his first year; α_2 would represent the log-odds that a man would leave in his second year, given that he did not leave in his first; and so on. We can then compute all other possible logit-hazard profiles, corresponding to other subgroups of the heterogeneous population, by setting some or all of the predictors to values other than 0.

Estimating the Parameters of the Discrete-Time Hazard Model

With event history data collected on a random sample of n individuals ($i = 1, 2, \ldots, n$), we can fit the discrete-time hazard model in Equation 3 and

estimate its parameters. In this section, we introduce notation to describe discrete times to event, construct the likelihood function for the discrete-time hazard process, and show how the parameters of the hazard model can be estimated.

Recording Discrete Event-History Data

In this article, we are assuming that every individual in the sample lives through each successive discrete time period until he or she experiences the event of interest or is censored by the end of data collection. For each individual i, therefore, the chronology of event occurrence (the event history) can be conveniently recorded using a sequence of dummy variables Y_{ij}, whose values y_{ij} are defined as:

$$y_{ij} = \begin{cases} 0 & \text{if individual } i \text{ does not experience the event in period } j, \\ 1 & \text{if individual } i \text{ does experience the event in period } j. \end{cases}$$

Because we are studying the occurrence of a single nonrepeatable event, this sequence of y values can only display one of two possible patterns. One, the Y_{ij} will take on the value zero in every time period that was observed during data collection, including the last, indicating that individual i never experienced the event of interest at any time during data collection and was ultimately censored. Or, two, if individual i was not censored, then she or he must have experienced the event in one of the time periods observed during data collection. Once the event is observed, it cannot be repeated, and data collection terminates for this individual. So for an uncensored individual, fewer y values may be required to describe the event history; they will consist of a sequence of zeros terminating in the value one.

In addition to describing each person's career history by a chain of y values, it is also useful to separately record whether each person's career was ended by the target event or by censoring. This information is stored in a censoring indicator, C_i, whose values c_i are defined as:

$$c_i = \begin{cases} 0 & \text{if individual } i \text{ is not censored}, \\ 1 & \text{if individual } i \text{ is censored}. \end{cases}$$

Finally, for each individual in the sample, we can record the last time period in which data were collected. Earlier, we referred to this terminal time period as j_i, the subscript indicating that it may differ from individual to individual.

These three sets of information are inextricably linked—given any two, we know the third. For example, for individual i, if we know the last time period and whether he or she was censored (i.e., if we have the values j_i and c_i), then we can derive all the values of y_{ij}. These interrelationships will prove useful during the construction of the likelihood function.

<div align="center">

Constructing the Likelihood Function

</div>

Estimators for the parameters $[\alpha_1 \alpha_2, \ldots, \alpha_J]$ and $[\beta_1 \beta_2, \ldots, \beta_p]$ of the logistic discrete-time hazard model in Equations 3 and 4, and therefore of the h_{ij}, can be obtained by the method of maximum likelihood. Because of censoring, however, we construct the likelihood function as a product of two distinct parts. An individual who is not censored ($c_i = 0$) experiences the event in time period j_i and cannot therefore experience it in any earlier (or later) time period. An individual who is censored ($c_i = 1$) will not experience the event in any time period up to, or including, the time period when he or she was last observed, j_i, but he or she may experience it after the end of data collection. So the likelihood of observing the data is constituted from two types of contributions: (a) for uncensored individuals, the probability that the event occurs in time period j_i and, (b) for censored individuals, the probability that the event occurs after time period j_i. Following Allison (1982), we derive expressions for each separately.

First, the probability that an uncensored individual will experience the event in time period j_i can be written as a product of terms, one per period, describing the conditional probabilities that the event did not occur in Periods 1 through $j_i - 1$ but did occur in period j_i:

$$Pr\{T_i = j_i\} = Pr\{T_i = j_i \mid T_i \geq j_i\}Pr\{T_i \neq j_i - 1 \mid T_i \geq j_i - 1\} \ldots$$

$$Pr\{T_i \neq 2 \mid T_i \geq 2\}Pr\{T_i \neq 1 \mid T_i \geq 1\}$$

where we suppress the parenthetical covariate dependency throughout the equation to economize on space. The equation can be reformulated in terms of the h_{ij} as:

$$Pr\{T_i = j_i\} = h_{ij_i}(1 - h_{i(j_i - 1)})(1 - h_{i(j_i - 2)}) \ldots (1 - h_{i2})(1 - h_{i1})$$

$$= h_{ij_i} \prod_{j=1}^{j_i - 1} (1 - h_{ij}). \tag{5}$$

Second, the probability that a censored individual will experience the event after period j_i can be constructed similarly as a product of terms, one per period, describing the conditional probabilities that the event did not occur in any observed period, 1 through j_i:

$$Pr\{T_i > j_i\} = Pr\{T_i \neq j_i \mid T_i \geq j_i\}Pr\{T_i \neq j_i - 1 \mid T_i \geq j_i - 1\}$$

$$\ldots Pr\{T_i \neq 2 \mid T_i \geq 2\}Pr\{T_i \neq 1 \mid T_i \geq 1\}$$

where we have again suppressed the covariate dependency throughout in order to save space. This can be reexpressed in terms of the h_{ij} as:

$$Pr\{T_i > j_i\} = (1 - h_{ij_i})(1 - h_{i(j_i - 1)})(1 - h_{i(j_i - 2)})$$

$$\ldots (1 - h_{i2})(1 - h_{i1}) = \prod_{j=1}^{j_i} (1 - h_{ij}). \tag{6}$$

which is the population survivor function.[6]

Then, assuming that individuals in the sample are independent (given their $z_{1ij}, z_{2ij}, \ldots, z_{Pij}$ values), the likelihood function is simply the product of the probabilities of observing the sample data, $Pr\{T_i = j_i\}$ in the case of the uncensored individuals ($c_i = 0$) and $Pr\{T_i > j_i\}$ in the case of those who were censored ($c_i = 1$):

$$L = \prod_{i=1}^{n} [Pr\{T_i = j_i\}]^{1-c_i}[Pr\{T_i > j_i\}]^{c_i}. \tag{7}$$

In Equation 7, the sample data appear in the likelihood function via the presence of the c_i and the j_i. Substituting from (5) and (6) into (7):

$$L = \prod_{i=1}^{n} \left[h_{ij_i} \prod_{j=1}^{j_i-1} (1 - h_{ij}) \right]^{1-c_i} \left[\prod_{j=1}^{j_i} (1 - h_{ij}) \right]^{c_i}.$$

Taking logarithms gives the log-likelihood function:

$$l = \sum_{i=1}^{n} \left[(1 - c_i) \log_e h_{ij_i} + (1 - c_i) \sum_{j=1}^{j_i-1} \log_e(1 - h_{ij}) + c_i \sum_{j=1}^{j_i} \log_e(1 - h_{ij}) \right].$$

Or, more simply:

$$l = \sum_{i=1}^{n} \left[(1 - c_i) \log_e\left(\frac{h_{ij_i}}{1 - h_{ij_i}}\right) + \sum_{j=1}^{j_i} \log_e(1 - h_{ij}) \right]. \tag{8}$$

Equation 8 can be modified to introduce the event-history indicators y_{ij}. If individual i is not censored ($c_i = 0$), the target event occurs in time period j_i; thus all y_{ij} equal zero except for the very last (when $j = j_i$), which equals one. If individual i is censored ($c_i = 1$), in contrast, the target event does not occur in any time period including the last (when $j = j_i$); so all the y_{ij}, including that for time period j_i, equal zero. Therefore, we can write:

$$\sum_{j=1}^{j_i} y_{ij} \log_e\left(\frac{h_{ij}}{1 - h_{ij}}\right) = \begin{cases} \log_e\left(\dfrac{h_{ij_i}}{1 - h_{ij_i}}\right) & \text{when } c_i = 0 \\ 0 & \text{when } c_i = 1 \end{cases}$$

$$= (1 - c_i) \log_e\left(\frac{h_{ij_i}}{1 - h_{ij_i}}\right). \tag{9}$$

Substituting from (9) for the first term inside the bracket in Equation 8 eliminates the censoring indicator from the log-likelihood, replacing it by the dichotomous realizations of the event-history process, the y_{ij}:

$$l = \sum_{i=1}^{n} \left[\sum_{j=1}^{j_i} y_{ij} \log_e\left(\frac{h_{ij_i}}{1 - h_{ij_i}}\right) + \sum_{j=1}^{j_i} \log_e(1 - h_{ij}) \right].$$

This can be rewritten as:

$$l = \sum_{i=1}^{n} \sum_{j=1}^{j_i} \left[\log_e\left(\frac{h_{ij_i}}{1 - h_{ij_i}}\right)^{y_{ij}} + \log_e(1 - h_{ij}) \right].$$

Collecting like terms and antilogging yields:

$$L = \prod_{i=1}^{n}\prod_{j=1}^{j_i} h_{ij}^{y_{ij}}(1 - h_{ij})^{(1 - y_{ij})}. \tag{10}$$

Equation 10 is the likelihood function for the discrete-time hazard process in terms of the data (the y_{ij}) and the hazard probability parameters (the h_{ij}). However, in Equation 3, we have reparameterized the h_{ij} as a logistic function of a smaller number of secondary parameters: $\alpha_1, \alpha_2, \ldots, \alpha_J, \beta_1, \beta_2, \ldots, \beta_P$. Maximizing the likelihood in (10), under the logistic reparameterization in (3), provides maximum likelihood estimates of $\alpha_1, \alpha_2, \ldots, \alpha_J, \beta_1, \beta_2, \ldots, \beta_P$ (and hence the h_{ij}). Notice that the likelihood function for the discrete-time hazard model in (10) is identical to the likelihood function for a sequence of $N = (j_1 + j_2 + \cdots + j_n)$ independent Bernoulli trials with parameters h_{ij}.

As demonstrated by Allison (1982), Brown (1975), and Laird and Oliver (1981), the equivalence of the likelihood functions of the discrete-time hazard model in (10) and the independent Bernoulli trials model allows us to treat the N dichotomous observed values y_{ij} as a collection of independent dichotomous variables with a hypothesized logistic dependence on predictors. We can regard them as the values of the outcome variable in a logistic regression analysis of the time-period indicators D and covariates Z. This provides a simple method of obtaining maximum likelihood estimates of $\alpha_1, \alpha_2, \ldots, \alpha_J, \beta_1, \beta_2, \ldots, \beta_P$ (and hence the h_{ij}) using nothing more than standard logistic regression analysis software. Because computer software for conducting logistic regression analysis is so widely available, we now illustrate the fitting of hazard models via this modified logistic regression approach, rather than via direct maximization of the likelihood in (10).

Creating the Person-Period Data Set

Because discrete event history data are not usually stored in the format necessary for the required logistic regression analysis, researchers must manipulate their data prior to analysis. In the typical person-oriented data set, each person in the sample has one record (line) of data. The ith individual's record usually contains information about:

- *Duration*. The length of time the individual was alive—usually recorded as the last time period in which the person was observed, j_i.
- *Censoring*. The value of the censoring indicator C_i, which indicates whether the person actually experienced the event of interest in the last time period in which he or she was observed or whether censoring occurred. The value of C_i is 0, if individual i was not censored in time period j_i, and 1, if he or she was.
- *Selected predictors*. Individual i's values on covariates are recorded in each time period j up to, and including, time period j_i. For time-invari-

ant predictors, only a single value is recorded because it applies uniformly across all time periods. Time-varying predictors, on the other hand, may take on a different value in each period.

Before conducting the requisite logistic regression analyses, this person-oriented data set must be converted into a new person-period data set in which each person has multiple records (lines of data), one per time period of observation. To distinguish these multiple records within person, new variables must be created identifying the time period to which each record corresponds—these variables are called the *time indicators*. The values of the covariates must be recorded so that they are appropriate to each time period. And the *event indicator* variable, Y, must be created using the duration and censoring information. Consequently, in the new person-period data set, the ith individual has j_i records, with the jth of these containing information about the jth time period:

- *The time indicators*. Following Equation 3, the set of dummy variables, $D_{1ij}, D_{2ij}, \ldots, D_{Jij}$, takes on values that identify the particular time period to which the record refers. In this case (the record corresponding to time period j), all of the time indicators take on value 0 except for the jth dummy, D_{jij}, which has value 1.
- *The predictors*. In the jth record, the covariates contain the ith individual's values of the P covariates appropriate for time period j, $Z_{1ij}, Z_{2ij}, \ldots, Z_{Pij}$. Time-invariant predictors have values that are identical in all time periods between Period 1 and period j_i. Time-varying predictors, on the other hand, have values that may differ from time period to time period.
- *The event indicator*. The variable Y records the values y_{ij} that indicate whether the event of interest occurred for individual i in time period j. Its value is 0, if the event of interest did not occur, and 1, if it did.

Because each person gets multiple records, the original n records (equal to the sample size) of the person-oriented data set become $N = (j_1 + j_2 + \ldots + j_n)$ records in the new person-period data set.

Figure 3 illustrates the conversion using a small subset of variables for four Michigan special educators. In the appendix, we present computer code (SAS, Version 6) for implementing the conversion and for carrying out statistical analyses.

In the original person-oriented data set (Fig. 3, top panel), each teacher has a single record that contains: an identification code (ID), the number of years he or she taught before the end of data collection (j_i, recorded as LASTPD), whether the career was censored (C_i, recorded as CENSOR), and the values of one time-invariant predictor and one time-varying predictor. The time-invariant predictor FEMALE indicates the teacher's sex (i.e., describes whether the teacher is female or not) and contains a single value for all time

Original Person Data Set

ID	LASTPD	CENSOR	FEMALE	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
01	01	0	1	0	.	.	.	.	.	.	.	.	.	.	.
02	03	0	0	1	1	1	.	.	.	.	.	.	.	.	.
03	12	0	1	1	1	1	0	0	0	1	1	1	1	1	1
04	12	1	1	0	0	0	0	0	0	0	0	0	0	0	0

Converted Person-Period Data Set

ID	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	D_{11}	D_{12}	FEMALE	SUPPORT	Y
01	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1
02	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
02	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
02	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1
03	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0
03	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0
03	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0
03	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
03	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
03	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
03	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0
03	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0
03	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0
03	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0
03	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0
03	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
04	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
04	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
04	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
04	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
04	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
04	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
04	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
04	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
04	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0
04	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0
04	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
04	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0

FIGURE 3. *Converting the person-oriented data set into the person-period data set as a precursor to discrete-time hazard modeling by logistic regression*

periods. The time-varying predictor SUPPORT, in contrast, can take on a different value in each time period and so is represented in the original person-oriented data set by a vector of 12 variables, $S_1, S_2, \ldots, S_{12}$, one per year. Its values indicate whether the teacher provided support services ($S_j = 1$) or direct instruction ($S_j = 0$) in each year. In general, each time-varying predictor is represented in the person-oriented data set by J vari-

ables, one per discrete time period observed, and the values of these variables are set to *missing* whenever j_i is less than J (as for Teachers 1 and 2).

In the person-period data set (Fig. 3, lower panel), each teacher has one record for each discrete time period that he or she was observed. Teacher 1, who was observed for 1 year, has 1 record. Teacher 2, who taught for 3 years, has 3, and so on.

All person-period records contain identical variables, with values appropriate for that person in that period. The ID variable identifies the teacher, its value being replicated in each period. The 12 time indicators identify the discrete time period being referenced in the record. For all teachers, D_1 is 1 in the record for the first period, D_2 is 1 in the second period, D_3 is 1 in the third period, and so forth, with all other values being set to 0. The time-invariant predictor, FEMALE, has the same value in the several records that describe each teacher; thus, Teacher 3, who is female, has the value 1 recorded in all 12 of her person-period records. The original 12 predictors $S_1, S_2, \ldots, S_{12}$, which represent the time-varying predictor SUPPORT, become a single column of values called SUPPORT, with values appropriate to each time period. Thus, in the three records of Teacher 2, the values of SUPPORT are 1, 1, and 1 corresponding to the values of S_1, S_2, and S_3 in the original person-oriented data set. Finally, the dichotomous event indicator, Y, describes whether individual i experienced the event of interest during time period j. For example, both Teacher 3 and Teacher 4 have 12 values of the event indicators (with the first 11 values being zero); the 12th value for Teacher 3 is 1 (indicating that she terminated employment in that year), whereas the value for Teacher 4 is 0 (indicating that she continued to teach beyond the end of data collection and was censored).

To estimate the parameters of the discrete hazard model in Equation 3, we use standard logistic regression software to regress the event indicator, Y, on the predictors in the person-period data set. The equivalence of the discrete-time and independent Bernoulli likelihood functions demonstrated earlier ensures that the obtained goodness-of-fit statistics, standard errors, and other inferential statistics are appropriate for testing the influence of the predictor on the risk of event occurrence. Having estimated the parameters (the αs and the βs), we can reconstruct estimates of the h_{ij} by substituting into Equation 3 to obtain a maximum likelihood estimate of the hazard function and, ultimately, estimates of corresponding fitted survivor functions and estimated median lifetimes, which we now describe.

Data-Analytic Strategies for Discrete-Time Survival Analysis

Doing data analysis with discrete-time hazard models exercises many of the same analytic skills that are used in regression analysis. A taxonomy of models must be fitted in order to investigate the effects of substantively interesting collections of predictors. Pairs of fitted models can be compared

to assess the influence of specific predictors, singly and in groups. Key assumptions must be checked. When interesting models have been identified, they can be used to support substantive interpretation.

In the remainder of the article, we use a specific data-based example to introduce and discuss general issues in data analysis for discrete-time survival analysis. In this section, we describe selected strategies for building, selecting among, and interpreting discrete-time hazard models. In the next section, we describe the three important assumptions implicit in these models, and we comment on simple methods for evaluating their validity in practice.

In Table 2, we present the parameter estimates, standard errors, and associated goodness-of-fit statistics for the simple taxonomy of selected discrete-time hazard models that we will use to support our discussion.

TABLE 2

Parameter estimates, standard errors, and goodness-of-fit statistics for three hazard models fitted to the Michigan special educator data: (A) the main effect of the time indicators; (B) the main effects of the time indicators and time-invariant special educator gender (0 = male, 1 = female); and (C) the main effects of the time indicators and time-varying job assignment (0 = classroom teacher, 1 = support staff)

Predictor	Model A Parameter estimate (*SE*)	Model B Parameter estimate (*SE*)	Model C Parameter estimate (*SE*)
D_1	−2.03 (.05)	−2.41 (.08)	−2.06 (.05)
D_2	−2.09 (.05)	−2.47 (.08)	−2.11 (.05)
D_3	−2.03 (.06)	−2.41 (.08)	−2.05 (.06)
D_4	−3.12 (.06)	−2.49 (.09)	−2.14 (.06)
D_5	−2.32 (.07)	−2.70 (.09)	−2.35 (.07)
D_6	−2.41 (.07)	−2.78 (.10)	−2.43 (.08)
D_7	−2.75 (.09)	−3.12 (.11)	−2.77 (.09)
D_8	−2.98 (.12)	−3.35 (.13)	−3.01 (.12)
D_9	−3.12 (.14)	−3.48 (.15)	−3.15 (.14)
D_{10}	−3.26 (.17)	−3.62 (.18)	−3.28 (.17)
D_{11}	−3.68 (.25)	−4.03 (.26)	−3.70 (.25)
D_{12}	−4.35 (.45)	−4.69 (.45)	−4.31 (.45)
FEMALE		0.44 (.07)	
SUPPORT			0.19 (.07)
−2LL	14583.74	14538.18	14576.31
Change in −2LL (*df*)		45.56 (1)	7.43 (1)
p		< .0001	.0064

These models were fitted using the entire Michigan special educator data set and the predictors introduced in Figure 3.

Fitting an Initial Discrete-Time Hazard Model: The Main Effect of Time

Model A is the simplest model in our taxonomy. Notice that it includes no unique intercept term (in the classic sense) but, instead, contains as predictors the 12 time indicators, D_1 through D_{12}, as a group:[7]

$$\text{logit}_e(h_{ij}) = [\alpha_1 D_{1ij} + \alpha_2 D_{2ij} + \cdots + \alpha_{12} D_{12ij}]. \tag{11}$$

We have bracketed together the terms involving the time indicators because, as a group, they represent the main effect of a single conceptual predictor, which we will refer to as TIME. We recommend that all discrete-time survival analyses begin with such an initial model.

By saying that our initial model includes only the main effect of TIME, we highlight a seeming paradox in discrete-time hazard modeling: TIME, the conceptual outcome, is the fundamental predictor of the hazard profile. This seeming anomaly occurs because, to make the problem of censoring amenable to analysis, we have reformulated the question "When does the event occur?" to "What is the risk of event occurrence in each time period?" This switch sacrifices nothing intellectually because we can, via summary statistics, interpret fitted models in the original metric of interest—time.

Beginning the data analysis with an initial time-only hazard model provides not only a simple model against which to compare more complex models but an interpretation of $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_{12}$ that provides direct information on the shape of the whole-sample hazard function.

Recovering the Whole-Sample Hazard Function

By substituting $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_{12}$ back into (11), we can recover the fitted risk of leaving teaching in each year of the career and answer questions like: As a group, are special educators more likely to leave early or late in their careers? Which years are the most hazardous? Because Model A contains no predictors other than TIME, this substitution yields the fitted hazard function for the entire sample, assuming a homogeneous population in which individuals are not distinguished by values of any covariates. For instance, in the first year ($j = 1$) where $D_1 = 1$ and all other time indicators have value zero, a maximum likelihood estimate of h_1 is:

$$\hat{h}_1 = \frac{1}{1 + e^{-(-2.0337 \times 1)}} = .1157,$$

where we have retained additional decimal places in the calculations to ensure precision. Therefore, among all special educators in their first year of service, we estimate that there is almost a 12% risk of leaving teaching.[8] Similar computations lead to estimates of $h_2, h_3, h_4, \ldots, h_{12}$.

The obtained estimates of discrete yearly hazard are identical to the hazard estimates introduced without substantiation in the sixth column of Table 1. But now we see that they are also maximum likelihood estimates, obtained here via logistic regression analysis. In addition, they constitute the discrete limit of the better known Kaplan-Meier estimate of continuous-time hazard rate (Kaplan & Meier, 1958). As Efron (1988) noted, these "logistic estimates . . . approach the Kaplan-Meier estimate[s] as the number of parameters grows large" (p. 414; i.e., as time becomes more finely discretized).

Recovering the Whole-Sample Survivor Function

Based on the fitted hazard profile, we can also compute maximum likelihood estimates of the population survivor function by the method of cumulation already described earlier in the article. Specifically, the estimated survivor probabilities are obtained by substitution into Equation 6:

$$\hat{S}_j = \prod_{k=1}^{j} (1 - \hat{h}_k).$$

The estimated values of $S_1, S_2, \ldots, S_j$ can be used to construct a fitted survivor function. Many authors display sample survivor functions as step functions (Cox & Oakes, 1984; Kalbfleisch & Prentice, 1980). We have adopted a different approach, plotting the sample survival probabilities at the times for which the periods are named: S_1 at time t_1, S_2 at t_2, and so forth (i.e., at the end of each interval). We have connected these points with a linear smooth, which decreases the bias of the estimated median lifetime (Miller, 1981). A maximum likelihood estimate of the median lifetime can then be obtained by geometric construction on the fitted plot (e.g., see the lower panel of Fig. 4).

Alternative Parameterizations for the Main Effect of Time

In using a set of indicator variables to distinguish among the time periods, we have adopted the most general parameterization of the main effect of time on hazard. Under this strategy, a separate parameter $\alpha_1, \alpha_2, \ldots, \alpha_{12}$ represents the population hazard probability in each time period under observation: α_1 captures the risk of leaving teaching in the first year; α_2 captures the risk of leaving in the second year given that the teacher did not leave in the first year, and so on.

Patterns in the values of these parameters represent the temporal shape of the overall population hazard function, and, because of the flexibility of the parameterization, the hazard profile is free to adopt whatever shape best describes the population pattern of risk. If the risk of event occurrence were independent of time, for example, the hazard function would be flat, and $\alpha_1, \alpha_2, \ldots, \alpha_{12}$ would all be equal. If the early risk of event occurrence were

high, the parameters corresponding to early time periods would be free to take on large values (i.e., less negative values, in this case).

Although general, parameterizing the hazard profile using time indicators lacks parsimony. Representation of the main effect of time requires the inclusion of many unknown parameters in the discrete-time hazard model, especially when data collection is lengthy and time finely discretized. And, as we will see later, when fitting hazard models to data, researchers should consider the possibility of interactions between TIME and other predictors, necessitating inclusion of even more additional parameters. In such circumstances, a more parsimonious representation for TIME can reap improvements in statistical power, coefficient stability, and time to convergence during estimation.

Although we do not demonstrate them here, there are two easy methods for rendering the representation of TIME more parsimonious: (a) by constraining (groups of) the αs to have specific values or (b) by adopting a particular algebraic form for the shape of the logit-hazard profile. Use of either strategy should be informed either by knowledge of the prospective shape of the hazard function from prior exploratory or published research or, less persuasively, by a post hoc inspection of $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_{12}$ that may suggest potential structure to be retroactively imposed on the αs.

In Model A, for instance, inspection of $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$, and $\hat{\alpha}_4$ suggests that a more parsimonious model might constrain the first four α parameters to be equal. Under the second strategy, the dummy variable representation can be replaced by a linear, polynomial, nonlinear, or even piecewise dependence of logit-hazard on TIME, now treated as a continuous variable. For instance, in Model A, the systematic decline in the $\hat{\alpha}$s across successive time periods suggests that a linear, or perhaps quadratic, time dependence might be appropriate. Efron (1988) provides examples in which both a cubic and a cubic-linear spline are evaluated as the potential time dependence for logit-hazard in a discrete-time analysis of head-and-neck cancer survival. Here, we continue to use Model A in Table 2 as the basis for subsequent analyses because the ratio of sample size to number of unknown parameters is large, thereby allowing us to preserve the flexibility of the most general parameterization of time.

Assessing the Effects of Additional Predictors

Additional influences on risk can be investigated by adding further predictors to the initial discrete-time hazard model. Their contribution to prediction can easily be assessed by comparing the goodness of fit of the original and extended models, using standard decrement-to-chi-square testing, based on the known asymptotic distributional properties of -2 times the log-likelihood statistic (-2LL). For instance, Model B in (12) differs

from Model A in (11) by the inclusion of the time-invariant predictor FEMALE, the influence of which is captured by the slope parameter β_1:

$$\text{logit}_e(h_{ij}) = [\alpha_1 D_{1ij} + \alpha_2 D_{2ij} + \cdots + \alpha_{12} D_{12ij}] + \beta_1 \text{FEMALE}_{ij}. \quad (12)$$

From the row labeled $-2LL$ in Table 2, we see that the model chi-square goodness-of-fit statistic has decreased from 14583.74 to 14538.18 on addition of FEMALE—a difference of 45.56 for the loss of 1 degree of freedom ($p < .0001$)—indicating that the addition of this new predictor has significantly improved the overall fit of the hazard model. We can reject the null hypothesis that the predictor FEMALE has no effect on the population hazard profile (H_0: $\beta_1 = 0$). When adding a single predictor, the same conclusion can be reached by comparing the magnitude of $\hat{\beta}_1$ to its asymptotic standard error (also provided in Table 2).

How does FEMALE affect the hazard profile? The positive coefficient associated with the predictor FEMALE (0.44) estimates the vertical elevation of the fitted logit-hazard function for women above that of men. It can be antilogged and interpreted as odds and odds-ratios in the usual way (Hosmer & Lemeshow, 1989). In Model B, when $\hat{\beta}_1$ is antilogged, for instance, we find that the estimated odds of leaving special education teaching in any given year are 1.55 times greater for women than men.

Interpreting the Effects of Important Predictors Using Fitted Hazard and Survivor Profiles

Although odds-based interpretations have intuitive appeal, we have found that the results of discrete-time hazard modeling are sufficiently esoteric to warrant the use of more direct methods of presentation and interpretation. The effects of important predictors on the hazard profile can be well displayed by plotting fitted hazard and survivor functions at substantively important values of the predictors. This is akin to using fitted plots to report findings of multiple regression analyses. We illustrate this approach in Figure 4, which displays fitted hazard and survivor functions for men and women computed from the parameter estimates for Model B in Table 2. Comparison of the two fitted hazard functions in Figure 4 demonstrates the large sex differential.

Notice that, unlike the sample hazard functions for men and women in Figure 2, these fitted functions have the same basic shape—in fact, one appears to be a magnification of the other.[9] Were we to replot the fitted logit-hazard functions, they would be parallel. They have been constrained to appear this way by the proportional-odds assumption built into the logistic hazard model of (3), (4), and (12). Figure 2, in contrast, showed hazard profiles estimated separately for subsamples of men and women, and, even though the proportional-odds assumption is a reasonable assump-
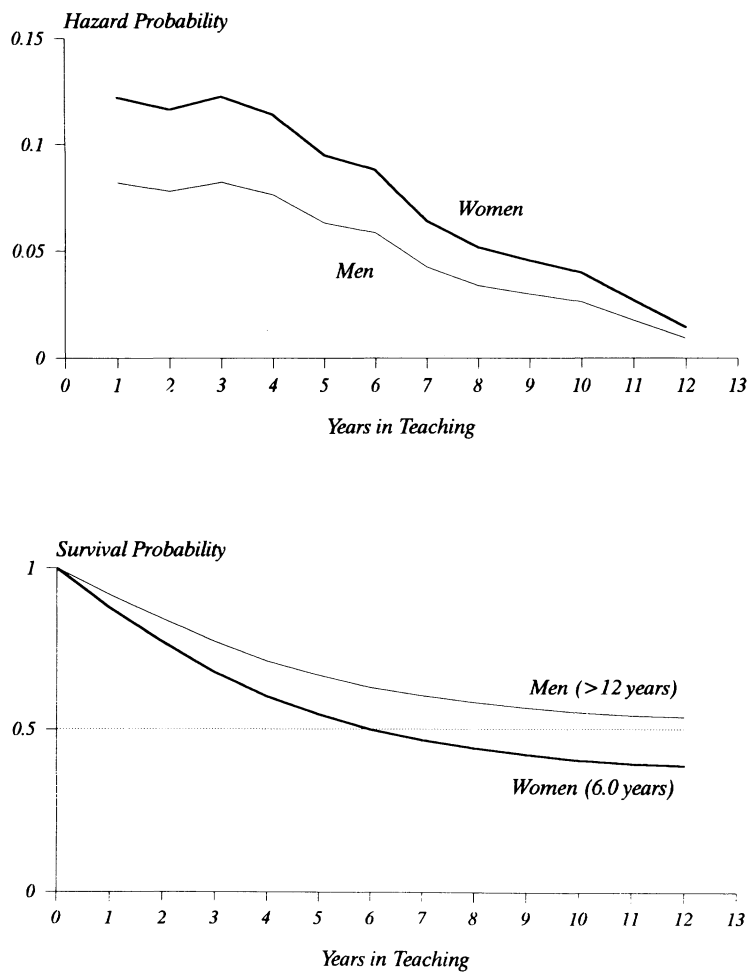
Hazard Probability

Years in Teaching

Survival Probability

Years in Teaching

FIGURE 4. *Fitted survivor and hazard functions describing the risks of leaving teaching by sex, from a hazard model containing the main effect of the time indicators and the main effect of time-invariant sex as predictors (estimated median lifetimes in parentheses)*

tion to make when modeling these data (see below), sampling variation ensures minor differences in shape such as those detected in Figure 2 where subsample profiles have been plotted.

The fitted survivor plots in Figure 4 show the cumulative effects of the yearly differentials in risk. At the end of 12 years, we estimate that the Michigan public schools are left with 54% of their male special educators versus only 39% of the women. Another perspective on the divergent

careers of the two groups comes from comparison of the estimated median lifetimes: 6.0 years for women versus more than 12 years for men. Even though censoring takes its toll (we cannot estimate a median lifetime precisely for men), the large difference between these "average" career durations is a meaningful way to communicate the results. In addition, we echo the advice of Gentleman and Crowley (1991), who present innovative methods for displaying survival data, but we believe that these different approaches are complements to, not replacements for, one another.

Testing and Interpreting the Effects of Time-Varying Predictors

These same methods of inference and display also permit the influence of time-varying predictors on the discrete-time hazard profile to be investigated. However, interpretation of detected effects is not quite so straightforward. We illustrate this by comparing Model A with Model C in Table 2—the latter model includes a time-varying predictor for the main effect of a teacher's job assignment SUPPORT:

$$\text{logit}_e(h_{ij}) = [\alpha_1 D_{1ij} + \alpha_2 D_{2ij} + \cdots + \alpha_{12} D_{12ij}] + \beta_2 \text{SUPPORT}_{ij}. \quad (13)$$

In each time period, the value of SUPPORT is set to 1, if the individual worked as a support staff member, or set to 0, if the individual worked as a classroom teacher. Because a person's job assignment can fluctuate from year to year, the values of SUPPORT may differ from year to year. In (13), β_2 captures the main effect of time-varying SUPPORT on the population logit-hazard profile.

Has adding SUPPORT to Model A improved prediction of the risk profile? A comparison of goodness-of-fit statistics for Models A and C confirms that SUPPORT is a statistically significant predictor of hazard ($\Delta\chi^2 = 7.43$, $\Delta df = 1$, $p = .0064$). The antilog of $\hat{\beta}_2$ is 1.21, indicating that the estimated odds of leaving in any given year are 1.21 times greater for support professionals than for classroom teachers.

This finding must be interpreted cautiously, however, because someone who provided support services at one time in his or her career may not necessarily be subjected to the additional risk of providing support services in every year—his or her job assignment may differ from year to year. Risk is only elevated in those years that individuals provide support services, not in the years they teach. The hazard model in (13) permits the extra risk associated with being a member of the support staff to begin in those years for which the individual has SUPPORT = 1. The model postulates that, although the values of SUPPORT may fluctuate over time, its effect on logit-hazard remains constant. The model is set up so that the time-varying predictor has a time-invariant effect.

Fluctuation in the values of a time-varying predictor from period to period complicates the use of fitted hazard and survivor profiles as an interpretive tool, because it is possible to plot a different fitted profile for every possible

temporal combination of values on the time-varying predictor. Unlike the case of a time-invariant predictor, where the investigator must select a single interesting predictor value at which to display the fitted profile, now a system of values must be selected to represent meaningful patterns of predictor fluctuation over time. Even in the case of our simple example of the dichotomous time-varying predictor SUPPORT, this can lead to many hundreds of potential fitted profiles.

To simplify display and interpretation, we recommend that a small number of fitted profiles be selectively displayed for substantively interesting temporal combinations of time-varying predictor values. For instance, in Figure 5, we present the two most extreme fitted hazard and survivor functions produced by different combinations of values of SUPPORT: one for individuals who were classroom teachers continuously throughout the entire period (SUPPORT set to 0 in every period) and the other for individuals who worked continuously as support professionals (SUPPORT set to 1 in every period). Because we have displayed the extremes, these hazard profiles comprise an envelope surrounding the fitted hazard profiles for special educators in all possible career combinations. People whose job assignment fluctuated over time will possess a fitted hazard function that falls intermittently at one or other of these two extremes. For example, the fitted hazard profile for individuals who worked 6 years in a support role and then became teachers for the rest of their careers would initially follow the top function (dotted line) and then skip to the bottom function (solid line) in Year 7.

Assumptions Implicit in the Discrete-Time Hazard Model

In postulating the discrete-time hazard model, we implicitly make three important assumptions. First, the linear-logistic model is a valid representation of reality—that equal differences in the values of a predictor correspond to equal vertical displacements of the logit-hazard profile (the *linearity* assumption). Second, the hazard model requires no error term—that all heterogeneity across individuals is accounted for by variation in the values of the covariates (the *no unobserved heterogeneity* assumption). Third, the logit-hazard profiles corresponding to all possible values of every predictor are distinguished only by their relative elevation (the *proportionality* assumption). We comment on each of these assumptions in turn.

The Linearity Assumption

The linearity assumption in the discrete-time hazard model resembles the linearity assumption made in linear regression analysis. However, here, the model assumes that vertical displacements in logit hazard are linear per unit of difference in each predictor. As in any linear model, we can check this assumption's tenability by: (a) exploratory data analysis, using graphical methods, or (b) statistical inference.
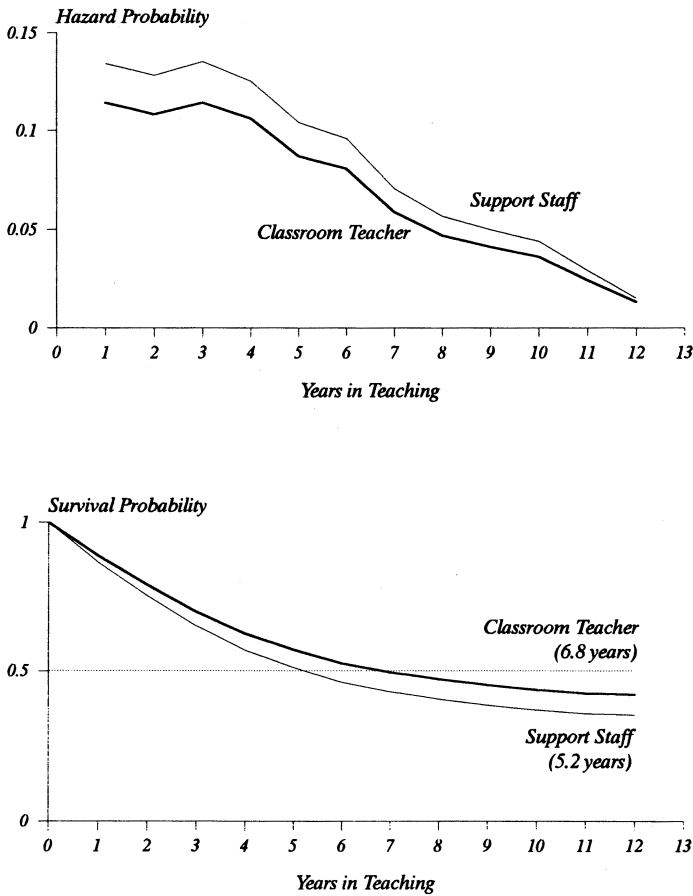
Hazard Probability

Years in Teaching

Survival Probability

Years in Teaching

FIGURE 5. *Fitted survivor and hazard functions describing the risks of leaving teaching by the teacher's job assignment (support staff vs. classroom teacher), from a hazard model containing the main effect of the time indicators and the main effect of time-varying job assignment (estimated median lifetimes in parentheses)*

Under the exploratory approach, in preliminary analyses, we simply divide the sample into strata by predictor values, estimate sample logit-hazard functions in each stratum, and display them simultaneously. The linearity assumption is met if equal differences in the predictor correspond to approximately equal vertical displacements of the subsample logit-hazard profiles; otherwise, the assumption is violated. Nonlinearity detected by this method usually can be resolved by transformation of the predictors or by judicious re-representation of a continuous predictor as a set of dummy

variables. When examining the effects of age at entry into teaching on the career lengths of special educators, for example, Singer (1993) originally treated age at entry as a continuous predictor. After estimating separate subsample logit-hazard functions for teachers in multiple entry-age strata (ages 20–21, 22–23, 24–25, 26–27, 28–29, 30–34 and 35+ years at entry), she concluded not only that the effect of age was nonlinear but that it behaved dichotomously. Teachers younger than 30 at entry shared one hazard function; teachers 30 and older at entry shared another. Because this cliff effect had been found in many other studies of teachers' careers, she abandoned the continuous specifications of entry age and reclassified the predictor as a dichotomy.

Alternatively, as in regression analysis, potential nonlinearity in the predictors can be checked by adding nonlinear predictor specifications to the hazard model. The linearity assumption is met if the addition of nonlinear terms does not improve fit; if the additional nonlinear term does improve fit, then it is retained in the model. When examining the relationship between a teacher's score on the National Teachers' Examination and his or her risk of leaving teaching, for instance, Murnane, Singer, and Willett (1989) and Murnane, Singer, Willett, Kemple, and Olsen (1991) found that the effect of test score on hazard had both a linear and a quadratic component. Very high-scoring teachers were disproportionately likely to leave.

The Assumption of No Unobserved Heterogeneity

The discrete-time hazard models presented in this article do not contain an error or random disturbance term. They simply reparameterize the multiple hazard parameters (the h_{ij}) as a deterministic function of a smaller subset of structural parameters (the αs and βs) common across individuals. The model specifies that individuals are distinguished only in their values on $Z_1, Z_2, \ldots, Z_P$. All variation in hazard profiles across individuals is hypothesized to depend solely on observed variation in the predictors. The survival analysis literature refers to such models as having no *unobserved heterogeneity*.

Omission of an important predictor from the model is tantamount to pooling the several hazard profiles for the heterogeneous populations defined by values of the ignored predictor. Some might argue that one can ignore the unobserved heterogeneity by simply regarding hazard modeling as an efficient method of simultaneously estimating hazard profiles in subgroups defined by the included predictors (subject to the proportional-odds assumption below). Although this rationalization seems reassuring, it ignores what Vaupel and Yashin (1985) have termed "heterogeneity's ruses" (pp. 176–183). In their article, they illustrate that mixing heterogeneous populations with different risk profiles can yield a pooled profile that may have a shape entirely different from the component profiles. This happens because selection occurs over time; individuals in high risk groups die out

early, and the surviving population comes to look less like the original population. For example, though risk profiles in the several unobserved subpopulations may each be flat over time, the aggregate profile may demonstrate diminishing risk over time as high risk individuals experience the event of interest and disappear quickly from the risk set (e.g., see Vaupel & Yashin, 1985, figs. 1–11). So the effects of selection due to unobserved heterogeneity directly impact the shape of fitted hazard profiles (i.e., the estimated time dependence of risk), and, when these shapes are interpreted, one can never be sure whether the obtained hazard profile legitimately describes the pattern of risk for a randomly selected member of the general population or whether it describes no one at all, being simply the mythical average of several vastly different profiles.

Allison (1982) suggests that a random error term ϵ_{ij} could be included in the hazard model. He notes, however, that these disturbances are likely to be autocorrelated over time, given the longitudinal nature of the process being modeled, and would therefore naturally undermine treatment of the discrete time periods as independent in analyses of the person-period data set. If the autocorrelation were ignored, then "by analogy with ordinary least squares regression, one would expect this dependence among the observations to lead to inefficient coefficient estimates and estimated standard errors that are biased downwards" (Allison, 1982, p. 83).

These problems, however, are not unique to discrete-time hazard modeling. The ruses of selection undermine interpretation of the shape of the hazard profile, regardless of whether it is obtained in discrete or continuous time. Random disturbances are also omitted from continuous-time hazard models, and, if included, ignored autocorrelation would "again lead to bias, inefficiency and inconsistent standard errors" here too (Allison, 1982, p. 83). Methods for dealing directly with these problems are in their infancy.

The Proportionality Assumption

Both continuous-time and discrete-time survival analysis usually involve a proportionality assumption. For instance, the discrete-time hazard model in (3) and (4) postulates that the entire family of logit-hazard profiles represented by all possible values of the predictors shares a common shape and is mutually parallel, being shifted only vertically for different values of the predictors. Because discrete-time hazard is a (conditional) probability, points on the logit-hazard function describe the log-odds of event occurrence in each time period (given no earlier occurrence of the event). Were we to antilog these logit-hazard profiles to yield profiles describing the conditional odds of event occurrence at different values of the predictors, then these new odds profiles would be magnifications or diminutions of one another—they would be proportional. This proportional-odds consequence of the discrete-time hazard model is referred to as the proportionality assumption.

When the magnitude of the hazard probabilities is small (e.g., as in Table 1), the odds of an event occurring are approximately equal to the probability of its occurring (i.e., $h_{ij} \approx h_{ij}/(1 - h_{ij})$ when h_{ij} is small). Thus, in many settings, the discrete-time proportional-odds assumption under-writes an approximate proportional-hazards assumption so that the hazard profiles corresponding to different values of the predictors also appear to be magnifications and diminutions of one another (see, as we have noted earlier, Figs. 4 and 5).[10]

We find it helpful to draw an analogy between the proportionality assumption in discrete-time survival analysis and the assumption of parallel regression slopes in the analysis of covariance. In analysis of covariance, the covariate adjustment is appropriate only if the regression lines describing the relationship between the outcome and the covariate are parallel in all subgroups, with each pair of regression lines having a constant vertical separation. In discrete-time hazard models that make the proportionality assumption, all possible logit-hazard functions are also assumed to be mutually parallel, with each pair also having a constant vertical separation.

There is a simple graphical method for verifying the proportionality assumption: In preliminary analyses, if logit-hazard profiles estimated separately within strata are all approximately parallel, then the assumption is met; if they are not, it is violated. But we hasten to add that we have found, in studies of a wide variety of phenomena—including teachers' careers, child mortality, duration of breast-feeding, time to undergraduate degree, time to doctorate, and age at first suicide ideation—that violations of the proportionality assumption are the rule, rather than the exception (Singer & Willett, 1991; Willett & Singer, 1991).

In many real world situations, logit-hazard profiles corresponding to different values of the predictors are not simply shifted versions of the same baseline shape; many predictors do more than vertically displace the logit-hazard profile—they also alter its shape. Ignoring such violations of the proportionality assumption challenges the integrity of the findings. To avoid this, we suggest that researchers routinely adopt a more skeptical approach from the outset. Assume the very real possibility of a nonproportional relationship between hazard and the chosen predictors, and test to confirm whether this is the case. Nonproportional relationships can be explored straightforwardly in discrete-time survival analysis by exploring statistical interactions between the substantive predictors and TIME, a procedure that we now describe in greater detail.

Nonproportional Hazard Models

The hazard models that we have postulated so far implicitly assume that a predictor has an identical effect in every time period. However, nature does not always operate so simplistically. The effects of both time-invariant

and time-varying predictors can themselves fluctuate over time. In our data example, for instance, the added risk attributed to being a member of the support staff may be larger early in the career than in the later years.

By virtue of the way they are formulated, discrete-time hazard models easily permit the exploration of whether the effect of any predictor varies over time; one simply includes the statistical interaction between TIME and the predictor in the hazard model. An interaction between a predictor and TIME implies that the effect of the predictor on the hazard profile differs from time period to time period—the vertical displacement in logit-hazard per unit difference in the predictor is no longer the same in every time period. When this occurs, the shapes of the logit-hazard functions corresponding to different values of the predictor are no longer identical, and the proportional-odds assumption has been violated.

Interactions between predictors and TIME are not simply methodological nuisances, they can lead to richer substantive interpretation. They are easily incorporated in discrete-time survival analyses by forming cross-products in the person-period data set between the time indicators and the chosen predictor and including these cross-products, along with the relevant main effects, as predictors in the hazard model. We use our data example to illustrate this by considering a discrete-time hazard model that includes a time-invariant predictor representing entering cohort and its interaction with TIME. Entry cohort is a substantively meaningful construct for special educators' careers because of the 1975 passage of Public Law 94-142, The Education for All Handicapped Children Act, which changed the demand for and responsibilities of special educators (Singer & Butler, 1987). Preliminary data analysis suggested that the effect of entry year was dichotomous (split at 1975), and so we used a dichotomy—COHORT—to indicate year of hire (0 for hire before 1975, 1 for hire in 1975 or later). The top panel of Figure 6 presents fitted hazard functions for a model that included only the main effects of TIME and COHORT. Special educators hired before 1975 seem to be more likely to leave in every year of their career. Because this is a main effects model (and because hazard probabilities are small), the upper hazard function is forced to be an (approximate) magnification of the lower one. Had we used a logit scale on the vertical axis, the two functions would be parallel and equidistant. This parallelism is a feature of the main effects model—it forces the shapes of the two hazard functions to be similar.

Now consider a discrete-time hazard model in which the effect of COHORT can vary across time periods:

$$\text{logit}_e(h_{ij}) = [\alpha_1 D_{1ij} + \cdots + \alpha_{12} D_{12ij}] + \beta_1(D_{1ij} \times \text{COHORT}_{ij})$$
$$+ \beta_2(D_{2ij} \times \text{COHORT}_{ij}) + \cdots + \beta_9(D_{9ij} \times \text{COHORT}_{ij}).$$

Notice that we have included only nine terms to represent the interaction. This is an idiosyncracy of the particular predictor used; because of censor-
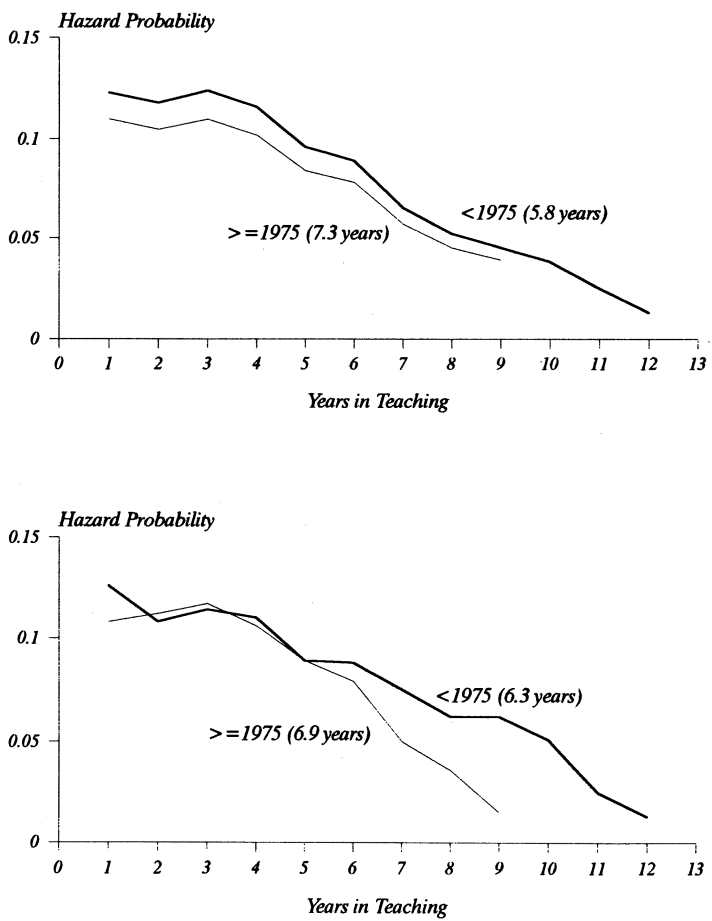
FIGURE 6. *Fitted hazard functions describing the risks of leaving teaching by year of entry into teaching (pre-1975 vs. 1975 and later), from a hazard model containing: (a) the main effect of the time indicators and the main effect of time-invariant entry cohort (top panel) and (b) the main effect of the time indicators, the main effect of time-invariant entry cohort, and the two-way interaction between the time indicators and time-invariant entry cohort (estimated median lifetimes in parentheses)*

ing, teachers in the later cohort have only 9 years of data. Statistical comparison of the COHORT main effects and interaction models reveals that the addition of the two-way interactions with TIME, as a group, significantly improves the prediction of hazard ($\Delta\chi^2 = 28.71$, $\Delta df = 8$, $p < .0001$). Thus, the proportional-odds assumption has been violated. The bottom panel of Figure 6 presents the fitted hazard functions for the interaction model.

Notice the failure of the proportional-odds assumption; the two fitted hazard functions now have very different shapes.

When the effect of a predictor—either time-invariant or time-varying—varies over time, proportional-odds models are inappropriate. Serious analytic consequences await those who blindly fit proportional-odds models without examining the tenability of the assumptions. Had we stopped at the main effects model here, for instance, we would have erroneously concluded that the estimated median lifetimes in the two cohorts differed by 1.5 years and that the effect of COHORT was constant over the career. The interaction model shows that the career paths of the two cohorts are nearly identical during teachers' early years on the job and that the estimated median lifetimes differ by only 0.6 years.

Further Advantages of Discrete-Time Survival Analysis

Discrete-time survival analysis offers an easily applicable framework for analyzing a type of event occurrence data that is frequently collected in educational research. These methods are comprehensible and convenient. In discrete-time, hazard is a probability (rather than a rate, as in continuous time). Interpretation of the parameters of the discrete-time hazard model is straightforward, and it can be fitted easily using standard logistic regression analysis.

We believe that the comprehensibility of discrete-time survival analysis is an important feature in its favor because, in a recent review of empirical applications of survival analysis in psychology (a field that has seen increasing applications of continuous-time methods), we found that many researchers did not understand that continuous-time hazard is a rate, rather than a probability (Singer & Willett, 1991). And, because the continuous-time proportional hazards model (Cox regression) is far removed from familiar statistical practice, researchers rarely use basic data analytic tools, such as searching for nonlinearity or interactions, in their work. Rather, the proportional-hazards assumption was accepted as fact, and researchers simply searched for predictors that had a "statistically significant" association with hazard. We believe that these problems are less likely to occur in discrete-time survival analysis.

Finally, the discrete-time approach facilitates examination of the shape of the hazard function. This is in sharp contrast to Cox regression, where the shape of the hazard function is ignored in favor of estimating only the shift parameters associated with covariates (our βs) under the assumption of proportionality. Because inspection of the shape of the hazard function indicates when an event is most likely to occur, and how these risks vary over time, we believe that descriptions of the shapes of hazard functions have an important role to play in the description of the educational career.

Notes

[1] We are referring here, and throughout the rest of the article, to the process of *right-censoring*, where the endpoints of the careers of a proportion of the sample are unknown. The reader must be careful to distinguish this from *left-censoring*, where the career start-points are unknown. This latter type of censoring is much more analytically intractable.

[2] Given, as we shall see, the values of selected covariates to be included as predictors of hazard in our statistical models.

[3] Our partition of continuous time most accurately captures the structure of these educational data and is consistent with the partition used by Miller (1981). Teachers are hired at Time 0. The first year of their career, Year 1, begins immediately after hire and extends to (and includes) the end of that year, which we write as $(0, t_1]$. Under our definition, the jth year, $(t_{j-1}, t_j]$, excludes t_{j-1} but includes t_j. Note that some authors (e.g., Cox & Oakes, 1984; Kalbfleisch & Prentice, 1980) partition time periods slightly differently as $[0, t_1), [t_1, t_2), \ldots, [t_j, t_{j+1}), \ldots$. Under their definition, the jth time period includes t_j but excludes t_{j+1}. The choice of partition has consequences for the definition and continuity of the cumulative distribution function and the survivor function (see footnote 7).

[4] As in ordinary regression, of course, this representation is most convenient when the predictors are defined so that 0 is a substantively interesting value (as in this example). Nevertheless, even when 0 is not a substantively interesting (or even legitimate) value, we can center the predictor or conceive of this baseline hazard profile much as we conceive of a y intercept.

[5] Other functional forms can be used as well, including the complementary log-log function and probit.

[6] Following Miller (1981, p. 2), we have defined the survivor function $S(t)$ as equal to $Pr\{T > j\}$. Under this definition, the cumulative distribution function (cdf), $F(t) = Pr\{T \leq j\}$, is equal to $1\text{-}S(t)$. This definition is consistent with our partition of time (see footnote 4) and ensures that: (a) the cdf is right continuous (as is customary, see Mood, Graybill, & Boes, 1974, Definition 2, pp. 54–56) and (b) the survivor function is left continuous. Some authors (e.g., Cox & Oakes, 1984) use an alternative definition of the survivor function, $S(t) = Pr\{T \geq t_j\}$, which, in conjunction with their alternative partition of the time axis (see footnote 3), leads to the left continuity of the cdf. While their definition "simplif[ies] slightly some subsequent formulae involving the hazard function" (p. 13), even Cox and Oakes (1984) note that it is based on an "unusual convention" (p. 13). Despite these distinctions, as long as each definition uses the appropriate time partition, the survivor functions cover identical regions of the time axis (assuming $Pr\{t \leq T < t + \Delta t\} \to 0$ as $\Delta t \to 0$), and the results of this article hold.

[7] We have used a "no intercept" formulation for interpretive convenience. If this option is not available to the reader, then the hazard models can be modified to include an intercept, α_0. Then, however, one of the time indicators—the first, D_{1ij}, say—must be dropped from the model to avoid complete linear dependence among the time indicators. Then, for instance, Model A becomes:

$$\text{logit}_e(h_{ij}) = [\alpha_0 + \alpha_2 D_{2ij} + \alpha_3 D_{3ij} \ldots + \alpha_{12} D_{12ij}],$$

with similar modifications to other models. In the reformulated models, interpretation of the α parameters differs: α_0 is the logit-hazard in the first time period, and α_2 through α_{12} represent deviations from this logit-hazard in each of the subsequent time periods.

[8] The subscript i has been suppressed in this expression for estimated first-year hazard due to the omission from the model of predictors that distinguish among individuals.

[9] Strictly speaking, this apparent magnification of one hazard profile to generate the other is only approximate in the discrete-time hazard model and only holds when h_{ij} is small. For further discussion of this point, see the next section.

[10] In continuous-time survival analysis, where log-hazard rate (rather than logit-hazard probability) is modeled as a linear function of predictors, the proportionality assumption is a true proportional-hazards assumption.

APPENDIX
Estimating discrete-time hazard models using SAS (Version 6.1)

```
*   CREATING THE PERSON PERIOD DATA SET *;

DATA PERSPER;
    SET PERSON;

    *   IN THE CODE BELOW, THE ARRAY "TIME" INDEXES THE TIME INDICATORS, "ASSIGN"
        INDEXES THE TIME-VARYING ASSIGNMENT INDICATORS, AND "FEMTIME" INDEXES
        THE INTERACTION BETWEEN FEMALE AND THE TIME INDICATORS. ALSO NOTE THAT
        WE HAVE CODED THE NON-EVENT CATEGORY OF Y TO BE 2 INSTEAD OF 0 SO THAT
        WE ARE CONSISTENT WITH SAS' MANUAL SPECIFICATION. *;

    ARRAY TIME[12] D01-D12;
    ARRAY ASSIGN[12] S01-S12;
    ARRAY FEMTIME[12] FD01-FD12;
    DO PERIOD = 1 TO MIN(LASTPD, 12);
        IF PERIOD = LASTPD AND CENSOR = 0 THEN Y = 1;
            ELSE Y = 2;
        DO INDEX  = 1 TO 12;
            SUPPORT = ASSIGN[PERIOD];
            IF INDEX = PERIOD THEN TIME[INDEX] = 1;
                ELSE TIME[INDEX] = 0;
            FEMTIME[INDEX] = FEMALE*TIME[INDEX];
        END;
        OUTPUT;
    END;

*   FITTING BASELINE MODEL WITH TIME INDICATORS ONLY * ;

PROC LOGISTIC DATA = PERSPER NOSIMPLE OUT = ESTIMATE;
    TITLE2 "BASELINE MODEL";
    MODEL Y = D01-D12/NOINT;

*   RECONSTRUCTING FITTED HAZARD AND SURVIVAL FUNCTIONS * ;

DATA ESTIMATE (REPLACE = YES);
    SET ESTIMATE;

    ARRAY TIME [12] D01-D12;
    SURVIVAL = 1;
    DO PERIOD = 1 TO 12;
        X = TIME[PERIOD];
        HAZARD = 1/(1 + (EXP(-X)));
        SURVIVAL = (1 - HAZARD)*SURVIVAL;
        OUTPUT;
```

```
    END;
    KEEP PERIOD SURVIVAL HAZARD;

PROC PRINT DATA = ESTIMATE UNIFORM NOOBS;
    ID PERIOD;
    VAR SURVIVAL HAZARD;
    FORMAT SURVIVAL HAZARD 6.4;

PROC PLOT DATA = ESTIMATE;
    PLOT (SURVIVAL HAZARD)*PERIOD;

*   MODEL WITH MAIN EFFECT OF FEMALE *;

PROC LOGISTIC DATA = PERSPER NOSIMPLE OUT = ESTIMATE;
    TITLE2 "MAIN EFFECTS OF FEMALE";
    MODEL Y = D01-D12 FEMALE/NOINT;

DATA ESTIMATE (REPLACE = YES);
    SET ESTIMATE;

    ARRAY TIME  [12] D01-D12;
    DO SEX = 1 TO 2;
        SURVIVAL = 1;
        DO PERIOD = 1 TO 12;
            X = TIME[PERIOD] + (SEX-1)*FEMALE;
            HAZARD = 1/(1 + (EXP(-X)));
            SURVIVAL = SURVIVAL*(1 - HAZARD);
            OUTPUT;
        END;
    END;
    KEEP SEX PERIOD SURVIVAL HAZARD;

PROC SORT DATA = ESTIMATE;
    BY SEX;

PROC PRINT DATA = ESTIMATE NOOBS UNIFORM;
    BY SEX;
    ID PERIOD;
    VAR SURVIVAL HAZARD;
    FORMAT SEX SEXFMT. SURVIVAL HAZARD 6.4;

PROC PLOT DATA = ESTIMATE;
    PLOT (SURVIVAL HAZARD)*PERIOD = SEX;

*   MODEL WITH INTERACTION BETWEEN FEMALE AND TIME *;

PROC LOGISTIC DATA = PERSPER NOSIMPLE OUT = ESTIMATE;
  TITLE2 "INTERACTION BETWEEN FEMALE AND TIME";
  MODEL Y = D01-D12 FD01-FD12/NOINT;

DATA ESTIMATE (REPLACE = YES);
    SET ESTIMATE;

    ARRAY TIME [12] D01-D12;
    ARRAY FEMTIME [12] FD01-FD12;
    DO SEX = 1 TO 2;
        SURVIVAL = 1;
        DO PERIOD = 1 TO 12;
            X = TIME[PERIOD] + (SEX-1)*FEMTIME[TIME];
            HAZARD = 1/(1 + (EXP(-X)));
            SURVIVAL = SURVIVAL*(1 - HAZARD);
            OUTPUT;
        END;
    END;
    KEEP SEX PERIOD SURVIVAL HAZARD;

PROC SORT DATA = ESTIMATE;
    BY SEX;

PROC PRINT DATA = ESTIMATE NOOBS UNIFORM;
    BY SEX;
```

```
ID PERIOD;
VAR SURVIVAL HAZARD;
FORMAT SURVIVAL HAZARD 6.4;

PROC PLOT DATA = ESTIMATE;
PLOT (SURVIVAL HAZARD)*PERIOD = SEX;
```

References

Abedi, J., & Benkin, E. (1987). The effect of students' academic, financial and demographic variables on time to the doctorate. *Research in Higher Education*, 27, 3–14.

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

Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data.* (Sage University Paper Series on Quantitative Applications in the Social Sciences, Series No. 07-046). Beverly Hills: Sage.

Blossfeld, H. P., Hamerle, A., & Mayer, K. U. (1989). *Event history analysis: Statistical theory and application in the Social Sciences.* Hillsdale, NJ: Erlbaum.

Boli, J., Katchadourian, H., & Mahoney, S. (1988). Analyzing academic records for informed administration: The Stanford curriculum study. *Journal of Higher Education*, 59, 54–68.

Brown, C. C. (1975). On the use of indicator variables for studying the time-dependence of parameters in a response-time model. *Biometrics*, 31, 863–872.

Charters, W. W., Jr. (1970). Some factors affecting teacher survival in school districts. *American Educational Research Journal*, 7, 1–27.

Civian, J. T. (1990). *Using proportional hazards models to examine time to doctorate.* Unpublished doctoral dissertation, Harvard University Graduate School of Education, Cambridge, MA.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B*, 34, 187–202.

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data.* London: Chapman & Hall.

Efron, B. (1988). Logistic regression, survival analysis, and the Kaplan-Meier curve. *Journal of the American Statistical Association*, 83, 414–425.

Frank, A. R., & Keith, T. Z. (1984). Academic abilities of persons entering and remaining in special education. *Exceptional Children*, 51, 76–77.

Gentleman, R., & Crowley, J. (1991). Graphical methods for censored data. *Journal of the American Statistical Association*, 86, 678–683.

Goldstein, R., Anderson, J., Ash, A., Craig, B., Harrington, D., & Pagano, M. (1989). Survival analysis software on MS/PC-DOS computers. *Journal of Applied Econometrics*, 4, 393–414.

Gross, A. J., & Clark, V. A. (1975). *Survival distributions.* New York: Wiley.

Harrell, F., & Goldstein, R. (in press). Survival analysis software for microcomputers: A review. *The American Statistician*.

Hosmer, D., & Lemeshow, S. (1989). *Applied logistic regression.* New York: Wiley.

Kahn, M., & Kulick, F. (1975). Relationship of drug involvement to dropping out of college. *Psychological Reports, 37*, 1095–1098.

Kalbfleisch, J. D., & Prentice, R. L. (1980). *The statistical analysis of failure time data*. New York: Wiley.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*, 457–481.

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

Miller, R. G. (1981). *Survival analysis*. New York: Wiley.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *An introduction to the theory of statistics* (3rd ed.). New York: McGraw-Hill.

Murnane, R. J., Singer, J. D., & Willett, J. B. (1988). The career paths of teachers: Implications for teacher supply and methodological lessons for research. *Educational Researcher, 17*(6), 22–30.

Murnane, R. J., Singer, J. D., & Willett, J. B. (1989). The influences of salaries and "opportunity costs" on teachers' career choices: Evidence from North Carolina. *Harvard Educational Review, 59*(3), 325–346.

Murnane, R. J., Singer, J. D., Willett, J. B., Kemple, J., & Olsen, R. (1991). *Who will teach? Policies that matter*. Cambridge, MA: Harvard University Press.

Ott, M. D. (1988). An analysis of predictors of early academic dismissal. *Research in Higher Education, 28*, 34–48.

Pierce, D. A., Stewart, W. H., & Kopecky, K. J. (1979). Distribution-free regression analysis of grouped survival data, *Biometrics, 35*, 785–793.

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

SAS Institute. (1989). *User's guide, version 6* [Computer program]. Cary, NC: Author.

Singer, J. D. (1993). Are special educators' career paths special?: Results from a 13-year longitudinal study. *Exceptional Children, 59*, 262–279.

Singer, J. D., & Butler, J. A. (1987). PL 94-142: The Education for All Handicapped Children Act: Schools as agents of social reform. *Harvard Educational Review, 57*, 125–152.

Singer, J. D., & Willett, J. B. (1991). Modeling the days of our lives: Using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events. *Psychological Bulletin, 110*(2), 268–290.

Stage, F. K. (1988). University attrition: LISREL with logistic regression for the persistence criterion. *Research in Higher Education, 29*, 343–357.

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

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

Vaupel, J. W., & Yashin, A. I. (1985). Heterogeneity's ruses: Some surprising effects of selection on population dynamics. *The American Statistician, 39*, 176–185.

Willett, J. B., & Singer, J. D. (1991). From whether to when: New methods for studying student dropout and teacher attrition. *Review of Educational Research, 61*(4), 407–450.

Yamaguchi, K. (1991). *Event history analysis*. Newbury Park, CA: Sage.
Zwick, R., & Braun, H. I. (1988). Methods for analyzing the attainment of graduate school milestones: A case study (ETS Research Report No. 88-30). Princeton, NJ: Educational Testing Service.

Authors

JUDITH D. SINGER is Associate Professor, Graduate School of Education, Harvard University, Appian Way, Cambridge, MA 02138. She specializes in quantitative methods.

JOHN B. WILLETT is Associate Professor, Graduate School of Education, Harvard University, Appian Way, Cambridge, MA 02138. He specializes in quantitative methods.