**Sample Size Simulation Study:**

**Clinical Trial for Acute**

**Graft-Versus-Host Disease**

Abstract

*Background:*

An allogeneic hematopoietic stem cell transplant is used to treat cancers of the blood. Acute graft-versus-host disease (aGVHD) is a side effect of the transplant and occurs in 40-50% of the patients who receive this transplant. A diet including an increase in potato starch, is thought to be preventative for aGVHD. A single-arm clinical trial is purposed to determine the effectiveness of the proposed diet on preventing aGVHD.

*Methods:*

In order to determine the sample size needed for the clinical trial a simulation study was used. Patients were simulated using binomial, truncated exponential, and truncated normal distributions. The log-rank test of survival was used to determine if the survival rate of simulated patients on the diet differs from the historical survival rate. This analysis was performed at different sample sizes to determine the lowest possible sample size with an 80% power. Both intention-to-treat and censored analyses were performed, with the option for an interim analysis. Three new rates for an aGVHD were used: 0.40, 0.35, and 0.30; compared to the historical rate of 0.50.

*Results*:

The intention-to-treat analysis without an interim analysis lead to the lowest sample size with 80% power (Rate-sample size: 0.30 – 41, 0.35 – 81, 0.40 – 191). To maintain 80%, power the sample size needed to be increased for censored data and for an interim analysis.

*Conclusions:*

Using an intention-to-treat analysis without and interim analysis and depending on the rate of interest (0.30, 0.35, 0.40), between 41 and 191 patients would need to be recruited for the diet high in potato starch to determine if there is a difference between the historical rate of aGVHD and the rate in dieted patients.

**Introduction**

Cancers of the blood are bad. There are many treatments, but if a patient is unresponsive to these they may be treated with an allogeneic hematopoietic stem cell transplant (alloHSCT). An alloHSCT is a treatment where they take healthy stem cells, from a genetically-matched donor, and place them in the patient. Matching a donor is not a perfect science. If the match is not ideal the immune system in the alloHSCT (graft) and the immune system of the patient start to fight each other. When this reaction occurs within the first 100 days of transplant it is called acute graft-versus-host disease (aGVHD). aGVHD is one of the leading causes of death in alloHSCT patients early on. Heavy doses of steroids are the common treatment for aGVHD, but they come with their own host of side-effects.

With new discoveries in the human microbiome, it is proposed that a diet high in potato starch could help prevent aGVHD in patients. The diet would consist of the patients ingesting 40g of starch a day for 100 days (the equivalent to 1.5 medium russet potatoes a day). To test the diet high in potato starch as a preventative treatment for aGVHD, a single-arm clinical trial is proposed. The historical rate of aGVHD in the alloHSCT patients is 0.50.

Other considerations for the clinical trial will include drop-out rates, and whether or not to include an interim analysis. Historically around 15% of alloHSCT patients will experience an event within 100 days and therefore will not be observable for an aGVHD (we are assuming dietary starch has no impact on these outcomes). It is predicted that approximately 20% of patients will discontinue the diet for a variety of reasons.

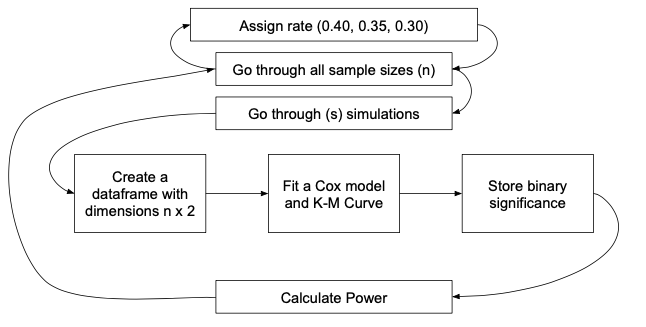
**Project Aims**

*Aim 1:*

How many patients need to be enrolled to reduce the 50% historical rate of aGVHD to each of 40%, 35%, and 30%? Assume power of 0.80 and Type I error rate of 0.05.

*Aim 2:*

Determine how the introduction of an interim analysis would affect the sample size calculations from Aim 1.

**Methods**

**Figure 1**. A diagram showing the simulation process.

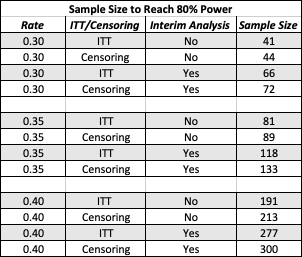
To answer the aims a simulation study was performed in R (v. 3.5.3). The first two steps were iterating through the rates (0.40, 0.35, 0.30) for different sample sizes. For each combination of rate and sample size a simulation was ran. A Bernoulli distribution with the given rate, was used to determine if a participant was going to have an aGVHD (event), the time of which would be determine by a truncated exponential distribution (see Appendix). If the patient had an event status of zero, they were assigned a time over 100 days (200 days for our purposes), so as not to influence the survival curve.

Next using the survival package in R (v. 2.44-1.1) a survival curve was fit and a confidence interval for the curve at time 100 was extracted. To maintain a Type I error rate of 0.05 and 94% confidence interval was used (see Appendix). If the lower bound of confidence interval was greater than the historical rate of 0.50, the significance of the simulation was set to one, and zero otherwise. This was then performed 1,000 times and the power for that rate and sample size was calculated by taking the sum of the significant simulations and dividing it by the total number of simulations. The sample size was divided by 0.85, to adjust for the 15% dropout expected.

The above analysis would be for an intention-to-treat (ITT) analysis. To censor those participants who did not comply with the diet, the analysis was changed slightly. For the censored analysis, two different event statuses and times were sampled. The first event status and time would be in reference to aGVHD and was done the same as before. Next an event two would be whether the participant was a non-complier. This event status was drawn from a Bernoulli distribution with rate 0.80, assigning an event status of zero to non-compliers. For compliers, there time was set to 200 days to not affect the analysis. For non-compliers, their time to event was drawn from a truncated normal distribution (see Appendix). The two event/time combinations were then compared and whichever event had the minimum time was used in the analysis. The rest of the analysis proceeded in the same fashion.

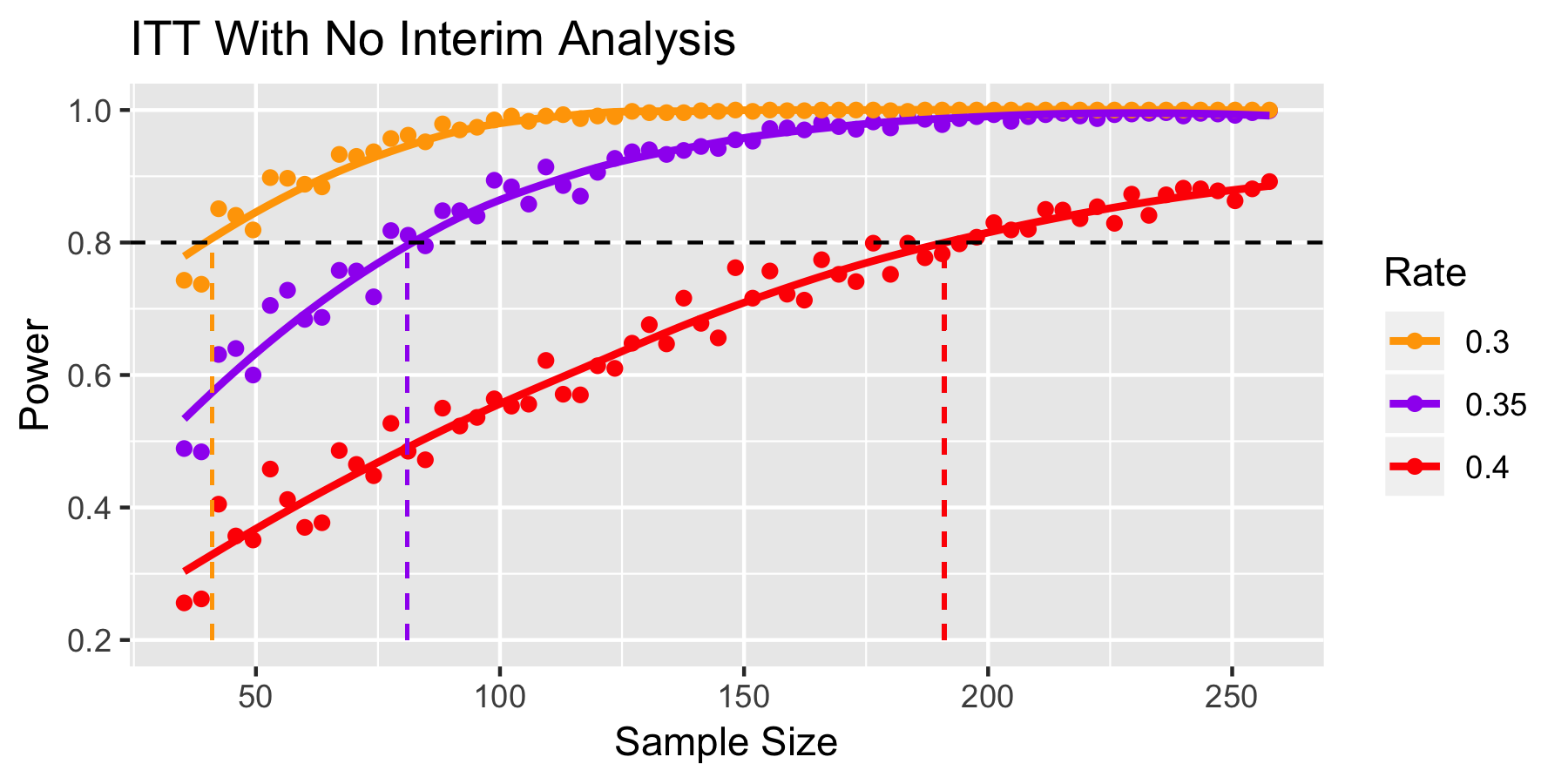
For Aim 2 we considered these two types of analyses (ITT and censored) with the inclusion of an interim analysis. We set the interim analysis to occur halfway through recruitment. The process was the same as before, but a different confidence interval was chosen for the interim and final analysis. After half of the patients were recruited and followed for 100 days, the analysis was performed the same as before but with an 85% confidence interval. If the lower bound of this confidence interval was below 0.50, the simulation was given a significance value of 0 and terminated. If the lower bound was above 0.50, the next half of the patients were simulated and a final analysis was done, also with an 85% confidence interval (see Appendix).

Also created as part of this analysis is a power calculator. The calculator includes an input for sample size, disease rate, and the ability to censor for non-compliers and add in an interim analysis. The output is the power achieved under those settings. An example of the calculator can be found in the appendix, and can be used online at <https://dfhannum.shinyapps.io/powerapp/>.

**Results**

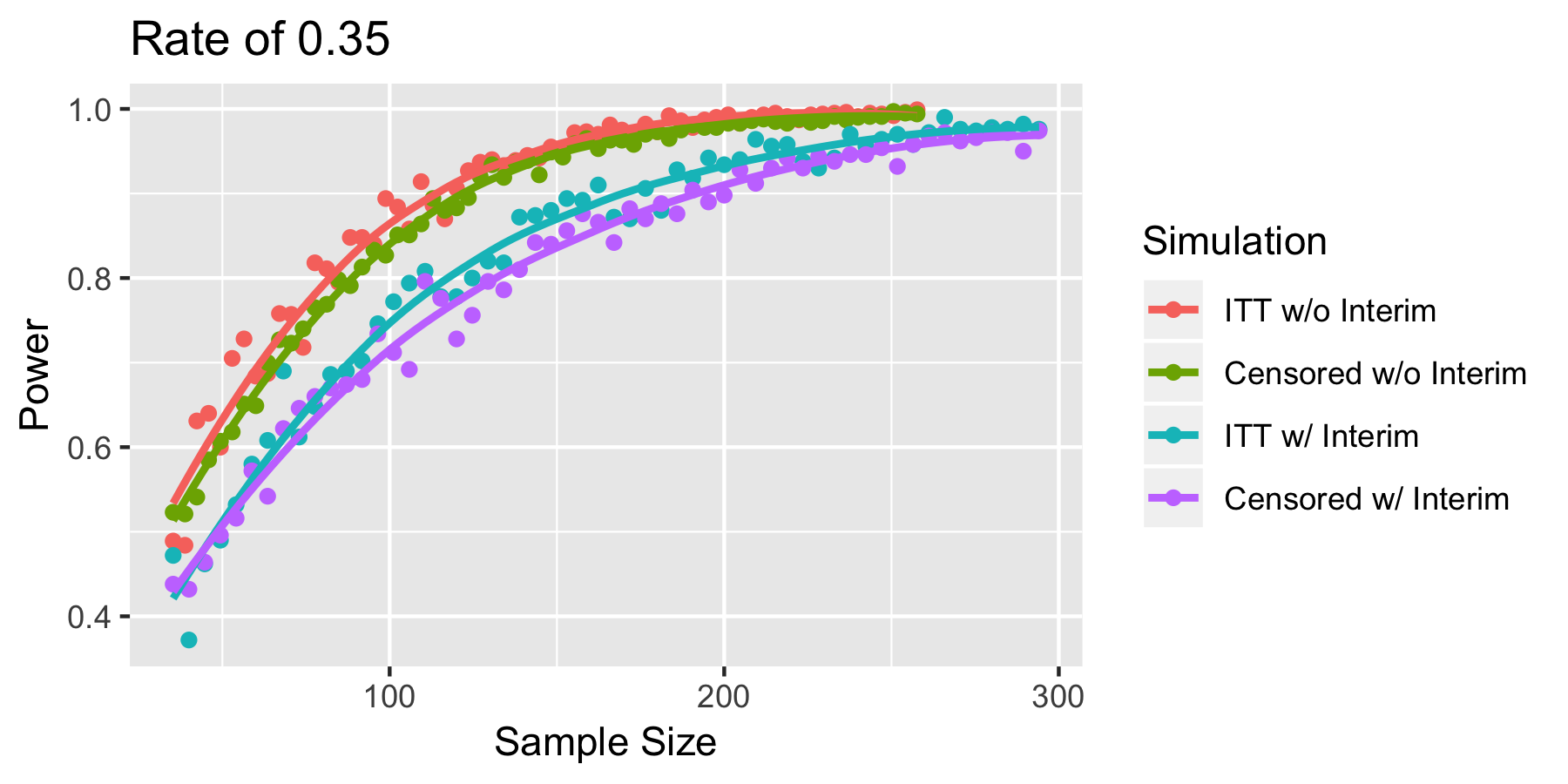
**Table 1**. The results from the four different simulation studies for the three different sample sizes. The sample size was the minimum size necessary to achieve 80% power.

The range of sample sizes necessary for 80% power range from 41 to 300, depending on the rate, analysis and whether or not there is an interim analysis. As would be expected the larger the difference between rates (historical vs new rate) the smaller the sample size. This can be seen in Figure 2, where the smoothed curves shift right as the difference between the historical rate and new rate get smaller.



**Figure 2**. A power plot for an intention-to-treat analysis without an interim analysis, for all three rates. The black horizontal line indicates the 80% power threshold, and the colored vertical lines indicate where on the x-axis for each smoothed curve crosses the 80% power threshold

Doing a censored analysis did not drastically change the sample size from the intention-to-treat analysis (~10% increase in sample size). Adding an interim analysis had a large effect on the sample size necessary for 80% power. Both of these can be seen in Table 1, and viewed graphically in Figure 3. The largest gap between curves is the difference of using an interim analysis. One benefit of the interim analyses is that most of the simulations that were rejected for insignificance occurred in the interim stage instead of the final stage, with a ratio of 14:1 (ITT) and 13:1 (Censored) in the two simulations.



**Figure 3**. A power plot for a single rate across the four different types of single-arm clinical trials. All four curves look proportional to each other. The largest difference between curves can be seen when an interim analysis is added.

**Conclusions**

The power necessary to reach 80% power varies drastically depending on the type of study initiated. We provided the output for many different types of single-arm clinical trials to provide the investigators with the information necessary to proceed with the study. Our recommendation would be an ITT analysis. The benefit of an ITT analysis is you see the effect of the treatment in a truer clinical setting. The underlying rate of the treatment effect may not be fully understood, but instead the rate of the treatment when applied to patients is known. The ITT analysis takes in to consideration the difficulty of the treatment program, by not adjusting for non-compliers.

We determined the interim analysis to occur halfway through the trial. For ITT analysis, the interim analysis increased the sample size by between 45% and 61%. This means that the interim analysis stage would occur at a sample size not drastically smaller than the sample size necessary for the non-interim final analysis for large differences in rate. For example, for a rate of 0.30 and ITT analysis, the interim analysis would occur at 33 patients whereas the full analysis without an interim would occur at 41 patients. If the investigator believes the new rate would be closer to 0.50 then the differences become greater.

In conclusion, we recommend doing an ITT analysis. If the difference in rates is believed to be large and interim analysis would not be very helpful, but if the difference may be smaller an interim analysis would be useful.

**Limitations**

Perhaps a more complex model could be used to model the drop-out of patients who cannot be observed for an aGVHD, instead of inflating the sample size at the end of the analysis. There was no clear model for these patients, which is why we decided to inflate at the end. Another limitation is we chose halfway through recruitment in the trial. This seemed a logical place to do the interim analysis, but perhaps there would be other times that would minimize the overall sample size. With this being said we do not believe either of these adjustments would have a drastic effect on the sample sizes calculated in this analysis.

**Appendix**

Truncated Exponential Distribution

The distribution for time to an aGVHD are known with a median of 28 days and an interquartile range of [7,56] days. The best distribution to model this data is a truncated exponential distribution with a lambda value ~ 0.02. This distribution gives us a median of 29 days and an interquartile range of [13,53] days.

Truncated Normal Distribution

We decided to use a truncated normal distribution to model non-compliance in this study. We hypothesized that most of the patients would start to non-comply with the diet in the middle of the study. Most patients would comply for the study at the beginning, and if the patient had complied for the majority of the study they would not likely non-comply towards the end of the study. These assumptions lead us to fit the bell shaped truncated normal distribution with mean 50 and standard deviation 25.

Type I Error Rate

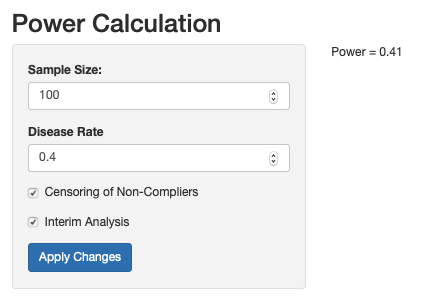
*Standard:*

Originally it was assumed that using a 95% confidence interval in the survival analysis stage would lead to a Type I error rate of 0.05. After running simulations with the null rate (0.50), the Type 1 error rate fluctuated below 0.05. We tested different percentages of confidence intervals and determined that a 94% confidence interval was most consistently near 0.05 for our simulations.

*Interim:*

We knew we would have to adjust our confidence interval for the interim analysis in order to not decrease our Type I error too much (which ended up making us reanalyze our Type I error assumptions in the standard analysis). After doing multiple simulations with different combinations of confidence intervals for the interim and final analysis on the null rate, setting both confidence intervals to 85% lead to the most consistent Type I error rates around 0.05.

Figure 1a



**Figure 1a**. A screenshot of the power calculator available online for this simulation study of a single-arm clinical trial. On the left in the grey box are all the inputs available and on the right the resulting power of that study is displayed.