JOURNAL OF CLINICAL RESEARCH BEST PRACTICES

Vol. 7, No. 6, June 2011 "Can You Handle the Truth?"

Making Sense of Biostatistics: Survival Analysis

By William Irish

Survival analysis is a special branch of statistics that encompasses a wide variety of methods for analyzing the timing of events. The prototypical event is death, which accounts for the name given to these methods. Survival analysis is also appropriate for many other kinds of events, such as divorce, remarriage, unemployment, detection of a tumor, and recurrence of disease after treatment.

Events are generally referred to as "failures," although the event may not have any negative connotations associated with it. For example, a counterintuitive event could be *stopping* of dialysis treatment in a diabetes trial.

Subjects are followed beginning at a certain starting point until the time at which the event of interest occurs. Because of time limits and restrictions on data collection, there is a good chance that not all the events can be observed for the subjects in a study. For example, in a heart disease study, some subjects will live past the end date of the study. As a result, the exact event times of these individuals are unknown. The end-date of the study has to be used as the "lower bound" of their time to death. For them, information on survival is said to be partial, incomplete or "censored."

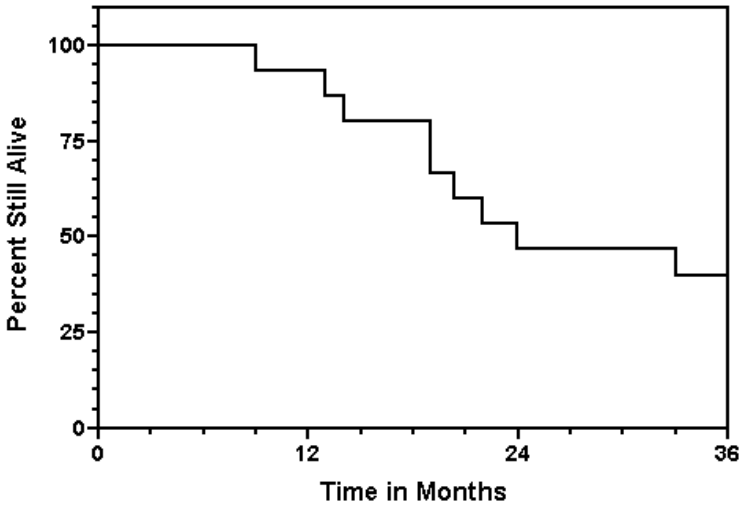
The process by which data can be incomplete is usually termed "censoring," as if a higher power blacked out the data. Censoring can arise in a variety of ways and creates interesting problems. For example, some subjects will survive to the end of a clinical trial. Others will withdraw from the study. In each case, the subject is removed from observation but is still at risk of failure — a subject does not cease to run the risk of heart disease simply because he or she has ceased to participate in a study. Such incomplete observation of the failure time is called right censoring (because time is plotted from left to right) and is common in prospective studies.

When individuals are removed from observation because they die from causes other than the one of interest, a special kind of right censoring occurs. This is known as "censoring by competing risks." For example, in a study of survival focusing on death from heart disease, subjects who die from lung cancer are considered to be right censored. This type of censoring is different from the one discussed previously because when a subject is removed, he or she is no longer at risk from the failure of interest.

The mathematical representation of failure times is of major importance to researchers, despite the complications posed by censoring. The most common statistic is the survival distribution, denoted by S(t). S(t) is defined as the probability that an individual will survive longer than time t. For example, S(12 months) = 0.50 means that an individual has a 50% chance of surviving more than 12 months. The properties of S(t) are such that the probability of surviving at time t=0 is 100% and of surviving an infinite time is zero. In other words, the probability that a subject is alive at the start of the study is one and, if followed forever, the probability that he or she will be alive is zero. The graphical representation of S(t) is often called the "survival curve."

The survival curve is usually estimated using the Kaplan–Meier estimator, also known as the product limit estimator. The Kaplan-Meier curve represents the proportion of the study population still surviving (or free of failure) at successive times. A Kaplan–Meier survival curve consists of a series of horizontal steps of declining height which, when a large enough sample is taken, approaches the true survival function for that study population. An example of a simple survival curve is presented in Figure 1. Fifteen subjects were followed for 36 months and nine of them died. At 24 months, the expected survival is 40% (6 out of 15).

Figure 1. A Simple Kaplan-Meier Survival Curve



The life-table method is another way to estimate the survival curve. This method is used when failure, such as death, is presented in calendar time units (quarterly, semi-annually, etc.). For example, when analyzing patient survival post-treatment in a cancer study, death may be known only to have occurred in the Month 6 to Month 12 follow-up time window, but the exact date of death is unknown. The life-table method can be thought of as the group-data analog to the Kaplan-Meier estimator. And like the Kaplan-Meier estimator, the life-table method incorporates all survival information accumulated up to the termination of the study. For example, in computing a five-year survival rate for breast cancer patients, one need not restrict oneself only to those subjects who have participated in a study for five or more years. Subjects who have entered for four, three, two or even one year contribute useful information to the evaluation of five-year survival. In this way, the life-table technique uses incomplete data like individuals withdrawn alive, as well as the death data.

The life-table method is often used to estimate the survival curve when the volume of data is quite extensive (e.g., tens of thousands of subjects). In these situations, it can be quite time consuming to tabulate and graph the Kaplan-Meier estimator of the survival curve, even when using statistical software packages such as SAS (SAS Institute, Cary, NC). When the volume of data is quite extensive, the survival times are grouped into intervals (e.g., three-month periods) and the life-table method used.

In most studies, the investigator is interested in estimating the survival curve, not just for one group of subjects but for subgroups in, for example, each arm of the study. Using the Kaplan-Meier method or the life-table method, the survival curve can be estimated for each group separately. For example, in a study of immunosuppressive drugs for prevention of acute rejection post-kidney transplantation, we can estimate the rejection-free survival curve for subjects taking one type of drug and the corresponding rejection-free survival curve for subjects taking a different type of drug. Or, the subjects can be grouped by age, gender, etc. Almost invariably, the survival times of the different groups will vary. These differences can be visualized by drawing multiple survival curves on the same chart. The three-year survival probability might be identical for two groups, but the shape of the curves might be quite different. However, this method gives only a rough idea of the difference between the distributions. It does not reveal whether the differences are statistically significant or merely chance variations.

Three commonly used statistical tests for comparing survival curves are the log-rank test, the Breslow test, and the Tarone-Ware test. All three tests are designed to handle censored

data and are based on computing the weighted difference between the observed and expected number of failures at each of the time points. Each test uses different weights for the failure times. The most common test is the log-rank test, which weights all failures equally (i.e., all the weights are equal to one). A significant p-value (i.e., p-value < 0.05) based on the log-rank test implies that the two survival curves are statistically different and, therefore, represent two different survival distributions.

Author

William Irish, PhD, is Vice President of Outcomes Research and Biostatistics at CTI Clinical Trial and Consulting Services. Contact him at 513.618.4057 or birish@ctifacts.com.