Group Randomized Trial Design For Targeted Agent – A Simulation Study

Zijiang Yang

March 30, 2017

Abstract

In this project, we conduct a simulation study in order to design a group randomized trial for a pharmaceutical company to test a targeted agent for lung cancer. We first set up the simulation structure by listing some assumptions that may reflect the true conditions, and then run the simulation with different combinations of study lengths and sample selections. In the end, we list several recommended options of design for the pharmaceutical company to choose from.

Introduction

A pharmaceutical company has a promising new drug and would like approval from the FDA to start selling the drug. The drug is a targeted agent, and the study will be run in lung cancer patients with advanced disease. The drug works better in patients who are biomarker positive (about 25% of the population) rather than biomarker negative, but there is considerable uncertainty about this finding. Under standard therapy, lung cancer patients with advanced disease have a five-year survival rate of 30%, and the new agent is expected to increase this survival rate to somewhere between 35% and 45% in the biomarker positive group, and somewhere between 30% and 40% in the biomarker negative group.

The two-arm trial that is being planned is of a standard therapy versus the combination of the standard therapy and the new agent. The objective of this project is to help the pharmaceutical company design this study by running simulations. Specifically, we are trying to optimize the sample size and power of the study, and answer the following questions:

- How long will the trial take to run?
- Which group of patients should be enrolled?

- How many patients should be enrolled?
- How will the randomization be conducted?
- How long should the accrual period be?
- How long should the additional follow-up period be?
- How long should each patient be followed for?
- What will be the power of the study?
- How will the data be analyzed?

Methods

The study comprises two main steps. First, we determine the randomization framework and the simulation structure by making several assumptions. At this step, we also determine which hypothesis test to use in calculating the power. After we are satisfied with the setup of our simulation structure, we run simulations with combinations of different length of accrual period, follow-up period, number of countries, etc. We save the simulation results, and select the best options with respect to sample size, length and power of the study.

Randomization framework

When a lung cancer patient comes in, we first test the biomarker status of the patient. After that, we randomly assign the patient to either treatment or control group. We then follow up each patient until the patient dies or the end of study. This same process is conducted simultaneously in the three or four countries.

Simulation structure

In order to build our simulation structure, we make the following assumption:

1. We assume that the biomarker status for each patient follows a Bernoulli distribution with p = 0.25:

$$B_i \sim Bernoulli(0.25).$$

2. We assume that the group assignment for each patient follows a Bernoulli distribution with p = 0.50:

$$G_i \sim Bernoulli(0.5)$$

3. We assume that the length of the gap of entry time (rate of accrual) follows an exponential distribution, with λ = (number of countries)*72/365. This is because the rate of accrual is 6 patients per month in each country, and we are using a unit of days:

$$S_i - S_{i-1} \sim Exp((number of countries) * 72/365)$$

4. We assume that the survival time for each patient follows an exponential distribution. Patients in the control group have a five year survival probability of 30% regardless of the biomarker status. This corresponds to a hazard rate of -ln(0.3)/5 = 0.24. For the treatment group, we propose two possible scenarios:

- Uniform scenario: for patients with biomarker positive, the five-year survival probability follows a uniform distribution between 0.35 and 0.45, while for patients with biomarker negative, the five-year survival probability follows a uniform distribution between 0.3 and 0.4:

$$\lambda_i \sim Unif(0.16, 0.21), if G_i = 1 \quad B_i = 1$$
$$\lambda_i \sim Unif(0.18, 0.24), if G_i = 1 \quad B_i = 0$$
$$X_i \sim Exp(\lambda_i)$$

- Best case scenario: the five-year survival probability is 0.45 for biomarker positive patients, and 0.4 for others.

5. In this study, we assume to have two types of censoring. First is the end-of-study censoring. Patients get right censored if they are still at risk by the end of study. We also treat dropouts as censored. We assume the rate of dropout is independent for each patient, and follows an exponential distribution with $\lambda = 0.05/365$. Both types of censoring are non-informative censoring, where each subject has a censoring time that is statistically independent of their failure time. Assume T is the end of study, then the follow up time for patient i is the minimum of his dropout time, death time, and end of study time:

$$C_i \sim Exp(0.05/365)$$
$$T_i^* = min(X_i, C_i, T)$$

Since the censoring is non-informative, we choose to use log-rank test to test the difference in the survival and hazard functions between the treatment and control group, and calculate the power. We also tried to use one-tailed t-test to test the difference in 5-year survival rate, but log-rank has a higher power in our case. We

calculate the power by calculating the proportion of simulations that have p-value smaller than or equal to 0.05, and our target is to achieve a power of 0.80. If our target is to maximize the power, we can maximize our sample size by utilizing all the resources given, that is we accrue patients from all four countries, and set the length of accrual period to five years and the length of study ten years, and follow up all patients accrued. However, that is not an efficient choice with respect to time and costs. It may be possible that we can achieve a relatively good power with a smaller sample size and shorter study length, and our goal is to find out the smallest sample size and shortest length of study we need to achieve our target of statistical power.

We run 1000 simulations each time with the following options: * # countries: {3,4} * only including biomarker positive patients: {True, False} * Survival function assumption: {Best case, uniform} * Accrual time T0 (years): {3,4,5} * Study total length T (years): {5,6,7,8,9,10} * Hypothesis test: log-rank test

Results

Table 1 lists some of the possible options under best case scenario. We also include a more detailed table of power under different scenarios and options in the appendix. We also list the cases that we have a power greater than or equal to 0.7.

Uniform Assumption

Since we have limited information about the effectiveness of the drug, the power of the study will largely be based on our assumption of the survival function, and uniform assumption is reasonable in this case, because all the values for the five-year survival rate will be equally probable in the specified range. Under the uniform assumption and significance criterion of 0.05, the best power we can achieve within ten years of study is 0.79. This is very close to our target of 0.80, and to achieve this power, we need to utilize all the resources given and maximize our study length. The mean total sample size in this case is 1438.97, and the mean follow-up time for each patient equals 1154.751 days. Of course, if we want a higher power, we can increase the sample size by making the accrual period longer than five years.

Best Case Assumption

Under best case assumption, if we utilize all the resources given, we can achieve a power close to 1. Therefore, if for some reason we are able to assume the best case scenario, we will be able to largely shorten the length of study and lower the sample size. For example, as listed in Table 1, we only need three years of accrual

Table 1: Simulation Results Under Best Case Scenario

Option	T0	T	Countries	Positive Only	Total Sample Size	Mean Follow-up Days	Power
1	3	5	4	FALSE	865	827	0.85
2	3	8	3	FALSE	649	1145	0.86
3	4	7	3	FALSE	865	1009	0.91
4	4	7	4	FALSE	1153	1009	0.96
5	4	9	4	TRUE	289	1203	0.81
6	5	10	4	FALSE	1441	1200	1

period with two additional years of follow-up period to get a power of 0.85. The sample size is 865, and the mean follow-up time is 827 days in this case. If we only include biomarker positive patients, then we only need a sample size of 289 to get a power of 0.81, as listed in option 5.

However, since it is very likely that the effectiveness of the drug is worse than the best case scenario, we need to make our design more conservative by making the target of the power higher to account for this uncertainty. If we set the target to 0.90, option 3 is the one that has the smallest sample size and shortest total study length that can achieve this target (with 3 countries). If we want to increase the power, we can accrue patients in four countries instead of three. For option 3, if we choose four countries, we will end up with option 4, where we have a power of 0.96, and a total sample size of 1153. Generally, if we want to increase the power without changing the length of study, we can increase the number of countries, and if we want to shrink the sample size, we can include only biomarker positive patients in the study, but the power will also be lower. The target of power we set really depends on our confidence that we can reach the best case scenario.

Conclusion

In conclusion, if our target is to achieve a power of 0.80, and at the same time finish the test as soon as possible, then we may want to accrue patients from four countries instead of three. If our target is not the time that we spend on the test but the follow-up costs, we may want to shrink the sample size by including only biomarker positive patients. If we are confident that the effectiveness of the new drug is promising, that it can increase the five-year survival rate to 45% for biomarker positive patients, and 40% for biomarker negative patients, then we only need three years of accrual period, with two additional years of follow-up period, and include all 865 patients accrued from four countries, in order to reach a power of 0.85. If we are relatively confidant but not 100% sure about the effectiveness of the drug, then four years of accrual period, with three additional years of follow-up and including all patients in three countries may be a good choice. In that case, we can reach a power of 0.91, with a sample size of 865, and mean follow-up time for each patient 1009 days. If we are totally not sure about the effectiveness of the drug, then we may assume

that the five-year survival rate follows a uniform distribution. In that case, we may need to set the accrual period greater or equal to five years, and total length of study equal to ten years, in order to get a power greater than 0.80.

Appendix

R Code and graphs are attached.