

# Survival Analysis Assignment 1

## Introduction

The objective of this assignment is to get you acquainted with a more practical analysis of a survival data set, to interpret the results from the analysis and explain their findings to substantive researchers.

The assignment consists of a presentation and a report.

## Presentation

Students prepare a 10-15 minutes presentation from the analyses of the “exercises” below, and present it for an imaginary meeting with researchers working on this topic. Speaking time should be approximately equally divided between students in the same group. After the presentation, the other students (“the researchers”) can ask questions for clarification and provide comments. After the presentation students can use the feedback to finalize their report.

## Report

Students prepare a concise and self-contained report from this assignment, where at least the answers to the exercises are provided, and where results of the analysis are explained for an imaginary substantive researcher. Preferably, a single R markdown file (Rmd) may be prepared. Please send the report (pdf, Word or html) and Rmd file to [h.putter@lumc.nl](mailto:h.putter@lumc.nl).

## Presentation, report and grading

Please make sure that the presentation and report are appropriate for a scientific audience of substantive researchers. For instance, when presenting survival curves, choose an appropriate time range, provide meaningful text in x- and y-axes.

The presentation will count towards 15% and the report towards 10% of the total grade. The remaining 75% comes from the exam/resit result.

## UNOS data

Consider a cohort of 9775 children who have undergone kidney transplantation. The file “UNOS.txt” contains data from the United Network for Organ Sharing (UNOS); for details concerning the variables look at the file “informationUNOS.doc”.

**Exercise 1** — Illustrate in a table the characteristics of the population (age, sex, race, donor, ...).

**Exercise 2** — Plot the Kaplan-Meier overall survival curve for pediatric kidney transplant recipients for the first 12 years after transplantation.

**Exercise 3** — We are going to compare mortality rates (hazard functions) between children whose transplanted kidney was provided by a living donor (in general a family member) and those whose source was recently deceased (variable donor type: `txtype`). Use the life table method to calculate the death rates for the first 5 years for each group (take in the first year intervals of 4 months and then look at each year) and show the results in a table. Estimate the hazard ratio in each time interval as the ratio between the mortality rates in the two groups. What do you notice?

**Exercise 4** — Show a plot with Kaplan-Meier survival curves for the two donor types.

**Exercise 5** — Fit a univariate Cox model with predictor donor type. Report the hazard ratio and 95% confidence interval and interpret the result obtained.

**Exercise 6** — Research shows that an important determinant of mortality after kidney transplant is the age of the recipient. Fit a Cox model with `age` as predictor and estimate the hazard ratio and its confidence interval. Consider age first as continuous variable and then divide into categories.

**Exercise 7** — Fit a multivariate Cox model by using other predictors and describe your results.

**Exercise 8** — Estimate the survival function for specific covariate patterns. Based on the previous results choose the best predictors.

**Exercise 9** — Check the proportional hazards assumption. You may use the function `cox.zph`. Discuss the result and possible implications.

**Exercise 10** — Plot the Schoenfeld residuals and comment.