

## EXPLORATORY METHODS IN SURVIVAL ANALYSIS

J. CROWLEY, M. LEBLANC, R. GENTLEMAN, AND S. SALMON

**ABSTRACT.** Despite considerable research in the past two decades, much of the practice of survival analysis retains a black box flavor. In this paper we review some of the available exploratory methods for survival data and indicate some directions for future research. Methods discussed include box plots, running median plots, nonparametric estimation of the Cox regression function, and tree-based approaches to forming prognostic groups and building regression functions. The methods are applied to data on patients with multiple myeloma treated on clinical trials conducted by the (US) Southwest Oncology Group, a multi-institutional organization dedicated to finding cures for cancer.

### 1. INTRODUCTION AND NOTATION

Exploratory methods for survival analysis remain underutilized despite having received increasing attention from methodologists over the past decade. There are two aspects of survival analysis which are particularly responsible for the lag in the development and use of exploratory methods. The first is the fact that in most applications most survival data are subject to censoring. This means that for some individuals only partial information is available on their survival time, making many ordinary plotting methods less informative than they would otherwise be. The second reason is that the regression model most used in survival analysis, the Cox proportional hazards model, does not lend itself easily to pictorial representations of the data.

By exploratory methods we shall refer to those methods that attempt to describe the relationship between the response and the covariates of interest, putting very few restrictions upon that relationship. The emphasis of such methods is to use the available data to derive some idea as to the relationships that exist, rather than to test hypotheses that certain relationships hold. These methods are particularly well suited to the situation in which the data of interest have been collected and analyzed for some specific purpose but are still available to explore other questions of interest. The relationships would then be used to suggest directions for further research.

Interest will be focused in two areas, graphical methods and recursive partitioning. The methods will be illustrated using data on patients with multiple myeloma, a malignancy affecting the plasma cells of the blood.

---

1991 *Mathematics Subject Classification.* Primary 62G07; secondary 62P10 .

*Key words and phrases.* Survival analysis, exploratory methods, graphics, recursive partitioning .