Circulation
JOURNAL OF THE AMERICAN HEART ASSOCIATION

Survival Methods

Sowmya R. Rao and David A. Schoenfeld

Circulation 2007;115;109-113
DOI: 10.1161/CIRCULATIONAHA.106.614859
Circulation is published by the American Heart Association. 7272 Greenville Avenue, Dallas, TX 72514

 Print ISSN: 0009-7322. Online ISSN: 1524-4539

Survival Methods

Sowmya R. Rao, PhD; David A. Schoenfeld, PhD

This article gives an overview of survival methods in medical studies. We briefly describe survival data and discuss the methods used for analysis of such data. We apply these methods to data from a clinical trial and discuss the results. Survival methods are applicable when the measure of interest is time to an event such as mortality or occurrence of disease. The concept of censoring makes survival methods unique. If a patient goes through the study without having the event, his time to the event is (right) censored, in the sense that we only know that the event happened after the last time we observed the patient. Thus, for each patient we have 2 pieces of data: The first is a time that is either the patient's event time or the time that the patient was last followed up, and the second is an indicator that denotes whether the time is an event time or a follow-up time. Another way of thinking of censoring is to assume that each patient has an event time and a censoring time after which the patient would no longer be observed. Whenever the censoring time is less than the event time, the event time is missing. Survival methods also assume that the censoring time is unknown when it is greater than the event time.

Survival distributions are usually described in terms of 2 functions: the survival function, S(t), defined as the probability that a person survives past a specified time t; and the hazard function, h(t), which is the instantaneous failure rate and is defined as:

$$h(t) = -\left\{\frac{\left[\frac{dS(t)}{d(t)}\right]}{S(t)}\right\}.$$

Suppose a patient has survived to time t; then the hazard function is the probability that the patient will have an event in the next instant. The hazard function is conceptually useful in describing survival distributions but is rarely published. The greater the hazard function, the shorter is the survival time.

The survival methods that we describe require that the censoring time is independent of the event time. This is called noninformative censoring. An example that illustrates when this assumption would always be met is a clinical trial in which patients enter the study over a period of time and there are no dropouts. If the patient does not have an event before the end of the study, the patient's event time will be censored. The distribution of the potential censoring time will only depend on when the patient entered the study. This time will be independent of the patient's time to event as long as there are no secular trends in the survival distribution. An extreme example of an instance when these assumptions would not be met is a study of time to death, where patients are no longer followed up after they recover from a disease. Patients who are lost to follow-up in a clinical trial or drop out of a clinical trial are problematic because the time to their last observation may or may not be related to their unobserved event time. For instance, if patients who feel better drop out of the study, the censoring may be informative. Approaches to this problem have been the focus of an extensive literature.[1]

Estimating the Survival Function

The most commonly used descriptive statistics for survival data are based on an estimate of the survival function. Often the median is reported, which is the value of t where the survival function, S(t), equals 0.5 (ie, 50% of the cohort is event free). Sometimes the value of S(t) is reported at t=1, 5, or 10. The mean survival time is rarely reported because, as we shall see, it cannot be estimated reliably. Each of these descriptive statistics starts with the estimation of the survival function. Often the estimate of the entire function is included in a report of a study. The advantage of this is that the behavior of the function over various time periods may be of interest. The curve may drop steeply at first because of early events or may level off if patients who survive past a certain point without an event are unlikely to have one in the future. As we see in the next section, the curve cannot be estimated past the longest follow-up time.

The Kaplan-Meier method[2] is frequently used to estimate the survival function when there are censored data. The best way to understand this method is to break up the time scale into intervals that end at each event time. Let $t_1, t_2, \ldots$ be ordered event times. Then, because no event occurs before time t_1, the value of S(t) is 1 from t=0 to just before $t=t_1$. Suppose that n_1 is the number of patients that are being observed at time t_1 (by convention, if a patient's censoring time is t_1, the patient is considered to be observed at t_1), and m_1 is the number of events at t_1. Then, S(t) is

$$\frac{(n_1 - m_1)}{n_1}$$

From Massachusetts General Hospital Biostatistics Center and Institute for Health Policy (S.R.R.), and Massachusetts General Hospital Biostatistics Center, Department of Biostatistics, Harvard School of Public Health (D.A.S.), Boston, Mass.

Correspondence to Sowmya R. Rao, PhD, Massachusetts General Hospital, Biostatistics Center, 50 Staniford St, Suite 560, Boston, MA 02114. E-mail srrao@partners.org

(*Circulation.* **2007;115:109-113.**)

Circulation **is available at http://www.circulationaha.org** **DOI: 10.1161/CIRCULATIONAHA.106.614859**

for $t=t_1$ up until just before t_2. Note that this is simply the proportion of patients that survive past t_1 among those who survived until t_1. At t_2, one estimates the probability that a patient survives past t_2 given that the patient lives up to t_2 by

$$\frac{(n_2-m_2)}{n_2}.$$

The estimate of S(t) is then

$$\left[\frac{(n_1-m_1)}{n_1}\right]\times\left[\frac{(n_2-m_2)}{n_2}\right]$$

from $t=t_2$ up to just before $t=t_3$. Most computer programs that compute the Kaplan-Meier survival curve start with 2 columns of data; the first is the survival or censoring time, and the second is a censoring indicator that is 0 if the time is a censoring time and 1 if the time is an event time.

The plots with survival time on the horizontal axis and the proportion surviving, S(t), on the vertical axis when there is censoring are called the KM curves. These start at 1 (because probability of survival beyond time 0 is 1) and step down toward zero. If there are subjects who survive beyond the study time, then the survival curve does not go to zero but stays horizontal. Because the survival curve does not change in intervals in which no events occur, one can calculate the curve at event times only. The mean time to an event is estimated by the area under the survival function. If the largest time is an event time, then the survival function goes to zero at that time, and this estimate is finite. Otherwise, the mean time cannot be estimated. This is why median survival is used more often than mean survival. The median survival time is estimable only if the survival curve drops to or below 0.5. The median survival time is estimated by the value on the horizontal axis at the intersection of a horizontal line drawn from the vertical axis to the survival line where $S(t)=0.5$. If the KM curve drops to or below 0.5 but does not equal 0.5, then the first event time when the curve falls below 0.5 is used.

Survival curves can be compared to assess differences in treatment effects. If ≥ 2 groups are being compared, survival curves are plotted for each group. If the survival curves are parallel to each other, then the group that consistently has a higher survival curve than the other has longer survival, and the treatment given to this group is concluded to be the better of the treatments being studied. Although visually the survival curves for the 2 groups might seem different from each other, we need to test whether the "true" survival curves are statistically different by using a formal test.

Hypothesis Testing

In most studies, one is interested in evaluating the effect of 1 or more treatments or exposures (eg, aspirin/placebo, smoking) on an outcome of interest (eg, first myocardial infarction) either by itself or adjusted for other covariates. In survival analysis, the outcome of interest is the survival time, and one is interested in comparing the survival times between groups or assessing the relationship of exposure/covariates to the survival time. Standard methods of data analysis (eg, *t* tests, linear or logistic regressions) cannot be applied to survival data because they do not account for censoring. If censored observations are excluded from the analysis, the results will be biased.

Proportional Hazards or Log-Rank Test

The log-rank test can be used to test the hypothesis of no difference in survival between the 2 groups. This test makes the assumptions that the observations are independent and that the censoring distribution is independent of the survival distribution; notably, the censoring distribution can be different in each group, so that the test can be used to compare a current treatment group with a historical control.

This tests the null hypothesis that there is no statistical difference between the survival curves in the 2 groups. The basic idea behind the test is that at each event time t_i there will be n_{1i} patients in group 1 and n_{2i} patients in group 2. Under the null hypothesis of no treatment effect, the probability that the treatment group of the patient who had the event will be in group 1 is

$$\frac{n_{1i}}{(n_{1i}+n_{2i})}.$$

Thus, if we define an indicator variable δ_i to equal 1 if the patient is from group 1 and zero if the patient is from group 2, then

$$\sum_i\left\{\delta_i-\left[\frac{n_{1i}}{(n_{1i}+n_{2i})}\right]\right\}$$

has a mean equal to zero. This is the numerator of the log-rank test. The denominator is the standard deviation of this quantity:

$$\left[\sum_i\frac{n_{1i}n_{2i}(m_{1i}+m_{2i})(n_{1i}+n_{2i}-m_{1i}-m_{2i})}{(n_{1i}+n_{2i})^2(n_{1i}+n_{2i}-1)}\right]^{1/2}.$$

The log-rank test can be used to compare ≥ 2 survival curves. It is preferable to use the multivariable regression method to assess the relationship of many risk factors to survival.

Cox Proportional Hazards Model

This is the most popular method to evaluate the relationship between covariates and survival with the use of a mathematical model. This is called a semiparametric model because it does not assume any distribution for the baseline hazard. The model is defined as

$$h(t;x_1,x_2,\ldots,x_k)=\lambda_0(t)\exp(\lambda_1x_1+\lambda_2x_2+\ldots+\lambda_kx_k)$$

where $\lambda_0(t)$ is the baseline hazard at time t and $x_1,x_2,\ldots,x_k$ are k independent covariates. No assumptions are made regarding the baseline hazard function.

We can test the association of each of the independent variables with survival time adjusted for other covariates. It is important to understand the meaning of the parameters in a proportional hazards model. Suppose first that the covariate, say x_1, has 2 values, 0 and 1. Then, $\exp(\lambda_1)$ is the hazard ratio for patients with $x_1=1$ versus those with $x_1=0$. That is the instantaneous probability of an event in one group divided by that probability in the other. Notice that we are modeling the hazard so that if patients for whom $x_1=1$ have longer survival times, then λ_1 will be negative. The model specifies that the

hazard ratio is constant over time and for the values of all the other covariates. When the covariate is continuous, then $\exp(\lambda_1)$ is the hazard ratio for a unit change in the value of x_1. It is often helpful to divide continuous covariates by their standard deviation so that the units for each covariate are comparable, and $\lambda_1,\lambda_2,\ldots$ have the same scale, which would be 1 standard deviation. Hazard ratios are approximately equal to the relative risk (ratio of risk in the exposed group to the risk in the unexposed group) and are used interchangeably.

Parametric Methods

Parametric methods assume that the survival times follow a specified distribution. Exponential, Weibull, Gompertz, Gamma, and log-normal distribution are often used for survival times. In the exponential model the hazard function is constant, which means a person's probability of an event in the future is independent of how long the person has gone without an event. Weibull, Gompertz, or Gamma can be used when the hazard function is monotonically increasing or decreasing. The log-normal distribution is assumed when the hazard increases in the beginning and then decreases.

Probably the most commonly used parametric survival model is the exponential model, which is defined as

$$h(t; x_1, x_2, \ldots, x_k) = \exp(\lambda_0 + \lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_k x_k) .$$

This model has some very useful properties. Without covariates, the survival function is $S(t)=\exp[-t\exp(\lambda_0)]$. The hazard function is constant $\exp(\lambda_0)$, the mean survival is $\exp(-\lambda_0)$, and the median survival is $-\exp(-\lambda_0)\times\log(0.5)$. To estimate λ_0, let n be the total number of events and let f be the total amount of follow-up; then λ_0 is estimated by log(n/f), which has a standard error of $1/\sqrt{n}$. The estimates from the exponential model are the same as those from the proportional hazards model when the data are exponential, which is why the exponential distribution is often used for sample size calculations even when the data are to be analyzed with a proportional hazards model or a log-rank test.[3,4]

Kalbfliesch and Prentice[5] provide more details on parametric methods.

A brief discussion of other issues related to survival analysis is presented in the next section.

Additional Topics

Competing Risks

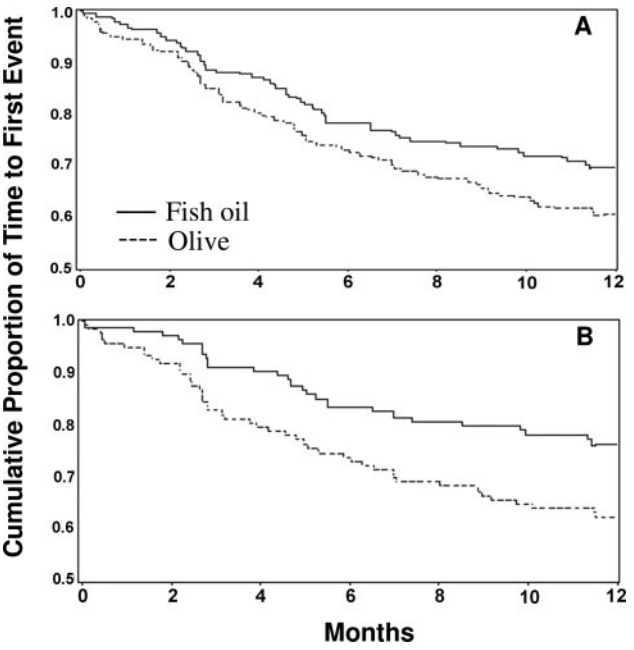
The concept of competing risks is another important issue to consider when survival over time is studied. For example, subjects in a cohort might be at risk of cardiovascular (other than myocardial infarction) mortality or dying because of myocardial infarction. The analysis approaches involve either computing the all-cause hazards, in which all events are taken into account, eg, cardiovascular mortality or death due to myocardial infarction (whichever occurs first is the outcome), and the cause-specific hazards, in which only the time to the event of interest is observed, and the times to other events are censored. For example, if cardiovascular mortality (other than a myocardial infarction) is of interest, then only time to this event is observed, and subjects who die due to myocardial infarction are censored. Plots of cumulative hazard functions are generally preferred when cause-specific hazards are used, and tests based on cumulative incidence have been developed because the Kaplan-Meier curves and log-rank tests may give biased results. For more information, see References by Gail through Allison.[6–10]

Interval Censoring

It is important to identify the time origin in a survival analysis. The usual time origin is the entry into the study. In that case, at time 0 the value of n_0 is the total number of patients. However, many survival techniques will also work without this restriction. For instance, one can analyze age at death among nursing home patients using survival theory. In that case, the time variable is age, and n_j will not decrease but will vary as people of different ages enter the nursing home. This is known as truncation. For those interested in this type of survival analysis, see Hyde[11] and Turnbull.[12] There are also a wealth of techniques for the situation in which the event time is only known up to an interval; for instance, if the event were the development of a condition that could only be diagnosed by an ultrasound or a laboratory test, all that one would know was that it occurred after the last negative test and before the next positive one. This is known as interval censoring.[13]

Time-Varying Covariates

The assumptions of proportional hazards may not hold for a given data set. The proportional hazards model can be made more general because one can add time-varying covariates to handle situations in which the hazard ratio is not constant over time or add interaction and quadratic terms when the hazard ratio is not constant over other covariates. For in-



Kaplan-Meier analyses of the time to ICD shock for ventricular tachycardia/ventricular fibrillation or death from any cause. A is based on the intention to treat. B is limited to only those compliant for at least 11 months. n at baseline=200 (fish oil)/202 (placebo); n at risk in month 4=165 (fish oil)/155(placebo); n at risk in month 8=132 (fish oil)/129(placebo).

stance, the covariate t×Z would model a situation in which the hazard ratio increases linearly when Z=1 compared with Z=0. Thus, issues of whether the model "fits" the data are actually issues about whether the model is correctly specified. Tests of fit are described by Schoenfeld,[3,14] and Wei.[15] and the consequences of misspecification on testing are described by Lagakos and Schoenfeld[16] and Gail et al.[17]

If the model is used for testing, the conclusion is that misspecification is not a great problem.[16] However, estimates are biased and represent a weighted average of the hazard over the duration of the study.

One of the advantages of covariate adjustment is that it can help to ameliorate the effects of informative censoring. In an analysis with covariates, the censoring distribution can depend on the covariates. Informative censoring in this case would be a dependency between the event time and the censoring time for patients with the same value of all covariates included in the model. Thus, for instance, if a covariate affected both the censoring time and the event time, then the censoring would be informative in a model without the covariate but noninformative in a model with the covariate.

Power Calculations

Power calculations are useful for designing studies. To calculate the power for survival analysis, one needs to know the total number of subjects in the trial, accrual time (time during which subjects are recruited in the study), failure time (time at which an event or death occurs), median survival ratio (or hazard ratio) or the minimum detectable hazard, and the significance level.[3,4,13,18] Standard software packages like SAS, SPSS, S-PLUS, and STATA have procedures that are simple to use for all methods described in this article.

Example

As an example of how these methods are used in practice, consider the report "Prevention of Fatal Arrhythmias in High-Risk Subjects by Fish Oil n-3 Fatty Acid Intake," which recently appeared in this journal.[19] We briefly summarize the uses of survival methods in that report.

The aim of the study was to evaluate whether n-3 fatty acids prevent potentially fatal ventricular arrhythmias in high-risk patients. A total of 402 patients with implantable cardioverter-defibrillators (ICDs) were randomly assigned to double-blind treatment with either a fish oil or olive oil daily supplement for 12 months. The primary end point was time to first ICD event for ventricular tachycardia or fibrillation confirmed by stored electrograms or death from any cause. Analyses were performed both according to the intention to treat and according to actual treatment. All randomized

Comparison of Baseline Characteristics of All Enrollees for the Placebo and Fish Oil Treatment Arms in the Fatty Acid Antiarrhythmia Trial: Analysis of Time to First Event

	Hazard Ratio	95% CI	*P*
Intention-to-treat analysis (n=402)			
Unadjusted			
Confirmed events	0.72	0.51–1.01	0.057
Including probable events	0.69	0.49–0.97	0.033
Multivariable analysis*			
Confirmed events	0.67	0.47–0.95	0.024
Including probable events	0.66	0.46–0.92	0.016
On-treatment analysis for all on treatment† (n=402)			
Controlling for baseline left ventricular ejection fraction			
Confirmed events	0.73	0.50–1.07	0.110
Including probable events	0.62	0.48–1.02	0.062
Multivariable analysis*			
Confirmed events	0.67	0.46–0.98	0.037
Including probable events	0.65	0.45–0.95	0.026
On-treatment analysis for at least 11 months‡ (n=236)			
Confirmed events	0.62	0.39–0.97	0.034
Including probable events	0.62	0.40–0.96	0.029
Multivariable analysis*			
Confirmed events	0.52	0.32–0.83	0.0060
Including probable events	0.53	0.34–0.84	0.0070

*Multivariable model controlled for gender, left ventricular ejection fraction (continuous), New York Heart Association class III congestive heart failure, history of myocardial infarction, history of prior defibrillator therapies for ventricular tachycardia/ventricular fibrillation, time from ICD implant (continuous), and sustained ventricular tachycardia as the indication for the ICD (all measured at baseline).

†On-treatment analysis for all subjects who had taken any of their prescribed oil supplements; the follow-up was censored at 2 months after medication was stopped.

‡On-treatment analysis only for those subjects who were on treatment at least 11 months.

subjects in this study were included in the intention-to-treat analysis. The primary analysis, based on confirmed events, was an intention-to-treat analysis of the survival free of appropriate ICD events for ventricular tachycardia/ventricular fibrillation and/or death from any cause, which included all ICD events that occurred during the 12-month period after the first dose of the study drug, irrespective of the duration of treatment. An "on-treatment" analysis was done that included all ICD events that occurred no later than 2 months after treatment was stopped. In this analysis, the date of cessation of treatment plus 2 months was used as the censoring variable. To ensure that the time to event was independent of time to noncompliance, conditional on covariates in the model, the authors tested for associations between baseline variables and time to noncompliance and used any that were significant as covariates in this analysis.[20]

Time to first event analysis was calculated by the Kaplan-Meier method, and survival time across the 2 groups was compared with log-rank tests. Cox proportional hazards models were also performed to calculate hazard ratios and to adjust for clinical covariates that were associated with noncompliance in the on-treatment analysis and with the primary end point in the multivariable analysis.

The survival plots displayed in the Figure and the results of the analysis displayed in the Table were published previously in *Circulation*.[19]

Time to First Event Analyses

In the primary analysis, according to the intention-to-treat principle, there was a trend toward a longer time to first ICD event for ventricular tachycardia/ventricular fibrillation confirmed by electrograms or death from any cause among patients randomized to fish oil compared with those randomized to the olive oil placebo ($P=0.057$). According to the KM estimates (Figure), 28% of patients in the fish oil arm (n=57) and 39% of patients in the olive oil arm (n=78) had reached the primary end point at 12 months. This difference corresponds to a hazard ratio of 0.72. The multivariable analysis that controlled for baseline clinical characteristics resulted in a hazard ratio of 0.67, 95% confidence limits of 0.47 to 0.95, and significance of $P=0.024$ (Table).

In the "on-treatment" analysis of "confirmed" events, which included all who had taken any prescribed oil supplements during the 12-month period, the hazard ratio was 0.73, which was not significant ($P=0.11$). This analysis controlled for baseline left ventricular ejection fraction, which was the only variable affecting time to noncompliance. After the investigators controlled for more baseline variables, the reduction in risk associated with use of fish oil became significant (hazard ratio, 0.67; $P=0.037$) (multivariable analysis, Table).

It is interesting that the P value decreases in the multivariable analysis compared with the analysis that only considers treatment. This is not necessarily due to confounding, which was not large in this study. The effect is rather due to the fact that a multivariable model that controls for factors that affect the time to event increases the power to see a treatment effect when the covariates are equally distributed between the treatment groups, which is usually the case in clinical trials. This effect is quantified in Schoenfeld et al.[21] When a study is designed, it is often a conundrum to decide whether the covariate analysis should be a primary or secondary analysis, but the decision regarding the primary analysis and the covariates to be used in the analysis should be made before data collection for a confirmatory type of study. It has more power at the cost of complexity.

Sources of Funding

This study was supported by research grant CA-74302 to Dr Schoenfeld.

Disclosures

None.

References

1. Scharfstein DO, Robins JM. Estimation of the failure time distribution in the presence of informative censoring. *Biometrika*. 2002;89:617–634.
2. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53:457–481.
3. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics*. 1983;39:499–503.
4. Schoenfeld DA. Find statistical considerations for a study where the outcome is a time to failure. Available at: http://biostatistics.mgh.harvard.edu/sample_size/quan_measur/para_time.html. 2001. Accessed December 13, 2006.
5. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. New York, NY: Wiley; 1980.
6. Gail M. A review and critique of some models used in competing risk analysis. *Biometrics*. 1975;31:209–222.
7. Gail M. Competing risks. In: Kotz S, Johnson NL, eds. *Encyclopedia of Statistical Sciences*. Vol 2. New York, NY: Wiley; 1982:75–81.
8. David HA, Moeschberger ML. Life tests under competing causes of failure and the theory of competing risks. *Bull Int Statist Inst*. 1969;43:267–269.
9. Moeschberger ML, David HA. Life tests under competing causes of failure and the theory of competing risks. *Biometrics*. 1971;27:909–933.
10. Allison PD. *Survival Analysis Using the SAS System: A Practical Guide*. Cary, NC: SAS Institute; 1995.
11. Hyde J. Testing survival under right censoring and left truncation. *Biometrika*. 1977;64:225–230.
12. Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. *JRSS, Series B: Methodological*. 1976;38:290–295.
13. Finkelstein DM. A proportional hazards model for interval-censored failure time data. *Biometrics*. 1986;42:845–854.
14. Schoenfeld DA. Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika*. 1980;67:145–153.
15. Wei LJ. Testing goodness of fit for proportional hazards model with censored observations. *J Am Stat Assoc*. 1984;79:649–652.
16. Lagakos SW, Schoenfeld DA. Properties of proportional-hazards score tests under misspecified regression models. *Biometrics*. 1984;40:1037–1048.
17. Gail MH, Wie S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*. 1984;71:431–444.
18. Schoenfeld DA, Richter J. Nomograms for calculating the number of patients needed for a clinical trial with survival as an endpoint. *Biometrics*. 1982;38:163–170.
19. Leaf A, Albert CM, Josephson M, Steinhaus D, Kluger J, Kang JX, Cox B, Zhang H, Schoenfeld DA, for the Fatty Acid Antiarrhythmia Trial Investigators. Prevention of fatal arrhythmias in high-risk subjects by fish oil n-3 fatty acid intake. *Circulation*. 2005;112:2762–2768.
20. Valerie M, Stanley KE, eds. *Statistics in Medical Research: Methods and Issues, With Applications in Cancer Research*. New York, NY: Wiley; 1982.
21. Schoenfeld DA, Borenstein M. Calculating the power or sample size for the logistic and proportional hazards models. *J Stat Computation Simulation*. 2005;75:771–785.

KEY WORDS: proportional hazards models ■ Kaplan-Meiers estimate ■ survival