Survival Methods: Additional Topics

David Oakes and Derick R. Peterson

The online version of this article, along with updated information and services, is located on the World Wide Web at:

http://circ.ahajournals.org/cgi/content/full/117/22/2949

Survival Methods
Additional Topics

David Oakes, PhD; Derick R. Peterson, PhD

Survival analysis concerns outcome variables that are the times to the occurrence of events such as myocardial infarction, hospitalization for heart failure, or death. Rao and Schoenfeld[1] give an excellent account of frequently used techniques in survival analysis, including Kaplan–Meier estimation of a single survival distribution, the log-rank test for comparing 2 survival distributions, and Cox's proportional-hazards model (PHM) for assessing the affect of multiple predictors on survival.

Our article delves more deeply into the methodology, emphasizing the use of time-dependent covariates both as devices for assessing the fit of a model and as risk predictors in their own right. We review the accelerated life model as an important alternative to the PHM. We address the analysis of event-time data when individuals may experience multiple events and the use of combined end points. We describe the use of cumulative incidence curves as an alternative to Kaplan–Meier estimation in situations involving competing risks. Finally, we mention some issues related to the analysis of quality of life and cost-effectiveness data in survival studies.

Review of Cox's PHM

Recall that the hazard function h(t) for an event at time t is the instantaneous event rate among t subjects who have not yet experienced the event. It is related to the survivor function (probability of not yet having experienced the event), S(t), by the expression $h(t) = -[1/S(t)]dS(t)/dt$ (Rao and Schoenfeld[1]). Cox's model states that the hazard function for an individual with covariates $x_1, x_2, \ldots, x_3$ takes the form

$$(1) \qquad h(t; x_1, x_2, \ldots, x_k) = h_0(t)\exp(\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k)$$

where $h_0(t)$ is the so-called baseline (or reference) hazard and the x_j are covariates. Model 1 implies that the ratio of the hazard functions for 2 individuals is constant over time because the term $h_0(t)$ cancels from the ratio and the other terms are free of t. When, as often happens, a covariate x is an indicator variable, say, in a clinical trial, and $x=0$ or $x=1$ for patients randomized to placebo or active therapy, respectively, then the quantity $\exp(\lambda)$ represents the hazard ratio for patients assigned to active therapy compared with patients assigned to placebo. When $k>1$, so that the model includes

≥ 2 covariates, the effect of each is interpreted as if all the others are "held constant." Software for fitting Cox's model provides estimates $\hat{\lambda}_j$ of the coefficients λ_j for each covariate in model 1 and their SEs, SE $(\hat{\lambda})$. The corresponding estimated hazard ratio is exp $(\hat{\lambda})$, and its 95% CI is $(\exp\{\hat{\lambda} - 1.96SE(\hat{\lambda})\}, \exp\{\hat{\lambda} + 1.96SE(\hat{\lambda})\})$. To be precise, such a CI, calculated from an estimate and its SE, is called a Wald interval; other approaches to the calculation are sometimes preferable, but the Wald approach is usually adequate. When x is numeric (eg, blood pressure in mm Hg or body mass index in kg/m^2), then λ and its associated quantities are interpretable analogously to regression coefficients. For example, if λ represents the coefficient of blood pressure (measured in mm Hg), then $\exp(\lambda)$ represents the hazard ratio associated with a 1-mm Hg increase in blood pressure.

Note that, depending on the units of x, a hazard ratio close to unity can represent a very substantial effect; as an extreme example, a hazard ratio of 1.001 for the covariate age expressed in days corresponds to a hazard ratio of $1.001^{365} = 1.44$ per year of age or $1.44^{10} = 38.3$ per decade of age. Likelihood ratio statistics with approximate χ^2 distributions can be used to compare 2 different models, provided that 1 is nested within the other. This happens, for example, if the covariates in a "smaller" model are a subset of the covariates in a "larger" model. As an illustration, the statistic for comparing a model for predicting mortality with age and gender as covariates with a model including only age answers the question of whether gender is associated with mortality, controlling for (or adjusting for) age. The related, but different, question of whether the relationship between mortality and gender depends on age can be addressed by including an additional term representing the statistical interaction between age and gender. This idea is illustrated below.

Application to the Multicenter Diltiazem Postinfarction Trial

We illustrate Cox's model with an analysis of the Multicenter Diltiazem Postinfarction Trial (MDPIT)[2] of the effect of diltiazem therapy on the risk of a recurrent cardiac event. In this study, 2466 post–myocardial infarction patients were randomized to placebo or diltiazem therapy. The primary outcome measure was the time from enrollment to the first

Table 1. MDPIT Study: Analysis Using the PHM

Hazard Ratio (PC Versus No PC) for Patients Assigned to Placebo	Hazard Ratio (D Versus P) for Patients Without Pulmonary Congestion	Hazard Ratio (D Versus P) for Patients With Pulmonary Congestion
1.82 (1.36–2.42)	0.80 (0.63–1.01)	1.25 (0.94–1.74)

Values in parentheses are 95% CI.

recurrent cardiac event, defined as cardiac death or nonfatal reinfarction. For patients who did not experience a recurrent cardiac event (the majority), the time to event was censored at the last follow-up visit (see Rao and Schoenfeld[1]). Although diltiazem therapy had no significant overall effect, further analysis revealed a significant interaction between diltiazem therapy and pulmonary congestion on x-ray. Among patients without pulmonary congestion on x-ray, diltiazem therapy reduced the risk of an event; however, among patients with pulmonary congestion, the risk was increased. Although such post hoc subgroup analyses must be interpreted very cautiously, this finding is made more credible by the statistical significance of the test for interaction or inequality of the 2 hazard ratios. The first entry in Table 1 represents the hazard ratio estimate with 95% CI for the effect of pulmonary congestion for patients allocated to placebo, ie, (PC, P) versus (No PC, P). The second and third entries give corresponding estimates for the effect of diltiazem therapy compared with placebo among patients without pulmonary congestion, (No PC, D) versus (No PC, P), and among patients with pulmonary congestion, (PC, D) versus (PC, P). We define 3 binary covariates. Although the reader need not dwell on this, an appropriate coding to obtain these hazard ratios is to set $x_1 = 1$ for patients with PC, $x_1 = 0$ for patients without PC, $x_2 = 1$ for patients in the (No PC, D) group, $x_2 = 0$ for patients in the other 3 groups, $x_3 = 1$ for patients in the (PC, D) group, and $x_3 = 0$ for patients in the other 3 groups.

A slightly different coding of the 3 covariates is more convenient for testing for statistical interaction, ie, that the true hazard ratio for diltiazem relative to placebo differs for patients without and with pulmonary congestion. We obtain $P = 0.03$, suggesting that diltiazem has different effects in the 2 groups.

Because Cox's model estimates ratios of hazards rather than absolute values, we must select 1 category of patient (ie, 1 set of covariate values) as a reference value for which the hazard ratio is set to unity. The software for fitting Cox's model can be used to construct survival curves corresponding to specific sets of values for the covariates. These curves are constructed under the proportional-hazards assumption, so they may differ from Kaplan–Meier estimates based on relevant subsets of the data because the latter curves are not constrained by this assumption. However, they are more variable if the assumption holds.

The prespecified primary analysis of MDPIT included adjustment for 3 further covariates: New York Heart Association classification, use of β-blockers at randomization, and days since index myocardial infarction. Adjustment means that covariates representing these factors were included in model 1, along with the treatment comparisons of primary interest.

The analysis also included stratification on enrolling hospital. This means that different and possibly nonproportional baseline hazard functions, $h_0(t)$, are allowed for each enrolling hospital. This forces the analysis to depend on comparisons of patients within the same stratum (hospital), eliminating effects of systematic differences between strata. When there are many strata (MDPIT had 38), this approach is generally preferable to the use of additional covariates to adjust for strata effects because it yields a more stable estimation procedure and provides greater flexibility in allowing nonproportional-hazards functions in different strata.

When the additional baseline covariates are strongly associated with survival, the adjusted hazard ratios for the treatment effect are generally more precise than the unadjusted estimates. In MDPIT, the hazard ratios and associated 95% CIs for the effect of diltiazem relative to placebo became 0.77 (95% CI, 0.61 to 0.98) for patients without pulmonary congestion and 1.41 (95% CI, 1.01 to 1.96) for patients with pulmonary congestion, with $P = 0.0042$ for the test of interaction. When adjustment covariates are both predictive of the outcome and highly correlated with the variables of interest, as often happens in observational studies, adjusted estimates can differ substantially from unadjusted estimates, not only in precision but also in magnitude and even direction.

The Accelerated Life Model as an Alternative Formulation

For complete (ie, uncensored) data, multiple linear regression is a standard technique for assessing the joint effect of predictors $x_1, x_2, \ldots, x_k$ on an outcome variable y. The response is expressed as a linear combination of covariates plus a random "error" (or deviation). These deviations typically are assumed to have zero mean and constant variance and to be uncorrelated. Because survival times are necessarily positive, it is usual to take their logarithms before fitting models to survival data. Because the logarithm of a product of numbers equals the sum of their logarithms, the parameters λ_j represent multiplicative rather than additive effects on the original time scale. This gives the following model:

$$(2) \qquad \log T = \lambda_0 + \lambda_1 x_1 + \cdots + \lambda_k x_k + \varepsilon$$

This model is called the accelerated life or accelerated failure time (AFT) model. It has some similarities with model 1 in the way that the covariates x_i are combined into a single linear predictor via the coefficients λ_i, but there also are some important differences. In the AFT model, the linear combination of covariates predicts the mean value of the response variable y, whereas in the PHM, it predicts the hazard function. As mentioned by Rao and Schoenfeld,[1] for exponential distributions, which have constant hazard functions, the entire right side of model 1 does not depend on t, and the AFT model (model 2) is equivalent to the PHM. The only

Table 2. MDPIT Study: Analysis Using the Accelerated Life Model

Scale Factor (PC Versus No PC) for Patients Assigned to Placebo	Scale Factor (D Versus P) for Patients Without Pulmonary Congestion	Scale Factor (D Versus P) for Patients With Pulmonary Congestion
0.33 (0.19–0.57)	1.52 (0.97–2.36)	0.65 (0.35–1.20)

Values in parentheses are 95% CI.

difference is in the sign of the coefficients; because increasing hazards correspond to shorter survival times and a lower mean survival, a positive λ in model 1 corresponds to a negative λ in model 2.

This equivalence extends to a wider class of distributions, namely those with a power law form for $h_0(t)$, ie, with $h_0(t)$ proportional to t^{k-1}. This class of distributions, known as the Weibull family, has increasing hazard if $k>1$, has decreasing hazard if $k<1$, and reduces to the exponential family with constant hazard function if $k=1$. Software for fitting AFT models assuming exponential or Weibull distributions, or other parametric families including the log normal and log logistic, to censored survival data is widely available. Various methods for fitting semiparametric AFT models to censored data (ie, models that do not assume a specific form for the distribution of the errors ε) have been developed, but their use is not yet common in applied work.

It typically happens that PHMs and parametric AFT models fitted to the same data set will give estimated coefficients of opposite directions but in similar numerical ratios and with similar levels of relative importance as judged by their corresponding Wald statistics (ratio of the estimated coefficient to its SE). However, this may not be the case if the true model does not have a monotone hazard (So-called "bathtub" hazard functions, which decrease initially, reach a minimum value, and subsequently increase, are an example of this).

We fit Weibull distributions to the data from the MDPIT study and contrast this fit with the fit from the PHM. The AFT model (model 2) asserts that the percentiles of the distributions of survival time corresponding to different values of the covariate are in a fixed ratio; ie, the ratio of the medians of the 2 distributions is the same as that of their 10th, 20th, etc, percentiles. This ratio is called the scale factor. Table 2 presents estimates of the scale factors in the same format as the hazard ratios in Table 1.

Comparison of Tables 1 and 2 shows the expected pattern of effect. Hazard ratios exceeding unity correspond to scale factors less than unity and vice versa. The estimated value of the index parameter k in the Weibull distribution is $\hat{\kappa} = 0.54$. This corresponds to a decreasing hazard function; the risk of an event decreases with longer time from entry.

Testing Proportional Hazards: Application to MDPIT

We now consider possible violations of the PHM. Covariates that are highly predictive of early failure may become less influential as time progresses. For example, a simple way to investigate this possibility is to allow separate hazard ratios for different time periods, thus making the proportional-hazards assumption only within each smaller time interval. For example, in the MDPIT study of the effect of diltiazem on the risk of a recurrent cardiac event, we fit separate coeffi-

cients for the time periods of 0 to 90 days, 90 to 365 days, and >365 days. Specifically, in equation 1 we may replace certain λ_j by λ_{j1}, λ_{j2}, or λ_{j3}, depending on whether $t\leq90$, $90<t<365$, or $t\geq365$. Such a model can be fit with most standard software. The proportional-hazards assumption can then be formally tested via the likelihood ratio test for the null hypothesis H_0: $\lambda_{j1}=\lambda_{j2}=\lambda_{j3}$ for all j. Overall, the fit of the PHM was satisfactory ($\chi^2_6=10.6$, $P=0.1$), although there was some suggestion that both the adverse effects (ie, among patients with pulmonary congestion) and the beneficial effects (among patients without pulmonary congestion) of diltiazem therapy were more evident early than later.

If Weibull distributions of survival time, as presented earlier, are assumed for each of the 4 groups of patients, then the proportional-hazards assumption is satisfied if and only if the 4 index parameters are the same. This null hypothesis can be tested by comparing the fit of a single model to the combined data with that of 4 different Weibull distributions fit to the data from each group. The likelihood ratio statistic is $\chi^2_3=3.52$ ($P=0.32$), indicating little evidence against the null hypothesis.

The Long-QT Syndrome Analysis: An Example of Nonproportional Hazards

We now turn to a recent observational study among long-QT syndrome (LQTS) adolescents 10 to 20 years of age.[3] The primary outcome was the occurrence of aborted cardiac arrest (ACA) or sudden cardiac death (SCD). Subjects were included in the study if they had survived to 10 years of age without suffering an event, and age was used as the primary time variable. Figure 1 shows Kaplan–Meier estimates of survival over the decade. The curves do not cross. Survival for female subjects is always higher than survival for male subjects at each age in this range, but the pattern of the curves suggests that the risk is lower for female subjects over the earlier years but then increases more rapidly than the risk for male subjects. The standard log-rank test applied to these data gives $P=0.31$. The reason for the nonsignificance is not lack of information—the analysis is based on a total of 126 events observed among 2772 adolescents—but violation of the proportional-hazards assumption.

In contrast, were follow-up for this analysis to have been concluded at 13 rather than 20 years of age, the log-rank probability value would have been highly significant ($P=0.0001$) despite being based on only 27 events. Analysis using time-dependent covariates suggested that in the range of 10 to 13 years of age the hazard for male subjects was 4.6 times that for females (95% CI, 2.0 to 10.4), that from 13 to 18 years of age the hazards for the 2 genders were approximately equal (hazard ratio, 1.1; 95% CI, 0.6 to 1.9), and that from 18 to 21 years of age the hazard for female subjects was somewhat higher than that for male subjects (hazard ratio,
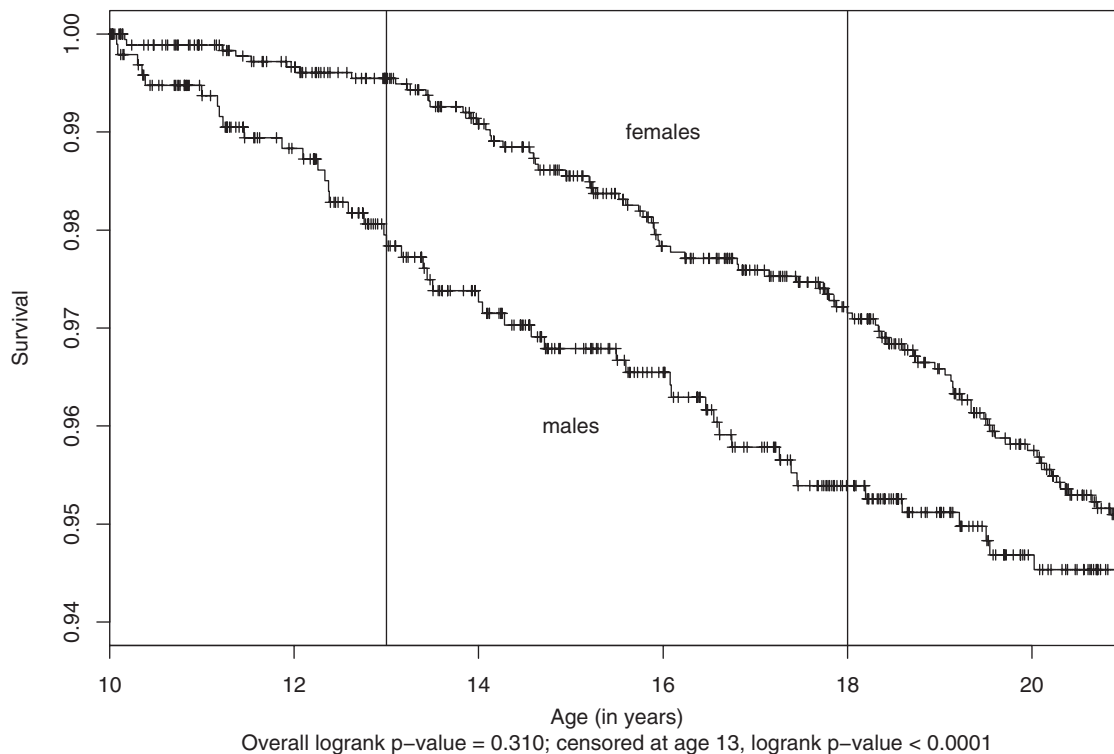
Figure 1. LQTS study: survivor functions for ACA/SCD by gender. Survivor functions were calculated for time from 10 years of age to ACA/SCD using separate Kaplan–Meier estimates for male and female subjects. See text for details.

2.3; 95% CI, 1.0 to 5.6). This pattern of time-dependent effects (nonproportional hazards) persisted when other covariates were included in the model.

Because of this nonproportionality, the authors chose to stratify on gender when examining the effect of other predictors, including the QTc interval. As indicated above, such an analysis is equivalent to fitting models of the form of equation 1 separately for male and female subjects, ie, allowing different functions $h_0(t)$ for the 2 genders, but requiring that the fitted coefficients λ_j be the same for the 2 genders. The estimated effect of QTc can then still be summarized by a single hazard ratio applicable both to male and female subjects. It is more difficult to summarize the effect of gender in a way that is both succinct and accurate. Simply quoting the single, highly nonsignificant probability value or associated hazard ratio would be grossly misleading because it would suppress the clear pattern of divergence followed by partial convergence of the 2 curves shown in Figure 1.

Figure 2 shows plots of estimated survival curves from a gender-stratified Cox model, adjusting now for the QTc interval, here dichotomized at 530 ms. The hazard ratio for subjects with a QTc interval ≥530 ms compared with those of the same gender and QTc interval <530 ms was 3.0 (95% CI, 2.0 to 4.2; *P*<0.0001). Figure 2 forces proportionality for the 2 QTc-specific curves within each gender, but the curves for male and female subjects reflect the actual data, separating at first and then converging, like the 2 curves in Figure 1. Alternatively, we could plot the individual Kaplan–Meier curves for all 4 groups; however, this strategy becomes problematic when many groups are to be compared because each curve would then be based on very small samples.

Time-Dependent Covariates for Dynamic Effects

Time-dependent covariates can be used to model dynamic effects. For example, in the MDPIT study, we examined the prognostic significance of nonfatal reinfarction as a predictor of cardiac death.[4] This was done by incorporating an (additional) covariate x(t) into the PHM model 1. This covariate took the value x(t)=0 for any patient who did not have a nonfatal reinfarction; for a patient who did suffer a nonfatal infarction, this covariate was x(t)=0 before the occurrence of this event and x(t)=1 after its occurrence. The hazard ratio exp(λ) associated with the corresponding coefficient λ then represents the risk of cardiac death among patients who have suffered a recurrent event relative to that among patients who have survived the same length of time since entry, have not suffered a recurrent event, and have the same values of any baseline covariates also included in the model. For our data, after adjustment for several baseline predictors, the occurrence of a nonfatal reinfarction (ie, subsequent to the index infarction that led to the patient being eligible for the study) increased the subsequent risk of cardiac death 3-fold among all patients and 5-fold among patients who had no infarction before the index infarction.

Somewhat similar issues arose in the early days of heart transplantation surgery. Because eligible patients have to wait for a suitable donor heart to become available, in evaluations of the effectiveness of heart transplantation at prolonging life, it is not valid just to compare the overall survival of patients who ultimately receive a heart transplant with those who do not. The inclusion of any such covariate that effectively looks ahead in time (eg, to the time at or after transplant) to
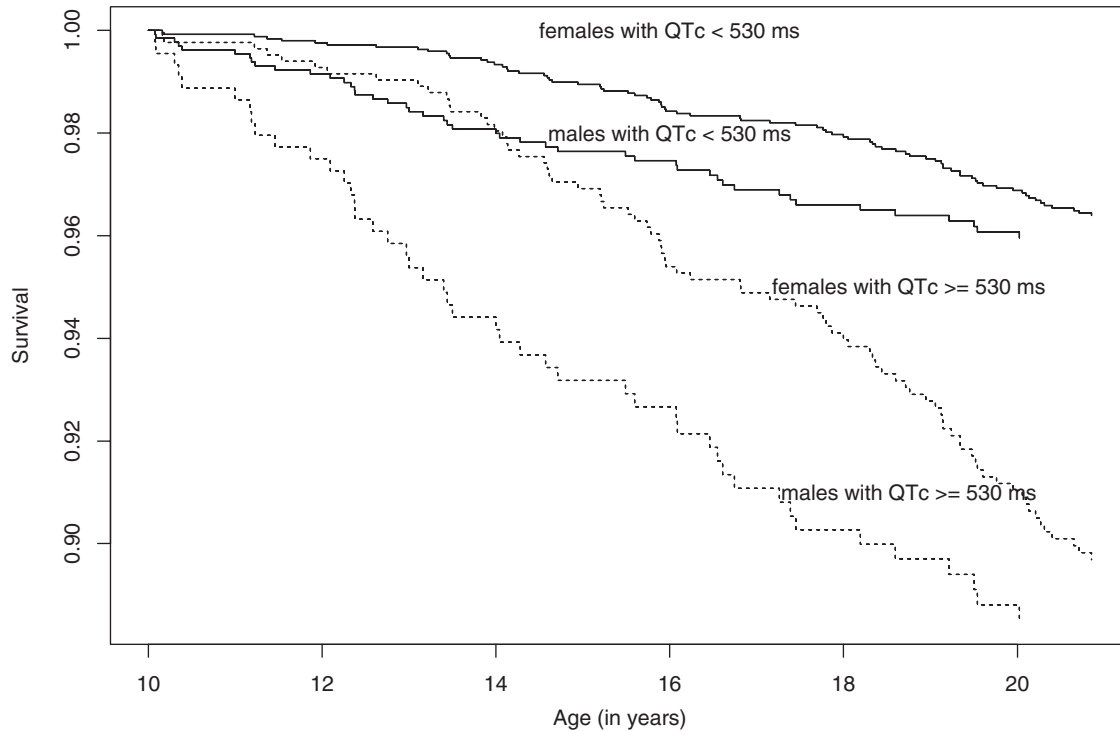
Figure 2. LQTS study: survivor functions from a gender-stratified model. Survivor functions for time from 10 years of age to ACA/SCD were calculated from a Cox model stratified by gender and including a binary indicator variable for QTc interval. Within each gender, the 2 QTc curves are constrained to follow the PHM but not vice versa. See text for details.

determine its value at earlier time points (eg, at the time origin) constitutes a critical violation of the fundamental principles and assumptions underlying all survival analysis methods. The pretransplant survival (measured from the point at which the patient is judged eligible for a transplant) of patients who receive a transplant should be credited to the nontransplanted group. This can be done by defining a time-dependent covariate x(t) that takes the value x(t)=0 for a patient until and unless that patient receives a transplant and subsequently changes to x(t)=1. When analyzed this way, the early data from some early studies did not show any benefit of heart transplantation in saving life. Of course, the picture is now very different because of the development of drugs such as cyclosporine that can control the risk of rejection. For further discussion of the statistical issues, see Turnbull et al,[5] Mantel and Byar,[6] and Cox and Oakes.[7]

In the LQTS study, it was decided to model β-blocker therapy as a time-dependent covariate. Note that this therapy was introduced as and when it was judged to be needed; the situation would be very different in a randomized clinical trial to evaluate the effectiveness of β-blocker therapy. Use of a time-dependent covariate to represent β-blocker therapy allows patients to switch from 1 risk group to the other (on/off β-blockers) over time in accordance with their actual use of this medication at that time. Similar reasoning applies to the effect of prior events such as syncopal episodes. Hobbs et al[3] recently demonstrated that considering both the number and timing of past syncopal episodes significantly improved their ability to estimate the risk of ACA/SCD. As expected, having >1 prior event or having more recent events (within 2 years) significantly elevated the risk of ACA/SCD compared with having fewer or more distant prior events.

Competing Risks and Composite Outcomes

As mentioned, the primary outcome for the MDPIT study was the time to the earlier of cardiac death or nonfatal reinfarction. One reason for using such composite outcome measures is to gain power. If the PHM is correct, the power to detect a treatment effect of a specified magnitude is determined by the number of events. For example, in a randomized study with equal allocation to active treatment and placebo, 66 events are required to provide 80% power to detect a true hazard ratio of 2.0 using a 2-sided significance level of 0.05. Detecting a hazard ratio of 1.5 under the same assumptions would require 191 events. Schoenfeld[8] provides a discussion. In MDPIT, 251 patients suffered cardiac death. Inclusion of the nonfatal reinfarctions increased the total number of primary outcome events to 428, thus substantially increasing the overall power of the study to detect treatment effects of moderate size. (Note that this calculation still counts each patient at most once. Only the first event is counted for patients who suffer multiple reinfarctions or a reinfarction followed by cardiac death.) In such situations, it is important to check, informally at least, that the effects of the treatment on the 2 components of the end point are in a consistent direction. Otherwise, a beneficial effect on 1 component might be balanced by an adverse effect on another component, leading to a possibly misleading conclusion of no overall effect. To avoid this problem, it is sometimes recommended that separate analyses be conducted for each component of the composite scale.

This approach has its own serious problems, however, especially if the different components are ordered by severity in some sense (clearly, a cardiac death is a more "severe" outcome than a nonfatal reinfarction). Fleiss et al[9] criticized the then-widespread practice of analyzing nonfatal reinfarctions separately, treating deaths as censoring events. Nonfatal reinfarction may be reduced by either a lower overall rate of reinfarction or a higher fatality rate among these reinfarctions. The 2 possibilities clearly have very different implications.

A related issue concerns the appropriate presentation of cumulative incidence rates and survival curves for different possible outcomes. Consider, for example, calculation of the cumulative incidence of nonfatal reinfarction. In computing the survival curve for cardiac death, we would not censor patients who suffered a nonfatal reinfarction. Further follow-up data on these patients are available, so there is no need to exclude these data from the analysis. Moreover, as we saw previously, patients who suffer a nonfatal reinfarction are subsequently at substantially increased risk of cardiac death, so censoring their follow-up time at the reinfarction would lead to a falsely optimistic estimate of overall survival. Presentation of the nonfatal reinfarction data is more difficult. We cannot say what the nonfatal reinfarction rate would have been among the patients who died if they had not died, and even if we could, that information would not be very useful. The "noninformative censoring" assumption (Rao and Schoenfeld[1]) implicit in the calculation of the Kaplan–Meier estimate would not be satisfied. Pepe[10] suggested calculating cumulative incidence rates, which estimate the marginal probability of the relevant event having occurred before time t, in contrast to the Kaplan–Meier estimate, which purports to estimate the conditional rate of the event of interest if all other types of event were removed. As Pepe stated, "Competing risks are acknowledged as such in the cumulative incidence function." More recently, Fine and Gray[11] proposed an analog of Cox's model for cumulative incidence functions.

Multiple Events

Throughout this article, we have assumed that the statistical analysis includes only a single event of interest for each patient. Multiple reinfarctions can occur, of course (1 patient in MDPIT had 4), and the question arises as to whether and how information on multiple events should be included in the analysis. The simplest approach, according to Andersen and Gill,[12] is to extend the Cox model for the hazard function to a similar model for the intensity function, now defined as the time-dependent rate of occurrence but without the condition that the event has not already occurred. However, this approach, at least in its simplest form, ignores the difference between multiple events happening to the same individual and single events happening to multiple individuals. Therefore, we do not recommend it for routine use.

Two popular approaches to the analysis of multiple events are the conditional approach of Prentice et al[13] and the so-called marginal approach of Wei et al.[14] In the conditional approach, the times between successive events are modeled sequentially, with the risk of each event expressed conditionally on the entire sequence of previous events experienced by that individual. It is usual to "reset" the time clock after each event, leading to modeling the sequence of "gap times" (intervals between events) rather than the event times themselves (although the original paper considered both possibilities).

In the marginal approach, separate PHMs of the form of model 1 are postulated for the occurrence of the first, second, etc, event to occur to each individual. Parameter estimates from the separate analyses are then combined, and an adjustment is made to the SE to account for the correlations between the estimates (which can be substantial). The conditional approach is more natural and mechanistically plausible. The marginal approach has the counterintuitive feature that individuals are considered "at risk" for their second and subsequent events even before they have experienced their first event. However, when the original time origin has a very specific meaning in terms of the analysis (when, for example, it represents the date of randomization in a clinical trial), the marginal approach has the advantage of linking the occurrence of each outcome to that same time origin and thus providing greater power to detect a given magnitude of treatment effect, supposing the effects on the different components are qualitatively consistent.

Both approaches are now available in standard software. The key message to the practitioner is that the different approaches can lead to different models and thus to different parameter estimates. In particular, it is important to be clear, when reviewing an analysis, whether the parameter estimates refer to total times to events or to gap times between events.

Statistical Analysis of Quality of Life and Medical Cost Data

There is increasing interest in the evaluation of the effect of medical and surgical interventions on both the quality of life experienced by the patient and the costs incurred or saved by the procedure. Suppose that we have devised measures $Q(t)$ of the total cumulative quality of life experienced by a patient up to time t from entry to the study and $C(t)$ of the total costs incurred. The definition and construction of such measures involve complex issues outside the scope of this article, but we may suppose that, for each patient, we have observed or calculated data on $C(t)$ and $Q(t)$ at intervals throughout the study or until follow-up on that patient ceases for any reason. For patients who die, we will have complete data, while for those who survive to the end of the study or who are lost to follow-up, we will have incomplete data, including observations only up to the point of last contact. Statistical analysis of such data must allow for the informative nature of such censoring. Patients who die will tend to have had a lower quality of life and higher rate of accumulating costs than patients who survive. The correct method of analysis involves a concept known as inverse probability weighting. Each individual with complete data is taken to represent a number $w = \pi^{-1}$ of individuals in the underlying population who might have been observed had there been no censoring, where π is the (estimated) probability that such an individual would not be censored.[15–18]

Disclosures

None.

References

1. Rao SR, Schoenfeld DA. Survival methods. *Circulation*. 2007;115: 109–113.

2. Multicenter Diltiazem Postinfarction Trial Research Group. The effect of diltiazem on mortality and reinfarction after myocardial infarction. *N Engl J Med*. 1998;319:385–392.

3. Hobbs JB, Peterson DR, Moss AJ, McNitt S, Zareba W, Goldenberg I, Qi M, Robinson JL, Sauer AJ, Ackerman MJ, Benhorin J, Kaufman ES, Locati EH, Napolitano C, Priori SG, Towbin JA, Vincent GM, Zhang L. Risk of aborted cardiac arrest or sudden cardiac death during adolescence in the long-QT syndrome. *JAMA*. 2006;296:1249–1254.

4. Benhorin J, Moss AJ, Oakes D. Prognostic significance of nonfatal reinfarction. *J Am Coll Cardiol*. 1990;15:254–258.

5. Turnbull BW, Brown BW, Hu M. Survivorship analysis of heart transplant data. *J Am Stat Assoc*. 1974;69:74–80.

6. Mantel N, Byar DP. Evaluation of response time data involving transient states: an illustration using heart transplant data. *J Am Stat Assoc*. 1974; 69:81–86.

7. Cox DR, Oakes D. *Analysis of Survival Data*. London, UK: Chapman and Hall; 1984.

8. Schoenfeld D. Sample size formula for the proportional hazards regression model. *Biometrics*. 1983;39:499–503.

9. Fleiss JL, Bigger JT, McDermott M, Miller JP, Moon T, Moss AJ, Oakes D, Rolnitzky LA, Therneau T. Nonfatal reinfarction is, by itself, an inappropriate end point in clinical trials in cardiology. *Circulation*. 1990; 81:684–685.

10. Pepe MS. Inference for events with dependent risks in multiple endpoint studies. *J Am Stat Assoc*. 1991;86:770–778.

11. Fine JP, Gray JP. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94:496–509.

12. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Ann Stat*. 1982;10:1100–1120.

13. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika*. 1981;68:373–379.

14. Wei LJ Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modelling of marginal distributions. *J Am Stat Assoc*. 1989;84:1065–1073.

15. Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. *Biometrics*. 1997;53:419–434.

16. Huang Y, Louis TA. Nonparametric estimation of the joint distribution of survival time and mark variables. *Biometrika*. 1998;85:785–798.

17. Zhao H, Tsiatis AA. Efficient estimation of the distribution of quality adjusted survival time. *Biometrics*. 1999;55:1101–1107.

18. Bang H, Tsiatis AA. Estimating medical costs with incomplete data. *Biometrika*. 2000;77:329–343.