Sample Size Determination for Clinical Trials

Paivand Jalalian
Advisor: Professor Kelly McConville

May 17, 2014

Abstract

An important component of clinical trials is determining the smallest sample size that provides accurate inferences. The Frequentist approach to determining sample size is most common; however there has been a recent shift towards using a Bayesian approach due to its flexibility. This paper will review both the Frequentist and Bayesian branches of statistics and their respective approaches to sample size determination. As well, the advantages and disadvantages of using each method will be considered. Finally, along with the Bayesian approach to sample size determination, we will also discuss a Bayesian adaptive design for clinical trials that allows for sample size adjustments during the trial.

1 Introduction

Clinical trials are research studies used to gain insight about new treatments or interventions, for example drugs, procedures, medical devices, etc. Clinical trials are an important part of the development process of new interventions because they determine and confirm the efficacy, as well as the safety, of an intervention.

Conducting a clinical trial requires a lot of preparation, and an important aspect of designing a clinical trial is determining the correct sample size for that trial. Having the correct sample size limits unneeded exposure to possible harmful treatments, while ensuring accurate results. Additionally, determining the correct sample size saves money and resources. There are many ways in which sample size can be calculated, and all these methods aim to find the "best" sample size, or the smallest sample size necessary to get accurate and inference worthy results.

A common approach to calculating sample size is the Frequentist approach because of its simplicity. However, in recent years, the Bayesian approach has become more popular due to its ability to incorporate existing information about the effect of a treatment, as well as its flexibility.

In this paper we will first introduce both the Frequentist and Bayesian statistical approaches and the differences between the two. Then, using that background, we will outline the basic process of clinical trials and then propose the sample size determination methods for each approach, as well as the limitations of using each approach. Lastly, we extend the Bayesian sample size approach to a two stage adaptive clinical trial.

2 Overview of Frequentist and Bayesian Statistics

There are two dominant statistical approaches that are commonly used, the Bayesian approach and the Frequentist approach. Here we will summarize the main ideas from each methodology, which we will later use to compare the sample size determination for each approach.

2.1 The Frequentist Approach

The Frequentist approach uses hypothesis testing and probabilities to make statistical inferences about unknown parameters.

Under the Frequentist approach the data is considered random, because if the study is repeated the data will be different with every repetition. On the other hand, the unknown parameter being tested, or the hypothesis, is believed to be fixed and is either true or false. As stated above, inference is made by looking at probabilities, p-values, where this probability is the expected frequency that a random event will occur, or the probability of the data given a hypothesis.

Before a given study begins, null and alternative hypotheses are stated, which are customarily 'no relationship or effect exists' versus 'there is some effect or relationship'. Next a significance level, typically $\alpha = .05$, is chosen. The significance level is the probability of rejecting a null hypothesis that is true, or the fixed probability that a observed result couldn't have occurred by chance alone. Data are collected and a statistical test is conducted to calculate a p-value, which in this case can be interpreted as the probability of getting results as extreme as the one observed assuming the null hypothesis is true. If p-value $\leq .05$ the results are thought to be "significant" and the alternative hypothesis is favored.

2.2 The Bayesian Approach

The Bayesian statistical approach uses existing beliefs and/or information, along with newly collected data, to draw inference about unknown parameters. More succinctly, this is summarized through Bayes' Theorem.

Definition 2.1 *Posterior Distribution ($\xi(\theta_i|\boldsymbol{x})$): The distribution that describes the posterior probability of θ given old information and newly acquired data.*

Theorem 2.2 *Bayes' Theorem: Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$ be events such that $0 < P(\theta_i) < 1$, for $i = 1, 2, 3, \ldots, m$. And let $\boldsymbol{x} = (x_1, x_2, \ldots, x_k)$. If $\boldsymbol{\theta}$ follows a continuous distribution, then,*

$$\xi(\theta_i|\boldsymbol{x}) = \frac{f_n(\boldsymbol{x}|\theta_i)\,\xi(\theta_i)}{\int f_n(\boldsymbol{x}|\theta)\,\xi(\theta)\,d\theta}, \tag{1}$$

and if $\boldsymbol{\theta}$ follows a discrete distribution, then for $j = 1, 2, 3, \ldots,$

$$\xi(\theta_i|\boldsymbol{x}) = \frac{f_n(\boldsymbol{x}|\theta_i)\,\xi(\theta_i)}{\sum_{j=1}^{m} f_n(\boldsymbol{x}|\theta_j)\,\xi(\theta_j)}.$$

The variables, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)$ are the unknown parameters of interest. Under Bayesian statistics these are random variables, and therefore we would like to find the distribution of these variables. As stated above, Bayesian statistics is unique in its ability to incorporate existing information about $\boldsymbol{\theta}$, and this is represented by the *prior distribution*, $\xi(\theta_i)$. Because this distribution is based on prior information, it is constructed before the experiment begins.

After determining the prior distribution, we use the observed data, $\mathbf{x} = (x_1, x_2, \ldots, x_k)$, where $\mathbf{x}$ is independent and identically distributed, and the prior distribution to construct the *likelihood function*, $f_n(\mathbf{x}|\theta_i)$. This likelihood function is the conditional probability distribution of the data $\mathbf{x}$ given the parameter θ_i, and is calculated as follows,

$$\begin{aligned} f_n(\mathbf{x}|\theta_i) &= f(x_1, x_2, \ldots, x_k|\theta_i) \\ &= f(x_1|\theta_i) \times f(x_2|\theta_i) \times \cdots \times f(x_k|\theta_i) \\ &= \prod_{j=1}^{k} f(x_j|\theta_i) \end{aligned}$$

In the denominator of (1), we have the normalizing constant

$$\int f_n(\mathbf{x}|\theta)\, \xi(\theta)\, d\theta,$$

which is a unique value that ensures that

$$\int \xi(\theta_i|\mathbf{x})\, d\theta = 1.$$

When using Bayes' theorem, it is common to leave out the normalizing constant to make calculations easier, and modify the theorem to say the posterior distribution is "proportional" to the product of the prior multiplied by the likelihood function,

$$\xi(\theta_i|\mathbf{x}) \propto f_n(\mathbf{x}|\theta_i)\xi(\theta_i).$$

Finally, using Bayes' Theorem we have derived the *posterior distribution*, which is the conditional distribution of $\boldsymbol{\theta}$ given $\mathbf{x}$. This posterior distribution can be analyzed and summarized by looking at its mean, standard deviation, etc. It can also be used in another experiment as the prior distribution, as we continue to gain inference about our parameters.

It is important to note that unlike Frequentist statistics, Bayesians consider the data to be fixed, they believe that there is a single set of data that we are continuously sampling from. Additionally, Bayesians use probability to represent beliefs that values of a parameter are true. More specifically, Bayesians define probability as the probability of our hypothesis given the data.[12]

2.2.1 Prior Distributions

Prior distributions summarize and express existing information about an unknown parameter and how much researchers believe in the possible values the parameter can take on. The researcher has the option of making the prior distribution *informative* or *non-informative*.

A non-informative prior has little impact on the posterior distribution. It is used when little to no information exists about a parameter, or when the researcher wants to take a more conservative approach to the data analysis. This approach is more conservative because when the prior is non-informative, the data will have more of an influence on the inference and posterior distribution. A common non-informative prior is the uniform distribution because it states that every value for the parameter is equally likely, however any distribution can be made relatively non-informative by setting the variance equal to a large value.

An informative prior incorporates existing information that will impact the resulting posterior distribution. There are two types of informative priors, *skeptical* and *optimistic* [9]. A skeptical prior distribution assumes that there is no difference between the effectiveness of both treatments. This distribution can be a normal distribution centered around the null. Conversely, an optimistic prior is centered around the alternative hypothesis and has a strong belief that the new treatment is effective.

The associated parameters of a prior distribution are called *prior hyper-parameters*. If these hyper-parameters are known, determining the posterior distribution is relatively easy using Bayes' Theorem. However, if some of these hyper-parameters are unknown, an estimation method or hierarchical model must be used. The hierarchical Bayes' model allows the researcher to create levels of prior distributions, or *hyperpriors*, for unknown hyper-parameters of the desired prior distribution. These added hyperpriors fill in missing information or elaborate about our prior distribution. [8]

Unfortunately, no single method is a panacea for picking a prior distribution, and computations using Bayes' Theorem can become computationally intensive. To make calculations and decisions easier, *conjugate prior families* were constructed.

Definition 2.3 *Let $X_1, X_2, \ldots, X_n$ be conditionally i.i.d. given $\boldsymbol{\theta}$ with a common distribution $f(x|\boldsymbol{\theta})$. Let ψ be the family of distributions for $\boldsymbol{\theta}$. If both the prior distribution, $\xi(\boldsymbol{\theta})$, and posterior distribution, $\xi(\boldsymbol{\theta}|\boldsymbol{x})$, belong to ψ, then ψ is called a conjugate family of priors for $f(x|\boldsymbol{\theta})$.*

Thus, conjugate prior families are distributions such that the prior and posterior distributions are the same, or, in other words, our likelihood functions multiplied by our prior distribution results in a posterior that is proportional to the same distribution as our prior. As an example, if our data follows a Binomial distribution, $X \sim \text{Binomial}(p)$, then the conjugate prior is a Beta distribution, $\xi(p) \sim \text{Beta}(a, b)$ where $a, b > 0$, and p given our data will also follow a Beta distribution, $\xi(p|x) \sim \text{Beta}(a_1, b_1)$. Thus, using conjugate prior families can make decisions and calculations simpler because it removes the need of finding the normalizing constant through integration.[12]

3 Introduction to Clinical Trials

Before we begin talking about adaptive clinical trials, the general overview of what clinical trials are and the process of how drugs are reviewed should be discussed.

Clinical trials are used to research whether a treatment or device is safe and effective for humans. After development, drugs may first be tested on animals first to help determine

toxicity levels and possible harmful side effects, and then moved on to humans. Thus, clinical trials usually occur in the final stages of the development process.

Before a clinical trial begins, a protocol is prepared that details the experiment in length and why those decisions were made, including number of participants, who is eligible, what will be measured, etc. In this protocol, the researcher will also outline how they will ensure that their study will not be biased. A *bias* is a systematic error that deviates our results from the true result and effects inferences made. To combat this researchers will typically include comparison groups, groups they can compare results with, randomization, participants are assigned to groups randomly to make sure that differences occurred are because of the treatment and not where participants are allocated, and/or masking, where they ensure participants don't know which group they are in as long as the safety of the participant is not compromised. This protocol must be approved by the Food and Drug Administration (FDA) before any research begins.

As stated above, comparison groups can be used as a safeguard against bias, but can also be used to compare the effects of the new drug with existing treatment or placebo. This comparison may not only uncover which treatment is better, but also whether a treatment is better for a specific patient.

Clinical Trials follow specific standards that protect the patients and help the researcher produce statistically sound results. The FDA has established the general steps to the process of clinical trials which will be detailed in the next section. The conclusion of a clinical trial is to determine whether a new treatment improves a patients condition, has no effect, or causes harm.

3.1 Phases of Drug Development

The following is the FDA approved process. After the drug is created, sponsors (companies, organizations, etc) must show the FDA results from animal testing and present their proposal for human testing. If the FDA believes the drug is safe and approves their proposal, testing is continued on humans. This clinical trial testing occurs in 3 phases:

Phase 1 Testing: The focus of Phase 1 is on safety. This testing is usually on healthy volunteers to determine what the drugs main side effects are and how the drug is metabolized. This phase is also used to evaluate what a safe dose range is.

Phase 2 Testing: If Phase 1 doesn't show high levels of toxicity, then researchers can continue on to Phase 2. The focus of Phase 2 is on effectiveness. The drug is administered to a larger sample who have the specific disease or condition. In this phase, the drug is compared to patients who receive a placebo treatment, or the common treatment. Some Phase 2 trials are broken up into two stages so trials can terminate earlier if no significant data is found.

Phase 3 Testing: If Phase 2 testing shows that the drug is effective, Phase 3 testing can begin. In this phase of testing, more information is gathered about the effectiveness and safety of the drug by looking at a different population, different dosages, or using the drug with a combination of others. This phase also determines any long-term effects from use.

After these phases, a meeting is set with the FDA and the sponsor before the New Drug Application is submitted. The NDA application is submitted and reviewed by the FDA, as well as the drug labeling and facility used to make the drug. Onces all these steps are completed and passed by the FDA, the drug will be approved. [10]

4 Sample Size Determination

Determining sample size is one of the most critical calculations in any study or experiment because it can directly influence results. Having the right sample size will make it more likely that results couldn't have occurred by chance alone but from a true effect or difference. Additionally, having the right sample size can ensure that if a statistically significant difference exists it will be uncovered in our data. If the effect size is small, a larger sample size is required to detect a difference, however, if the effect size is large a smaller sample size is needed. Lastly, having the correct sample size, and correct participant pool, can assure that the data are representative of the targeted population, and not just those who participated in the study. All of these points can be solved by having a sufficiently large sample size, however many don't have unlimited recourses to make a larger participant pool feasible. Thus, having a sample size that is just large enough can help save money and make sure resources are allocated as efficiently and effectively as possible.

This section will cover the Frequentist approach and two methods of the Bayesian approach to sample size determination, and review some limitations to each approach.

4.1 Frequentist Sample Size Determination

The following is a derivation of the Frequentist approach of determining the appropriate sample size of a comparative experiment with a desired power.[6] [7]

Suppose the experiment is looking at the mean difference between two treatments where μ_c is the average response of the control group and μ_t is the average response of the new treatment group. Also suppose that all responses, X_{ci} = responses from the control group and X_{ti}= responses from the treatment group, are normally distributed. Then we would like to test $H_0 : \mu_c = \mu_t$ and $H_a : \mu_c \neq \mu_t$.

First, a value for $\delta = \mu_t - \mu_c$ is selected, which represents the minimal size of difference between the treatment and control groups the researcher would consider important. This is different from the Bayesian approach because now δ is fixed. Since δ is initially unknown, a value is chosen that is obtainable but also enough to distinguish between groups. It can be determined by looking at previous experiments. Similarly, the variances of the treatment and control group responses, σ_c^2 and σ_t^2, are unknown but estimated from prior data.

Next, the *power* of the test, P, is determined. In Frequentist hypothesis testing, there are four outcomes that can occur once a decision about H_0 is made (Table 1).

As the table shows, power is the probability of rejecting a false null hypothesis. If the actual difference between the two treatments is greater than δ, the researcher would like to have a strong probability (0.8, 0.9, or 0.95) of showing a statistically significant difference, a difference that could not be due to chance alone. In other words, if the actual difference is

Decision	H_0 True	H_0 False
Fail to Reject H_0	no error	Type II Error
Reject H_0	Type I Error	no error (Power)

Table 1: There are four possible outcomes for every decision made from a hypothesis test.

δ, the power is the probability of actually observing a difference δ. Determining the power is important because the power also decides the probability of a *Type II error*, $\beta = 1 - P$, or the probability of failing to reject a false null hypothesis. Lastly, the researcher should determine the significance level, α. This value determines the probability of rejecting a true null hypothesis, which is also known as a *Type I error*.

It is important to note that the Type I and Type II errors are inversely related, see Figure 1, therefore a decrease in one results in an increase in another (when the sample size remains constant). This, however, will not effect our calculations because we are determining a sample size for selected α and β.

As stated above, the two groups follow normal distributions,

$$X_{c1}, X_{c2}, \ldots, X_{cn} \sim N(\mu_c, \sigma_c^2)$$

$$X_{t1}, X_{t2}, \ldots, X_{tn} \sim N(\mu_t, \sigma_t^2).$$

Normal distributions have the unique property of being closed under addition. Let $Y_1, Y_2, \ldots, Y_m$ be normal random variables with means μ_i and variances σ_i^2 for $i = 1, 2, 3, \ldots, m$. Let W= $c_1Y_1 + c_2Y_2 + \cdots + c_mY_m$ where $c_1, c_2, \ldots, c_m$ are real numbers. Then $W \sim$ Normal(μ_W, σ_W^2) where,

$$\mu_W = c_1\mu_1 + c_2\mu_2 + \cdots + c_m\mu_m$$

and

$$\sigma_W^2 = c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \cdots + c_m^2\sigma_m^2.$$

We can now calculate the sample mean distributions for both groups.

$$\bar{X_c} = \frac{X_{c1} + X_{c2} + \cdots, X_{cn}}{n}$$

then,

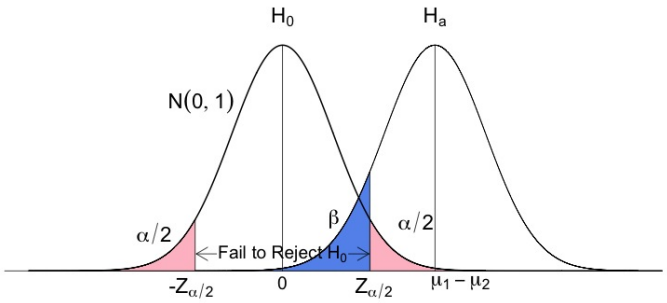
$$\mu_{\bar{X_c}} = \frac{\mu_c + \mu_c + \cdots + \mu_c}{n} = \mu_c$$

and

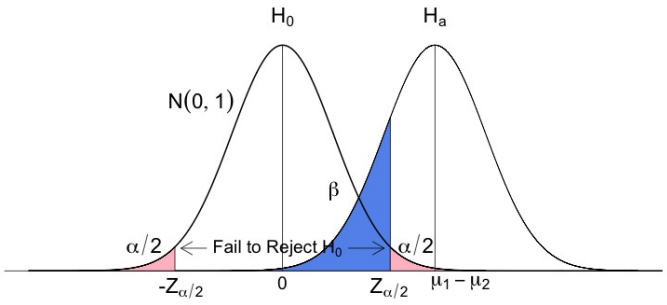
$$\sigma_{\bar{X_c}}^2 = \frac{\sigma_c^2 + \sigma_c^2 + \cdots \sigma_c^2}{n^2} = \frac{\sigma_c^2}{n}.$$

A similar calculation can be made for the sample mean of the treatment group.

Let $\mathbf{D} = \bar{X_t} - \bar{X_c}$ be the observed mean difference. Using the closure property of normal distributions, $\mathbf{D}$ is normally distributed about $\mu_t - \mu_c$ with a variance of $(\sigma_t^2 + \sigma_c^2)/n$ or standard deviation of $\sqrt{(\sigma_t^2 + \sigma_c^2)/n}$. We will use the previously devised *test statistic*, an



(a) The pink represents the Type I error and the blue is a Type II error.



(b) When α is smaller, the Type I error decreases, but the Type II error increases showing their inverse relationship.

Figure 1: Type I and Type II Error

equation that standardizes the mean difference, for this hypothesis test to make calculations simpler,

$$Z_0 = \frac{(\bar{X}_t - \bar{X}_c) - (\mu_t - \mu_c)}{\sqrt{\frac{(\sigma_t^2 + \sigma_c^2)}{n}}}.$$

Recall that $(\mu_t - \mu_c) = 0$, therefore we can rewrite the test statistic for this specific example,

$$Z_0 = \frac{\mathbf{D} - 0}{\sqrt{\frac{(\sigma_t^2 + \sigma_c^2)}{n}}} = \frac{\mathbf{D}}{\sqrt{\frac{(\sigma_t^2 + \sigma_c^2)}{n}}}.$$

We see that under H_0, the test statistic follows a standard normal distribution, $Z_0 \sim N(0, 1)$. Under H_a, $\mathbf{D} = \delta$, thus $Z_0 \sim N(\delta\sqrt{n}/\sqrt{\sigma_t^2 + \sigma_c^2}, 1)$.

Looking at [Figure 2] we see that if H_a is true, a Type II error, β, will be made if $-Z_{\alpha/2} \leq Z_0 \leq Z_{\alpha/2}$. Thus, the probability of a Type II Error β is the probability that Z_0 will fall between those critical values. Thus,

$$-Z_{\alpha/2} \leq Z_0 \leq Z_{\alpha/2}$$
$$-Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sqrt{\sigma_t^2 + \sigma_c^2}} \leq 0 \leq Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sqrt{\sigma_t^2 + \sigma_c^2}}.$$

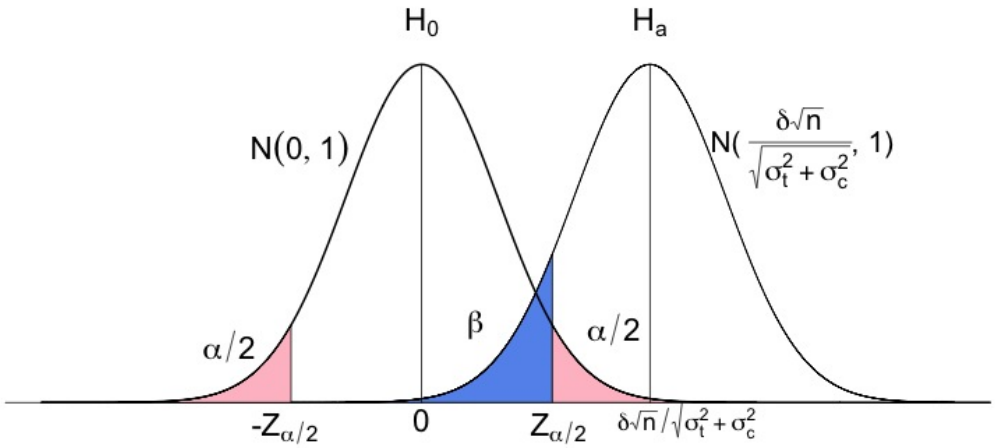


Figure 2: This figure graphically shows what our null and alternative hypotheses are, and where their distributions overlap.

From here, we can use the the the function $\Phi(x)$, which calculates the probability to the left of x under the standard normal distribution. Then,

$$\beta = \Phi\left(Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sqrt{\sigma_t^2 + \sigma_c^2}}\right) - \Phi\left(-Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sqrt{\sigma_t^2 + \sigma_c^2}}\right),$$

and

$$\beta \approx \Phi\left(Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sqrt{\sigma_t^2 + \sigma_c^2}}\right)$$

because when $\delta > 0$, $\Phi(-Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sqrt{\sigma_t^2+\sigma_c^2}}) \approx 0$. Let Z_β be the higher percentile of the standard normal distribution, which means $\beta = \Phi(-Z_\beta)$. Our previous equation for β can be simplified to,

$$-Z_\beta \approx Z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sqrt{\sigma_t^2 + \sigma_c^2}}.$$

Thus, for a two-sided hypothesis test with a significance level α and power $1 - \beta$,

$$n \approx \frac{(Z_{\alpha/2} + Z_\beta)^2 (\sigma_t^2 + \sigma_c^2)^2}{\delta^2}.$$

4.1.1 Example

Suppose we would like to test the success of a new drug that lowers systolic blood pressure, pressure in the arteries when the heart is beats. Let $\delta = 10$mm Hg be the minimally significant difference between the treatment and control groups . We want to detect this difference with a probability (or *power*) of at least 90%, and a significance level, α, of 0.05 ($Z_{\alpha/2} = 1.96$). If we have a power of 90%, then $\beta = .1$ and $Z_\beta = 1.28$. Finally, let $\sigma_c = \sigma_t = 15$ It follows that,

$$n = \frac{(Z_{\alpha/2} + Z_\beta)^2 (\sigma_t^2 + \sigma_c^2)^2}{\delta^2} = \frac{(1.96 + 1.28)^2 (15^2 + 15^2)^2}{10^2} \approx 47.23.$$

Thus, a sample size of approximately 48 people is needed for our test to have a power of 90%.

4.1.2 Limitations of the Frequentist Approach

The equation for calculating the sample size using the Frequentist approach is straight forward and easy to use, which is why many clinicians use this method. It is, however, important to indicate some of the limitations to using this method.

A major limitation to this approach is the inability to incorporate and allow past experience and information to inform our expectation of what will occur. Additionally, in this approach we are estimating some values, like σ_c and σ_t, however our approach does not take into account the uncertainty involved with estimating these values. Thus, our predictions are less realistic.

Some limitations also arise from thinking of the unknown parameter as a fixed value. Because of this, Frequentists cannot describe any uncertainty about our unknown parameter. For example, while Bayesians can state that an interval contains an unknown parameter with 95% probability, Frequentists can only state that the parameter is either in the interval (100% probability) or not (0%). Rather, they will use probability to state that, in the long run, 95% of such intervals will cover the parameter being estimated, and this statement will no longer be relevant once the data is collected. Additionally, it is rare that exact fixed values will apply to situations in which this hypothesis testing is being used for.

The Frequentist approach has a strong dependence on sample size. When sample sizes are very large or small, we begin to run into problems. Holding everything the same, larger sample sizes can detect small differences and can bias the p-values to reject the null hypothesis which could be true. Similarly, when sample sizes are small, it becomes more difficult to reject the null hypothesis and therefore it is difficult to detect small differences.

In most situations, under the Frequentist approach, the null will be false because it is rare that two means will be exactly equal, or that a value will be exactly zero. It becomes pointless to test H_0 because if the null is a fixed value, it will always be false.

Lastly, the Frequentist approach is not very flexible to possible changes that could occur during a study. For example, it is statistically difficult to account for trials that needed to be stopped early. [8]

4.2 Bayesian Sample Size Determination - Average Power Method

There are several different methods of determining sample size using the Bayesian approach. Here we will use show how to use the Average Power method when our data follow a logistic regression and when our data follow a Poisson distribution. Both methods will be comparing the effects of two treatments, the newly developed treatment and an existing one. Then we will review some of the limitations of the Bayesian approach.

4.2.1 Logistic Regression Sample Size Calculation

In this section we will cover how to calculate sample size for a logistic regression model.

In Bayesian sample size calculations, instead of a single hypothesized value for H_0, we allow H_0 to include a range of values. This can be written as, $H_0 = \Delta$ where $\Delta \in [\delta_U, \delta_L]$. The value δ_L is the threshold in which the new treatment would be considered inferior to the control, and δ_U will be the threshold where the new treatment will be considered clinically superior to the control.

Let Y be a categorical binary random variable where $Y_i = 1$ if the disease progresses for the i^{th} patient and $Y_i = 0$ if the disease does not, for $i = 1, 2, \ldots N$. Then, $Y_i \sim \text{Bernoulli}(\pi_i)$ where π_i is the probability that the disease will progress, or $\pi_i = P(Y_i = 1)$. A possible model to estimate π_i is the logistic regression,

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_i, \tag{2}$$

where x_i is a binary categorical variable indicating if the patient was in the control ($x_i = 0$) or in the treatment group ($x_i = 1$), β_0 is how much effect the control has on the progression of the disease, and β_1 indicates how much the new treatment has an effect compared to the control. Thus, if we wanted to calculate the effect of the new treatment, then we would look at $\beta_0 + \beta_1$. Equation (2) can be used to find π_i,

$$\begin{aligned} \log\left(\frac{\pi_i}{1-\pi_i}\right) &= \beta_0 + \beta_1 x_i \\ \frac{\pi_i}{1-\pi_i} &= e^{(\beta_0+\beta_1 x_i)} \\ \pi_i &= \frac{e^{(\beta_0+\beta_1 x_i)}}{1+e^{(\beta_0+\beta_1 x_i)}}. \end{aligned} \tag{3}$$

The value $\frac{\pi_i}{1-\pi_i}$ is the odds of "success", or odds of the disease progressing. This value is used to calculate the odds ratio, or the progression of the disease given that a patient was in the treatment or control. As stated above x_i is an indicator variable, thus, for $x_i = 1$

$$\text{odds}_{\text{treatment}} = \frac{\pi_i}{1-\pi_i} = e^{\beta_0+\beta_1},$$

and for $x_i = 0$

$$\text{odds}_{\text{control}} = \frac{\pi_i}{1-\pi_i} = e^{\beta_0}.$$

Therefore, Δ, or the odds ratio of disease progression if in the treatment as opposed to the control group, is

$$\Delta = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{\beta_1}.$$

Thus, e^{β_1} has become our new Δ from above and is our parameter of interest. If our odds ratio $\Delta = 1$, that means there is no difference between treatment and control group. If $\Delta > 1$, then the odds of disease progression are Δ times more likely in patients in the treatment group. Conversely, if $\Delta < 1$ then the odds of disease progression are Δ times less likely in patients in the treatment group.

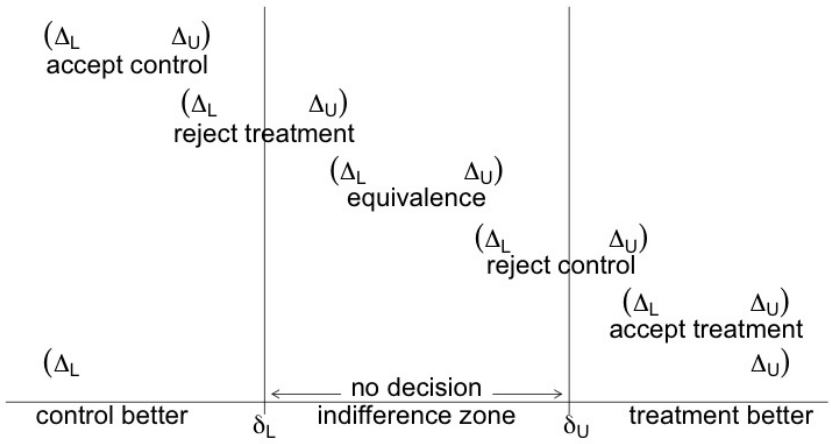


Figure 3: This figure shows the indifference zone (δ_L, δ_U), and the six possible outcomes for a clinical trial.

To find the correct sample size will test different sample sizes and decide based on where the 95% posterior confidence interval for Δ is with respect to the indifference zone [Figure 3]. The values for β_0 and β_1 are fixed such that the control is accepted, treatment is accepted or they are equivalent. Along with these $\beta_i's$, the treatment each patient x_i receives is also fixed, for $i = 1, 2, \ldots N$. These values are used to calculate π_i using (3). With $\hat{\pi}_i$ we can generate random fake data values for Y_{ij} from the binomial distribution, for $j = 1, 2, \ldots, N_{rep}$. The prior, likelihood, and each generated data vector $Y_j = (Y_{1j}, Y_{2j}, \ldots, Y_{Nj})$, can be plugged

into Baye's Theorem (1) and we can find the posterior distribution for Δ. Lastly, using this posterior distribution, we can construct a 95% credible interval for Δ and see where in the indifference graph it lies. This is repeated for the N_{rep} datasets, and we can compute the probability for each of the six outcomes by changing β_0 and β_1. We see how many times our confidence interval lies in the correct zone of our graph, and the proportion of times this occurs is our power. Thus N is selected based on our desired power. For example, if we set $\beta_1 = 0$ we are stating that there is no difference between the control and treatment. Then, we create our fake data, posterior, and confidence interval and choose a sample size such that the 95% confidence interval falls outside the indifference zone 5% of the time. Thus, we have chosen a sample size for a specific Type I error, α.

4.2.2 Poisson Sample Size Calculation

The following is a summary of Hand (2011).

We will now assume the data follow a Poisson distribution.

Definition 4.1 *Let $\mu > 0$. A random variable Y follows a Poisson distribution if the probability mass function of Y is,*

$$f(y|\mu) = \frac{e^{-\mu}\mu^y}{y!} \quad \textit{where } y = 0, 1, 2, \ldots$$

The Poisson Distribution is a probability distribution for the counts of events that occur randomly in a given interval of time when the mean is μ [12].

Let X_i be the number of adverse events in n_i patient time. Also let λ_i be the rate for the events for population i, where $0 < \lambda_i < 1$ and $i = 1, 2$, because we will be looking at two different treatments groups. Then, $X_i \sim \text{Poisson}(n_i\lambda_i)$, where

$$P(X_i = x_i|\lambda_i) = \frac{(n_i\lambda_i)^{x_i}}{x_i!}e^{-n_i\lambda_i} \quad \text{where } x_i = 0, 1, 2, \ldots \text{and } \lambda_i > 1.$$

We must first decide on and define our prior distribution for λ_i. There are many options for prior distributions for λ_i, however the gamma distribution is chosen because it is the conjugate prior, and, thus, will make calculations easier. Thus, the prior for λ_i will be,

$$f(\lambda_i|a_i, b_i) = \frac{b_i^{a_i}}{\Gamma(a_i)}\lambda^{a_i-1}e^{-b_i\lambda_i} \quad \textit{where} \quad a_i, b_i > 0,$$

or $\lambda_i \sim \text{Gamma}(a_i, b_i)$. It follows that, using (1), the posterior density of λ_i is,

$$\begin{aligned}
\xi(\lambda_i|x_i) &= \frac{f_n(x_i|\lambda_i)\,\xi(\lambda_i)}{\int f_n(x_i|\lambda_i)\,\xi(\lambda_i)\,d\lambda_i} \\
&\propto f(x_i|\lambda_i)\xi(\lambda_i) \qquad (4) \\
&\propto \frac{b_i^{a_i}}{\Gamma(a_i)}\lambda^{a_i-1}e^{-b_i\lambda_i}\frac{e^{-n_i\lambda_i}(n_i\lambda_i)^{x_i}}{x_i!} \\
&\propto \lambda^{x_i+a_i-1}e^{-\lambda(n_i+b_i)} \qquad (5) \\
&\propto \frac{(n_i+b_i)^{(a_i+x_i)}}{\Gamma(a_i+x_i)}\lambda^{x_i+a_i-1}e^{-\lambda(n_i+b_i)}, \qquad (6)
\end{aligned}$$

and $\lambda_i|x_i \sim \text{Gamma}(a_i + x_i, n_i + b_i)$. We are able to eliminate the constants during our calculations, as seen in (4) and (5), because we know our resulting posterior distribution must integrate to 1, and can thus add in the necessary constants, (6).

We have calculated our posterior distribution (6), however we don't know what the values a_i and b_i are. To determine these values we will use the *Maximum Likelihood Estimation Method*, which is used to estimate unknown parameters. This method finds the values of our unknown parameters that maximizes our likelihood function, because these values maximize the likelihood of observing the data.

However, before we can calculate our likelihood function, we must first find the marginal density of X_i. To find the marginal density of X_i, we integrate λ_i out of the joint density of X_i and λ_i. We want to use the marginal density of X_i because λ_i is also an unknown parameter and, thus, we don't want our a_i and b_i to be in terms of λ_i. It follows that the marginal density of X_i is,

$$\begin{aligned}
m(x_i|a_i,b_i) &= \int_0^\infty P(X_i = x_i|\lambda_i)f(\lambda_i|a_i,b_i)\,d\lambda_i \\
&= \int_0^\infty \frac{b_i^{a_i}}{\Gamma(a_i)}\lambda^{a_i-1}e^{-b_i\lambda_i}\frac{e^{-n_i\lambda_i}(n_i\lambda_i)^{x_i}}{x_i!}\,d\lambda_i \\
&= \frac{b_i^{a_i}n_i^{x_i}}{\Gamma(a_i)\Gamma(x_i+1)}\int_0^\infty \lambda^{x_i+a_i-1}e^{-\lambda(n_i+b_i)}\,d\lambda_i \\
&= \frac{b_i^{a_i}n_i^{x_i}}{\Gamma(a_i)\Gamma(x_i+1)}\frac{\Gamma(a_i+x_i)}{(n_i+b_i)^{(a_i+x_i)}}\int_0^\infty \frac{(n_i+b_i)^{(a_i+x_i)}}{\Gamma(a_i+x_i)}\lambda^{x_i+a_i-1}e^{-\lambda(n_i+b_i)}\,d\lambda_i \\
&= \frac{\Gamma(a_i+x_i)}{\Gamma(a_i)(x_i+1)}\left(\frac{b_i}{n_i+b_i}\right)^{a_i}\left(\frac{n_i}{n_i+b_i}\right)^{x_i}.
\end{aligned}$$

Given k independent trials with x_{ij} events, where $i = 1, 2$ and $j = 1, 2, \ldots, k$, we can now calculate the marginal likelihood,

$$\begin{aligned} M(a_i, b_i) &= m(x_{i1}, x_{i2}, \ldots, x_{in}|a_i, b_i) \\ &= m(x_{i1}|a_i, b_i) \times \cdots \times m(x_{in}|a_i, b_i) \\ &= \prod_{j=1}^{k} m(x_{ij}|a_i, b_i). \end{aligned}$$

Instead of evaluating this product, it is common to use the log of the likelihood function to make calculations easier. As stated above, the marginal maximum likelihood estimators for a_i and b_i are the maximum values of the marginal likelihood, however the program used to carry out this method minimizes rather than maximizes a given function. For that reason, we must find the negative log of the marginal likelihood,

$$\begin{aligned} L(a_i, b_i) &= \sum_{j=1}^{k} -\log\left[m(x_{ij}|a_i, b_i)\right] \\ &= -\log\left[\prod_{j=1}^{k} m(x_{ij}|a_i, b_i)\right] \\ &= -\sum_{j=1}^{k} \log\left[m(x_{ij}|a_i, b_i)\right] \\ &= -\sum_{j=1}^{k} \log\left[\frac{\Gamma(a_i + x_i)}{\Gamma(a_i)(x_i + 1)} \left(\frac{b_i}{n_i + b_i}\right)^{a_i} \left(\frac{n_i}{n_i + b_i}\right)^{x_i j}\right] \\ &= -\sum_{j=1}^{k} [\log(\Gamma(a_i + x_i)) - \log(\Gamma(a_i)) - \log(x_i + 1) + a_i \log(b_i) \\ &\qquad - a_i \log(n_i + b_i) + x_{ij} \log(n_i) - x_{ij} \log(n_i + b_i)]. \end{aligned} \tag{7}$$

Finally, using the equation above, we can use R or other mathematical programs to find the marginal maximum likelihood estimators, or best values for $\hat{a_i}$ and $\hat{b_i}$ such that,

$$(\hat{a_i}, \hat{b_i}) = \arg\min\{L(a_i, b_i)|x_{i1}, \ldots, x_k\}.$$

We now have all the necessary information to calculate our sample size.

When conducting Phase 2 clinical trials, we are interested in comparing the difference between two treatments. Let λ be the rate for the placebo or control group, then let $\lambda\theta$ be the rate for the treatment group. It follows that if $\theta > 1$, then our treatment is worse than the control since there is a higher rate of adverse events, if $\theta = 1$ then the two are equal, and if $\theta < 1$ then our treatment is better than the control. The variable θ serves the same function as δ used in the logistic regression model, it represents the difference in the two groups. In this method of determining sample size, we are looking to find the smallest sample size n, given our parameter of interest θ, such that,

$$E\{I[P(\theta > \theta_0|x_1, \ldots x_n) > 1 - \alpha]\} \geq 1 - \beta. \tag{8}$$

This equation is the expectation, or expected probability, that we will reject the null hypothesis. The function $I()$ is an indicator function, a function that equals either 1 or 0 based on certain criteria. Here, $I()$ is 1 if we reject our null hypothesis, and 0 if we fail to reject our null hypothesis.

Similar to the logistic regression sample size calculation done above, this calculation is also simulation based. First, we must select a value for θ assuming that the alternative hypothesis is true, and denote that value as θ^*. Using our value for θ^* and prior distribution for λ, (6), we generate fake data for our control and treatment group, where $x_1 \sim \text{Poisson}(n\lambda)$ and $x_2 \sim \text{Poisson}(n\lambda\theta^*)$. We then use the Monte Carlo simulation to produce a posterior probability $P(\theta > \theta_0|x_i)$. The Monte Carlo simulation method randomly samples from our generated data and calculates the proportion of times $\theta > \theta_0|x_i$. To use this method priors must be determined for all parameters. Recall that $\lambda \sim Gamma(a_i, b_i)$, and we will select a non-informative prior for θ, $\theta \sim \text{Gamma}(0.01, 0.01)$.

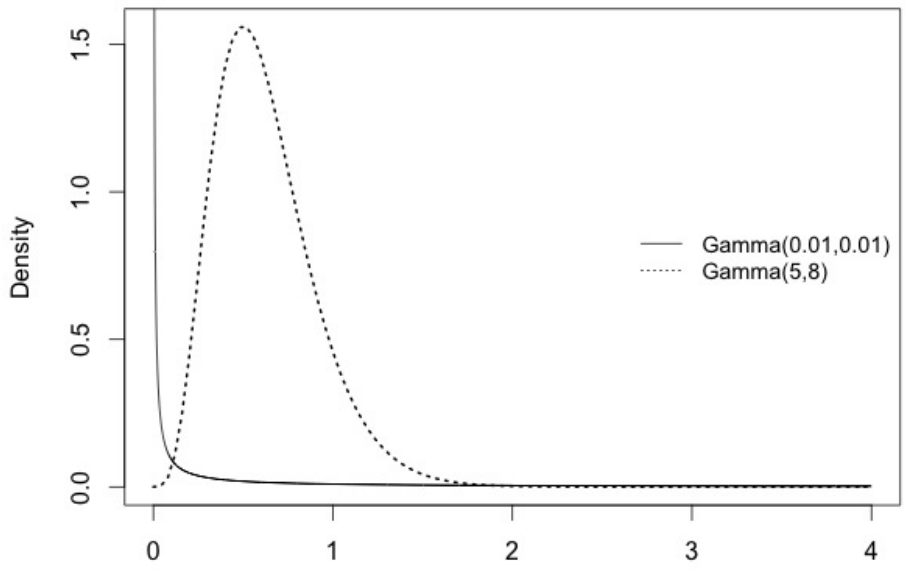


Figure 4: This figure shows the difference between an informative and non-informative gamma prior.

This gamma distribution has a mean of 1 and variance of 10. From the graph [4] we see that there is a lot of weight at zero however the rest of the prior is relatively flat, indicating that we do not have a lot of information about our parameter because the probability of any value above 0 occurring is relatively equal. Comparing it to a more informative prior, it is easier to see why this gamma distribution doesn't provide much information.

Using the posterior probability, those with $P(\theta > \theta_0|x_1, \ldots, x_n) > 1 - \alpha$ are assigned a value of 1, and others a 0 in a vector created, $\mathbf{v}$. Taking the average over vector $\mathbf{v}$ we have

an estimate of the average power. This process is repeated for different selected values of n, and then a spline fit is used to find the sample size n that satisfies equation 8.

4.2.3 Example

In this example we will be looking at thromboembolisms, the formation of a clot in a blood vessel that is released and plugs another blood vessel. We are specifically looking at this occurrence in the mitral valve, which is between the left atrium and left ventricle in the heart. CarboMedics mitral valves can be used to replace damaged mitral valves. The study done was a case series, a study that follows a group of patients who have a similar diagnosis and are undergoing the same treatment. We will only be looking at the data for the first four series completed.

Using the data given in Table 2, we can optimize the marginal log-likelihood (7) to determine the MMLEs. Thus, the priors are $\lambda_1 \sim \text{Gamma}(0.68, 4.22)$ and $\lambda_2 \sim \text{Gamma}(1.75, 8.72)$. Note that Time (n) are expressed in 0.1*patient*years.

First Four Seres	
Events (x)	Time (n)
16	43.1
12	58
0	67.7
6	80

Table 2: Thromboembolisms Data

We will now go over a calculation of the average power sample-size determination. We would like to find the smallest sample size such that we can compare old and new CarboMedics mitral valves. If the new mitral valves are better than the old, then $\theta < 1$. Thus, the following are our hypotheses,

$$H_0 : \theta = 1 \quad and \quad H_a : \theta < 1.$$

The following will be our prior distributions for the clinical trials,

$$\theta \sim \text{Gamma}(0.01, 0.01),$$
$$\lambda \sim \text{Gamma}(0.68, 4.22).$$

Let $\theta^* = .8$, a possible value for theta assuming the alternative hypothesis is true. Lastly, we choose a power of 80%, this is a common power to choose because it is strong, but doesn't require as big of a sample size for a power of 90% or 99%. Now, we can use simulation methods to test the average power of several sample sizes, n, and will select the smallest n for which we get an average power of 80%. Using R code we find that the smallest sample size with a power of 80% is $n \approx 188.2$ person years.

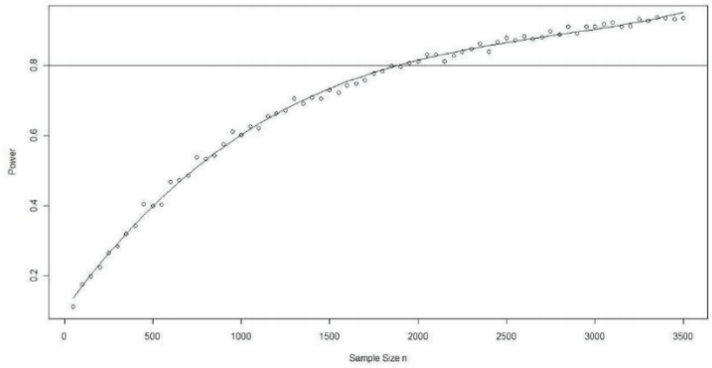


Figure 5: The graph resulting from simulations showing the power on the y-axis, and person×years×0.1 on the x-axis.

4.2.4 Limitations to the Bayesian Approach

The greatest limitation to using the Bayesian approach is having to design a prior distribution for all unknown parameters. Using different priors can produce different results, meaning that choosing the incorrect prior can compromise all analysis of results. Similarly, because there are no set steps on how to create a prior, the one chosen and used by a researcher could be considered unfit by another researcher. Thus, although the Bayesian approach is useful in it's ability to incorporate preexisting information, the process of actually creating a distribution from that preexisting data can be difficult and could result in incorrect results.

There are times when many priors might be necessary, like when using a hierarchical model, which makes not only choosing priors more labor intensive, but can also make analysis computationally extensive. With more complex models, calculations are not only difficult, but the difficulty can give room to more possible errors. There are, however, methods and techniques being formulated to assist with these calculations. For example, the Markov Chain Monte Carlo algorithm is a commonly used method.

5 Bayesian Adaptive Clinical Trials

Adaptive clinical trials are characterized by the adaptive features incorporated in their design. These adaptive features are planned changes in the design of the trial that occur at interim points and are guided by the acquired data thus far. Interim analysis is the analysis of data while the study is still in progress. It is important that any changes that happen during a clinical trial are planned before any data are collected. Unplanned changes are questioned because they may introduce additional bias and compromise the legitimacy of results. It is also important that changes made during a clinical trial are based on information from the specific study, and not solely on external events. Some possible design modifications that may be included in the design are dosage levels, total sample size of the study including early termination, schedule of patient evaluation for data collection, etc.

There are clear advantages to adaptive experimental designs, and the most significant

is that the allowance of mid-study adjustments decrease the time and resources needed to determine the efficacy or futility of a treatment. Currently, it can take about 10 years, and millions of dollars to bring a new drug into the market, however the amount of drug development failures far outnumber the amount of successes. Thus, the costs of failed drug development is far greater. Adaptive designs allow researchers to terminate unsuccessful candidates earlier, while spending less money than required for other experimental designs. Researchers can have predetermined stopping rules to protect patients from dangerous or pointless treatments. Usual stopping rules include analysis that confirms superiority: there is a clear benefit for the experimental group, inferiority: there is clear harm to an experimental group, or futility, where no difference is seen. These are usually predefined statistically. Additionally, adaptive clinical trials may reach a similar conclusion as non-adaptive clinical trials but more efficiently, and may provide a better understanding of the drug being tested. For example, dosages may be adjusted during a study, rather than having to design and conduct a separate study for another dosage level.

There are also, however, disadvantages to the use of this experimental design. A major concern is the abuse of the decreased amount of resources needed and speed of these studies, as well as a lack of understanding of these new techniques. The worry is that researchers will conduct statistically unsound adaptive studies to gain fast results, which could produce incorrect results. The interim changes can introduce bias which can lead to false conclusions. Similarly, the conduct flaws could also potentially end up hurting the participants or future patients. The misuse of adaptive clinical trials can also waste additional resources because trials that had flawed data would need to be conducted again. [3] [14]

5.1 Adaptive Clinical Trial with Poisson Outcomes

There are many different types of adaptive clinical trials, here we will be focusing on trials that allow for sample size adjustments and early termination in Phase II Trials. The following will be a continuation of a summary of Hand (2011).

Similar to the example in Section 4.2.3, count data of negative events will be used and the data will follow a Poisson distribution. Thus the greater number of events indicates a lack of efficacy in the treatment. Let y_i be the number of occurrences of the event of interest for the i^{th} patient such that $y_i \sim \text{Poisson}(\lambda)$ where λ represents the mean number of adverse events for the i^th patient. The parameter λ is a function of both the time each participant was in the study and event rate θ.

This Phase II study will have two stages. After the first stage, the data will be analyzed to determine possible changes in the second stage. The design will be outlined in the following way; a max number, or threshold, of events will be predetermined for the first and second trials as a way to measure efficacy and if the trial should be continued.

More formally stated, let r_1 be the threshold number of events for the first stage, and r the threshold number of events across the two stages. Then, r_1 will determine if the trial should continue to the second stage, and r will determine the efficiency of the treatment after both stages. The first stage of this Phase II clinical trail could conclude in one of two ways,

1. if $s_1 = \sum_{i=1}^{n_1} y_i \geq r_1$ then the trial is stopped and we conclude that the treatment

wasn't effective,

2. if $s_1 = \sum_{i=1}^{n_1} y_i < r_1$ we would continue on to stage 2 of the treatment,

where s_1 is the observed number of events over a period of patient time t_1 and n_1 is the number of patients in the first stage. If the clinical trial continues on to the second stage, n_2 patients are enrolled. Then, again, the two possible conclusions are,

1. if $s = \sum_{i=1}^{n} y_i \geq r$ then the trial is stopped and we conclude that the treatment wasn't effective,

2. otherwise we continue on to Phase III,

where $s = s_1 + s_2$ is the total number of observed events over time $t = t_1 + t_2$, and $n = n_1 + n_2$ is the total sample size across both stages.

In a two stage Phase II adaptive clinical trial, we can specify two rather than one prior for θ, a *design prior* and an *analysis prior*. The analysis prior is used in the actual analysis of the data, and reflects the amount of uncertainty about the treatment and some prior knowledge. The analysis prior is usually noninformative to allow the data to provide the most influence on the posterior distribution. On the other hand, the design prior is more informative and incorporates more prior knowledge. This prior is used to calculate the prior predictive distribution of the data, and is usually centered around the alternative hypothesis value(s). The design prior is used to calculate a sample size that would result in a large posterior probability of rejecting the null hypothesis, under the assumption that the alternative is true. Care must be taken when creating the design prior because they are very influential on the sample sizes. If the design prior strongly supports the alternative hypothesis, r and t will be smaller. And vice versa, the less support the design has for the alternative, the greater r and t will be.

Just as above, we will use the conjugate gamma distribution to make calculations simpler. A more informative gamma will be used for the design prior, $\xi_D(\theta)$ with the parameters (a_D, b_D), and a noninformative gamma will be used for the analysis prior, $\xi_A(\theta)$ with the parameters (a_A, b_A). The analysis prior will be centered around the placebo, or control treatment, rate (θ_0). Thus, $S_1 \sim Poisson(t_1\theta)$ where

$$f(s_1|t_1) = \frac{(t_1\theta)^{s_1}}{(s_1)!}e^{(-t_1\theta)} \qquad \text{for} \qquad s_1 = 1, 2, \ldots, \infty.$$

The variable S_1 is the number of observed events in the first stage.

Using this equation we can calculate the respective posterior distributions for the analysis prior. The calculations will be similar to (6) and will yield another gamma distribution; $\theta|S_1 \sim \text{Gamma}(s_1 + a_A, t_1 + b_A)$. Using the design prior, the prior predictive density of S_1 can be calculated. Note that this is similar to finding the marginal density of S_1,

$$
\begin{aligned}
m_D(s_1) &= \int_0^\infty \xi_D(\theta) f(s_1|\theta)\, d\theta \\
&= \int_0^\infty \frac{(b_D)^{a_D}}{\Gamma(a_D)} \theta^{a_D-1} e^{-\theta b_D} \frac{e^{-t_1\theta}(t_1\theta)^{s_1}}{s_1!}\, d\theta \\
&= \frac{(b_D)^{a_D}(t_1)^{s_1}}{\Gamma(a_D)\Gamma(s_1+1)} \int_0^\infty \theta^{s_1+a_D-1} e^{-\theta(t_1+b_D)}\, d\theta \\
&= \frac{(b_D)^{a_D}(t_1)^{s_1}}{\Gamma(a_D)\Gamma(s_1+1)} \frac{\Gamma(a_D+s_1)}{(t_1+b_D)^{(a_D+s_1)}} \int_0^\infty \frac{(t_1+b_D)^{(a_D+s_1)}}{\Gamma(a_D+s_1)} \theta^{s_1+a_D-1} e^{-\theta(t_1+b_D)}\, d\theta \\
&= \frac{\Gamma(a_D+s_1)}{\Gamma(a_D)(s_1+1)} \left(\frac{b_D}{t_1+b_D}\right)^{a_D} \left(\frac{t_1}{t_1+b_D}\right)^{s_1}
\end{aligned}
$$

where $\xi_D(\theta)$ is the design prior and $f(s_1|\theta)$ is the Poisson likelihood of S_1.

The above are all used during the first stage of the clinical trial. Suppose that the first stage has ended, therefore s_1 and t_1 are known. We have determined that we may continue on to the second stage of the clinical trial, or $s_1 < r_1$. We can update our Poisson probability mass function using the new data,

$$
f(s|S_1 = s_1) = \frac{((t-t_1)\theta)^{(s-s_1)}}{(s-s_1)!} e^{-(t-t_1)\theta} \quad \text{for} \quad s = s_1, (s_1+1), \ldots, \infty. \tag{9}
$$

as well as our analysis prior to a gamma with the parameters $s_1 + a_A$ and $t_1 + b_A$. Note that the gamma distribution is the calculated posterior distribution for the original analysis prior. The posterior distribution for the second stage analysis prior is a gamma distribution with parameters $s + a_A$ and $t + b_A$. The prior predictive distribution of S using the design prior is

$$
\begin{aligned}
m_D(s) &= \int_0^\infty \xi_D(\theta|s_1) f(s|\theta, s_1)\, d\theta \\
&= \frac{\Gamma(a_D+s)}{\Gamma(a_D+s_1)(s-s_1)!} \frac{(b_D+t_1)^{(s_1+a_D)}(t-t_1)^{(s-s_1)}}{(n+b_D)^{(s+a_D)}}.
\end{aligned}
$$

where $\xi_D(\theta|s_1)$ is the improved design prior and $f(s|\theta, s_1)$ is the likelihood of the second stage sampling distribution (9).

We have derived all of the necessary equations needed to carry out the analysis once data is collected, however before the first stage of this clinical trial begins, we must find t_1, patient time, and r_1, the first stage threshold. To find t_1, we will find the smallest patient time such that for all $\tilde{t_1} \geq t_1$

$$
P_D[F(\theta_A; a_A + s_1, b_A + \tilde{t_1}) \geq \lambda_1] \geq \gamma_1. \tag{10}
$$

The function P_D is the predictive density calculated above, and $F(\cdot)$ is the distribution function for the posterior distributions for the first stage. Let λ_1,$\gamma_1 \in (0,1)$, where λ_1 is comparable to a confidence level, and γ_1 is comparable to power. This equation is looking at the predicted probability that the probability of θ_A being outside the confidence level λ_1 is greater than or equal to γ_1. We can see that (10) bears a resemblance to (8).

Once we have determined t_1, r_1 can be found by maximizing

$$F(\theta_A; a_A + s_1, b_A + t_1) \geq \lambda_1$$

over S_1, or $r_1 = s_1^* + 1$ where

$$s_1^* = \max(s_1 \in 0, 1, 2, \ldots, \infty : F(\theta_A; a_A + s_1, b_A + t_1) \geq \lambda_1).$$

Once the values for these parameters are found, the first stage of the clinical trial may begin. At the end of the first stage, the observed number of events s_1 will be known, and calculations to determine whether we should continue with the second stage can be computed.

Similar to t_1 and r_1, to find the t, the total patient time, involves finding the smallest t, for all $\tilde{t} \geq t$ such that

$$P_D[F(\theta_A; a_A + s - s_1, b_A + \tilde{t} - t_1) \geq \lambda_2] \geq \gamma_2 \tag{11}$$

and r is found by maximizing

$$s^* = \max(s_1 \in 0, 1, 2, \ldots, \infty : F(\theta_A; a_A + s - s_1, b_A + t - t_1) \geq \lambda_2)$$

when $r = s^* + 1$ and t is known. Depending on the calculated r and t values, the second stage of the clinical trial may now begin.

5.2 Example

In this example we will not analyze any data, but simulate how we would find the desired sample size and threshold values.

Primary humoral immunodeficiency is a life threatening condition that results from impaired antibody production because of a defect in the number or function of B-cells or antibody-mediated immunity. Individuals with this disease often have serious upper and lower respiratory bacterial infections. To treat this disease, Intravenous Immune Globulin is commonly used. It is a blood product with blood plasma that contain antibodies. We would like to demonstrate the effectiveness of this treatment. The FDA recommends the study have a significance level of 0.01 that shows a serious infection rate of less than 1 per person-year in adults. Using these guidelines, our adaptive two stage Phase II design will have the following hypotheses,

$$H_0 : \theta_0 = 1,$$
$$H_A : \theta_a = 0.5.$$

First we must define our design and analysis priors. Instead of just using one, we will demonstrate the effect of several on our parameter estimation calculations. In order to make

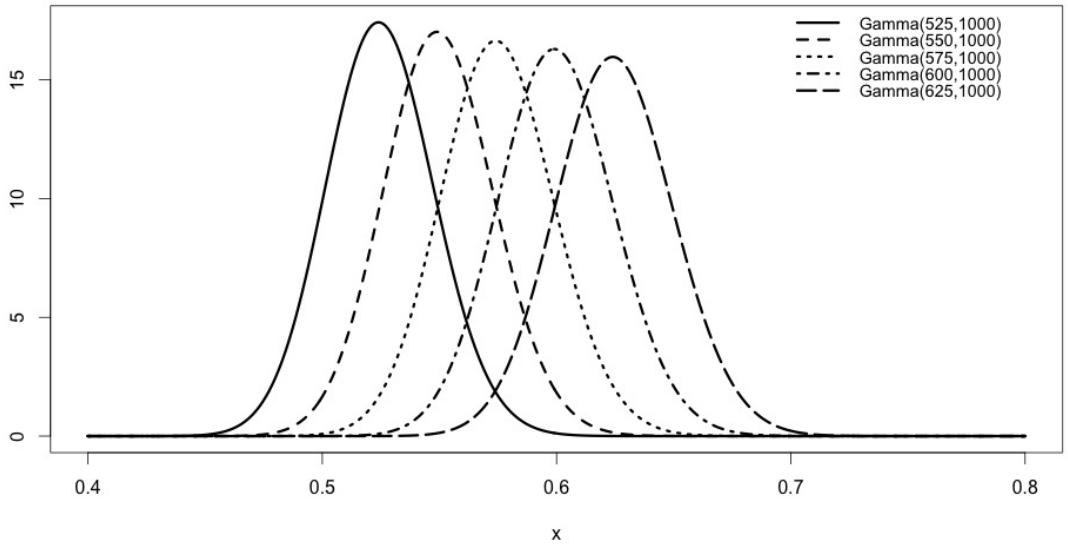


Figure 6: This graph shows the possible design priors that will be used in our calculations

sure the design prior is informative, the variance will be kept small and fixed, $b_D = 1000$. Although this design prior does support the alternative hypothesis, it does not do so very strongly as to let our conclusions be more influenced by the data. The first stage parameter values must be estimated first, and we will let $\lambda_1 = .99$, our confidence level, and $\gamma_1 = .8$ be our power for our calculations.

	$a_A = 0.04, b_A = 0.01$		$a_A = 0.4, b_A = 0.1$	
a_D	t_1	r_1	t_1	r_1
525	34.4	23	36	24
550	39.4	27	41	28
575	44.2	31	55.8	32
600	51.1	37	53	38
625	59.8	44	61.3	45

Table 3: Calculated first stage sample sizes for different analysis and design priors, for $b_D = 1000$, $\lambda_1 = .99$, and $\gamma_1 = .8$

From [Table 3], we can see that increasing a_D increases the sample size needed. This makes sense because if we compare the graphs of the different design priors, [Figure 6], we see that increasing a_D moves the mean closer to $\theta = 1$, and it will require a greater sample size to detect a smaller difference between the two treatments.

Given these values for the first stage parameters, we will calculate our second stage parameter values for various values of s_1 and for the various priors shown in [Table 3]. This will allow us to see the effect that the first stage values have on the second stage of our

clinical trial. Table 4 shows the total sample size t and threshold r.

	$a_A = 0.04, b_A = 0.01$					$a_A = 0.4, b_A = 0.1$				
				99% CI					99% CI	
a_D	s_1	t_1	r_1	Lower	Upper	s_1	t_1	r_1	Lower	Upper
	22	39.4	27	0.299	0.945	23	41	28	0.305	0.939
	23	39.4	27	0.318	0.977	24	41	28	0.323	0.970
550	24	39.4	27	0.336	1.009	25	41	28	0.341	1.000
	25	39.4	27	0.355	1.041	26	41	28	0.359	1.031
	26	46.7	33	0.374	1.073	27	48.2	34	0.377	1.061
	27	44.3	31	0.349	0.982	28	45.8	32	0.355	0.977
575	28	44.3	31	0.367	1.010	29	45.8	32	0.371	1.004
	29	44.3	31	0.384	1.038	30	45.8	32	0.388	1.031
	30	52.7	38	0.401	1.066	31	54.2	39	0.404	1.058
	33	51.5	37	0.390	0.989	34	53	38	0.393	0.983
600	34	51.5	37	0.405	1.012	35	53	38	0.408	1.006
	35	51.5	37	0.420	1.036	36	53	38	0.423	1.029
	36	61	45	0.435	1.059	37	62.5	46	0.438	1.052

Table 4: This table shows the stage two sample sizes and thresholds, as well as the 99% confidence intervals on θ, for $b_D = 1000$, $\lambda_1 = .99$, and $\gamma_1 = .8$

We see that for $a_A = 0.04$, $b_A = 0.01$, and $a_D = 550$, that for $s_1 \leq 25$, the sample size t is equal to the first stage sample size t, and the total threshold r is equal to the first stage threshold r_1. This indicates that the clinical trial should terminate after the first stage. The reason why we would terminate early even though $s_1 < r_1$ is because we have significant evidence that supports the alternative hypothesis, and the second stage of the clinical trial is not needed. On the other hand, looking at $a_d = 575$, for $s_1 \geq 30$, the second stage of the clinical trial will continue with the indicated r and t values.

6 Conclusion

Although many methods for sample size determination exist, we can see that the Bayesian methods provide more flexibility and variety in the clinical trial design process. This approach allows the incorporation of information that is already known, as well as the uncertainty of that information. Even if no information exists, the Bayesian method should still be used because no values are being estimated, and the information from the study conducted can be later used to form a prior distribution. Additionally, converting from a Frequentist inference to Bayesian can be difficult, thus sticking to Bayesian will allow the results of one study to be made into a prior for the next.

We can also see that useful components of the Frequentist approach, like significance level and power, can be incorporated in our analysis. Additionally, with the incorporation of adaptive elements, the trial can be adjusted based on information that may not have been completely known before the trial began. And the added ability to terminate trials early not only saves time but also resources. The greatest reservation to using this approach is having to determine a prior distributions, however as Bayesian methods continue to become more common, more methods will be produced to help with this process.

References

[1] Murphy, S. A. (2005). An experimental design for the development of adaptive treatment strategies. Statistics in medicine, 24(10), 1455-1481.

[2] Collins, L. M., Murphy, S. A., & Bierman, K. L. (2004). A conceptual framework for adaptive preventive interventions. Prevention Science, 5(3), 185-196.

[3] Park, M., Nassar, M., Evans, B. L., & Vikalo, H. (2012, August). Adaptive experimental design for drug combinations. In Statistical Signal Processing Workshop (SSP), 2012 IEEE (pp. 712-715). IEEE.

[4] Cirulli, J., McMillian, W. D., Saba, M., & Stenehjem, D. (2011). Adaptive trial design: its growing role in clinical research and implications for pharmacists. American Journal of Health-System Pharmacy, 68(9), 807-813.

[5] Hobbs, B. P., & Carlin, B. P. (2007). Practical Bayesian design and analysis for drug and device clinical trials. Journal of biopharmaceutical statistics, 18(1), 54-80.

[6] Snedecor, G., & Cochran, W. (1937). Statistical methods. (6th ed.). Ames, Iowa: The Iowa State University Press.

[7] Montgomery, D., & Runger, G. (2003). Applied statistics and probability. (3rd ed.). New York, NY: John Wiley & Sons, Inc.

[8] Quinn, G. P., & Keough, M. J. (2002). Experimental design and data analysis for biologists. Cambridge University Press, United Kingdom: Cambridge University Press.

[9] Cook, J. D., Jairo, A., & Pericchi, L. R. (2011). Skeptical and Optimistic Robust Priors for Clinical Trials.

[10] The fda's drug review process: Ensuring drugs are safe and effective. (2012, May 05). Retrieved from `http://www.fda.gov/drugs/resourcesforyou/consumers/ucm143534.htm`

[11] Hand, A. L. (2011). Bayesian sample-size determination and adaptive design for clinical trials with poisson outcomes (Doctoral dissertation)

[12] DeGroot, M., & Schervish, M. (2012). Probability and Statistics (4 ed.). : Addison Wesley Publishing Company Inc.

[13] (August 3, 2012). *Explore Clinical Trials* Retrieved from `https://www.nhlbi.nih.gov/health/health-topics/topics/clinicaltrials/`

[14] Food and Drug Association. (2010) *Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics*. Washington, DC: U.S. Government Printing Office.