Analyzing outcomes measured as a "time to event" including survival

Statistical concepts for clinical investigators

David M. Thompson
Training Unit
Oklahoma Shared Translational and Clinical Resource
January 5, 2015

OKLAHOMA
Clinical and Translational Science Institute
Located at The University of Oklahoma Health Sciences Center

Time to event outcomes include:

Overall or progression free survival

the time required for a patient with GB to improve by one grade on the Hughes scale

the time from diagnosis to resumption of independent ambulation

the time from initiation of study to the first date of a registered diagnosis of dementia

Censored observations

If every subject experiences the event of interest, groups' mean "time to event" is comparable using t-tests or other simple statistics.

Often, many subjects don't experience the event. Investigators follow them for a period of time, but don't observe improvement, disease progression, or an endpoint of interest. These subjects' observations are "censored."

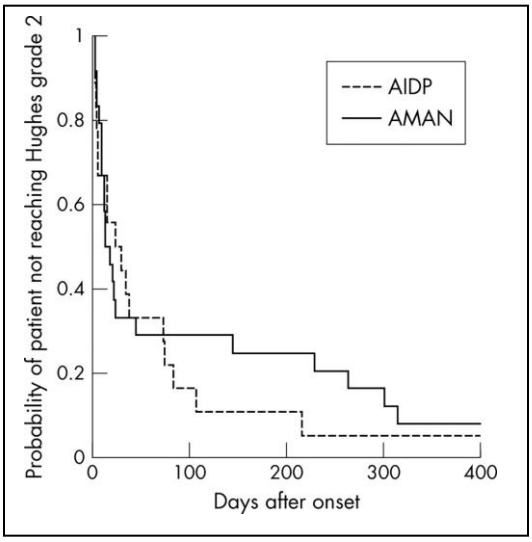
Censored information is incomplete but still very useful. We lose information and efficiency if we ignore the time period over which subjects are observed even if they never experience the event. So, special "time to event" techniques are used to analyze data when a subset of observations are censored.

Survival curves or survival functions

A Kaplan-Meier (KM) "curve" shows the calculated probability of experiencing the event at time t, given that one hasn't yet experienced the event at time t-1.

Because no subjects have yet experienced the event at time 0, survival curves begin at 1, signifying 100 percent "survival" and trend downward throughout the study period.

The "survival curve" may be a step function that traces horizontally during times between events, then traces vertically when an event occurs to one or more people in the sample.



Example of KM curves from Hiraga A et al. J Neurol Neurosurg Psychiatry 2005;76:719-722

Hazards and hazard rates

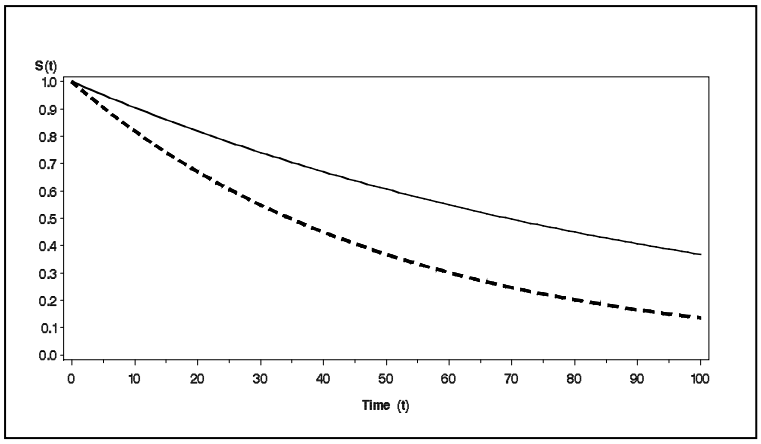
Widely used statistical approaches, including Cox regression, estimate the "hazard" of experiencing the event. Although some publications speak of "hazards" as if they are similar to "risks," hazards are not probabilities, and hazard ratios are not relative risks. Cumulative hazards have been likened to an actuarial concept, the "force of mortality," which reflect the number of events one might experience were the event repeatable (Clark et al., 2003). Spruance and colleagues (2004) note that a hazard ratio (HR) is equal to the *odds* that, between a randomly selected pair of subjects, one of whom was treated and the other of whom was not, that the treated member of the pair will experience the event before the untreated member. The *probability* that the treated member experiences the event first is equal to HR/(1+HR).

Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: Basic concepts and first analyses. Br J Cancer. 2003 Jul 21;89(2):232-8.

Spruance SL, Reid JE, Grace M, Samore M. Hazard ratio in clinical trials. Antimicrob Agents Chemother. 2004 Aug;48(8):2787-92.

Proportional hazards and hazard ratios

A widely used multivariable technique for time to event data, Cox regression, is founded on the assumption of proportional hazards. This assumption is met when groups' survival functions have a form, similar to those depicted to the right.



Survival functions with this quality have the additional property that their corresponding "cumulative hazard" functions are such that, at any given point in time, the ratio between the cumulative hazards is a constant quantity; at any given time, the hazards are proportional.

The assumption that cumulative hazards are proportional is a strong one whose viability investigators should check. However, if the complex survival functions estimated in Kaplan-Meier curves can be justifiably approximated by survival curves whose corresponding cumulative hazard functions are proportional, the simplification affords major advantages

First, we can use a single quantity, the hazard ratio, to compare groups.

Second, like all regression techniques, Cox or proportional hazards regression can adjust the estimate of the hazard ratio for other factors that might influence time to event, like sex, age or the existence of comorbidities.

However, if the proportional hazards assumption is not tenable, the Cox model's estimates of the hazard ratio are not trustworthy.

Inference for time to event data
(Hypothesis testing and estimation of confidence intervals)

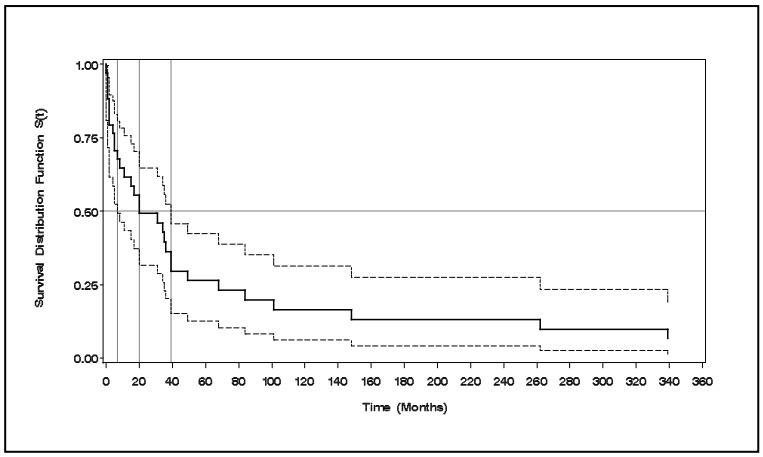
Estimating the population's true **survival function** on the basis of the observed samples

A standard estimator is the Kaplan-Meier (KM) product limit estimator, hence the common reference to Kaplan-Meier curves.

Kaplan EL, Meier P. Nonparametric estimation from incomplete observations.J Am Stat Assoc. 1958;53:457–481.

Placing a **confidence interval on the survival function**

Extracting features of the survival function, including **median time to event** (median survival time).



Comparing survival functions between "strata" or groups based on diagnosis or treatment.

The log rank test is the statistical test most frequently used to compare groups' survival functions. The log rank test evaluates the (null) hypothesis that the survival functions don't differ. If the log rank statistic generates a small p-value, we reject the null hypothesis; we conclude that the observed difference in survival functions is too large to have occurred due to natural sampling variability or "chance." Instead, we conclude that the survival functions truly differ between groups.

Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemotherapy Reports—Part 1. 1966; 50:163–170.

Estimating the **hazard ratio** on the basis of observed samples of subjects

Cox regression, often called "proportional hazards regression," generates estimates of the ratios of hazards between groups of subjects defined by, for example, risk factors or medical treatments.

Accounting for multiple sources of differences in survival or time to event

Kaplan-Meier estimates and log rank tests permits comparisons between survival functions among strata or levels defined by just one variable, such as a treatment group.
The KM estimator and log rank test cannot simultaneously compare functions among groups defined by more than one variable, nor can they adjust estimates for confounders. To do that, a multivariable regression approach, suitable for censored data, is required. The standard multivariable approach for time to event data is the **Cox regression model.**

Cox regression is also called "proportional hazards regression" because it relies on assumptions, described earlier, that relate to the shape of the groups' survival functions. Survival functions that meet these assumption will diverge steadily, so investigators can check the assumption of proportional hazards by examining the shape of the observed Kaplan-Meier curves.

Additionally, some studies will report examining a "log minus log plot" -- a plot of log(-log(survival) versus log(time) -- which produces parallel lines if the assumption of proportional hazards is met.

Classic reference

> Cox DR. Regression models and life tables. J R Stat Soc [B]. 1972;34:187–220.

Predictive power of Cox regression models
Most regression techniques feature a statistic that describes a model's "predictive power," the extent to which its predictions fit the data that were actually observed. The R2 statistic is a familiar measure of predictive power that accompanies many reports of linear regression models. The "Harrell's c statistic," which can take values from roughly 0.5 to 1, can perform this function in Cox (proportional hazards) regression and for logistic regression models.

> In general, c statistic reflect the proportion of instances in which we can expect the model to accurately predict the outcome when we apply it to individuals with real data.
>
> Harrell's c statistic represents the proportion of pairs of individuals, randomly drawn from the population such that both experienced the outcome (neither were censored), where the statistical model would correctly identify the member of the pair who actually survived longer.

Competing risks

In studies of time to event data, observations of the event might be "censored" because an individual leaves the study, is lost to follow up, or simply has not experienced the event at the study's end. Common analytic approaches, like Kaplan-Meier estimator of median survival time, or the log rank test, assume that censoring does not differ between groups.

However, certain outcomes are termed "competing risks" because they can alter the chance of observing the outcome of primary interest. Moreover, their effect on the chance of observing the primary outcome may differ among the study groups or strata. For example, death from another cause can be a competing risk when it occurs prior to and so prevent the observation of the event of interest.

Observations that involve events due to competing risks should not be treated as censored with respect to the event of interest. Doing so can bias the usual estimates of the survival function, like the Kaplan-Meier estimator. When competing risks are present, the validity of a log rank test, which compares groups on their time-to-event functions, are not valid.

Special techniques exist to analyze time to event data when competing risks are present. These typically compare "cumulative incidence" to estimate between-group time to event so that competing risks are appropriately accounted for. Gray's test is a standard alternative to the log rank test for this situation. It generates a chi-square statistic that tests the null hypothesis that the groups' cumulative incidence functions do not differ. If the chi-square statistic is large, it will be associated with a small p-value, leading to the conclusion that the groups differ with respect to the incidence of the event of interest.

Classic references

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

Gray RJ. A class of k-sample tests for comparing the cumulative incidence of a competing risk. Ann Stat. 1988;16:1141-54.

Pintilie M. [An introduction to competing risks analysis]. Revista Espanola Cardiologia. 2011 Jul;64(7):599-605. doi: 10.1016/j.recesp.2011.03.017. Epub 2011 May 31.[Article in Spanish; English translation downloaded October 26, 2013 from http://www.revespcardiol.org/en/an-introduction-to-competing-risks/articulo/90023477/]

Software

Gray RJ. <gray@jimmy.harvard.edu> (2011). cmprsk: Subdistribution Analysis of Competing Risks. R package version 2.2-2.
http://CRAN.R-project.org/package=cmprsk
http://cran.r-project.org/web/packages/cmprsk/