Proportional Hazards (Cox) Regression

MITCHELL H. KATZ, MD, WALTER W. HAUCK, PhD

PROPORTIONAL HAZARDS (COX) REGRESSION is a powerful analytic tool for testing whether several factors (e.g., cigarette smoking, hypertension) are independently related to the rate (over time) of a specific event (e.g., heart attack yes/no). It can also be used to control for baseline differences between groups in nonrandomized studies and randomized clinical trials (RCTs).

The availability of desktop computers and user-friendly software has resulted in a marked increase in the use of proportional hazards regression by clinical researchers. However, most detailed reviews of the technique[1-5] cannot be understood by non-statisticians. In this article we begin with a review of simpler types of survival analyses, highlighting the concepts of rate of outcome and censored observations. Building on these two concepts, we describe the statistical properties, underlying assumptions, interpretation, and application of proportional hazards regression. Also we describe time-dependent covariates, the use of proportional hazards regression versus logistic regression, and other technical aspects of proportional hazards regression. Finally, we illustrate the applications of this technique by reviewing 80 articles from the *New England Journal of Medicine* and the *Annals of Internal Medicine* that used proportional hazards regression during 1984, 1987, and 1990. Our goal is to enable non-statisticians to interpret these models and to provide guidelines for clinical researchers performing this type of analysis.

SURVIVAL ANALYSIS

Proportional hazards regression belongs to a family of analytic techniques referred to as survival analysis. Survival analysis measures the risk (occurrence) of an outcome (an event such as death) over time. Although the term is "survival analysis," the method applies to analysis of any time-to-event data, whether the event is death or something else.

The starting point ("zero time") for survival analysis of a RCT is generally the time of randomization. In observational studies, the choice of starting point is not always straightforward. For example, in determining survival times for patients with coronary artery disease, the starting point could be the date a patient first developed chest pain, the date the patient was first diagnosed with coronary artery disease by a physician, or the date angiography demonstrated significant coronary artery stenosis. Because there is no "right" starting point for observational studies, investigators should choose the starting point most appropriate for their research question and state their choice clearly in the methods section of their report.

The endpoint for survival analysis is the date of the outcome of interest. The outcome must be dichotomous (e.g., heart attack yes/no) rather than ordinal (e.g., small, moderate, or large areas of the heart not functioning) or continuous (e.g., cardiac enzyme levels). In some studies the outcome may be a composite event. An example of a composite event is development of a cardiac complication: a heart attack, an episode of pulmonary edema, or a cardiopulmonary arrest. With a composite outcome, the date of the first event is used. A study may have both primary (e.g., heart attack) and secondary (e.g., mortality) outcomes.

Survival analysis assesses the risk of an event "over time." This is in contrast to other statistical techniques, which assess the cumulative risk of an event at a particular point in time. For example, a simple proportion can be used to describe survival of a group of patients three years after a heart attack. We can say that after three years, 60% of the patients are still alive. However, we would still not know whether most of the deaths occurred within days of the heart attack, years after the heart attack, or at a constant rate during the study period.

A focus on rate of outcome is particularly important in studies of life-threatening diseases, such as cancer and AIDS. With these diseases, we are interested in treatments that slow the rate of death, even if most subjects ultimately die from that disease. For example, in a study of patients with brain metastases, patients were randomized to surgery followed by radiotherapy versus radiotherapy alone. At two years none of the subjects who had received surgery were alive and only 4% of the subjects who had received radiotherapy alone were alive.[6] However, survival was significantly longer for those who had received surgery than for those who had received only radiotherapy (median survival, 40 weeks versus 15 weeks).

Using a simple proportion to describe the survival of a group suffers from a second major limitation: it cannot accommodate subjects with variable durations of follow-up. If a third of subjects have been followed for a year, a third for two years, and a third for three years, the proportion of subjects alive at three years

Received from the AIDS Office, Department of Public Health (MHK), and the Departments of Medicine (MHK) and Epidemiology and Biostatistics (MHK, WWH), University of California, San Francisco, California.

Address correspondence and reprint requests to Dr. Katz: Director, AIDS Office, 25 Van Ness, Suite 500, San Francisco, CA 94102-6033.

could be based only on those with a three-year follow-up and so would not represent the experience of the entire group.

Variable durations of follow-up occur in clinical trials for a variety of reasons. Usually, clinical trials accrue subjects over a period of months to years, but end follow-up at a common point prior to all subjects' developing the outcome of interest. Therefore, at the end of the trial some subjects will have been followed for longer periods of time than others. Variable lengths of follow-up may also occur because subjects move, lose interest in the study, or become too ill to participate. On occasion, subjects must be withdrawn prematurely from a trial because of a non-endpoint illness that precludes them from continuing in the study or prevents them from being evaluated for the outcome in question. For example, in a RCT comparing warfarin with no treatment or aspirin (control group) in preventing stroke,[7] subjects were withdrawn if they died from causes other than stroke.

Survival analyses overcome the limitations of simple proportions, allowing us to describe the experience of a group over time and to accommodate variable lengths of follow-up in our studies.

Estimation of Survival Curves

Because there are numerous descriptions of how to construct survival curves,[8-11] we only briefly review these methods. Survival curves are graphic representations of the cumulative proportion of subjects remaining event-free at each point in time. The two commonly used methods for estimating the survival time of a group are the life-table (sometimes called actuarial or Cutler-Ederer) and the product-limit (Kaplan-Meier) methods. With the life-table method, the probability of survival is estimated for discrete time intervals (e.g., three months, six months). With the Kaplan-Meier method, an interval is defined by each occurrence of the outcome (e.g., death) in the sample.

To allow for variable lengths of follow-up, these methods "censor" subjects who have not experienced the specified outcome at the time of the last observation date. In effect, a subject is removed from the analysis at the censor date. The advantage of this approach is that a subject who is lost to follow-up at three years of a four-year study can contribute three years of event-free time to the analysis.

The advantages of being able to incorporate information from censored observations is illustrated by a four-year RCT comparing fluoride with placebo for the prevention of fractures among women with osteoporosis.[12] Of the 202 women who began the study, only 135 (67%) completed the four years of treatment. Had the investigators chosen a method of analysis that did not allow for censoring, a third of the women would have been dropped from the analysis. By using survival analysis the investigators could include observations from the time that these women entered into the study until the time they withdrew, thus making fuller use of available information.

Survival analysis assumes that censoring occurs independent of the rate of outcome. In other words, if subjects could be followed beyond the censored point, they would have the same rate of outcome as those not censored. When censoring is not independent from the rate of outcome, survival analysis is not valid. Censoring due to ending follow-up of all subjects at a common point is generally assumed to be independent of rate of outcome. An illustration of nonindependent censoring is a trial in which subjects in one treatment arm are more likely to drop out than subjects in the other arm due to a side effect of treatment that would ultimately result in an outcome event (e.g., death). In the example discussed above of the four-year RCT comparing fluoride with placebo for the prevention of fractures, the investigators reported that the dropout rates were similar for the treatment group and the placebo group. This finding is consistent with censoring occurring independent of the rate of outcome and strengthens the validity of their findings.

In a clinical trial, observations are not necessarily censored when treatment is stopped. In many studies, patients are followed for years after treatment has ended in order to observe long-term outcomes. Also, analysis is often conducted following the intention-to-treat principle. The principle is that time at risk and outcome events are allocated to the initially assigned treatment group whether or not the subject consumes that treatment.

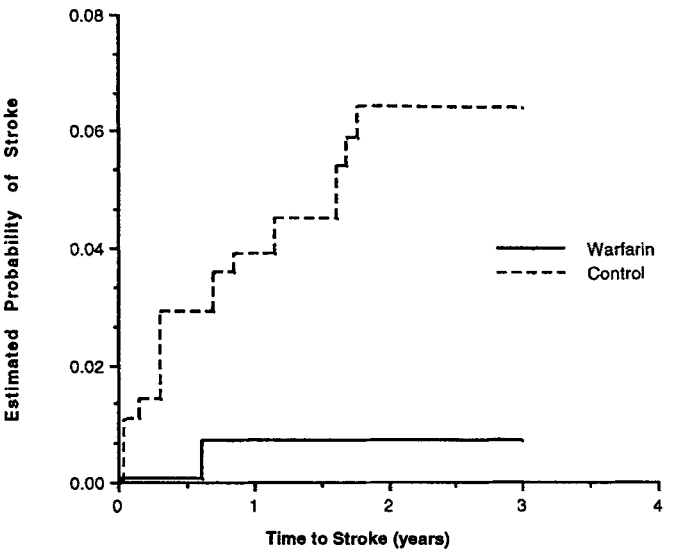

FIGURE 1. Kaplan-Meier curve showing estimated probability of stroke. The broken line represents the control group, and the solid line the warfarin group. Adapted with permission from information appearing in: Boston Area Anticoagulation Trial for Atrial Fibrillation Investigators. The effect of low-dose warfarin on the risk of stroke in patients with nonrheumatic atrial fibrillation. N Engl J Med. 1990;323:1505-11.

Based on the probability of survival at different points in time, survival curves can be graphed. Figure 1 shows two Kaplan-Meier curves representing the survival experience of subjects in the RCT comparing warfarin with no treatment or aspirin (control group) in preventing stroke.[7] Survival curves for two or more groups may be compared using the log-rank (also referred to as the Savage, Mantel-Cox, or Mantel-Haenszel) test or the Wilcoxon (also referred to as the Breslow or Gehan) test. The Wilcoxon test places greater weight on early events in a study than on late events, while the log-rank test gives equal weights to events that occur throughout the study.[4] These statistics test whether the curves are significantly different from one another in time to outcome. Because these statistics compare the entire curves, they should not be used to conclude that the curves differ at a particular point in time. In the case of the comparison between warfarin and control, the log-rank test showed that there were significantly more strokes over time in the control group than in the warfarin group ($p = 0.01$).

Assessing the Effect of Covariates on the Rate of Survival

Comparison of survival curves with statistical testing answers the question of whether two or more groups differ in their time to outcome. However, proportional hazards regression is needed to assess the simultaneous effect of multiple covariates on survival. (The term covariate is used throughout this review; the reader should be aware that covariates are also referred to in the literature as variables, predictors, prognostic factors, risk factors, and confounders. The distinction between a "risk factor" and a "confounder" is important to the interpretation of a study. However, in a statistical model, risk factors and confounders are treated mathematically in the same way, as covariates.)

The need for proportional hazards regression is best appreciated when considering an alternative method of studying the effect of multiple prognostic factors on the rate of an outcome. For example, Elwes et al., using three separate Kaplan-Meier analyses with log-rank tests, reported that the predictors of seizures included a neurologic handicap, a social handicap, and a psychiatric handicap.[13] The problem with this analysis is that the reader cannot evaluate whether these covariates are *independently* related to rate of seizure recurrence. Because these covariates are not mutually exclusive, it is likely that only one or two of them would have independently predicted rate of seizure recurrence in a proportional hazards model.

Proportional hazards regression produces a coefficient (a number) that provides a measure of the association between a covariate and rate of outcome after controlling for other covariates. The coefficient for each covariate estimated by the proportional hazards model can be used to predict the rate at which outcomes will occur for groups of subjects.

PROPORTIONAL HAZARDS MODEL

The model is specified in terms of the *hazard*. The hazard is the probability that a subject will experience an outcome (e.g., death) in the next unit interval of time given that the subject has not yet had the outcome (e.g., is still alive). It is also referred to as the instantaneous risk or force of mortality.

In a proportional hazards model with three covariates, the hazard (h) for a given subject is:

$$h = h_0(t)e^{b_1x_1 + b_2x_2 + b_3x_3} \quad (1)$$

where h_0 is the underlying hazard at time t. A strength of the model is that h_0 is arbitrary; that is, the underlying relationship of the hazard to time need not be specified (modeled). The symbols b_1, b_2, and b_3 represent the regression coefficients (the measures of effect for each of the three covariates), and x_1, x_2, and x_3 represent a particular subject's values for the three covariates.

For example, in a study of rate of heart attack, three important covariates are diastolic blood pressure at randomization ≥ 90 mm Hg (yes/no), history of diabetes (yes/no), and current cigarette smoker (yes/no). For ease of interpretation, dichotomous covariates are generally coded as 0 (no or absent) or 1 (yes or present). Thus, if a subject's value for all the covariates is 0 (not hypertensive or diabetic, nonsmoker), then the formula reduces to $h = h_0$, indicating that the hazard for this subject is the same as the underlying hazard. The hazard of a hypertensive, diabetic smoker (covariates all = 1) would be higher than the underlying hazard. How much higher is found as $e^{(b_1 + b_1 + b_3)}$ and depends on the magnitude of the coefficients.

Interpretation of Regression Coefficients

The values of the regression coefficients are initially unknown and are estimated by fitting the model to the data. Part of Cox's contribution to the field was to provide a method for estimating the best value for the coefficients and their associated standard errors (a measure of the variation of the estimated coefficient due to chance). The regression coefficients associated with a particular covariate can be positive or negative. A positive coefficient indicates that as the covariate (e.g., total cholesterol level) increases, the hazard increases (e.g., higher rate of heart attack). A negative coefficient indicates that as the covariate (e.g., high-density lipoprotein cholesterol) increases the hazard decreases (e.g., lower rate of heart attack).

A chi-square statistic can be used to test the null hypothesis that all the regression coefficients are simultaneously equal to zero (covariates are not associated with rate of outcome). A significant chi-square indi-

cates that the null hypothesis should be rejected. This global chi-square test is nonspecific[14]; it does not indicate which covariates are associated with rate of outcome.

The regression coefficients produced by the model are estimates of the population coefficients. The quotient of the regression coefficient divided by its standard error (called the Wald test) can be used to derive a p value using the normal (z) distribution (this is an approximation that assumes a large sample size). Other tests of the statistical significance of the regression coefficients are the likelihood ratio rest and the score test.

One can estimate the increase or decrease in the hazard due to a one-unit change in the covariate by computing the exponentiated coefficient (i.e., e^b, the mathematical constant e raised to the power of the coefficient's value). This is called the relative hazard. In the case of a dichotomous variable, such as gender (male/female coded as 1/0), the relative hazard is a measure of the relative difference in rates between men and women. A relative hazard greater than 1 indicates that men have a higher rate of outcome, while a relative hazard less than 1 indicates that women have a higher rate of outcome.

Confidence intervals for relative hazards — which indicate the precision of the estimate and the interval of plausible values — can also be calculated. The formula for determining the confidence interval for the relative hazard is $e^{b \pm z\ (\text{standard error})}$, where z is the standard normal deviate for alpha (e.g., 1.96 for 95% confidence intervals).

Multiplicative Assumption

Implicit in the proportional hazards model is the assumption that the covariates have a multiplicative effect on the hazard. Examination of equation 1 indicates that the effect of the covariates are additive on a logarithmic scale — the same as being multiplicative on an arithmetic scale ($\log(a) + \log(b) = \log(a \times b)$).[9] If two covariates, each of which are associated with an increase in rate of outcome, are entered into a proportional hazards analysis, the increase in the hazard for patients with both factors, relative to patients with neither factor should be the product of the increases in rate (as determined from the analysis of the two covariates together) associated with each of the two covariates.

In general, when modeling event rates, we would expect independent prognostic factors to influence their joint effect multiplicatively. In clinical research there may be situations when the joint effect of two factors is less than additive (antagonistic), additive, more than additive and less than multiplicative, or more than multiplicative.[15] In these cases an interaction is said to be present, and adding an "interaction term" (usually done by adding the product of the two variables) to the model may help reveal the nature of the relationship. For example, a study in New York City of survival time following an AIDS diagnosis found that persons who had Pneumocystis pneumonia and were intravenous drug users had a risk of death greater than the multiplicative risk of having Pneumocystis pneumonia and being an intravenous drug user.[16] This was demonstrated statistically by showing that the coefficient of the product term (Pneumocystis pneumonia × intravenous drug use) was positive and statistically significant.

Exponential Increase in Rate with Increases in Continuous Covariates

When continuous (rather than dichotomous) covariates are used in proportional hazards regression, the model assumes that a unit change in the covariate at any point in the scale is associated with a fixed change in the hazard. As the covariate increases, there is an exponential increase in the hazard. For example, in a model that uses diastolic blood pressure as a continuous variable to predict rate of coronary artery disease, the proportional increase in the hazard of coronary artery disease as blood pressure increases from 80 to 89 mm Hg should be the same as the proportional increase in the hazard as blood pressure increases from 90 to 99 mm Hg. Assuming that a ten-point change in blood pressure is associated with a twofold increase in the hazard, then the increase in the hazard for a 30-point increase in blood pressure would be 2^3, or an eightfold increase. Because increasing blood pressure is generally associated with an exponentially increasing hazard of coronary artery disease, blood pressure is frequently left as a continuous covariate in proportional hazards regression.[17,18]

Many clinical covariates are not associated with an exponential increase in hazard and should not be kept continuous. The relationship between levels of a covariate and rate of outcome may be U-shaped, J-shaped, or with a threshold. For example, the investigators of a nonrandomized study reported that for patients with myelodysplasia, the longer after a diagnosis a transplant was performed the higher the rate of relapse.[19] However, as can be seen in Figure 2, the association between time after a diagnosis and rate of relapse was U-shaped, with higher rates of relapse when transplants were performed soon and long after diagnosis. This effect was obscured in the proportional hazards model because the variable (time from diagnosis to transplantation) was entered into the analysis as a single variable measured on a four-level scale (<12 months, 12–30 months, 31–60 months, > 60 months).

When changes in a continuous covariate are not associated with exponential increases in the hazard, then the covariate can be accommodated by dividing the variable into multiple variables and assigning each

subject a yes/no (numerically 1/0) on each variable (referred to as dummy variables). One less variable than the number of categories is needed, and the reference group is represented by a "no" category for each of the other variables. For the study of transplantation with myelodysplasia, length of time from diagnosis could have been represented as three dummy variables: 12–30 months (yes/no), 31–60 months (yes/no), and > 60 months (yes/no). A subject who had received transplantation 40 months post diagnosis would have the values for the three dummy variables of 0, 1, and 0, respectively.

Although dummy variables allow incorporation of covariates that are not associated with exponential increases in the hazard, there are several problems. The choice of cutoff points is often arbitrary and does not necessarily reflect the biologic relationship of the variable to the rate of outcome. Also, the increase in the number of variables often decreases the power of the analysis.[20] An alternative method for incorporating continuous covariates is to perform a mathematical transformation (e.g., logarithm scale) to produce a new scale on which changes in the covariate are associated with exponential changes in the hazard. For example, Cello et al.[21] found that the log of creatinine was significantly associated with survival in patients with cirrhosis. Mathematical transformations will not help with U-shaped or J-shaped distributions.

Proportional Assumption

The term proportional in proportional hazards regression refers to the fact that in these models the ratios

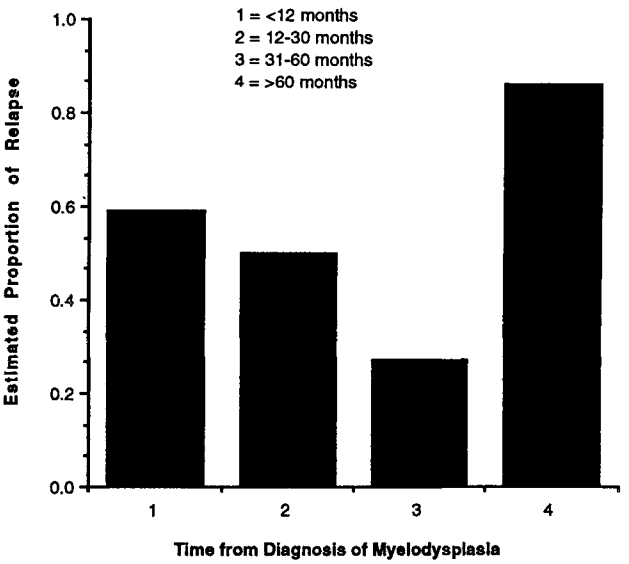

FIGURE 2. Bar graph illustrating that the covariate "time from diagnosis of myelodysplasia to bone marrow transplantation" (as a four-level variable) has a U-shaped relationship with the likelihood of relapse. Based on data obtained from: Appelbaum FR, Barrall J, Storb R, et al. Bone marrow transplantation for patients with myelodysplasia. Ann Intern Med. 1990;112:590-7.

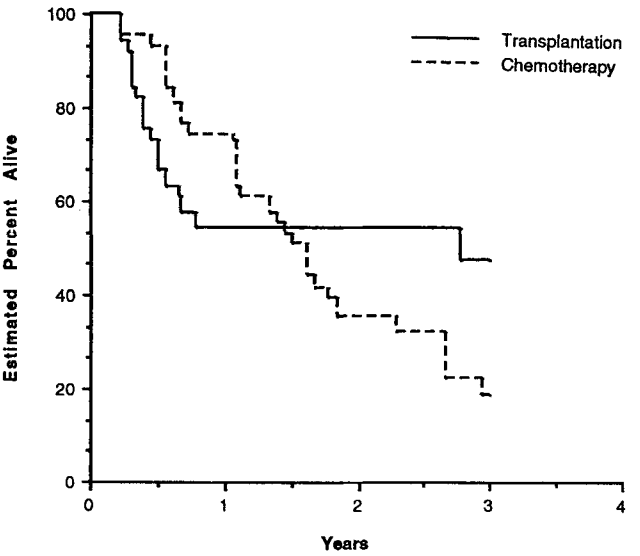

FIGURE 3. Kaplan-Meier curve showing estimated probability of survival. The broken line represents the chemotherapy group and the solid line represents the transplantation group. Reproduced with permission from: Appelbaum FR, Dahlberg S, Thomas ED, et al. Bone marrow transplantation or chemotherapy after remission induction for adults with acute nonlymphoblastic leukemia. Ann Intern Med. 1984;101:581-8.

of the hazards for persons with different patterns of covariates are constant over time. When there is only one covariate, the validity of the proportionality assumption can be verified by examination of the survival curves. The curves should initially show a steadily increasing difference between the two curves.[22] Ultimately, if the outcome is death or a condition that occurs in a very high proportion of exposed persons (e.g., AIDS in persons infected with HIV), the two curves will come together when most subjects have experienced the outcome of the study. Figure 1 shows a steadily increasing benefit for the warfarin group over the control group in preventing strokes.[7] In contrast, Figure 3 shows a study of bone marrow transplantation versus chemotherapy for patients with acute nonlymphoblastic leukemia where the differences are not constant over time. Early on, the rate of death is greater for those undergoing transplantation than for those receiving chemotherapy, and then it reverses, resulting in the curves' crossing.[23]

When the differences in the hazards between groups with different patterns of covariates are not proportional, then proportional hazards regression is not valid. Graphic methods, referred to as log-minus-log survival plots, have been developed to help assess whether the hazards of subjects with different covariate patterns are proportional over time after adjusting for other covariates.[1,24] An example is shown in Figure 4. We note that the difference between the two groups is proportional over time after adjusting for other covariates.

A second test of whether the data fit the proportionality assumption is to create an interaction variable that

is the product of the covariate and survival time (or the logarithm of survival time). Enter the interaction variable, along with the covariate itself, into a proportional hazards model. If the proportionality assumption is valid, the effect of the interaction term will be near zero.[4, 24] This method inherently creates a time-dependent variable (see below). It works best for dichotomous covariates, and when it is expected that the effect of the covariate varies monotonically (stays the same or consistently increases or decreases) with increasing survival time.

If there is an indication that the data do not fit the proportionality assumption, it is sometimes still possible to use proportional hazards methods by performing a stratified analysis.[1] The covariate that does not fit the proportionality assumption is used to stratify the sample. In a stratified model, the baseline hazard (h_0) is distinct in each stratum but the regression coefficients are calculated across strata.[24] By giving each stratum its own baseline hazard, the hazard for each stratum has a component that can vary over time differently than in the other strata.

Applications of Proportional Hazards Analysis

Proportional hazards regression may be used for both nonrandomized and randomized studies. Although we focus on the application of proportional hazards regression for medical studies, these techniques are equally useful for the analysis of behavioral data.[25]

Nonrandomized studies. Proportional hazards regression has two major purposes in nonrandomized studies: to determine prognostic factors while controlling for confounders, and to compare treatment regimens after adjustment for imbalances between the treatment groups.

In nonrandomized studies of prognosis, proportional hazards regression is used to determine whether a covariate is related to an outcome after controlling for other covariates. For example, Rubin et al.[18] found an increased rate of death among elderly men with elevated cholesterol levels after controlling for other covariates known to be associated with mortality, including age, smoking, and hypertension. Without controlling for confounding, their findings would be less convincing; perhaps elevated cholesterol levels were associated with an increased rate of death only because men with high cholesterol levels were older and more likely to smoke or be hypertensive.

In uncontrolled treatment trials proportional hazards regression may be used to determine subgroups of patients who have a particularly good prognosis. For example, a study using proportional hazards regression determined that a poor human leukocyte antigen (HLA) match was associated with a higher rate of kidney graft rejection.[26] The use of proportional hazards regression was important to exclude confounding factors; it could have been that degree of HLA match was important only because it was associated with other covariates (such as prior transplant loss), which, in turn, affected rate of graft rejection. After controlling for these covariates, degree of HLA match remained significant. The demonstration of an independent association between HLA match and kidney rejection has fueled efforts to use organ-sharing networks to optimize the match between donor and recipient.

Finally, in nonrandomized, controlled trials, proportional hazards regression may be used in an attempt to adjust for baseline differences between the treatment and control groups that occurred because treatment was not randomly assigned. For example, the efficacy of angiotensin-converting enzyme (ACE) inhibitors was studied in patients with systemic sclerosis and renal crisis.[27] Subjects who received ACE inhibitors had a higher rate of survival. Because treatment was not randomized, the investigators used proportional hazards regression to adjust for two covariates that were known to be associated with rate of survival: hypertension and age. After statistical adjustment, use of ACE inhibitors was still associated with a higher rate of survival.

The ability to adjust for measured baseline differences in nonrandomized treatment trials is one of the strengths of proportional hazards regression. In situations when randomized clinical trials are not practical —because of expense, logistic or ethical difficulties in randomizing patients, or need for a rapid answer[10, 28] —nonrandomized trials analyzed with proportional hazards regression may be useful.[28, 29] For example, it would have taken years to enroll enough patients to establish the efficacy of ACE inhibitors for patients with systemic sclerosis and renal crisis. More importantly, it

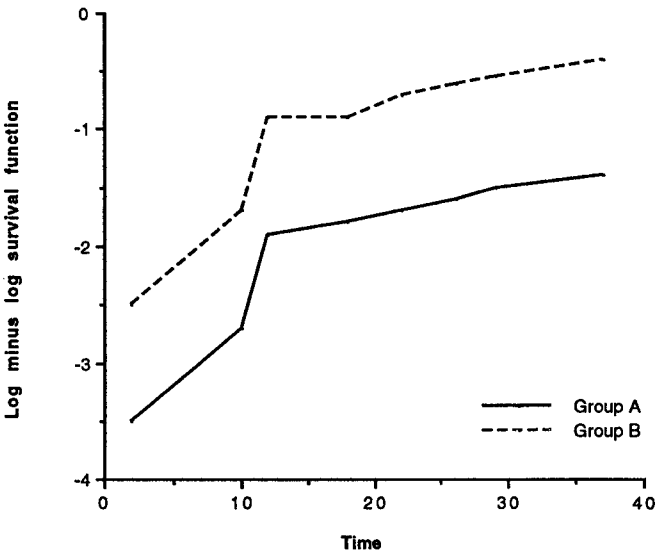

FIGURE 4. Log-minus-log survival plot showing a constant difference between group A (solid line) and group B (broken line).

would have been unethical to randomize patients to placebo because of the high mortality for the disease when patients go untreated and the high probability that ACE inhibitors would be helpful given clinical experience with similar diseases. However, statistical adjustment for baseline differences can be successful only when the other prognostic factors of a disease are known and measured; only randomization has the capability of controlling for the influence of unmeasured confounders.

Randomized Controlled Trials. Although randomization usually creates study groups that are similar with respect to baseline characteristics, randomization may, by chance, result in an unequal distribution of characteristics, especially in smaller studies. These random differences may confound the results of a study[29] if the covariate differing between the groups is casually related to the outcome. Proportional hazards regression can be used to adjust statistically for these differences. For example, in one RCT, patients with tuberculosis were randomized to either a six-month or a nine-month regimen of antituberculosis treatment.[30] The six-month regimen was as efficacious as the nine-month regimen. However, the group randomized to the six-month regimen appeared to be less ill at baseline: they were significantly less likely to have cavitary and pleural disease than were the nine-month regimen group. These baseline differences raised the possibility that the six-month regimen was less effective and only appeared to be as effective as the nine-month regimen because the subjects randomized to the six-month regimen were less sick. However, the investigators adjusted for the presence of cavitary and pleural disease in proportional hazards analysis and found that the two regimens still had similar levels of efficacy.

In addition to adjusting for baseline differences, proportional hazards regression may be used in RCTs to uncover factors other than the randomized treatment that are associated with rate of outcome. For example, in a RCT of corticosteroids versus placebo for adjunctive therapy of Pneumocystis pneumonia, hypoxia, in addition to receipt of placebo, was associated with a higher rate of respiratory failure.[31]

Proportional hazards regression may also be used to investigate subgroup effects. Using a product term (covariate × treatment assignment), investigators can detect subgroups for whom treatment is particularly effective or demonstrate a consistent ineffectiveness of treatment across subgroups of subjects. For example, in a RCT of beta carotene versus placebo to prevent skin cancer,[32] the investigators found that the rates of cancer did not differ by treatment group. The lack of significant interaction between treatment group and covariates (such as plasma beta-carotene level) strengthened their conclusion that beta carotene use was not efficacious in preventing skin cancer.

Time-dependent Covariates

Proportional hazards regression can accommodate covariates that change their values during the course of a study.[4, 33] For example, during a study a subject may start a new medicine or develop a new symptom. Time-dependent covariates were used in one study to test whether certain symptoms were associated with development of Pneumocystis pneumonia in patients with low CD4 lymphocyte counts.[34] In the model the covariates changed their values from absent to present at the time they appeared in a subject with a CD4 count less than 200 cells. The investigators found that the presences of oral candidiasis and fever, but not of fatigue or weight loss, were associated with an increased relative hazard of developing Pneumocystis pneumonia.

Another use of time-dependent covariates is to determine whether sicker subjects are more likely to be lost to follow-up (thereby violating the assumption of survival analysis that censoring be independent of rate of outcome). This was done in a study of zidovudine for asymptomatic HIV infection.[35] The last CD4 lymphocyte count, modeled as a time-dependent covariate, was used as an indicator of degree of illness. The count was not, however, significantly related to the hazard of being lost to follow-up. If subjects were more likely to drop out when they became sick, then this analysis would have found that CD4 counts were related to the hazard of being lost to follow-up. Use of a simple t-test to compare the last CD4 count of subjects lost to follow-up with the last CD4 count of subjects who stayed in the study would not have been valid. HIV infection causes a progressive decline in CD4 counts; therefore even if sicker patients were more likely to drop out, this finding would have been obscured because subjects who stayed in the trial the longest would tend to have the lowest CD4 counts.

Comparison with Logistic Regression Analysis

Logistic regression is a multivariable technique for assessing the independent effect of multiple covariates on the occurrence of a binary event.[9, 10, 15] It is similar to proportional hazards regression, and investigators may be unsure which method would be better to use. A detailed comparison of the statistical properties of the two models has been presented elsewhere.[36] Here we note a few general differences in the two techniques.

Proportional hazards analysis is based on rate of outcome, while logistic regression analysis is based on cumulative risk of outcome at a particular point in time (e.g., three years after the randomization date). Using cumulative risk means that early events have the same weight in the analysis as events that occur later. Since the goal of many medical treatments is extending life rather than curing disease, the timing of outcomes is important.

Observations cannot be treated as censored with

TABLE 1

Time Trends in the Use of Proportional Hazards Regression (PHR)*

	Articles Using PHR			Total Articles			% Articles Using PHR		
Year	*NEJM*	*Annals*	Both	*NEJM*	*Annals*	Both	*NEJM*	*Annals*	Both
1984	11	4	15	501	294	795	2%	1%	2%
1987	18	6	24	449	268	717	4%	2%	3%
1990	33	8	41	476	259	735	7%	3%	6%
TOTAL	62	18	80	1,426	821	2,247	4%	2%	4%

*$p < 0.001$ for chi-square for trend for *NEJM* alone and the two journals together; $p = 0.17$ for *Annals* alone. *NEJM* refers to the *New England Journal of Medicine* and *Annals* refers to the *Annals of Internal Medicine.*

logistic regression. Therefore a subject lost to follow-up a few days before three years cannot contribute to a logistic analysis comparing subjects who died within three years with subjects still alive at three years. Also, a subject who has an outcome (e.g., death) a few days after the chosen analysis point is treated in the analysis as belonging to the group that did not experience the outcome (e.g., still alive). On the other hand, an advantage of logistic regression is that the proportionality assumption does not have to be valid (because the model does not take into account when events occur). Both models assume a multiplicative relationship of the covariates to outcome, though in different scales.

In general, the use of logistic regression in place of proportional hazards models works best if the number of subjects lost to follow-up is small[37, 38] and the proportion of subjects experiencing an outcome is small.[39] If the proportion of subjects experiencing an outcome is large, then the differences between the relative hazards (from the proportional hazards model) and the odds ratios (from the logistic model) will be large.

Technical Issues

Several user-friendly software packages perform proportional hazards regression.[24, 40-42] These programs allow for forward and backward (stepwise) selection of covariates (instead of entering all variables simultaneously into the model). As with other types of regression analysis, there are advantages and disadvantages to using these selection techniques.[14, 15, 43] Guides for determination of sample size calculations for proportional hazards regression are available,[44] and the use of this technique with matched data is an area of active research.[45]

REVIEW OF THE LITERATURE

To assess the frequency and manner in which proportional hazards regression is being used in the literature, we reviewed all articles (excluding editorials and letters) in the *New England Journal of Medicine* and the *Annals of Internal Medicine* during the years 1984, 1987, and 1990. Eighty articles that used proportional hazards analysis were identified. Excluded

TABLE 2

Characteristics of Articles Using Proportional Hazards Regression (n = 80)*

Use of proportional hazards regression	
Nonrandomized studies	51 (64%)
Randomized clinical trials	29 (36%)
Term used for exponent of coefficient	
Relative risk	34 (43%)
Risk	5 (6%)
Relative hazards	5 (6%)
Risk ratio	2 (3%)
Hazard ratio	2 (3%)
Rate ratio	2 (3%)
Relative fertility	1 (1%)
Odds ratio	1 (1%)
Not reported	28 (35%)
Reported p values	61 (76%)
Reported confidence intervals of coefficients	38 (48%)
Did not report p values or confidence intervals	7 (9%)
Assessed multiplicative assumption	8 (10%)
Assessed proportionality assumption†	12 (15%)

*Percentages may not add up to 100% because of rounding off.
†One article, which incorporated time into the model, is excluded.

were six articles that used proportional hazards regression only for computing bivariate statistics, and one article for which the authors stated that they considered using proportional hazards regression but did not because their data did not fit the proportionality assumption.

There was a significant increase in the proportion of articles from these two journals reporting proportional hazards analysis, from 2% in 1984 to 3% in 1987, and 6% in 1990 ($p < 0.001$ for trend) (Table 1). The increase in the proportion of articles from the *Annals of Internal Medicine* alone was not statistically significant ($p = 0.17$).

Of these 80 articles, 51 (64%) were nonrandomized studies and 29 (36%) were RCTs (Table 2). We found enormous variability in the terms used to describe e^b. Relative risk, risk, and risk ratio were commonly used but have the disadvantage of also being used

to describe results from models (e.g., logistic regression), which focus on cumulative risk of outcome rather than rate of outcome. Relative hazard is a more specific term and therefore more appropriate. Of the articles reviewed, 61 (76%) gave the p value of the coefficients, 38 (48%) reported confidence intervals, and seven (9%) reported neither. Only eight studies (10%) reported that they assessed the multiplicative relationship of the covariates. And only 12 of 79 articles (excluding one article that incorporated time into the model) indicated that they assessed the validity of the proportionality assumption. Of these 12 articles, four (33%) noted that the data did not fully fit the proportionality assumption.

CONCLUSIONS

Proportional hazards regression is an increasingly used tool of clinical researchers. The main purposes of proportional hazards regression are to evaluate rate of outcome, to accommodate variable lengths of follow-up, and to assess the effect of multiple covariates on the rate of outcome.

To evaluate proportional hazards analysis, readers need to know whether the underlying assumptions of the model were fulfilled. Our review, as well a review of articles using logistic and proportional hazards regression from the *New England Journal of Medicine* and *Lancet*,[46] indicates that investigators do not consistently report whether they tested the multiplicative and proportional assumptions. Researchers should report how they tested these assumptions and whether there were departures from them. Use of the term relative hazard makes it clear to readers that the estimates are based on a proportional hazards model. Reporting the confidence intervals of the relative hazards helps readers to evaluate the precision of the findings.

Given the flexibility of proportional hazards regression for handling clinical data, its use in clinical research is likely to continue to increase. With an appreciation of the strengths and weaknesses of these models, readers can better interpret the validity of studies that use these techniques.

The authors appreciate the following people who thoughtfully reviewed earlier versions of the manuscript: Daniel Berrios, MPH, Warren Browner, MD, MPH, Susan Buchbinder, MD, Steven Cummings, MD, Virginia Ernster, PhD, Nancy Hessol, MSPH, Steven Hulley, MD, MPH, Karla Kerlikowske MD, Alan Lifson, MD, MPH, Walter Mebane, PhD, Anthony So, MD, Janice Westenhouse, MPH, and two anonymous reviewers.

REFERENCES

1. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data. New York: Wiley, 1980.
2. Cox DR. Regression models and life tables. J R Stat Soc. 1972;34:187-220.
3. Miller R. Survival Analysis. New York: Wiley, 1981.
4. Lawless J. Statistical Models and Methods for Lifetime Data. New York: Wiley, 1982.
5. Cox D, Oakes D. Analysis of Survival Data. London: Chapman and Hall, 1984.
6. Patchell RA, Tibbs PA, Walsh JW, Dempsey RJ, Maruyama Y, Kryscio RJ. A randomized trial of surgery in the treatment of single metastases to the brain. N Engl J Med. 1990;322:494-500.
7. Boston Area Anticoagulation Trial for Atrial Fibrillation Investigators. The effect of low-dose warfarin on the risk of stroke in patients with nonrheumatic atrial fibrillation. N Engl J Med. 1990;323:1505-11.
8. Muenz LR. Comparing survival distributions: a review for nonstatisticians. I. Cancer Invest. 1983;1:455-66.
9. Kahn HA, Sempos CT. Statistical Methods in Epidemiology. New York: Oxford, 1989.
10. Anderson S, Auquier A, Hauck WW, Oakes D, Vanaele W, Weisberg HI. Statistical Methods for Comparative Studies. New York: Wiley, 1980.
11. Reznick RK, Guest CB. Survival analysis: a practical approach. Dis Colon Rectum. 1989;32:898-902.
12. Riggs BL, Hodgson SF, O'Fallon WM, Chao EYS, Wahner HW, Muhs JM. Effect of fluoride treatment on the fracture rate in postmenopausal women with osteoporosis. N Engl J Med. 1990;322:802-9.
13. Elwes RDC, Johnson AL, Shorvon SD, Reynolds EH. The prognosis for seizure control in newly diagnosed epilepsy. N Engl J Med. 1984;311:944-7.
14. Greenland S. Modeling and variable selection in epidemiologic analysis. Am J Public Health. 1989;79:340-9.
15. Rothman KJ. Modern Epidemiology. Boston: Little, Brown, 1986.
16. Rothenberg R, Woelfel M, Stoneburner R, Milberg J, Parker R, Truman B. Survival with the acquired immunodeficiency syndrome: experience with 5833 cases in New York City. N Engl J Med. 1987;317:1297-302.
17. Levy D, Garrison R, Savage D, Kannel W, Castelli W. Prognostic implications of echocardiographically determined left ventricular mass in the Framingham Heart Study. N Engl J Med. 1990;322:1561-6.
18. Rubin SM, Sidney S, Black DM, Browner WS, Hulley SB, Cummings SR. High blood cholesterol in elderly men and the excess risk for coronary heart disease. Ann Intern Med. 1990; 113:916-20.
19. Appelbaum FR, Barrall J, Storb R, et al. Bone marrow transplantation for patients with myelodysplasia. Ann Intern Med. 1990; 112:590-7.
20. Gill R. Understanding Cox's regression model. Experientia Suppl. 1982;41:187-99.
21. Cello J, Grendell J, Crass R, Trunkey D, Cobb E, Heilbron D. Endoscopic sclerotherapy versus portacaval shunt in patients with severe cirrhosis and variceal hemorrhage. N Engl J Med. 1984;311:1589-94.
22. Tibshtrani R. A plain man's guide to the proportional hazards model. Clin Invest Med. 1982;5:63-8.
23. Appelbaum FR, Dahlberg S, Thomas ED, et al. Bone marrow transplantation or chemotherapy after remission induction for adults with acute nonlymphoblastic leukemia. Ann Intern Med. 1984;101:581-8.
24. Hopkins A. Survival analysis with covariates—Cox models. In: Dixon WJ (ed). BMDP Statistical Software Manual. Berkeley, CA: University of California Press, 1985;576-91.
25. Singer JD, Willett JB. Modeling the days of our lives: using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events. Psychol Bull. 1991; 110:268-90.
26. Sanfilippo F, Vaughn WK, Spees EK, Light JA, Lefor WM. Benefits of HLA-A and HLA-B matching on graft and patient outcome after cadaveric-donor renal transplantation. N Engl J Med. 1984; 311:358-64.
27. Steen VD, Constantino JP, Shapiro AP, Medsger TA. Outcome of renal crisis in systemic sclerosis: relation to availability of angiotensin converting enzyme (ACE) inhibitors. Ann Intern Med. 1990;113:352-7.
28. Hlatky MA, Lee KL, Harrell FE, et al. Tying clinical research to patient care by use of an observational database. Stat Med. 1984;3:375-84.
29. Rosati RA, Lee KL, Califf RM, Pryor DB, Harrell FE. Problems and

advantages of an observational data base approach to evaluating the effect of therapy on outcome. Circulation. 1982;65(suppl II):27-32.
30. Combs DL, O'Brien RJ, Geiter LJ. USPHS tuberculosis short-course chemotherapy trial 21:effectiveness, toxicity, and acceptability. Ann Intern Med. 1990;112:397-406.
31. Bozzette S, Sattler F, Chiu J, et al. A controlled trial of early adjunctive treatment with corticosteroids for *Pneumocystis carinii* pneumonia in the acquired immunodeficiency syndrome. N Engl J Med. 1990;323:1451-7.
32. Greenberg E, Baron J, Stukel T, et al. A clinical trial of beta carotene to prevent basal-cell and squamous-cell cancers of the skin. N Engl J Med. 1990;323:789-95.
33. Andersen PK. Time-dependent covariates and Markov processes. In: Moolgavkar SH, Prentice RL (eds). Modern Statistical Methods in Chronic Disease Epidemiology. New York: Wiley, 1986; 82-103.
34. Phair J, Munoz A, Detels R, et al. The risk of *Pneumocystis carinii* pneumonia among men infected with human immunodeficiency virus type 1. N Engl J Med. 1990;322:161-5.
35. Volberding P, Legakos S, Koch M, et al. Zidovudine in asymptomatic human immunodeficiency virus infection. N Engl J Med. 1990;322:941-9.
36. Feinstein AR, Wells CK, Walter SD. A comparison of multivariable mathematical methods for predicting survival—I. Introduction, rationale, and general strategy. J Clin Epidemiol. 1990;43:339-47.
37. Seigel DG. Cox regression analysis in clinical research. Arch Ophthalmol. 1990;108:888-9.
38. Hauck WW. A comparison of the logistic risk function and the proportional hazards model in prospective epidemiologic studies [letter]. J Chron Dis. 1985;38:125-6.
39. Green M, Symons M. A comparison of the logistic risk function and the proportional hazards model in prospective epidemiologic studies. J Chron Dis. 1983;36:715-24.
40. SAS Institute Inc. SAS User's Guide: Statistics, Version 6. Cary, NC: SAS Institute Inc., 1991.
41. Norusis MJ. SPSS Advance Statistics User's Guide. Chicago: SPSS, 1990.
42. Statistics and Epidemiology Research Corporation. Seattle, WA: Egret Manual. Statistics and Epidemiology Research Corporation, 1990.
43. Muenz LR. Comparing survival distributions: a review of nonstatisticians. II. Cancer Invest. 1983;1:537-45.
44. Lachlin JM. Introduction to sample size determination and power analysis. Controlled Clin Trials. 1981;2:93-113.
45. Wei LJ, Lin EY, Weissfeld L. Regression analysis of incomplete failure time data by modelling marginal distributions. J Am Stat Assoc. 1989;84:1065-73.
46. Concato J, Feinstein AR, Holford TR. The risk of determining risk with logistic and Cox regression. Ann Intern Med. 1993; 118:201-10.

■

REFLECTIONS

Medicine is a science which hath been, as we have said, more professed than laboured, and yet more laboured than advanced; the labour having been, in my judgment, rather in a circle than in progression. —SIR FRANCIS BACON (1561–1626), English philosopher, lawyer, and politician (*Advancement of Learning*, Bk. II)