Straphylocous Infection Project Proposal

Adam Sztonyk, Angel Alcantara Luna, Matthew Aydin

2025-05-01

The data that we will be using is the `burn` data frame from the `kmsurv` package in R. It has 154 rows (observations) and 17 columns. The dataset contains clinical data about burn patients, such as their age, gender, burn areas, and other important characteristics. The dataset has data about the time to excision, prophylactic antibiotic treatment, and staphylocous aureaus infection or when individual patients were censored. The scientific questions we are trying to answer are, "How do factors, such as age, race, gender and burn site indicators affect the risk and timing of straphylocous aureus infection? Is there multiple factors or only one that is significant?". The covariates in the data matrix are:

`Z1`: The type of treatment given (0 = routine bathing, 1 = body cleansing)

`Z2`: Gender of the burn patient (0 = male, 1 = female)

`Z3`: Race of the burn patient (0 = nonwhite, 1 = white)

`Z4`: Percentage of the total surface area burned on the patient's body

`Z5`: Burns in the head area of patient (0 = no, 1 = yes)

`Z6`: Burns in the buttock area of patient (0 = no, 1 = yes)

`Z7`: Burns in the trunk area of patient (0 = no, 1 = yes)

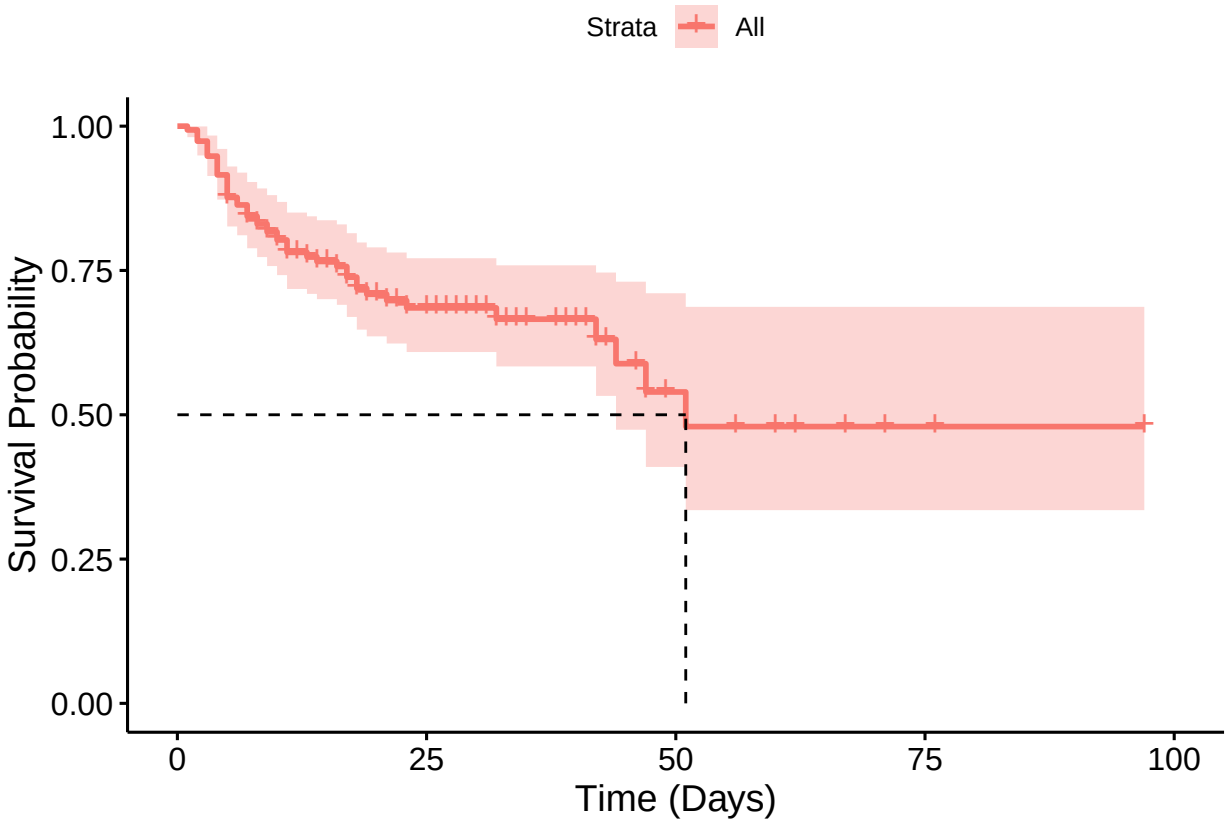
`Z8`: Burns in the upper leg area of patient (0 = no, 1 = yes)

`Z9`: Burns in the lower leg area of patient (0 = no, 1 = yes)

`Z10`: Burns in the respiratory tract area of patient (0 = no, 1 = yes)

`Z11`: The type of burn the patient suffered (1 = chemical, 2 = scald, 3 = electric, 4 = flame)

Kaplan Meier Estimate of Time to Staphylococcus Aureus Infection



Some statistical tools we can use are creating separate Kaplan-Meier curves for the time to Staphylococcus Aureus Infection based on the different factors in the dataset such as gender, race, where the burn is present, and the type of burn to see if we can identify any differences by looking at the Kaplan-Meier curve. However, some complications are that there are only 154 rows but around 5 different burn sites and additionally, there can be multiple burn sites at the same time which means that looking at the burn sites individually through Kaplan-Meier curves could not be accurate since they are not mutually exclusive. So, after the Kaplan-Meier curves we will use Cox Proportional Hazards models on the different factors to see which factors have a significant effect on the survival probability. We will also use the Cox Proportional Hazards model function coxph with multiple covariates to see if we notice if any factors are significant using the likelihood ratio test with the Cox model.

Bibliography

Klein and Moeschberger (1997) *Survival Analysis Techniques for Censored and Truncated Data*, Springer. Ichida et al. Stat. Med. 12 (1993): 301–310.