Hannah Chiu
Mentor: Nina Galanter
DRP Reflection: Introduction to Survival Analysis

This winter, I was lucky enough to participate in the SPA's Directed Reading Program, where I worked with my mentor, Nina Galanter, to learn about survival analysis. Through readings, I learned what survival analysis is and several methods that are used to analyze survival data in health. I then used what I learned to implement the final project, where I studied the time until death of breast cancer patients and presented my results to several other participants of the program.

Survival analysis is the collection of statistical procedures for the analysis of data—typically in the health sector—in which the outcome variable of interest is the time until an event occurs. In survival analysis, the initiating event could be the start of a study, or another event, like the diagnosis of a disease. The specified event—known as a failure—is usually death, but can also be another event, such as full recovery from a surgery. In addition, the data used in survival analysis is unique in that it can be censored. Censoring occurs when we have information about an individual's survival time, but don't know the exact survival time. Right censoring is the most common form of censoring; in this situation, it means that we know the specified event happened after some point in time. Examples of right censoring may include: a patient withdrawing from a study, being lost to follow-up, or not experiencing the specified event before the end of a study.

When conducting survival analysis, many models can be used to learn more about the data, but the main two that I learned about were Kaplan-Meier curves and the Cox Proportional Hazards (PH) model. The former is a popular way of visualizing probability of survival over time. Since Kaplan-Meier curves are non-parametric, the data doesn't need to meet any assumptions about the distribution of the data. Due to this, Kaplan-Meier curves are not smooth; instead, the curves are graphed as step functions. To see if there's a significant difference between two Kaplan-Meier curves, a log-rank test can be used, which results in a p-value. On the other hand, the Cox PH model can be used to simultaneously evaluate the effect of several factors on survival time. However, this model focuses on hazard, which is the risk of failure at any given time among people who survived until that point. The model has both parametric and non-parametric components, so one assumption it must meet is the proportional hazards (PH) assumption, which assumes that the hazard ratio is constant over time. The Cox PH model results in a hazard ratio: if the ratio is not 1, this suggests that the hazards differ between the two groups.

At the end of the quarter, I applied these concepts and models to analyze the time until death of breast cancer patients. Through my real data analysis, I found that there was an association between the type of treatment patients received and their survival time; specifically, patients treated with chemotherapy typically had a higher survival than those treated with hormone therapy. Overall, I am so grateful for this experience, as it was interesting to learn about an application of statistics in medicine that I probably wouldn't have been taught in class. I

finally want to say thank you to my mentor, Nina, for answering my questions about readings, clarifying concepts I didn’t understand, and helping me with creating my final presentation.