Journal of Consulting and Clinical Psychology
1993, Vol. 61, No. 6, 952-965


0022-006X/93/$3.00

# Investigating Onset, Cessation, Relapse, and Recovery: Why You Should, and How You Can, Use Discrete-Time Survival Analysis to Examine Event Occurrence

John B. Willett and Judith D. Singer

In this article, we show how discrete-time survival analysis can address questions about onset, cessation, relapse, and recovery. Using data on the onset of suicide ideation and depression and relapse into cocaine use, we introduce key concepts underpinning the method, describe the action of the discrete-time hazard model, and discuss several types of main effects and interactions that can be included as predictors. We also comment on practical issues of data analysis and strategies for interpretation and presentation.

Investigations of onset, cessation, relapse, and recovery ask questions about the occurrence and timing of critical events. When examining the course of affective illness, for instance, researchers ask when the disease first strikes, how much time passes between onset and treatment, how long the initial illness lasts, how long treated patients remain well, whether people with a history of illness relapse, how long treated people remain well, and how long second and subsequent illness spells last. Similar questions about event occurrence arise in studies of other illnesses, undesirable behaviors, and treatment efficacy.

Until recently, appropriate methods for studying event occurrence were unavailable; now, however, the methods of survival analysis offer the required tools. Statistical packages have begun to include survival analysis routines (Harrell & Goldstein, 1992). Nontechnical introductions to the topic are available (e.g. Allison, 1984; Singer & Willett, 1991), in addition to the classic references (e.g. Kalbfleish & Prentice, 1980). Continuous-time survival methods, which require that time-to-event be recorded on a continuous scale, are popular. Discrete-time methods, which only require knowledge of the particular time period in which an event occurs, are less prevalent.

In this article, we introduce discrete-time survival analysis to clinical research. We emphasize this methodology for five reasons. First, it is suited to much event history data collected in clinical settings where, for logistical and financial reasons, observations are often made in discrete time. Second, its simplicity provides an intuitive arena in which to begin learning about survival analysis. Third, time-varying predictors—whose values fluctuate over time—can be more easily included in discrete-time models. Fourth, common model violations are easily tested and remediated in a discrete-time framework. Finally, discrete-time survival analysis does not require special computer software—all estimation can be carried out within a standard statistical package.

Here we describe the foundations of discrete-time survival analysis and review how various effects can be modeled. We offer data-analytic advice and show how to communicate findings to the layperson. Our approach is conceptually oriented, and we use real data to motivate our arguments. Further technical details can be found in a companion article (Singer & Willett, 1993).

## Why Are Traditional Methods Unsuitable for Studying Event Occurrence?

When event occurrence is studied, there are usually some people in the sample that do not experience the target event during data collection. The event histories of these people are *censored*.[1] For example, in Glasgow, Klesges, Klesges, and Somes' (1988) study of smoking relapse, 15% of the sample did not relapse during the half-year of observation. Because we do not know how long these 15% ultimately remained abstinent, their smoking cessation histories are censored. The rarer the event, the more censoring occurs. When Kelly, Lawrence, and Brasfield (1991) studied whether gay men relapsed into high-risk sexual behavior after counseling, two thirds of the sample were censored because they were still "safe" 16 months later.

If an event history is censored, the researcher has incomplete information about event occurrence and knows only that, if the

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

The order of the authors was determined by randomization. We thank the American Statistical Association for support during their research fellowship at the National Center for Education Statistics, where this article was completed. We are indebted to Niall Bolger, Sharon M. Hall, and Susan B. Sorenson for providing empirical data.

Correspondence concerning this article should be addressed to John B. Willett or Judith D. Singer, Harvard University, Graduate School of Education, Appian Way, Cambridge, Massachusetts 02138.

[1] Technically, this is right-censoring. It must be distinguished from left-censoring, where the beginning of the career or lifetime is unknown.

Table 1
*What Do Survival Data Look Like? Age at First Suicide Ideation Among 417 College Students*

| Age (years) | No. of students who: Had not yet thought about suicide at the *beginning* of the year | No. of students who: Onset during *the* year | No. of students who: Were censored at *the end of* the year | Proportion of: *All* students who had not onset *by the end of* the year | Proportion of: Students who had not yet thought about suicide who *onset* during this year |
|---|---|---|---|---|---|
| 6 | 417 | 2 | 0 | .9952 | .0048 |
| 7 | 415 | 3 | 0 | .9880 | .0072 |
| 8 | 412 | 13 | 0 | .9568 | .0316 |
| 9 | 399 | 8 | 0 | .9376 | .0201 |
| 10 | 391 | 24 | 0 | .8801 | .0614 |
| 11 | 367 | 9 | 0 | .8585 | .0245 |
| 12 | 358 | 45 | 0 | .7506 | .1257 |
| 13 | 313 | 44 | 0 | .6451 | .1406 |
| 14 | 269 | 31 | 0 | .5707 | .1152 |
| 15 | 238 | 37 | 0 | .4820 | .1555 |
| 16 | 201 | 21 | 2 | *.4317* | .1045 |
| 17 | 178 | 17 | 11 | *.3904* | .0955 |
| 18 | 150 | 18 | 23 | *.3436* | .1200 |
| 19 | 109 | 11 | 31 | *.3089* | .1009 |
| 20 | 67 | 3 | 23 | *.2951* | .0448 |
| 21 | 41 | 1 | 40 | *.2879* | .0244 |

*Note.* Data were kindly supplied by Niall Bolger.

person ever experiences the event, it is after data collection ends. The researcher knows neither when nor whether the event will occur for this person—the very questions of interest. Censoring is evident in Table 1, which presents data on the ages at which students reported first thinking about suicide. Bolger, Downey, Walker, and Steininger (1989) collected these data retrospectively from 417 undergraduates. The first column lists student age in years. The next three columns tally the number of people who had not yet thought of suicide at the beginning of each age period, the number who contemplated suicide during each period, and the number whose histories were censored at the end of the period. Notice that all censoring occurs in the later age periods (age 16 and older). This pattern is common in retrospective studies of age-heterogeneous samples in which event histories are censored by data collection. In Table 1, 287 adolescents had considered suicide by the date of data collection, and 130 had not; the latter are censored.

Researchers react to censoring with several strategies. Seeking to use standard statistical methods, some focus exclusively on those people with known event times and set aside censored cases. In Table 1, this leads to an estimated average age at first suicide ideation of 13.5 years. However, of course, excluding 130 censored cases distorts the real distribution of age at first onset; the true average age at first suicide ideation must be greater than 13.5 years because this biased estimate ignores everyone who will first contemplate suicide later in life.

Other investigators impute the missing duration data, assigning to censored cases the duration value they possess when data collection ends so that they can be included in traditional analyses. This strategy is common in studies of relapse into undesirable behavior, such as smoking, alcohol abuse, and drug use. In our example, this strategy yields an estimated age at first suicide ideation of 15.4 years. Although larger than the original underestimate, this estimate, too, is incorrect. Imputing event times for censored cases changes a nonevent into an event and makes the "new" events occur at the earliest time possible.

Other researchers are more conservative. Rather than imputing event occurrence, they sacrifice some of their data on duration: They dichotomize the event histories at a particular time and ask *whether* the event has occurred by that time. In Table 1, for instance, there is a 35% chance that a student first thinks about suicide before age 14. Unfortunately, dichotomization eliminates much variation by pooling everyone into one of two groups. Dichotomizing at age 14 in Table 1, for example, throws away the ages at suicide ideation for the 108 students who first contemplated suicide after age 14 but before data collection. Also, any particular dividing criterion is arbitrary: Why choose age 14 rather than 13? Contradictory conclusions can result from different dichotomization times.

### *New Methods for Summarizing Event History Data*

Answering questions about event occurrence demands ways of summarizing discrete-time event history data that do not sacrifice data richness to arbitrary methods of sample aggregation. One such summary is the sample *survivor function*, a chronologically arranged list of estimated survival probabilities—the proportions of the initial sample that do not experience the event through each of several successive time periods. Simplicity and interpretability have fostered the appearance of sample survivor functions in the clinical research literature to summarize whether and when former alcohol abusers relapse (Cooney, Kadden, Litt, & Getter, 1991), former smokers relapse (Stevens & Hollis, 1989), released prisoners are reconvicted (Hart,

Kropp, & Hare, 1988), and violent individuals are arrested (Weisz, Walter, Weiss, Fernandez, & Mikow, 1990). Sample survivor probabilities for the suicide ideation data are presented in the fifth column of Table 1, indicating the proportion of the original sample that had not contemplated suicide by the end of each period: 88% had not first thought about suicide by the end of their 10th year, 48% had not done so by the end of their 14th year. Sadly, in our example (and in most survival studies), censoring prevents direct estimation of the sample survivor function for the entire time period under study. Why? Because we do not know, for example, whether the two adolescents censored at 16 ultimately contemplated suicide at 17 or 18. However, as we show below, the remaining sample survival probabilities can be estimated indirectly (giving the italicized values in column 5).

The sixth column of Table 1 presents a very useful summary of discrete-time event history data—the sample hazard probabilities. Defining each time period's risk set as those people "at risk" of experiencing the event in that period, the hazard probability is the proportion of the risk set who experience the event in that period. The hazard function is a chronological profile of these probabilities. In Table 1, everyone was a member of the age-6 risk set because no one considered suicide before then. Of the 417 children at risk of experiencing the event during the age-6 time period (i.e., those who had not yet thought about suicide for the first time), only 2 did so, leaving 415 children remaining in the age-7 risk set. The sample hazard probability at age 6 is therefore 2/417, or .0048. Of the 415 children remaining in the risk set, 3 contemplated suicide at age 7, yielding a sample hazard probability for this period of 3/415, or .0072, and so on.

Unlike the survivor probabilities, the sample hazard probabilities can be computed in every time period regardless of censoring—censored observations are simply removed from the risk set at the appropriate juncture, reducing the denominator of the hazard quotient. Consider the transition between ages 17 and 18 in Table 1. Of the 178 adolescents who had not yet first thought of suicide by the beginning of their 17th year, 17 did so during this year and 11 were censored. Thus, the sample hazard probability at age 17 was 17/178, or .0955. However, neither the 17 ideators nor the 11 censorees continue as members of the risk set into the 18th year because they had either experienced the event or their fates were unknown at age 17. Thus, only 150 people (178 − 17 − 11 = 150) were known to be at risk at age 18, providing the denominator of the age-18 hazard computation. Figure 1 (top panel) presents the sample discrete-time hazard function (interpolating linearly between adjacent hazard probabilities).

The hazard function is the cornerstone of survival analysis for several reasons. First, it tells us exactly what we want to know—whether and, if so, when events occur. Its magnitude summarizes the risk of event occurrence in each period. The sample hazard function of Figure 1, for example, suggests that adolescents were most likely to first contemplate suicide between ages 12 and 16. Second, the hazard function involves both noncensored and censored cases. In Table 1, for instance, every adolescent features equally in the hazard computation until he or she either contemplates suicide or is censored. Third, the sample hazard probabilities are computed in every time period that an event occurs—no information is ignored or pooled. Finally, the sample hazard function can be used to estimate the sample survivor function indirectly in time periods that censoring precludes its direct computation. Consider the age-15 and age-16 rows in Table 1. From the sample survival probability at age 15, we see that 48.2% of all children will not first contemplate suicide up to, and including, their 15th year. From the age-16 row, where the sample hazard probability is 10.45%, we estimate that 89.55% (=100% − 10.45%) of the entering-16th-year risk set will not first contemplate suicide in this year. Putting these two summaries together, we estimate that, at the end of the 16th year, 89.55% of the entering 48.2% will still not have experienced the target event. Thus, the survival probability at the end of the 16th year is simply .8955 multiplied by .4820, or .4317. The sample survival probability for any time period is just 1 minus the hazard probability for that period multiplied by the sample survival probability from the previous period (provided that censoring is independent of event occurrence). In Table 1, this "method of cumulation" provides estimated survival probabilities for ages 16 through 21. Figure 1 (lower panel) presents the sample survivor function.

The "center" of the sample survivor function can be summarized by the *median lifetime*, the time at which half the sample have experienced the target event, and half have not (i.e., when the sample survivor function equals .5). From Figure 1, the estimated median age that adolescents first contemplated suicide is 14.8 years. The median lifetime answers the descriptive question "On average, when did students first contemplate suicide?" and it uses data from noncensored and censored cases alike.

## The Statistical Underpinnings of Discrete-Time Survival Analysis

Statistical analysis does more than achieve appropriate sample summarization, it uses sample data to make inferences about the population. To achieve this, we must assume that a random sample of people are available from the population. Each population member is at risk of experiencing the selected target event—onset of a disease or behavior, relapse into drug abuse, ultimate recovery, and so on.[2] The population hazard function describes the risk of event occurrence in each time period—the probability that a randomly selected population member will experience the event in that period, given that it has not already occurred.[3] Our goal is to estimate this population profile of risk.

Usually, we also want to know if the population hazard function differs systematically for different types of people. In the case of first onset of major depression, for instance, we might ask, Is the hazard function describing the risk of depression

[2] We assume that each person can experience the target event once. Analytic methods for the single-event case are generalizable to the repeated occurrence of events (see Willett & Singer, in press).

[3] In continuous time, we know exactly when an event occurs and we define hazard differently because the probability that an event occurs at a single "infinitely thin" instant of time is zero (by definition). Continuous time hazard is the instantaneous rate of event occurrence, given no prior event occurrence (see Kalbfleisch & Prentice, 1980). It can assume any value greater than or equal to zero.

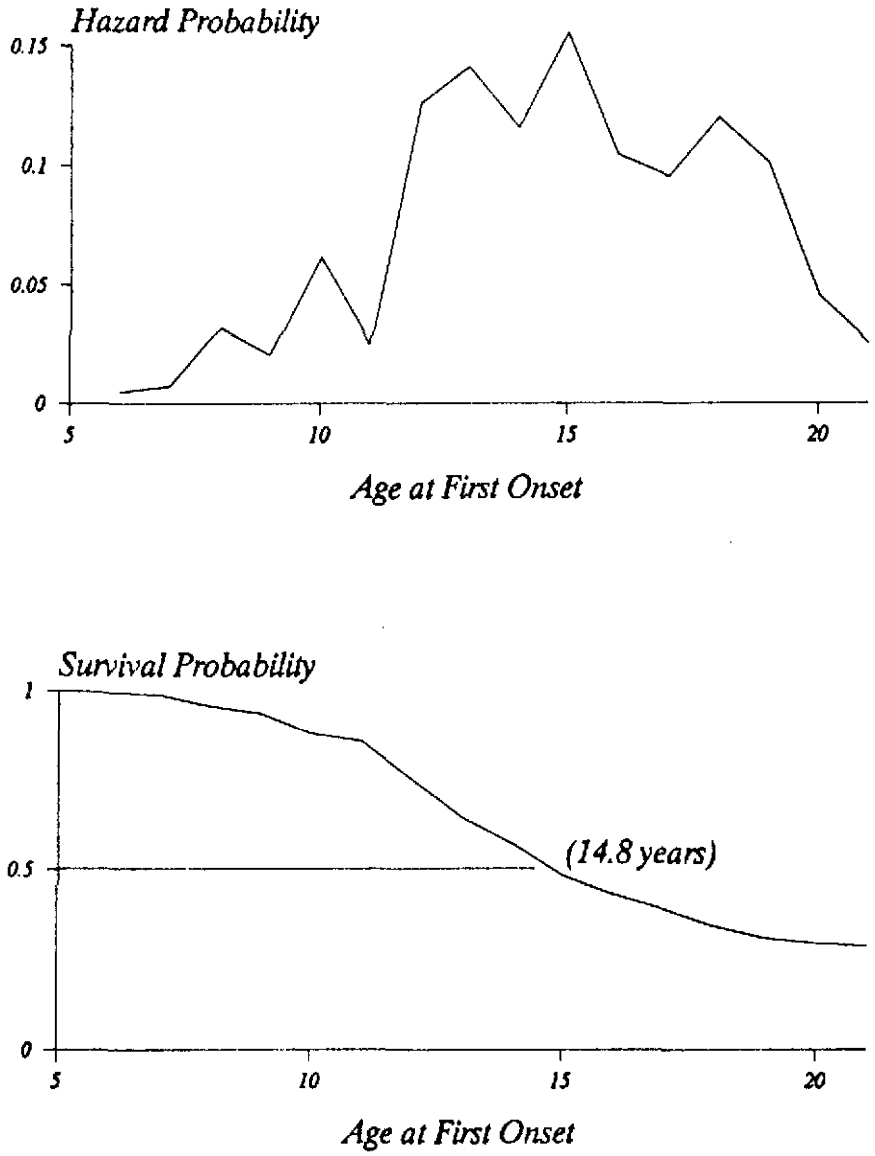


*Figure 1.* Sample survivor and hazard functions for age at first suicide (in years) ideation among 417 college-bound adolescents (estimated median lifetime in parentheses).

different for men and women? Does it differ among participants in different counseling programs or different types of therapy? Such questions ask about the relationship between the population hazard function and attributes of a person's background, *training, treatment, and so on. Conceptually, they treat the entire* hazard function as an outcome and variables like gender, type of counseling program or drug therapy as predictors of that outcome. We use a statistical model to represent the hypothesized relationship between outcome and predictors. Once the model has been specified and its parameters defined, we can test its fit to data and estimate its parameters.

But what statistical model best represents the population relationship between hazard profile and predictors? To motivate model specification, we examine the age of first onset of a major depressive disorder using retrospective data collected from a random sample of 3,131 adults between the ages of 18 and 97 (Sorenson, Rutter, & Aneshensel, 1991). For simplicity, we will consider the relationship between risk of onset and one predictor—respondent gender. If the risk profile for onset of major depression is related to gender, then perhaps we can learn about the nature of that relationship by inspecting hazard functions estimated separately for subsamples of men and women (Figure 2, top panel).

There are other ways of presenting the same information. For example, the hazard profiles could be displayed as odds, not probabilities. When hazard probability in a time period is .2, there is a 20% chance that the event of interest will occur in the period and an 80% chance that it will not (given no prior occurrence) and so the odds of event occurrence in this period are .2 to .8, usually written as 1/4 or .25. Odds can be computed

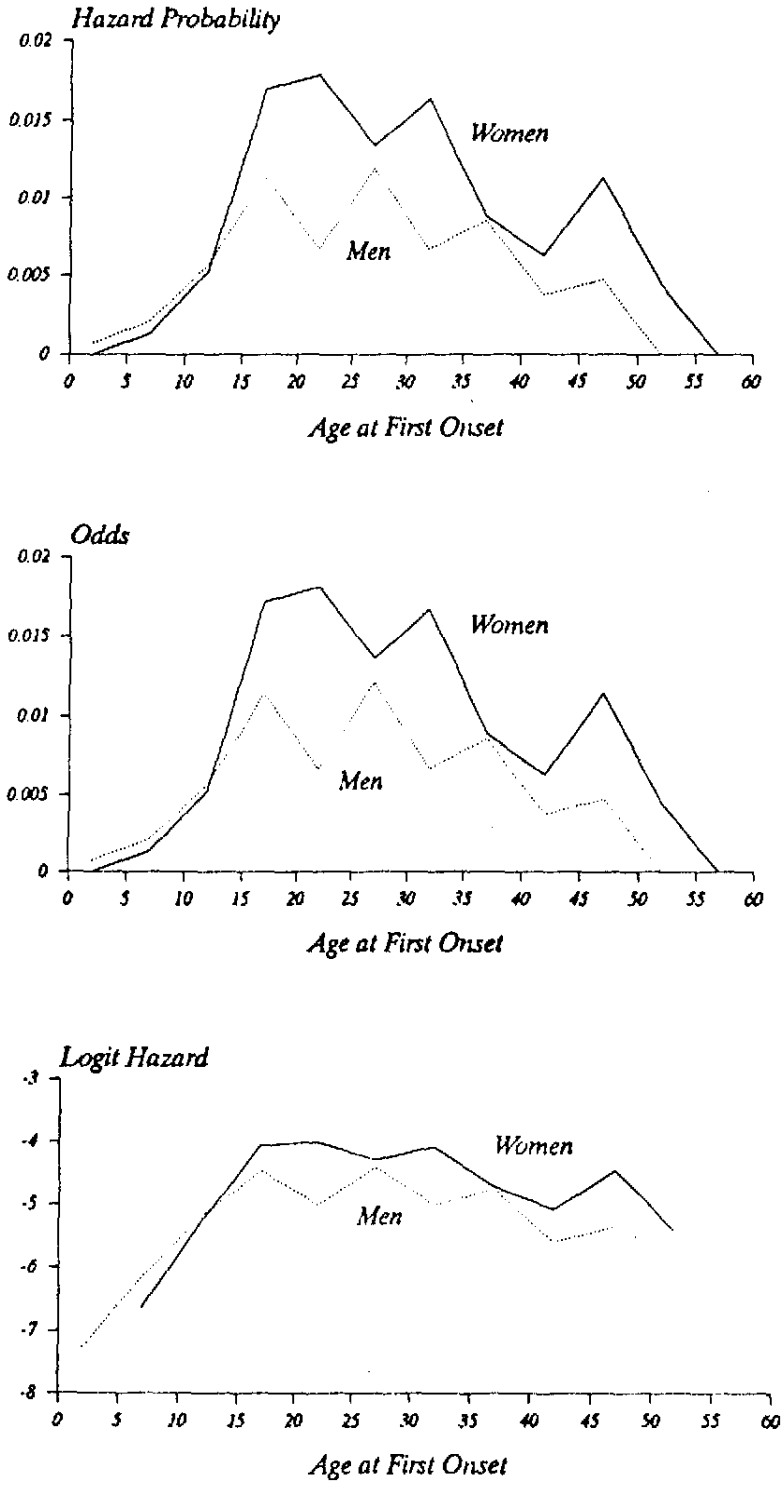


*Figure 2.* Sample risk functions describing the risks of onset of depressive disorder for men (dotted lines) and women (solid lines), on three scales: raw hazard probability (top panel), odds (middle panel), and log-odds, or logit hazard (bottom panel). Data were kindly supplied by Susan B. Sorenson.

by the formula odds = hazard/(1 − hazard). Figure 2 (middle panel) presents subsample odds profiles. Notice that the hazard and odds profiles are almost identical, an identity that holds only when hazard is small because a small number divided by 1 minus itself is approximately equal to the number. Figure 2 (lower panel) shows the natural logarithm of the subsample odds profiles, the logarithm being chosen as a standard transformation because it symmetrizes the potential range of the odds statistic. In discrete-time survival analysis, the label *logit-hazard* refers to the log-odds of event occurrence in any time period (given no earlier occurrence).

All three panels of Figure 2 tell the same story. Within any panel, men's and women's risk profiles are somewhat similar. The risk of first onset of depression is highest in the 20s and early 30s. Despite a crossover in preadolescence and haphazard zigzagging that subsequent analysis shows is sampling variation, there seems to be a small vertical separation between the male and female profiles. The sample hazard function for women is almost always higher than for men, as are the corresponding odds and logit-hazard profiles. Throughout most of their lives, women run a greater risk of depression.

Statements about the general shape of the risk profile and the difference in the net level of risk between the two groups conceptually decompose hazard into two parts: a baseline profile of risk and a shift in profile associated with gender. These ideas can be systematized by defining a predictor *FEMALE* (coded 0 = male, 1 = female). Then, treating the hazard function for men (when *FEMALE* = 0) as a baseline, we can imagine estimating the shift in profile when *FEMALE* switches from 0 to 1, generating the risk profile for women. In other words, ignoring minor variations in the shapes of the two profiles for the moment, we are drawn toward a hazard model specification in which variation in a predictor acts to vertically displace a baseline hazard profile, yielding another profile. For mathematical convenience, Cox (1972) framed the discrete-time hazard model in terms of logit-hazard rather than hazard. Similarities among the panels of Figure 2 confirm that this is reasonable. Therefore, a statistical model for the dependence of the risk profile for onset of depression on *FEMALE*, is as follows:

$$\log_e\left(\frac{h_j}{1-h_j}\right) = \text{logit}_e(h_j) = [\alpha_1 D_1 + \alpha_2 D_2 + \cdots + \alpha_J D_J] + \beta_1 FEMALE \quad (1)$$

The discrete-time hazard model in Equation 1 represents the hypothesized population relationship between logit-hazard and several predictors, including *FEMALE* and some that have yet to be explained ($D_1, D_2, \ldots, D_J$). Each predictor is multiplied by a slope parameter ($\beta_1$ and $\alpha_1, \alpha_2, \ldots, \alpha_J$), which we estimate and interpret. First, examine the left side of the model, where the outcome—the logit-hazard profile—is found. Subscript $j$ indexes discrete time periods, with $J$ demarking the last period observed. So, $h_1$ is the population hazard probability in time period 1, $h_2$ in time period 2, and so on. Together, the parameters $h_1, h_2, h_3, \ldots h_J$ make up the entire discrete-time population hazard function.

Now examine the terms in parentheses to the right of the equals sign in Equation 1. Notice that the model contains no single stand-alone intercept. Instead, using dichotomous dummy variables $D_1, D_2, \ldots D_J$ to reference the different time periods, the alpha parameters $\alpha_1, \alpha_2, \ldots \alpha_J$ act as multiple intercepts, one per period. In the first time period ($j = 1$), $D_1$ equals 1 and all other time dummies are 0, so the model reduces to this equation.

$$\text{logit}_e(h_1) = \alpha_1 + \beta_1 FEMALE \quad (2)$$

In other words, $\alpha_1$ is an intercept for time period 1. In the second time period ($j = 2$), $D_2$ equals 1, all other time dummies are 0, and $\alpha_2$ becomes the intercept. Similarly, $\alpha_3$ is the intercept in the third period, and so forth.

$$\text{logit}_e(h_2) = \alpha_2 + \beta_1 FEMALE$$
$$\text{logit}_e(h_3) = \alpha_3 + \beta_1 FEMALE \ldots$$
$$\text{logit}_e(h_J) = \alpha_J + \beta_1 FEMALE \quad (3)$$

Because of the way the time dummies are defined, the multiple intercept parameters represent the baseline hazard function, on a logistic scale. When *FEMALE* is set to zero (representing men) in these equations, the $\alpha_1, \alpha_2, \ldots \alpha_J$ parameters represent the period-by-period population logit-hazard probabilities for men and therefore provide the baseline logit-hazard function, because men are the baseline group. In this example, $\alpha_1$ is the log-odds that a man will first become depressed in time period 1, $\alpha_2$ is the log-odds that a man will first become depressed in time period 2 (given that he "survived" time period 1), and so on.

The final term at the right of the parentheses in Equation 1 represents the influence of respondent gender on the population logit-hazard function. We can illustrate its function by reexamining Equations 2 and 3 and evaluating the impact of changing *FEMALE* from value 0 (representing men) to value 1 (representing women) in every time period. The presence of the additional slope parameter permits an identical vertical shift of magnitude $\beta_1$ in logit-hazard in every time period that *FEMALE* equals 1. It is as though the logit-hazard function for men (denoted by $\alpha_1, \alpha_2, \ldots \alpha_J$ alone) has been shifted upward in every period by $\beta_1$, yielding the logit-hazard profile for women. $\beta_1$ is the difference in population logit-hazard per unit difference in *FEMALE*. If the model were fit to the data summarized in Figure 3, the estimated value of $\beta_1$ would be positive because women are generally at greater risk of the onset of depression than men throughout their lives.

Like a regular regression model, the hazard model in Equation 1 can include multiple predictors. To examine the effect of ethnicity on hazard while controlling for gender, for example, we could include simultaneously the predictors *FEMALE* and *HISPANIC* (coded 0 = non-Hispanic, 1 = Hispanic). Other regressionlike extensions are also possible. The model can incorporate both dichotomous and continuous predictors and, finally, the synergistic effect of several variables can be examined by also including interactions among them as predictors.

When studying events that unfold over time, one may find that predictor values also fluctuate over time. In a study of the first onset of depression, for instance, an individual might undergo therapy in some time periods but not in others, causing the value of a *THERAPY* variable to fluctuate across periods. Such time-varying predictors can also be included in discrete-time hazard models:

| ID | $D_1$ | $D_2$ | $D_3$ | $D_4$ | $D_5$ | $D_6$ | $D_7$ | $D_8$ | $D_9$ | $D_{10}$ | $D_{11}$ | $D_{12}$ | ROUTE | MOOD | RELAPSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 01 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 23 | 0 |
| 01 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 18 | 0 |
| 01 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13 | 1 |
| 02 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 27 | 0 |
| 02 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 30 | 0 |
| 02 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 35 | 0 |
| 02 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 29 | 0 |
| 02 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 36 | 0 |
| 02 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 32 | 0 |
| 02 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 27 | 0 |
| 02 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 22 | 0 |
| 02 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 28 | 0 |
| 02 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 30 | 0 |
| 02 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 24 | 0 |
| 02 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 26 | 0 |
| 03 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 38 | 0 |
| 03 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 48 | 0 |
| 03 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 48 | 0 |
| 03 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 48 | 0 |
| 03 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 44 | 0 |
| 03 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 47 | 0 |
| 03 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 48 | 0 |
| 03 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 43 | 0 |
| 03 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 42 | 0 |
| 03 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 44 | 0 |
| 03 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 46 | 0 |
| 03 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 14 | 1 |

*Figure 3.* The person-period dataset.

$$\text{logit}(h_j) = [\alpha_1 D_1 + \alpha_2 D_2 + \cdots + \alpha_J D_J] + \beta_3 THERAPY_j \quad (4)$$

To understand the action of a time-varying predictor, one must break down the model again into period-by-period components.

$$\begin{aligned} \text{logit}(h_1) &= \alpha_1 + \beta_3 THERAPY_1 \\ \text{logit}(h_2) &= \alpha_2 + \beta_3 THERAPY_2 \\ \text{logit}(h_j) &= \alpha_j + \beta_3 THERAPY_j \end{aligned} \quad (5)$$

The value of logit-hazard in each period again has two parts: a baseline part (the alphas) and a *THERAPY* shift of magnitude $\beta_3$. But *THERAPY* is time-varying. If the patient is in therapy in time period 1, then $THERAPY_1$ will equal 1; if not, then $THERAPY_1$ will equal 0. If the patient is in therapy in time period 2, then $THERAPY_2$ will be 1, otherwise 0, and so on for the other time periods. Notice how this affects the logit-hazard profile in Equation 5. In each time period that the patient is in therapy, the value of *THERAPY* is 1, and the effect of therapy "kicks in" for this period only, shifting the logit-hazard function vertically by $\beta_3$. This differs from the action of *GENDER* in Equation 3. Time-invariant *GENDER* retains its value across time, shifting the logit-hazard function by $\beta_1$ unilaterally in all periods. The effect of time-varying *THERAPY*, on the other hand, mirrors the scheduling history of patients as they pass in and out of therapy by shifting the logit-hazard function only in those periods that the patient is in treatment.

Of course, all logit-hazard models can be reexpressed in terms of raw hazard. Simple algebraic manipulation allows the hazard probability—rather than logit-hazard—to become the outcome. For instance, Equation 1 is identical to this equation.

$$h_j = \frac{1}{1 + e^{-[(\alpha_1 D_1 + \alpha_2 D_2 + \cdots + \alpha_J D_J) + \beta_1 FEMALE]}} \quad (6)$$

resembling a logistic regression model with a modified intercept term.

## Fitting Discrete-Time Hazard Models to Data

Standard logistic regression software can be used to fit discrete-time hazard models to data. Before processing, however, all data must be rearranged in a "person-period" format. We show this in Figure 3, which presents data on three cases from a study of relapse to cocaine abuse among 104 former addicts after treatment (Havassy, Hall, & Wasserman, 1991). Participants were observed weekly for 12 weeks or until relapse. Here, we focus on two predictors: (a) time-invariant *ROUTE* indicates the primary mode of cocaine ingestion before treatment (1 = intranasal, 0 = all other routes), and (b) time-varying *MOOD* indicates the former addict's weekly score on a positive mood scale. We ask whether and when relapse is likely to occur, and if the risk of relapse is related to the predictors.

In the person-period data set, each person has as many records as data collection periods. Case 01 was followed for 3 weeks and has 3 records, cases 02 and 03 were followed for 12 weeks, and each has records. Each record contains information on three types of variable: (a) the time indicators, (b) the predictors, and (c) the event indicator. There are as many time indicators as time periods under study; here 12—$D_1$, $D_2$, . . . $D_{12}$. The time indicators denote the particular time period that the record references: $D_1 = 1$ in the record for the 1st time period, $D_2 = 1$ in the 2nd time period, and so on. All other values are

set to 0. The predictors take on values corresponding to the time period referenced by the record. The values of time-invariant predictors such as *ROUTE* are constant across records within person; case 03, an intranasal user, has *ROUTE* coded 1 in all 12 of her records. The values of time-varying *MOOD* may differ across a person's records; in the three records for case 01, values of *MOOD* were 23, 18, and 13, for weeks 1, 2, and 3, respectively. The event indicator, *RELAPSE*, records whether the event of interest occurred in the time period referenced by the record (0 = no event, 1 = event). Because we are treating relapse as a single, nonrepeatable event, *RELAPSE* can equal 1 only in the last time period recorded for each person and only then if the event occurred in that period. If censoring occurs, *RELAPSE* remains at 0. For the first 11 records of cases 02 and 03, *RELAPSE* equals 0. Case 02 did not relapse during data collection (he was censored) and so *RELAPSE* also equals 0 in week 12. Case 03, in contrast, was not censored and relapsed in week 12; so in this record *RELAPSE* equals 1.

Discrete-time hazard models are fitted to data using standard logistic regression software by regressing the dichotomous event indicator *RELAPSE* on all the time dummies and selected predictors in the person-period data set. In a companion article (Singer & Willett, 1993), we show that this yields maximum-likelihood estimates of model parameters and that the obtained standard errors and goodness-of-fit statistics are appropriate for testing our hypotheses. We also provide computer code (SAS, Version 6) for creating the person-period data set, for fitting discrete-time hazard models, and for reconstructing prototypical hazard and survivor profiles from the fitted models.

## Interpreting Fitted Discrete-Time Hazard Models

Table 2 presents three discrete-time hazard models fitted to the cocaine relapse data. We used PROC LOGISTIC in SAS to regress *RELAPSE* on combinations of predictors—$D_1$ through $D_{12}$, *ROUTE*, and *MOOD*—in the person-period data set. Other software, as in SPSS-X and BMDP, provides equivalent output.

### *An Initial Hazard Model*

Model A in Table 2, the simplest possible discrete-time hazard model, is fitted as an analytic first step. It contains no substantive predictors (e.g., *ROUTE* or *MOOD*) to distinguish among sample members and so it describes the behavior of the entire sample assuming a homogeneous population. The model contains 12 intercept terms, one per time period under study, $\alpha_1$, $\alpha_2, \ldots \alpha_{12}$.

$$\text{logit}[h_j] = [\alpha_1 D_1 + \alpha_2 D_2 + \cdots + \alpha_{12} D_{12}] \qquad (7)$$

We bracket the 12 intercept terms because, together, they describe the overall temporal profile of risk and therefore represent the effect of a single conceptual predictor—the main effect of *TIME*. By referring to them as the main effect of *TIME*, we highlight an apparent paradox: *TIME*, the conceptual outcome in our questions, is the fundamental predictor in our analyses. This seeming anomaly arises because we have reformulated the question "When does the event occur?" as "What is the risk of event occurrence in each time period?" However, we can still answer both questions, as we show later.

Table 2
*Parameter Estimates, Standard Errors, and Goodness-of-Fit Statistics for Three Hazard Models Fitted to the Cocaine Relapse Data*

| | Parameter Estimates | | | | | |
|---|---|---|---|---|---|---|
| | Model A | | Model B | | Model C | |
| Predictor | *M* | *SE* | *M* | *SE* | *M* | *SE* |
| $D_1$ | −1.86 | 0.29 | −1.59 | 0.30 | −2.02 | 0.30 |
| $D_2$ | −2.08 | 0.34 | −1.79 | 0.34 | −2.31 | 0.35 |
| $D_3$ | −2.15 | 0.37 | −1.83 | 0.38 | −2.45 | 0.40 |
| $D_4$ | −2.58 | 0.46 | −2.26 | 0.47 | −2.92 | 0.49 |
| $D_5$ | −2.69 | 0.52 | −2.37 | 0.52 | −2.99 | 0.54 |
| $D_6$ | −2.23 | 0.43 | −1.86 | 0.44 | −2.66 | 0.46 |
| $D_7$ | −3.26 | 0.72 | −2.86 | 0.73 | −3.55 | 0.74 |
| $D_8$ | −2.51 | 0.52 | −2.10 | 0.53 | −2.96 | 0.54 |
| $D_9$ | −2.73 | 0.60 | −2.31 | 0.61 | −3.17 | 0.63 |
| $D_{10}$ | −3.83 | 1.01 | −3.38 | 1.02 | −4.10 | 1.03 |
| $D_{11}$ | −3.09 | 0.72 | −2.64 | 0.73 | −3.63 | 0.75 |
| $D_{12}$ | −3.04 | 0.72 | −2.62 | 0.73 | −3.69 | 0.76 |
| INTRANASAL ONLY | — | — | −1.05 | 0.32 | — | — |
| MOOD-30 | — | — | — | — | −0.08 | 0.01 |
| −2LL ($\chi^2$) | 412.17 | | 400.06 | | 381.52 | |

*Note.* Data were kindly supplied by Sharon M. Hall.

We fit the initial hazard model for three reasons. First, it is a benchmark to which more complex models (e.g., Models B and C) can be compared. Second, $\hat{\alpha}_1$, $\hat{\alpha}_2$, through $\hat{\alpha}_{12}$ describe the shape of the overall fitted logit-hazard profile. If risk of event occurrence were unrelated to time, the hazard function would be flat and the $\hat{\alpha}$s approximately equal. If event risks declined over time, values of the $\hat{\alpha}$s for later periods would be smaller than for earlier periods. This pattern is seen in Table 2 in which the Model A parameter estimates suggest that overall risk of relapse declines as time passes. Third, by substitution into Equation 7 and rearranging, the $\hat{\alpha}_1$, $\hat{\alpha}_2, \ldots \hat{\alpha}_{12}$ give maximum-likelihood estimates of the population hazard probabilities in each period. In the first period, $D_1$ equals 1 and $D_2$ through $D_{12}$, equal 0, thus,

$$\hat{h}_1 = \frac{1}{1 + e^{-(-1.8608 \times 1)}} = .1346 \qquad (8)$$

This fitted hazard probability suggests that 13.46% of treated cocaine abusers relapse during the first week posttreatment. In the second week, $D_2 = 1$ ($D_1$ and $D_3$—$D_{12} = 0$) and so an estimate of $h_2$ is $1/(1 + e^{-(-2.0808 \times 1)})$ or .1111; among those abstinent for 1 week, 11.11% relapse the second week. The hazard function in Figure 4 (top panel, left) was estimated like this; risk of relapse is highest right after treatment ends and, over time, successful abstainers become less and less likely to relapse.

A maximum-likelihood estimate of the survivor function can be created by the method of cumulation used in Table 1. Immediately after treatment, all former addicts are abstinent and the initial survivor probability, $S_0$, equals 1 by definition. Subsequent estimates are formed by multiplying the estimated survivor probability in the previous time period by 1 minus the esti-

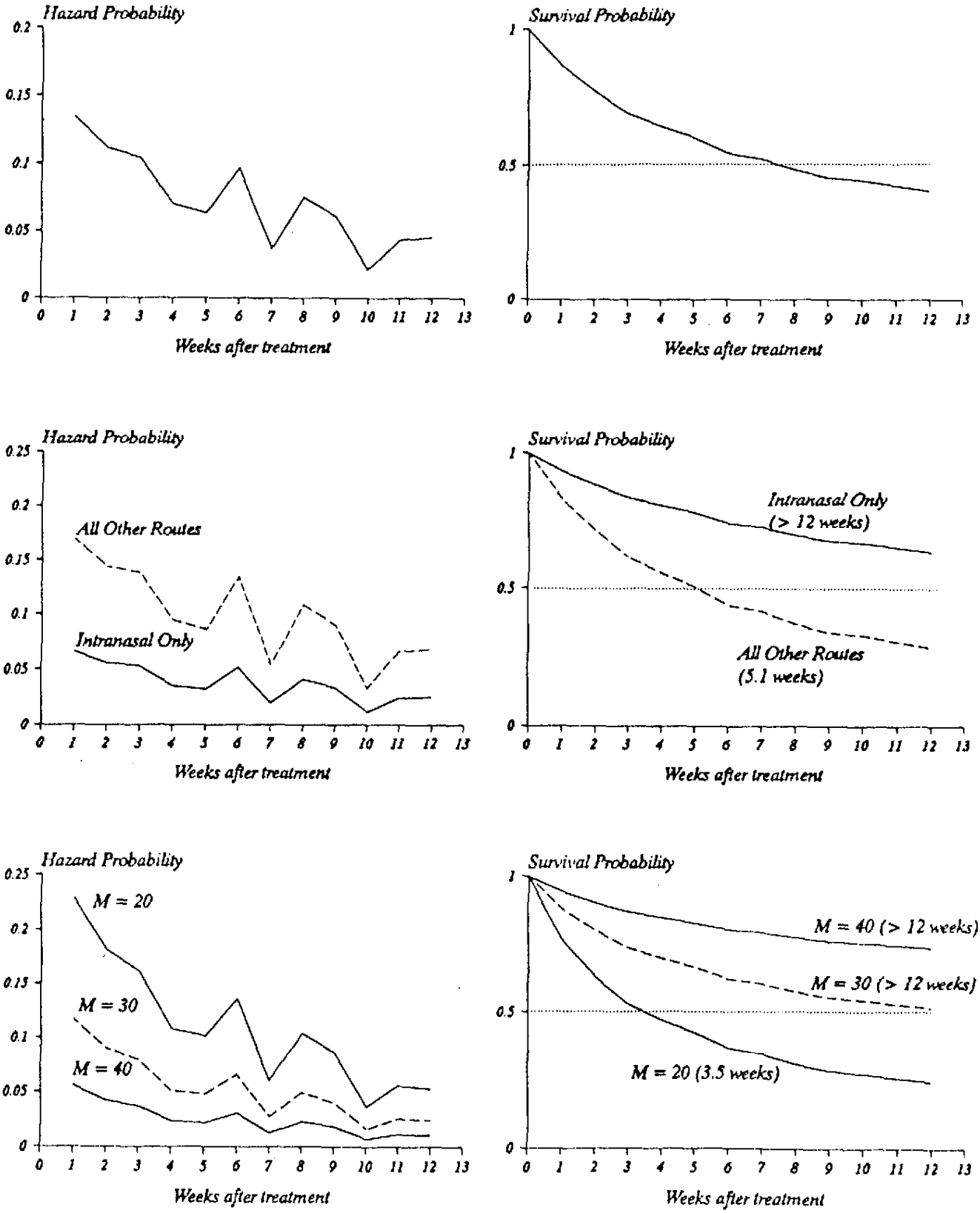


*Figure 4.* Fitted hazard functions (left panels) and survivor functions (right panels) describing the risks of relapse among 104 former cocaine abusers following treatment. Plots derived from hazard models containing the main effect of *TIME* (top panels), the main effect of *TIME* and the main effect of time-invariant *ROUTE* (middle panels), and the main effect of *TIME* and the main effect of time-varying *MOOD* (bottom panels). (Estimated median lifetimes are given in parentheses.)

mated hazard probability in the current period. Thus, 86.54% [= 1 × (1 − .1346)] of treated addicts were abstinent for more than 1 week, 76.93% [= .8654 × (1 − .1111)] were abstinent for more than 2 weeks, and so forth. After 12 weeks, when data collection ended, 40% were still abstinent. Figure 4 (top panel, right) displays the fitted survivor function for Model A.

The median lifetime can be estimated by linearly interpolating between time periods in which the fitted survivor function brackets 50%. In this study, half the former addicts relapsed within 7.6 weeks of treatment. Of course, there may be substantial variation about this average in the sample. In our example, only a further 10% of the sample relapsed by the end of the 12th week, and the remaining 40% may be abstinent to this day. Thus, estimates of the median lifetime should always be supplemented by the sample survivor function.

### *Main Effect of a Time-Invariant Predictor*

Model B adds the main effect of *ROUTE* to the main effect of *TIME*:

$$\text{logit}_i(h_j) = [\alpha_1 D_1 + \alpha_2 D_2 + \cdots + \alpha_{12} D_{12}] + \beta_1 ROUTE \quad (9)$$

When we add a substantive predictor to our model, we evaluate its effect by examining its contribution to the prediction of risk. Model goodness-of-fit is summarized by the "−2 log-likelihood" statistic ("−2LL" in Table 2) which is asymptotically distributed as $\chi^2$. Thus, we can use standard decrement-to-chi-square testing to check whether an extended model fits better than a reduced model. If added predictors are not associated with risk, the extended model will fit no better than the reduced model; if added predictors are associated with risk, then fit will improve. Compare, for example, Models A and B. Model A is nested within Model B because the sole difference between the two is the inclusion of *ROUTE*. When *ROUTE* is added to a model containing only the main effect of *TIME*, the model $\chi^2$ statistic decreases from 412.17 to 400.06, denoting better fit for Model B. Because one predictor was added, this decrement of 12.11 comes with a loss of 1 degree of freedom. We thus reject the null hypothesis that the population hazard profile is unaffected by *ROUTE* ($p < .001$).

How large is this effect? The parameter estimate ($\hat{\beta}_1 = -1.05$) can be interpreted like similar estimates from regular logistic regression models (Hosmer & Lemeshow, 1989). Because *ROUTE* is dichotomous (1 = former intranasal users, 0 = others), its sign (negative) tells us that intranasal users are less likely to relapse, in every week after treatment. Its magnitude indicates that the vertical separation of fitted logit hazard functions for the two groups is 1.05. Because a logit scale may be unfamiliar, parameter estimates can be antilogged, becoming odds-ratios. Antilogging $\hat{\beta}_1$ in Model B ($e^{-1.05} = 0.35$) says that the estimated odds of relapse for former intranasal users are about one third the odds of relapse for former addicts who administered cocaine by other means.

We can also describe a predictor's effect by displaying fitted hazard and survivor functions computed at interesting values of the predictor. The second panel of Figure 4 presents fitted hazard functions (left) and survivor functions (right) for both values of *ROUTE*. As with Model A, we obtained fitted hazard functions by substituting predictor values and parameter estimates into Equation 9 and fitted survivor functions by the method of cumulation. Median lifetimes were estimated by linear interpolation (censoring prevents us from estimating a median lifetime precisely for intranasal users—we know only that it exceeds 12 weeks). Graphics like these are also helpful when studying the effects of continuous predictors, because fitted hazard and survivor functions can illustrate risk differences that are due to substantively interesting differences in predictor value.

Each summary provides a different perspective on the link between risk of relapse and *ROUTE*. We advise using the one that best describes the substantive finding to the chosen audience. Fitted hazard functions document the large weekly differentials in risk between the groups (1 to 3). Fitted survivor functions pool these weekly risks into vastly different cumulative relapse rates (after 12 weeks, for example, only 28% of the intranasal users have relapsed versus 63% for all others). Estimated median lifetimes convey the risk differential between the groups in the original metric of time (5.1 weeks versus 12 weeks or more).

### *Main Effect of a Time-Varying Predictor*

Similar strategies can be used to examine the effects of time-varying predictors. Model C (Table 2) adds the main effect of time-varying *MOOD* to an initial model containing the main effect of *TIME* (*MOOD* has been centered at 30, the approximate sample mean, for interpretive ease).

$$\text{logit}_i(h_j) = [\alpha_1 D_1 + \alpha_2 D_2 + \cdots + \alpha_{12} D_{12}] + \beta_2(MOOD_j - 30) \quad (10)$$

*MOOD* has been subscripted by *j* to indicate that it is time-varying.

The model in Equation 10 permits many possible combinations of mood levels over time to lead to a different risk profile because risk of relapse in each week depends on a person's mood that week. The protective benefit of a high score accrues in high-scoring weeks, the liability of a low score takes its toll in low-scoring weeks. The model also specifies that, although the values of *MOOD* may fluctuate over time, its effect on logit-hazard is the same in all time periods. A 1-point difference on the positive mood scale always "kicks out" the same amount of risk whenever mood changes. Because the effect of *MOOD* does not vary over time, we say that the time-varying predictor (*MOOD*) has a time-invariant effect.

As before, we evaluate the impact of *MOOD* by comparing Models A and C. Adding *MOOD* reduces the model $\chi^2$ statistic by 30.63 for a loss of 1 degree of freedom, indicating an improvement in fit ($p < .0001$). We reject $H_0: \beta_2 = 0$ and regard *MOOD* as a statistically significant predictor of risk. How large is this time-varying effect? In Table 2, $\beta_2$ is estimated as −.0806. The negative sign indicates that more positive moods are associated with lower risk of relapse. Antilogging to form an odds ratio, $e^{-.0806} = 0.92$, we find that, when two former addicts who are 1 point apart on the mood scale are compared, the estimated odds of relapse for the more positive individual are .92 times the odds of relapse for his colleague.

When evaluating the effect of a continuous predictor such as *MOOD*, it can be hard to interpret odds ratios because differences of 1 point on the continuous scale may not correspond

to large differences in risk. We recommend that the effects of continuous predictors be displayed by plotting fitted hazard and survivor functions computed at substantively interesting values of the predictors. But what values should be chosen? We usually use well-spaced values suggested by the sample distribution of the predictor (unless there are other values widely accepted as meaningful). In this case, we selected three *MOOD* values—20, 30, and 40—corresponding roughly to the 10th, 50th, and 90th percentiles of the sample distribution of *MOOD*. Figure 4 (bottom panel, left) presents the fitted hazard functions for former addicts who had consistently low scores (20), average scores (30), and high scores (40). The fitted hazard functions have the same basic shape. In every week, those who report more positive moods are less likely to relapse than those who report fewer. The two extreme profiles envelop the hazard profiles for treated cocaine users in various mood combinations over time, ranging from the 10th percentile (with a score of 20) to the 90th percentile (with a score of 40). These three fitted hazard functions may not represent specific people; rather, they may circumscribe a region within which most of the possible hazard profiles fall. The fitted hazard profile for addicts who first report many positive moods and then report few, for example, would first follow the bottom function and then move up as positive mood declined.

Fitted survivor functions and estimated median lifetimes (Figure 4, bottom panel, right) summarize these risk differentials. The top survivor function describes what the fitted model predicts would happen to people whose mood levels were consistently high; 12 weeks after treatment, for example, an estimated 74% would still be abstinent. The middle survivor function represents our prediction for people with moderate mood levels; 12 weeks after treatment ends, an estimated 51% would still be abstinent. Those who consistently report few positive moods (20) have the grimmest prospects—their estimated median time to relapse is 3.5 weeks! Of course, these are predictions for prototypical former addicts whose *MOOD* values remain constant over time. A different temporal pattern of mood would generate different summary statistics.

## The Proportionality Assumption

Most standard hazard models invoke a *proportionality* assumption, as does the model in Equation 9. It postulates that logit-hazard profiles represented by both values of *ROUTE* have a common shape and are mutually parallel, being shifted only vertically for different values of *ROUTE*. Because discrete-time hazard is a (conditional) probability, the proposed logit-hazard profiles describe the (conditional) log-odds of event occurrence in each time period. If logit-hazard profiles are antilogged and redisplayed, profiles describing the conditional odds of event occurrence at different values of the predictors are produced. By the very nature of the logarithmic transformation, these odds profiles are now magnifications or diminutions of one another—they are proportional. However, in practice, many predictors affect both the vertical displacement and shape of the logit-hazard profile. Then, the proportionality assumption will not hold, and a model such as Equation 9 would be inappropriate.

Violations of proportionality are not simply methodological nuisances—they can lead to substantively meaningful findings. They indicate that a predictor's effect fluctuates over time. Some predictors may primarily affect early hazards, causing hazard profiles to be widely separated initially and to converge over time. Other predictors will primarily affect late hazards, causing initially coincident hazard profiles to diverge over time. Because the effects of many predictors of human behavior differ by developmental stage, we suspect that, as hazard models become more popular in psychological research, violations of the proportionality assumption will cease to be the exception and become the rule.

Researchers must become proactive—assume that nonproportionality exists until proven wrong. Preliminary graphical analysis is a good first step. If logit-hazard profiles estimated separately within strata are approximately parallel, the assumption is probably met; if they are not, it is violated. Visual examination, however, is more art than science. Differences in the shapes of men and women's hazard profiles in Figure 2, for example, may be due to sampling variation. We can tell by augmenting visual analysis with tests for proportionality violations based on the inclusion of interactions with time in the hazard model.

## Interactions With Time

In discrete-time survival analysis, the proportionality assumption can be tested by including statistical interactions between predictors and *TIME* in the hazard model. These interactions permit the effect of the predictors (on hazard) to differ by time period. This allows the predictor-attributable vertical displacement in logit-hazard to differ across time periods, ensuring that the shapes of the logit-hazard functions will differ at different values of the predictor.

We can test for an interaction between a predictor and *TIME* by (a) forming cross-products between the time indicators and the selected predictor in the person-period data set; (b) including all these cross-products, along with the relevant main effects, in the hazard model; and (c) testing their contribution to model fit. We illustrate this by returning to our data on age at first onset of suicide ideation (Bolger et al., 1989) to examine whether there is an interaction between the time-invariant predictor *WHITE* (0 = non-White, 1 = White) and risk of suicide ideation. The same procedure is suitable for testing the interaction between a time-varying predictor and time.

Figure 5 presents the results of fitting a "main effect of time" and an "interaction with time" model to these data. The main effects model (top) suggests that the odds that a White college student had considered suicide are 22% higher than for a non-White student (however, this effect is not statistically significant at conventional levels, $p = .2961$). A different story emerges from the interaction with time model (bottom). The interaction could have been specified using many cross-product terms, one per time dummy, but a consideration of developmental stages led Bolger et al. to suspect that it acted more simply, with an impact that varied only across two broad time periods: preadolescence (6 to 12 years) and adolescence (13 and above). Therefore, we created the predictor *ADOLESCENT* (0 = preadolescence, 1 = adolescence) to distinguish the two time periods and our interactive hazard model is as follows:

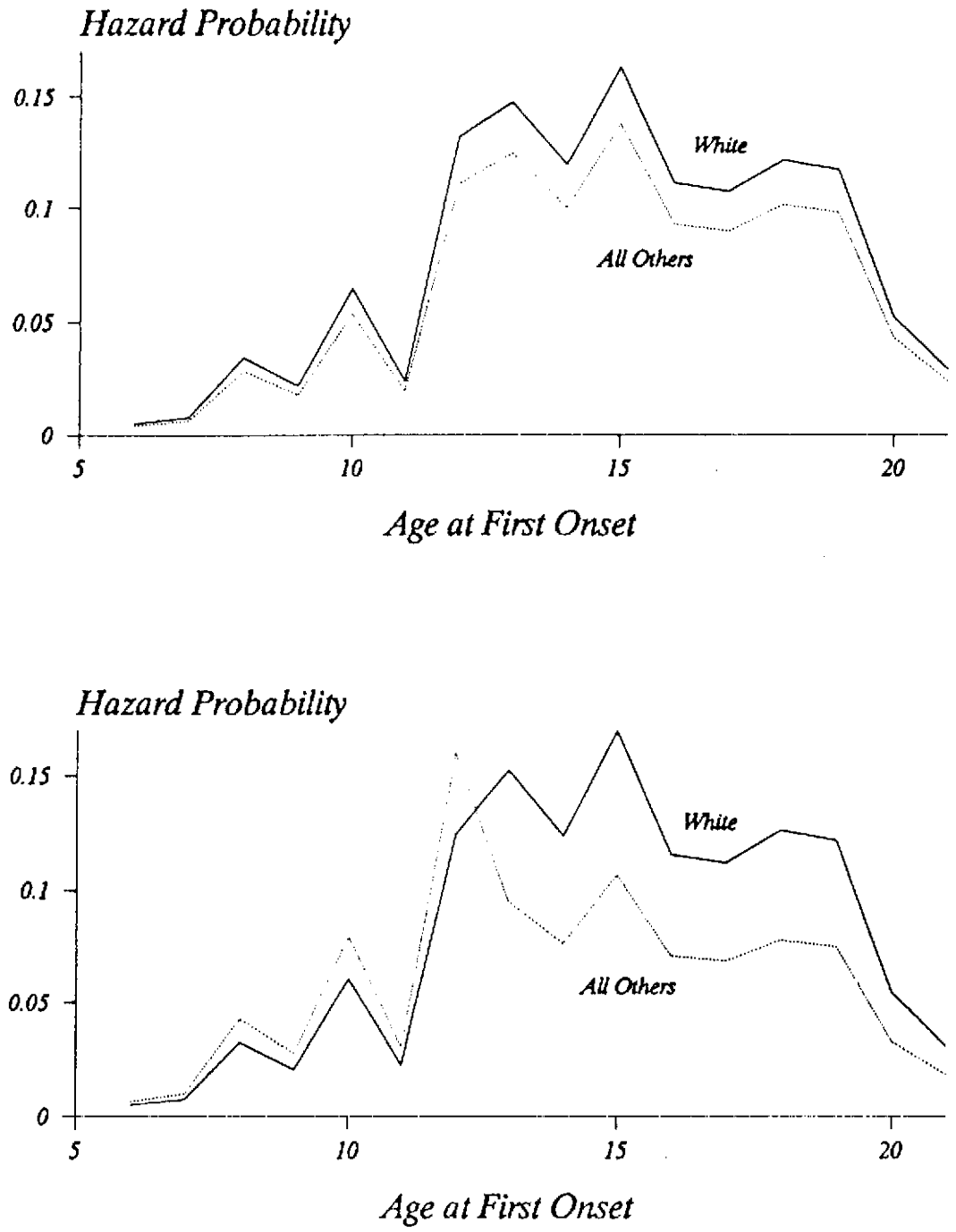


*Figure 5.* Fitted hazard functions describing the risks of onset of first suicide ideation by age (in years) for 417 college-bound adolescents, by ethnicity, from hazard models containing the main effect of *TIME* and the main effect of ethnicity (top panel) and the main effect of *TIME*, the main effect of ethnicity, and the two-way interaction between *TIME* and ethnicity (bottom panel). (Estimated median lifetimes in parentheses.)

$$\text{logit}_i(h_j) = [\alpha_6 D_6 + \cdots + \alpha_{21} D_{21}] + \beta_1 WHITE + \beta_2(ADOLESCENT \times WHITE) \quad (11)$$

The interaction allows the effect of *WHITE* to differ by developmental stage.

Comparing Equation 11 with a main effects model reveals that the inclusion of the interaction with time improves model fit; the model chi-square statistic declines by 4.64 for a loss of 1 degree of freedom. We reject the null hypothesis that the addition of the *WHITE* × *ADOLESCENT* interaction is of no conse-

quence ($p < .05$), the proportionality assumption is violated, and the interactive model is preferred. Figure 5 (bottom panel) presents fitted hazard functions for the interactive model. Comparing top and bottom panels highlights the untenability of the proportionality assumption and the degree to which a researcher can be misled by its unthinking acceptance. In the top panel, the two hazard functions have similar shapes, with the hazard function for Whites consistently higher than that for non-Whites. In the bottom panel, the hazard functions cross. In preadolescence, Whites are *less* likely than non-Whites to have considered suicide but in adolescence, they are *more* likely. Interestingly, because of the close coincidence in hazard in preadolescence, median lifetimes estimated from the interaction with time model are virtually identical for White children (14.6 years) and non-White children (15.4 years).

## Is Discrete-Time Survival Analysis for You?

There are five ways in which traditional analyses obscure important information about event occurrence, information that is assuredly revealed by survival analysis. First, traditional statistical summaries are inextricably linked to the particular time frame chosen for data collection and analysis, yet these frames are rarely motivated by substance. Researchers comparing 6-month, 1-year, or 5-year relapse rates, for example, are just describing cumulative differences in behavior until these times. All other variation over time in risk of relapse is lost. Just because two groups have identical relapse rates at one time point does not imply that they followed identical paths to get there—most in one group might have relapsed in the first month, whereas those in the other might have been equally likely to relapse at all points in time. Most cutpoints are convenient, not purposeful. By documenting variation in risk over time and by discovering what predicts it, we can better understand when events occur.

Disregard for variation in risk over time leads to a second problem with traditional methods; contradictory conclusions can result from variation in the particular time frames studied. Had Coryell, Keller, Lavori, and Endicott (1990) computed only 40-week, 80-week, or 200-week recovery rates when comparing patients with psychotic major depressive disorder and patients with schizoaffective depressive disorder, they would have reached three different conclusions: the 40-week rates would suggest that patients with schizoaffective depressive disorder were more likely to recover, the 80-week rates would have shown no difference, and the 200-week rates would suggest that patients with psychotic major depressive disorder were more likely to recover. Fortunately, Coryell et al. used survivor functions to disentangle the effects. Users of traditional methods must constantly remind themselves that their conclusions may change as the time frame changes. In survival analysis, the time frame is an integral part of the answer; it highlights, not obscures, variation over time.

Third, traditional analytic methods offer no systematic mechanism for dealing with censoring. If all censoring occurs at the same point in time, traditional data analysis can collapse the sampled individuals into two groups: those who experienced the event before the censoring point and those who did not. Using a sample of adults with major depressive disorder who had been treated with one of three different modes of psychotherapy, Gallagher-Thompson, Hanley-Peterson, and Thompson (1990), for example, compared patients who relapsed and patients who did not at 1 (and 2) years after cessation of treatment. But if the first weeks following cessation are the hardest, individuals who relapse soon after cessation may differ systematically from those who relapse subsequently. Dichotomization conceals such differences; survival methods bring such differences to light.

If censoring does not occur at the same time for everyone under study (as when researchers follow cohorts of patients admitted over time until a single fixed time point), traditional methods create a fourth problem: If censoring times vary across people, the risk periods vary as well. People followed for longer periods of time have more opportunity to experience the target event than do those followed for shorter periods of time. Observed differences in rates of event occurrence may be attributable to nothing more than research design. In Goldstein, Black, Nasrallah, and Winokur's (1991) study of suicidality among Iowans with affective disorders, the follow-up period ranged from 2 to 13 years. As they noted, "the highly variable period of follow-up is also a potential limitation, because those patients followed up for the shortest periods may not have been given the opportunity for their suicidal outcome to emerge" (1991, p. 421). Had they used survival methods, they could have addressed this concern because each person who did not commit suicide would have been censored at follow-up.

Fifth, traditional analytic methods have difficulty representing the effects of time-varying predictors or with permitting the effects of predictors to fluctuate over time. Researchers studying the impact of time-varying predictors may ignore information either by using the value of the predictor at a single time point or by pooling predictor values over time, perhaps as an average or a rate of change. Survival methods obviate this ad hoc approach. The effort is identical whether including time-invariant or time-varying predictors; so, too, it is easy to determine whether the effects of predictors are constant over time or whether they differ over time. Traditional methods build static models of dynamic processes; survival methods model dynamic processes dynamically.

We recommend the use of survival methods. When the methods were new and statistical software was unavailable, researchers reasonably adopted other approaches. Recent experience shows, however, that these methods lend themselves naturally to the study of social phenomena. Researchers who may have modified their questions about event occurrence and timing because of the lack of a suitable methodology in the past can now address them in their original form.

## References

Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data*. Beverly Hills, CA: Sage.

Bolger, N., Downey, G., Walker, E., & Steininger, P. (1989). The onset of suicide ideation in childhood and adolescence. *Journal of Youth and Adolescence, 18*, 175–189.

Cooney, N. L., Kadden, R. M., Litt, M. D., & Getter, H. (1991). Matching alcoholics to coping skills or interactional therapies: Two-year follow-up results. *Journal of Consulting and Clinical Psychology, 59*, 598–601.

Coryell, W., Keller, M., Lavori, P., & Endicott, J. (1990). Affective syndromes, psychotic features, and prognosis. *Archives of General Psychiatry, 47,* 651–662.

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

Gallagher-Thompson, D., Hanley-Peterson, P., & Thompson, L. W. (1990). Maintenance of gains versus relapse following brief psychotherapy for depression. *Journal of Consulting and Clinical Psychology, 58,* 371–374.

Glasgow, R. E., Klesges, R. C., Klesges, L. M., & Somes, G. R. (1988). Variables associated with participation and outcome in a worksite smoking control program. *Journal of Consulting and Clinical Psychology, 56,* 617–620.

Goldstein, R. B., Black, D. W., Nasrallah, A., & Winokur, G. (1991). The prediction of suicide: Sensitivity, specificity, and predictive value of a multivariate model applied to suicide among 1,906 patients with affective disorders. *Archives of General Psychiatry, 48,* 418–422.

Harrell, F., & Goldstein, R. (in press). A survey of microcomputer survival analysis software: The need for an integrated framework. *The American Statistician.*

Hart, S. D., Kropp, P. R., & Hare, R. D. (1988). Performance of male psychopaths following conditional release from prison. *Journal of Consulting and Clinical Psychology, 56,* 227–232.

Havassy, B. E., Hall, S. M., & Wasserman, D. A. (1992). Social support and relapse: Commonalities among alcoholics, opiate users, and cigarette smokers. *Addictive Behaviors, 16,* 235–246.

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

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

Kelly, J. A., St. Lawrence, J. S., & Brasfield, T. L. (1991). Predictors of vulnerability to AIDS risk behavior relapse. *Journal of Clinical and Consulting Psychology, 59,* 163–166.

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,* 268–290.

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

Sorenson, S. B., Rutter, C. M., & Aneshensel, C. S. (1991). Depression in the community: An investigation into age of onset. *Journal of Consulting and Clinical Psychology, 59,* 541–546.

Stevens, V. J., & Hollis, J. F. (1989). Preventing smoking relapse, using an individually tailored skills-training program. *Journal of Consulting and Clinical Psychology, 57,* 420–424.

Weisz, J. R., Walter, B. R., Weiss, B., Fernandez, G. A., & Mikow, V. A. (1990). Arrests among emotionally disturbed violent and assaultive individuals following minimal versus lengthy intervention through North Carolina's Willie M Program. *Journal of Consulting and Clinical Psychology, 58,* 720–728.

Willett, J. B., & Singer, J. D. (in press) It's deja-vu all over again: Using multiple-spell discrete-time survival analysis. *Journal of Educational Statistics.*

Received May 15, 1992
Revision received January 29, 1993
Accepted May 12, 1993 ■