Introduction to Survival Analysis

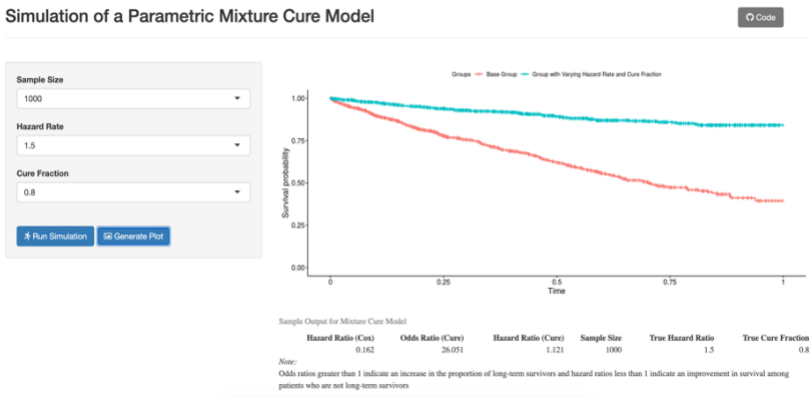
Final Project: Simulation of Mixture Cure Models

Howard Baek (Mentor: Subodh Selukar)

I reviewed Othus (2012) and learned that mixture cure models model the population for the possibility of 2 types of patients: cured and not cured. Cured patients will not fail or experience the event of interest such as cancer or HIV, during the study, whereas non cured patients will *eventually* experience the event of interest, but they can still be subject to administrative censoring. A cure model might be appropriate if the Kaplan-Meier plot has a long plateau at the right tail. A population with a cure fraction can violate certain assumptions in conventional survival analysis methods, including proportional hazards.

This final project is a simulation of a parametric mixture cure model. We generated two groups of data. In the first group, we fixed the base hazard rate to be 1 and base cure fraction to be 0. In a nested for loop, we varied the sample size (N = 10, 100, 1000), the hazard rate (λ = 1, 1.5, 2.5), and the cure fraction (0, 0.2, 0.5, 0.8) in the second group. Then, we used the inverse cdf method to generate times drawn from an exponential mixture cure distribution for two separate groups. By introducing censoring and finding the minimum of the event times and censoring times, we generated observed time and event status. This allowed us to model the data with a Cox proportional hazards model, from which we obtained a hazard ratio, and a mixture cure model, from which we extracted out the odds ratio and the hazard ratio.

When a dataset has a cure fraction, the Cox model is not correctly specified. The hazard ratio from the Cox proportional model is not giving a clear target because we have included a cure fraction in the data. When we have cure fraction in the data, we need to account for the heterogeneity between cured patients and non-cured patients, but the Cox model doesn't do this. Therefore, we might see the Cox model have unexpected power or type 1 error. Instead, the cure model is correctly specified: the uncured distribution is an exponential distribution with one parameter, the hazard rate. Since we are fitting an exponential mixture cure model on exponential mixture data, our exponential mixture cure model is correctly specified.



I've created a Shiny app that supplements this final project. Users can select different values of sample size, hazard rate, and cure fraction and view the corresponding output and plot.

Link: https://howiebaek-2.shinyapps.io/mixture-cure-model-sim/