EXPLORATORY PLOTTING OF CENSORED SURVIVAL DATA
AGAINST A TIME-DEPENDENT COVARIATE

Paul R Rosenbaum
Departments of Human Oncology and Statistics
University of Wisconsin-Madison

Douglass C Tormey
Departments of Human Oncology and Medicine
University of Wisconsin-Madison

This work was supported in part by grant P30-CA-14520 from the National Cancer Institute to the Wisconsin Clinical Cancer Center.
Summary

A method for plotting censored survival data against a time-dependent covariate is proposed and illustrated using serial data on a breast cancer marker.
1. The At-Risk Boxplot

Several methods exist for plotting censored survival data, including probability plots (e.g. Nelson (1974), Cox (1979), Kalbfleish and Prentice (1980, pp. 22-24, 42-43)), residual plots (e.g. Cox and Snell (1968), Kay (1977), also the "residual score" plots of Lagokos (1980)), and plots to examine proportionality of hazards in several groups (Kalbfleish and Prentice (1980, p. 91-95). However, there is no generally used method for directly plotting censored survival data against a time-dependent covariate.

The at-risk boxplot is motivated by the partial likelihood associated with the proportional hazards model (Cox (1972), Cox (1975)). Suppose n patients are under study, and that patient j has covariate value \( z_j(t) \) at time t. The Cox partial likelihood is made up of one contribution from each death: if patient i dies at time \( t_i \), then the likelihood compares the covariate value \( z_i(t_i) \) for the deceased patient with the covariate values \( z_j(t_i) \) at time \( t_i \) of all patients who are at risk at time \( t_i \), that is, all patients who were alive and uncensored immediately before \( t_i \). In parallel, at each death time \( t_i \), the at-risk boxplot compares the covariate value \( z_i(t_i) \) of the deceased patient with the covariate values \( z_j(t_i) \) at time \( t_i \) of all patients at risk at time \( t_i \) excluding the deceased patient. At each death time \( t_i \), a boxplot
(Tukey (1977)) is constructed displaying the current covariate values \( z_j(t_i) \) for patients alive and at risk at time \( t_i \). The covariate value \( z_i(t_i) \) for the deceased patient is added to boxplot by a suitable symbol, such as a triangle. Following McGill, Tukey and Larsen (1978), the width of the boxplot is taken to be proportional to the square root of the number of patients currently at risk.

The at-risk plot serves several purposes: first, it displays systematic relationships that exist between survival and covariate values; second, it displays deviations from the systematic relationships; third, it identifies units with unusual covariate values that may have a strong influence on model fitting; finally, it can help to identify data errors.
2. A Cancer Marker Example

Several investigations (e.g. Powles, et. al. (1976)) have suggested that the ratio of hydroxypoline to creatine (OHP/CR) in urine anticipates the recurrence of breast cancer. Here, the association between the time-dependent covariate OHP/CR and time to failure is studied in 47 nonmetastatic breast cancer patients treated at the Wisconsin Clinical Cancer Center.

Figure 1 is an at-risk boxplot of time to failure and current OHP/CR values. Four patients failed; the failures occurred at 172, 364, 405, and 541 days after surgery. At these four times there were, respectively, 19, 16, 8; and 2 patients at risk who did not fail and had OHP/CR measured within 20 days of the failure; these current OHP/CR values are represented by the boxplot. The covariate value \( z_\parallel(t_\parallel) \) for the patient who failed is added to the boxplot as a triangle.

The plot leads to the following observations. The patient who failed at 172 days after surgery had, at that time, an OHP/CR value higher than all but one of the 19 patients then at risk; this isolated observation is consistent with the hypothesis that OHP/CR is an effective marker. However, the other three failures had lower OHP/CR at their times of failure than most patients then at risk, which sheds some doubt on the stated hypothesis. One OHP/CR value at 172 days is seen to be atypically high, and may have high leverage if a Cox model is fitted by maximum partial likelihood.
3. Variations on the At-Risk Boxplot

Several variations on the at-risk boxplot are useful, including the following:

(1) If deaths are numerous, it will be impossible to display one boxplot for each death. There are two alternatives:
   (a) Divide the time axis into short intervals; construct one boxplot for each interval; use one triangle for each death. If deaths are very numerous, the single boxplot can be replaced by two contiguous plots, one for surviving patients and one for the deceased patients.
   (b) Boxplots can be replaced by narrower stickplots, that is, the box portion of the boxplot can be replaced by a single vertical line with short horizontal bars at the quartiles.

(2) Instead of plotting current covariate values, $z_j(t_1)$, other functions may be plotted, such as covariate values lagged by two weeks; i.e., $z_j(t_1 - 14)$. 
References


Biometrika 66:188-190.


Figure 1: Plot of Time to Failure Against Current OHP/CR

Δ = Failure