**University of
Zurich**ᵁᶻᴴ

Competing Risks Survival Analysis

Kelly Reeve, MSc (*kelly.reeve@uzh.ch*)
PD Ulrike Held (*ulrike.held@uzh.ch*)

August 31, 2018

1   Introduction

Survival analysis is a set of statistical techniques to describe and quantify time to event data. This data comprises the time it takes an individual to reach an endpoint. Although the term "survival data" may have a negative connotation, the end point of survival analysis does not have to be a negative event (e.g., death or disease progression). It can also be a positive event, like recovery [Sedgwick, 2010].

Competing risks often occur in the analysis of survival data [Austin et al., 2016], as other events may prevent the outcome of interest. Multiple outcomes may also be of interest. Examples include:

1. death prevents observation of time until relapse;

2. hospital discharge prevents observation of time until hospital-acquired infection;

3. primary interest in time until recovery but more than 10% of the patients die before recovery;

4. time to local reccurrence or time to metastasis are both of interest.

In all of these situations, competing risks are present. Competing risks are defined as events whose occurrence precludes the occurrence of the primary event of interest. In the first example, the researcher is interested in some event, but death may occur first in many patients, preventing the event of interest from happening to these patients. The second example shows that the competing risk does not have to be fatal. Here, leaving the hospital healthy prevents the researcher from observing hospital-born infections in these patients. The third example highlights a more subtle issue. The researcher may be interested in some event, but if enough events of some other type occur that prevent this from happening, competing risk methods must be considered. In the last example, more than one time-to-event outcome is of interest. This can also be the case when researchers have already studied a composite outcome such as disease-free survival, but would like to do a more detailed analysis of some covariate on both time until reccurrence and time until death.

In the above examples patients start in one state (e.g., cancer-free) and have the ability to move into one of several competing states (e.g., relapse, death by cancer, non-cancer death). It is important to note that here we **only focus on time until the first event**, otherwise even more specialized models would be required (multi-state models).
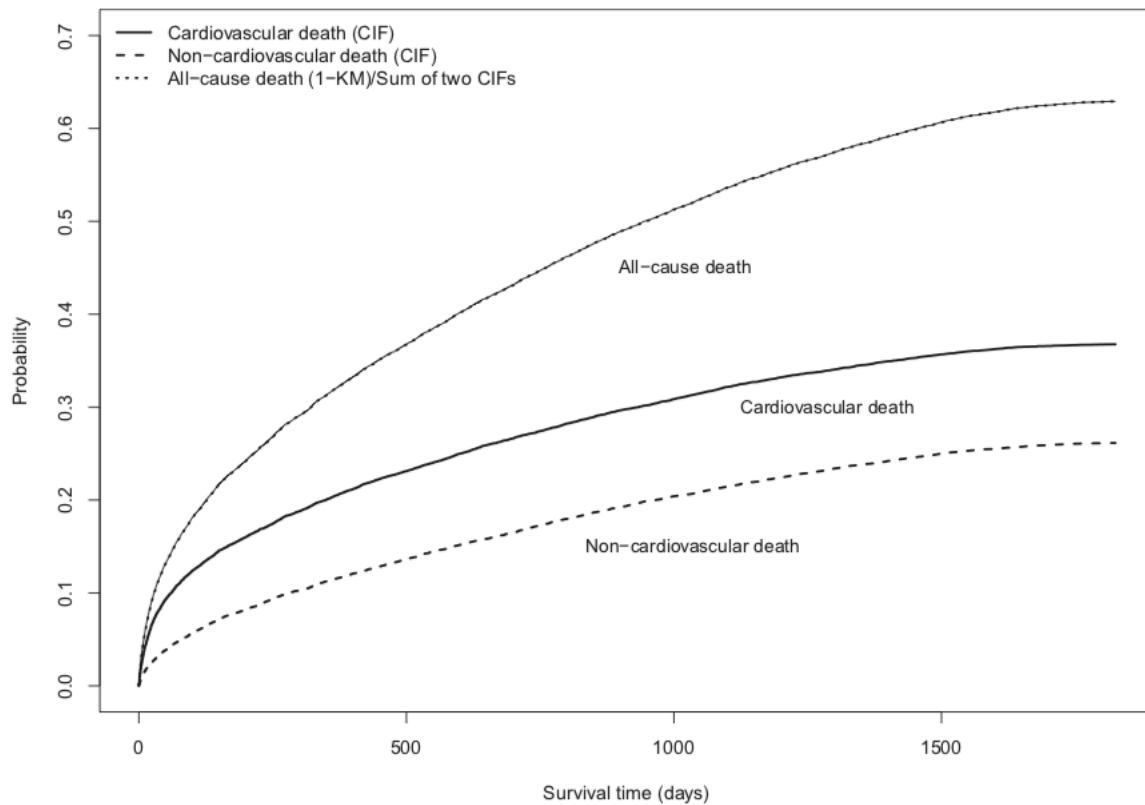
Figure 1: Cumulative incidence functions for cardiovascular, non-cardiovascular death, and all-cause death in the EFFECT study. This image comes from Austin et al. [2016].

2   Data Visualization

In a standard survival analysis, we would first look at Kaplan Meier plots. However, in the presence of competing risks, Kaplan Meier plots will be biased and misleading, as they underestimate the survival probability for the event of interest.

Instead, cumulative incidence functions (CIFs) are to be estimated and plotted. While Kaplan Meier plots are of survival (1-KM gives incidence when no competing risks are present), CIF plots, as implied by the name, are of incidence and therefore they increase over time instead of decrease. The CIF for a single event of interest takes overall survival into account (not just the survival due to that event) and so CIF plots are not biased like the Kaplan Meier plots. CIF plots for all competing risks should be included in the report for such an analysis (see Figure 1 for an example).

3   Covariate Effects

Usually clinicians are interested in the effects of covariates on the time to some defined event. In the standard survival analysis, Cox proportional hazard models would be used. In the presence of competing risks, the model must be modified. There are two different methods for this: cause-specific hazard regression and subdistribution hazard regression.

The cause-specific proportional hazards model treats patients experiencing competing events as censored. This may be more informative if the aim of the research is to study the etiology of disease. The cause-specific method leads to interpretations of how covariates effect the hazard rate of the event of
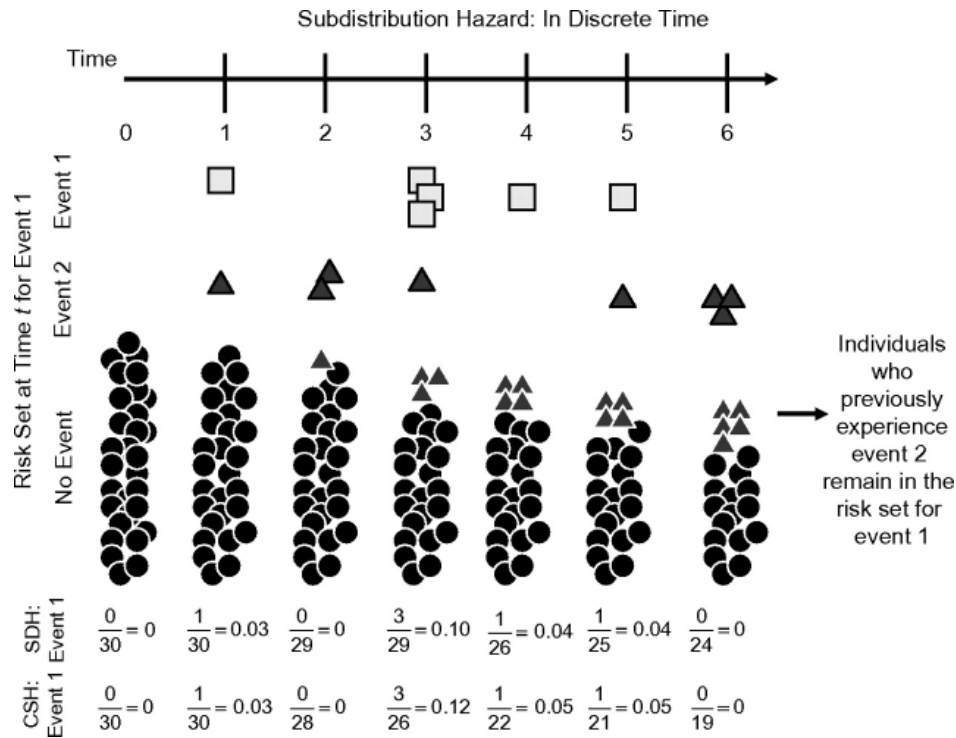
Figure 2: This image is meant to explain the difference in risk sets according to method - the sub-distribution (SDH) and the cause-specific (CSH) methods. The black circles represent everyone at risk at a given time. The black triangles and white squres represent patients experiencing event type 2 and event type 1 at a given time. Notice how the denominator differs over time for the different methods. This image comes from Lau et al. [2009].

interest.

The subdistribution proportional hazards model keeps patients experiencing competing events in the risk set for the event of interest (see Figure 2). This may be more useful if the aim is to study patient prognosis or to create prediction models. The subdistribution method allows one to make statements on how covariates effect incidence and to compare risks in different groups.

From the interpretation point of view, it is often beneficial to use both methods in order to get a better understanding. When reporting results, hazard rates should be accompanied by their confidence intervals. Interpretation of competing risk results can be quite tricky and requires patience and care. The cause specific method does not directly lead to interpretations of how covariates affect survival. In fact, covariates can have different effects on incidence and cause-specific hazards.

4   Software

Several R packages exist for the analysis of competing risks. The package survival (Therneau T (2015). *A Package for Survival Analysis in S*. version 2.38) can be used for survival analysis in general, and the cause-specific methods described here. The package cmprsk (Bob Gray (2014). *cmprsk: Subdistribution Analysis of Competing Risks*. R package version 2.2-7) can be used for the subdistribution methods described here. Of course, several other packages are also available.

5   References

If you would like to read more about competing risks analyses and to learn how such analyses are done in practice, please see Pintilie [2006]. This book is written from the practical perspective and illustrates several clinical science examples with R code and output. It also includes an introductory chapter to refamiliarize yourself with the basic concepts of standard survival analyses.

Peter C.. Austin, Douglas S. Lee, and Jason P. Fine. Introduction to the analysis of survival data in the prescence of competing risks. *Circulation*, 133(6):601–609, 2016.

J. P. Fine and R. J. Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509, 1999.

Bob Gray. *cmprsk: Subdistribution Analysis of Competing Risks*, 2014. URL https://CRAN.R-project.org/package=cmprsk. R package version 2.2-7.

Bryan Lau, Stephen R. Cole, and Stephen J. Gange. Competing risk regression models for epidemiologic data. *American Journal of Epidemiology*, 170(2), 2009.

Melania Pintilie. *Competing Risks: A Practical Perspective*. Statistics in Practice. Wiley, 2006.

R. L. Prentice, J. D. Kalbfleisch, A. V. Peterson, N. Flournoy, V. T. Farewell, and N. E. Breslow. The analysis of failure times in the presence of competing risks. *Biometrics*, 34(4):541–554, 1978.

H. Putter, M. Fiocco, and R. B. Geskus. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, 26(11):2389–2430, 2007.

Philip Sedgwick. Survival (time to event) data i. *BMJ*, 341:c3537, 2010.