

# Chapter 13

## Clinical Trial Design and Sample Size Calculation



In Chapter 12, we considered the many limitations of the currently available treatments for heart disease; because these invasive procedures are expensive and have serious side effects, there is an important global need to develop more cost-effective ways to prevent heart disease. In the early 1990s, a series of small, highly publicized studies offered hope that a simple intervention - taking high doses of vitamin E (**Figure 13.1**) - might reduce the risk of developing cardiovascular disease by as much as 40%. [1, 2] A subsequent randomized clinical trial compared the rates of myocardial infarction and cardiovascular death in a group of 1,035 patients taking vitamin E to those in a group of 967 patients taking a placebo; results were reported in 1996 and also showed that vitamin E provided a protective effect. [3] Yet, a pivotal clinical trial involving 9,541 patients in 2000 indicated that there might be no reduction in the risk of cardiovascular disease for those who take vitamin E supplements and when these patients were followed for more than 7 years, it appeared that the use of vitamin E supplements may actually increase the risk of developing heart failure. [4, 5]

This series of studies provides a good illustration of both the process and challenges of clinical research. Generally, early studies involving small numbers of patients provide data which allow us to generate hypotheses; these hypotheses must then be tested rigorously in larger clinical studies. Because of inherent biological variability, it is not uncommon for results of early studies to be contradicted by larger, more rigorously designed studies. In fact, a recent study compared conclusions presented in highly cited articles to those of subsequent studies with larger sample size or better controlled design. [6] Results showed that nearly 1/3 of highly cited studies were later contradicted and that this was most likely for studies where patients were not randomly assigned



**Figure 13.1.** There has been considerable disagreement among different epidemiologic and clinical studies designed to determine whether taking vitamin E supplements can reduce the risk of developing cardiovascular disease.

[http://4spectrum.us/catalog/images/E\\_travel\\_upright\\_spill.jpg](http://4spectrum.us/catalog/images/E_travel_upright_spill.jpg)

**Descriptive Statistics:**

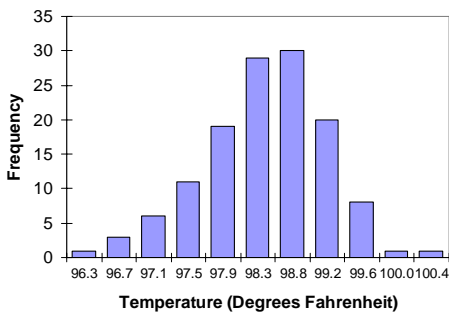
We can use descriptive statistics to analyze the average core body temperature for normal adults. The following dataset contains the core body temperature from a group of 65 men and 65 women. We can calculate parameters which measure the central tendency of the data as well as the dispersion in the data.

96.3 97.6 98.2 98.7 97.2 98.1 98.6 99.0  
 96.7 97.7 98.2 98.7 97.2 98.2 98.6 99.0  
 96.9 97.8 98.2 98.8 97.4 98.2 98.6 99.1  
 97.0 97.8 98.3 98.8 97.6 98.2 98.7 99.1  
 97.1 97.8 98.3 98.8 97.7 98.2 98.7 99.2  
 97.1 97.8 98.4 98.9 97.7 98.2 98.7 99.2  
 97.1 97.9 98.4 99.0 97.8 98.2 98.7 99.3  
 97.2 97.9 98.4 99.0 97.8 98.3 98.7 99.4  
 97.3 98.0 98.4 99.0 97.8 98.3 98.7 99.9  
 97.4 98.0 98.5 99.1 97.9 98.3 98.8 100.0  
 97.4 98.0 98.5 99.2 97.9 98.4 98.8 100.8  
 97.4 98.0 98.6 99.3 97.9 98.4 98.8  
 97.4 98.0 98.6 99.4 98.0 98.4 98.8  
 97.5 98.0 98.6 99.5 98.0 98.4 98.8  
 97.5 98.1 98.6 96.4 98.0 98.4 98.8  
 97.6 98.1 98.6 96.7 98.0 98.5 98.8  
 97.6 98.2 98.6 96.8 98.0 98.6 98.9

Using the definitions in the text, we find the following values:

Mean core temperature: 98.25 F  
 Median core temperature: 98.3 F  
 Mode core temperature: 98 F  
 Standard Deviation: 0.73 F

Using a bin size of 0.4 F, starting at a minimum value of 96.3 and ending at a maximum value of 100.4, we find the frequency histogram shown below.



[http://www.amstat.org/publications/jse/jse\\_data\\_archive.html](http://www.amstat.org/publications/jse/jse_data_archive.html)

Journal of Statistics Education, Volume 4, Number 2 (July 1996).

Source: [7].

to a treatment or control group!

In this chapter, we will examine how clinical studies and clinical trials are designed. Our goal is to develop the tools to properly interpret and use the results of clinical research. We begin by examining how to characterize biological data and its associated variability. Next, we provide an overview of the different types of clinical studies and clinical trials used to generate and test hypotheses. Finally, we examine how to calculate whether a study to test a hypothesis includes sufficient numbers of patients to ensure that any resulting conclusions are based on real differences in the data and not just statistical variability.

As we consider these important issues, we will find that the process of designing clinical research studies and clinical trials has many similarities to the scientific method that we learned about in Chapter 7. In order to help form hypotheses, we begin by carrying out observational and epidemiologic studies. Ultimately, we must carry out carefully controlled, randomized clinical trials designed to isolate the effect of just one factor and determine whether it has the predicted impact. The challenge in this process is that because of biological variability, we must always ensure that studies include a large enough group of patients to be sure that any differences we see are real and not just due to chance.

**Descriptive Statistics**

In clinical research, we are constantly concerned with determining whether differences observed between groups of patients are due to an intervention given only to one group or are simply due to biological variability. Generally speaking, we carry out clinical trials in groups of patients that we believe are representative of the greater population. How do we characterize these groups? An important first step is to use descriptive statistics to assess demographic variables (e.g. gender, age) for the patients in our sample and to assess clinical parameters for these patients (e.g. blood pressure, serum nicotine levels).

When characterizing data like this, we generally need to assess three main factors: the central tendency of the data (e.g. mean, mode), the spread in the data (e.g. variance, standard deviation), and how the data are distributed across the range of possible values (e.g. normally distributed).[8]

There are several methods to characterize central tendency of the data. If we are interested in a single, continuous variable,  $x$ , we define the mean to be the average of measured values for our population.[8] The median is that value which occupies the middle rank when the values are rank ordered

from least to greatest. The mode is the most commonly observed value.

Similarly, there are several measures of the dispersion of the data. Two of the most common are the variance and the standard deviation. The variance is the sum of the squared deviations from the mean.

$$\text{Variance} = \sum_{i=1}^N (X_i - \bar{X})^2$$

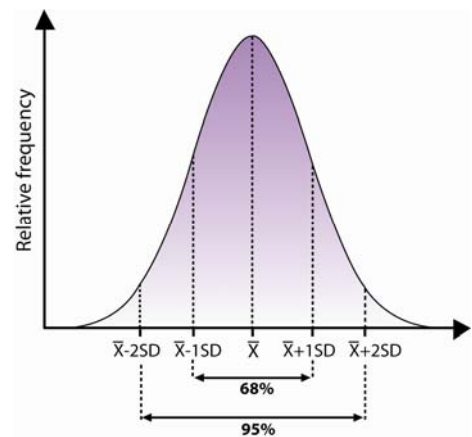
Equation 13.1

The standard deviation is the square root of the variance.

In order to examine the distribution of our data, we can display the data graphically in a histogram, in which we divide the data into bins and count the number of values that fall into each bin. A relative frequency histogram is a plot of the fraction of the observations that fall within that bin. The area under a relative frequency histogram is unity.

The distributions of data often tend to follow similar curves. One of the most commonly encountered distributions in biological data is the normal distribution (Figure 13.2). In normally distributed data, the relative frequency histogram is determined completely by the mean and the standard deviation, and the mean, mode and median are identical. One standard deviation on either side of the mean contains 68% of the area under the relative frequency histogram for a normal distribution. 95% of the data are contained within  $\pm 1.96$  SD of the mean

Figure 13.2. Normal distribution



### Types of Clinical Studies

There are two major types of clinical studies: those conducted with the goal of hypothesis generation and those designed to test a specific hypothesis. Hypothesis generating studies are often referred to as epidemiologic studies, while hypothesis testing studies are called clinical trials.

**Hypothesis Generation:** There are several types of studies which can be used to generate clinical hypotheses. For example, in a case study, we identify one patient and attempt to understand the factors related to their illness, either by reviewing their history or through physical examination or clinical testing. Similarly, in a case series, we identify a group of patients with a similar illness and try to ascertain similar factors that may be responsible for their disease. We have already seen a good example of a case series. Recall that in 1981, a series of four previously healthy homosexual men were identified who developed pneumocystis carinii pneumonia and mucosal candidiasis—the first description of AIDS and its association with be-

**Infant Diagnosis: June 13, 2007**

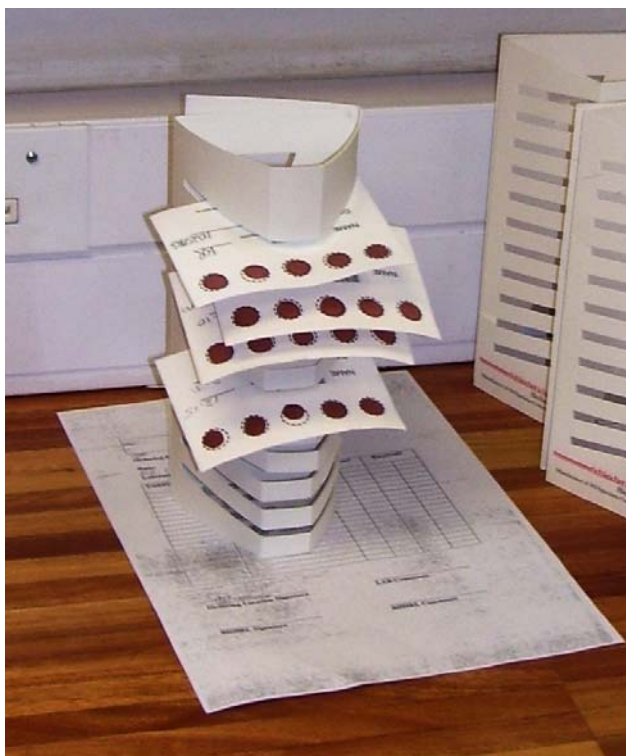
**Kim**

**Malawi**

The project that I'm going to be spending the rest of my summer working on is a Malawi Ministry of Health pilot project that is supported by a variety of NGOs, including BIPAI (Baylor International Pediatric AIDS Initiative), UNICEF, the Clinton Foundation, etc. It is a part of their "HIV-Free Generation" goal.

This project aims to identify HIV exposure or infection as early as possible to begin providing care as early as possible. Ideally, all pregnant women will be routinely screened for HIV at their prenatal care visits (ANC). Those that are pregnant will get CD4 counts, be clinically staged and given a single dose of Nevirapine (NVP) to take at the onset of labor. If they qualify (by CD4 count or WHO stage) they can be referred to an ARV clinic and started on ARVs. The single dose of NVP isn't needed for women on full ARV therapy, but does reduce mother to child transmission by about 50% if the women aren't otherwise on therapy. (The baby is also supposed to receive a dose when it is born for this to be most effective.)

Since this is obviously not ideal, right now the program is also involved in identifying infants below the age of 18 months, which is an age group that cannot be identified by the currently available rapid HIV tests. Those tests check for anti-HIV antibodies in the blood, but babies can have their mother's antibodies (passed through the placenta) for up to 18 months. But, we are now able to do DNA PCR testing on dry blood spots. These tests identify viral DNA in the blood, which you only have if you are infected. It can identify babies as early as 6 weeks old. The only problem is that while a positive is a definite positive, a negative is not a definite negative in infants that age. Continued breastfeeding is continued exposure, so infants have to be retested 6 weeks after they stop breastfeeding.



And breastfeeding is a huge problem. It would be best to not do it at all, but babies would almost all starve to death. Formula is far too expensive. So most women breastfeed. In that situation, it would be best to breastfeed exclusively for six months and then rapidly wean to solid foods/juice. But they also give other foods (mixed feeding). The problem with mixed feeding is that it actually increases the chances of transmission through a couple mechanisms. First, anything other than breast milk acts as an irritant and makes micro perforations in the baby's intestines, essentially breaking a barrier between the baby's body and maternal HIV. Secondly, the baby's body recognizes anything but mother's milk as foreign, which initiates an immune response, drawing CD4 cells to the intestinal tract, and thereby bringing the HIV's host cells into close proximity to the virus. Many mothers start mixed feeding as young as 2 months old because they are culturally pressured to

do so (if you don't give your baby other food as young as possible, you clearly don't love them enough) and continue with the mixed feeding until nearly 2 years old.

Right now our job is to train health care providers at these 7 pilot sites about the new testing algorithm and procedures. We then provide a week to two weeks of onsite support and coaching to make sure they get it right. We are also attempting to identify potential HIV+ kids in their pediatric wards, under 5 clinics and outpatient clinics.

So, that's what I'm up to. It's very interesting and very exhausting.

## Clinical Trials

behavioral risk factors.

**Hypothesis Testing:** Once we have formulated a hypothesis, there are a number of different types of clinical studies which can be used to test the hypothesis. Hypothesis testing studies are divided into two major groups: observational and experimental. In an observational study, we identify a group of patients with disease and a similar group of patients without disease. We collect data from these two groups of patients and compare them to test our hypothesis. Observational studies have the advantage of being relatively simple and inexpensive to carry out, but they can suffer from important problems. One of the most common is bias—suppose we identify a group of patients with lung cancer and a matched group of healthy controls. We notice that the lung cancer patients report an increased frequency of consuming alcohol, and we conclude that alcohol consumption may be related to the development of lung cancer. It is also possible that alcohol consumption is strongly correlated with cigarette smoking and that smoking is actually the factor which is responsible for the increase in lung cancer. In an observational study, we can't control the administration of the intervention to decisively show cause and effect—we can only retrospectively analyze differences in our two groups of patients.[9]

To avoid this and other sources of error, we can design experimental approaches to test our hypothesis. A clinical trial is a research study to evaluate the effect of a new diagnostic or therapeutic in a group of patients.[11] Clinical trials help us to determine whether new interventions are safe and effective. Clinical trials may involve a single group of patients or multiple groups of patients, but in general, their goal is to isolate all but a single variable and measure the effect of that variable. Clinical trials are done in a prospective way: they are first planned, then data are collected; this enables the investigators to manipulate the intervention in a controlled setting to show cause and effect.[9]

A single arm study involves a single group of subjects; subjects receive an intervention and then we monitor them to see if their condition improves.[12] Usually, we compare the change in their condition before and after the intervention (change from baseline) so that each patient serves as their own control. One problem with a single arm study is that improvements following an intervention may be partially due to the placebo effect—where patients or their physicians think they are getting better simply because they believe they are receiving an intervention—and this can change their perception of symptoms.

### How To Find a Clinical Trial:

The National Institutes of Health maintains a publicly accessible database of clinical trials to test experimental therapies for serious or life-threatening diseases conducted with federal and private support.

<http://clinicaltrials.gov>

The database provides a summary of the purpose of the study, patient recruiting status, criteria for patient participation, the location of the trial, the research study design, the phase of the trial, and the disease or condition and drug or therapy under study. Currently, information can be found for more than 36,000 ongoing trials.

Source: [10].

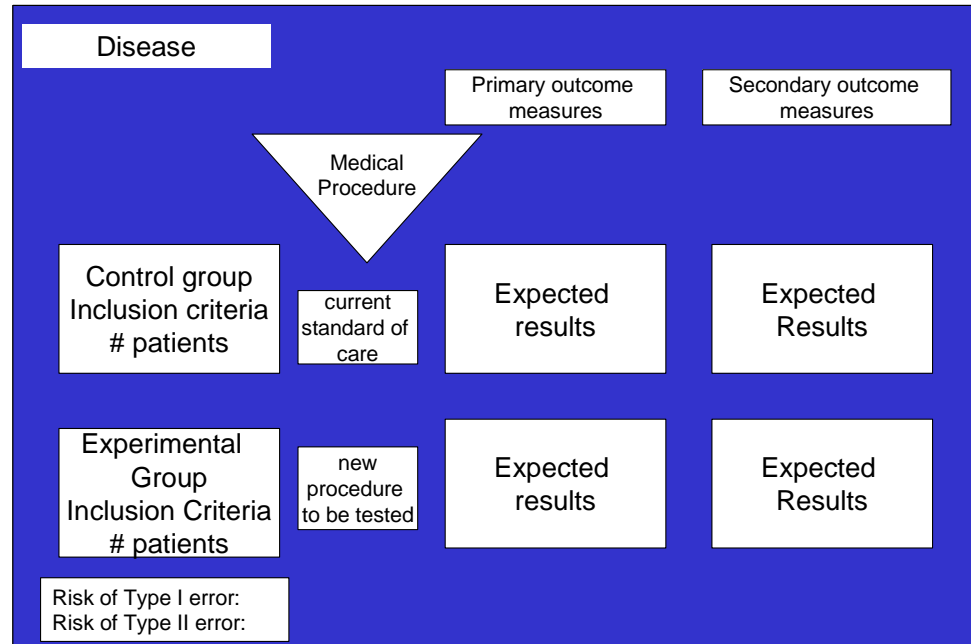
### Clinical Trials: Drugs

Clinical trials of new drugs are carried out in different phases to assess the safety and efficacy of the new drug.

In a phase I trial, researchers assess the safety of the drug in a series of 20-80 normal volunteers.

In a phase II trial, the experimental drug is given to a larger number of volunteers (100-300) to assess both its safety and efficacy.

In a phase III trial, the new drug is given to a large number of patients in order to compare its effectiveness to standard therapies and to monitor for side effects. Typically, phase III trials are randomized, placebo controlled, double-blinded studies. It is especially important to carry out a rigorous calculation of the sample size required for phase III trials.



**Figure 13.3:** To design a clinical trial, you must be able to fill out all the information in the template shown.

To avoid this problem, clinical trials can be carried out using two groups of patients, one of which receives the intervention and the other which receives a placebo. Randomized clinical trials, in which different subjects are randomly assigned to different treatment and control groups, are generally considered to be the strongest type of clinical evidence. [12] In a double-blinded, randomized clinical trial, participants are randomized in a way so that neither the participants nor their physician knows which group they are part of. This type of study design eliminates conscious bias. Subjects must be randomized in a way so that the treatment and control groups are similar on average; in this way we can attribute any differences in the outcomes between the groups to the intervention and not any other factors that may have differed between the groups.

**Sample Size:** The sample size is the number of subjects in each arm of a study needed to detect a predetermined difference.[13] Determining the necessary sample size for a clinical trial is a careful balancing act—as the sample size becomes larger, we reduce the risk that differences between the two groups arise due to chance, but the cost and complexity of doing the study increase. Our goal in setting a sample size for a randomized clinical trial is to appropriately balance these factors.

In general, setting a sample size is a complex statistical calculation. However, for data which can be described by the normal distribution, there is a relatively straightforward process to estimate the number of patients required. Let's examine the process to determine the sample size for a two-

**Weekend in Mochudi and Watching the Zebras (soccer team): June 18, 2007**

**Rachel**

**Botswana**

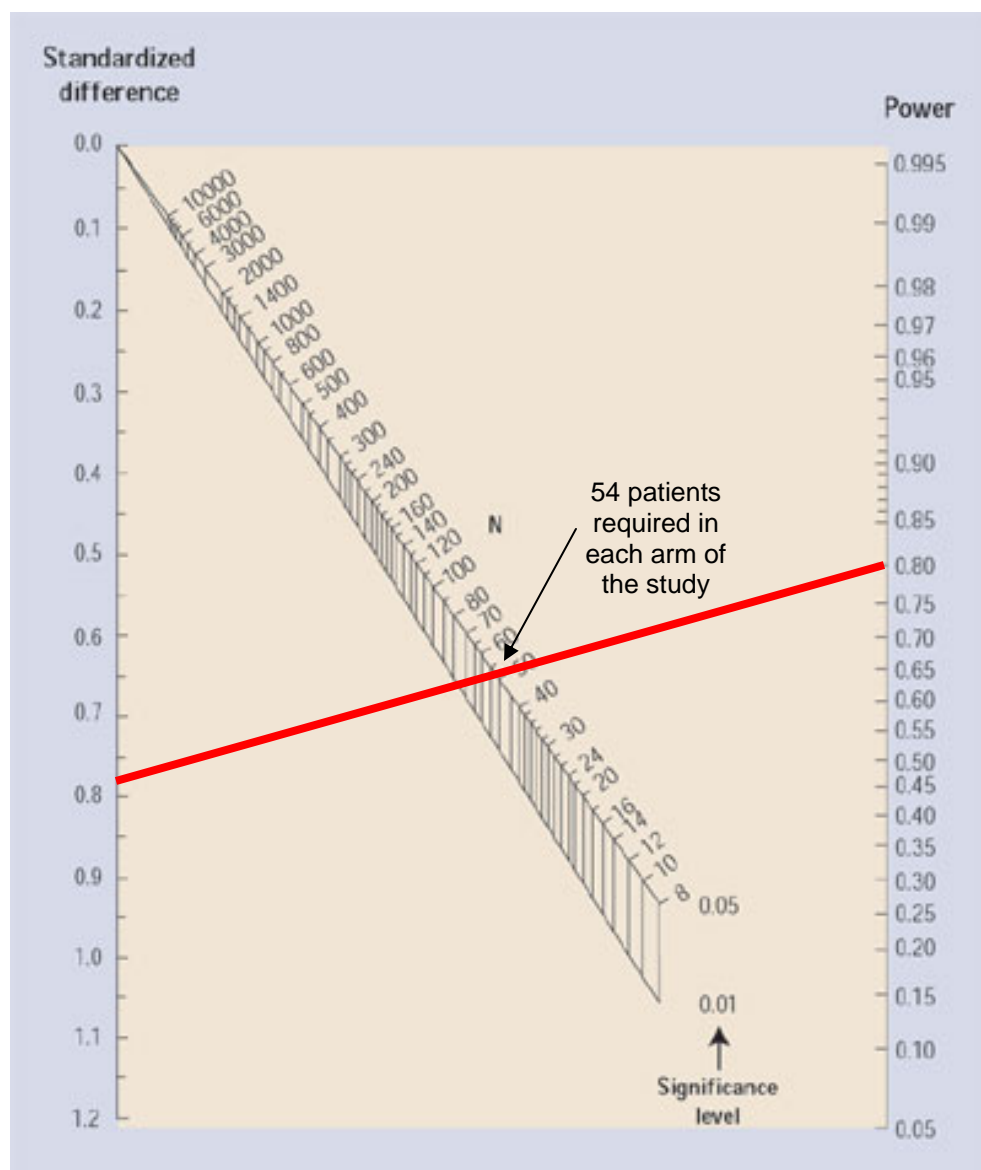
After a PAC doctor, Heather's suggestion and Liz's approval, we went ahead and started screening patients in hopes of casting a larger net to catch the patients with less than perfect adherence ( although for the most part, all the patients in Gabarone have good adherence, especially when compared to Francistown, a clinic where the pharmacist constructed a list of 31 reasons why he would not do pill counts, and is described as eccentric...) which brings me to an idea for another project at Rice which would not be life changing here, but would be a pleasant convenience- a pill counter!

I am thinking something akin to those change counters where you dump the change on top and it sorts it and counts it. The most important part would be cost, simplicity and ease of use so that at the Francistown clinic, there would be no excuse for not pill counting, a practice which reminds me of DOTS (directly observed therapy) for TB. While not being exactly direct, DOTS is helpful for making sure people continue to take their pills and be counseled when they miss doses.

After talking to the nurse head, Mamapula ( which means Mother of Rain) I found out that we probably should've been communicating better. She said she was a little in the dark as to what exactly we were doing, but after about 5 minutes explaining, she was good to go and helped show us how the system works at reception - where to pick up the binders of the patients to be seen-and how to do the pill counts - where to find the pharmacy sheet of how many pills the patient left with. I guess we hoped to expedite the process in the exam rooms with the doctors, but we weren't able to 'catch' all the patients that we had hoped to, only ending up with a couple. On the other hand, considering that the patients are taking their pills so well, this is great news for the clinic!

Tomorrow we will go back to the data extraction project in an effort to see the trends in the resistance tests ( RT) that patients who were failing on line 1 or 2 of the ARVs. RTs are expensive and usually only done in extreme circumstances but the results of collecting the mutations of the HIV are pivotal in letting the doctor know which drugs will be ineffective and which ones will work. <http://hivdb.stanford.edu/pages/algs/HIVdb.html>





**Figure 13.4:** Altman's nomogram, used to calculate the sample size required in each arm of a study, given the expected standardized difference, desired power and significance level. Note that for the results obtained using Altman's nomogram to be valid, we assume that the continuous variable follows a normal distribution in our population, the binary variable follows the normal approximation to the binomial distribution and the two groups being compared are of equal size.

arm, randomized clinical trial. In this type of trial, we have both a treatment group (receives the intervention) and a control group (receives a placebo). We will monitor the clinical outcomes in each group. We must select a primary outcome—this is the outcome we are most interested in comparing between the two groups. In selecting our sample size, we want to ensure that any differences between the treatment and control group are real, knowing that there will be some statistical uncertainty associated with the primary outcome. We will choose our sample size so that this uncertainty is sufficiently lower than the difference in primary outcome we wish to detect between the control and treatment groups. **Figure 13.3** provides a template which illustrates all of the information needed to calculate the sample size for such a trial.

## Clinical Trials

What happens if we choose our sample size to be too small? Essentially, there are two types of mistakes that can result. The first is that we mistakenly conclude that there is a difference between the two groups, when in reality there is no difference. This is called a Type I error, and can result in adopting an intervention, even when it is not truly effective. The second type of error that we can make is to mistakenly conclude that there is not a difference between the two, when in reality there is a difference. This is called a Type II error and can result in us discarding a potentially effective intervention. Our sample size will be dictated by the risk that we are willing to accept for making each of these types of errors.

The risk of making a Type I error is called the significance level of the study. Typically, studies are designed so that the risk of making a Type 1 error (significance level) is between 1% - 5%. The risk of making a Type II error is usually denoted by the variable  $\beta$ . Typically, studies are designed so that the risk of making a Type II error is between 10% - 20%. Often, we refer to the power of a study, where the power is defined to be 1—the probability of making a Type II error. The power of a study usually ranges from 0.8 to 0.9.

With these definitions, we can now outline the process of calculating a sample size.[14] First we must select a primary outcome. Our primary outcome can be either a binary measure (did the patient have a heart attack?) or a continuous variable (what is the patient's systolic blood pressure?). If the outcome variable is binary, we must estimate the expected rate of the primary outcome in the treatment group and the control group. If the primary outcome variable is continuous, we must estimate the size of the difference in the average outcome for the two groups and the expected standard deviation associated with this variable. Next we set acceptable levels of Type I and II errors. This determines the p-value and the power of the study. With this information, we can use a graphical tool called Altman's nomogram to estimate the sample size required in each group.

To use Altman's nomogram, we must first calculate the standardized difference we expect to find in the primary outcome variable. For a binary outcome variable, the standardized difference is calculated as:

$$\text{Standardized Difference} = \frac{P_1 - P_2}{\sqrt{p(1-p)}}$$

Equation 13.2

### The importance of the minimum clinically important difference:

"Samples which are too small can prove nothing; samples which are large enough can prove anything."

Source: [14].

$$\bar{p} = \frac{P_1 + P_2}{2}$$

Equation 13.3

where  $P_1$  is the fraction of patients in the treatment group who experience the primary outcome and  $P_2$  is the fraction of patients in the control group who experience the primary outcome.

For a continuous outcome variable, the standardized difference is calculated as:

$$\text{Standardized Difference} = \frac{\text{Effect Size}}{\text{Standard Deviation}}$$

Equation 13.4

where the Effect Size is the difference in the average value of the primary outcome we expect between the treatment and control groups and the Standard Deviation is that expected for the primary outcome variable.

We use Altman's nomogram ([Figure 13.4](#)) to determine the number of subjects required in each group by drawing a straight line to connect the standardized difference we wish to detect and the desired power of the study. The point at which this line intersects the line drawn at the desired significance level gives the sample size required for each group.

Suppose we wish to carry out a randomized clinical trial to compare the effectiveness of a new drug eluting stent to that of the standard uncoated stent. We design a study where patients requiring treatment for coronary artery disease are randomized to receive either the standard stent or the new drug eluting stent. We decide that our primary outcome will be to follow the patients and determine whether or not they experience restenosis during the first 6 months following the procedure. This is an example of a binary outcome variable, since each patient either does or does not experience restenosis. A pilot clinical trial with the new drug-eluting stent suggests that only 10% of patients experience restenosis following treatment with this device. A review of the literature shows that, on average, the number of patients who experience restenosis within 6 months following treatment with a standard stent is approximately 45%. We wish to design a trial with a 5% significance level and a power of 0.8, meaning that we are willing to take a 5% chance of

**Clinic: Dosing Guide: June 22, 2007**

**Lindsay**

**Botswana**

As we still attempt to fashion some sort of trial for our dosing guide, we are working on another project Dr. Lowenthal presented to us. For our “trial”, we just don’t have enough patients with poor adherence to create any kind of pool big enough to make our results somewhat reliable. For now, we’re for the most part just giving everybody dosing guides and not creating a control group (those w/o guides). As unscientific as this is, the doctors and nurses are referring patients to us whom they think really need the guide, at a rate of 1-2 per day. Given the current situation, it makes more sense to just allow the people who need guides to get them—we won’t be able to bring together a “scientific” study. There are many confounding factors in creating a “scientific” study, mainly that all patients referred to us will also have to go through adherence class again, which itself may raise adherence rates. I would rather rely on the patients’ comments when they come back to the clinic for their next appointment to see if the guide is helpful. I think this may end up being a lot more subjective than we all expected, but I’m okay with that.

We had an 11-year-old patient who came in the other day with her mother, who had recently taken custody of her due to the death first of a nanny and then of an aunt, who had been caring for the girl very well and making sure she took her meds. After I created a dosing guide for the mother (the girl wasn’t with us), she went back to the waiting room as usual. As I went back to the nurses to make sure the guide was correct, they told me about the family’s situation—the mother was quite irresponsible, and the 11-yr-old girl was taking the ARVs on her own! She had been doing a very good job despite the complexity of her regimen, but she still needed help, which she wouldn’t get from her mother. Her eyes lit up as I explained to her the dosing guide and how she could use it, as the mom’s eyes wandered around the room in boredom. I think we are better off helping the 5 or 6 children needing help who come in each week, rather than spending our time creating a trial. A doctor at clinic suggested that we try to set up a visit at one of the SOS Villages (orphanages) to see if we would be able to give all the “mothers” there dosing guides for all the kids who need them.



making a Type I error and a 20% chance of making a Type II error.

We begin by calculating the standardized difference:

$$\text{Standardized Difference} = \frac{P_1 - P_2}{\sqrt{\bar{p}(1-\bar{p})}} = \frac{0.45 - 0.1}{\sqrt{0.275(1-0.275)}} = 0.78$$

Equation 13.5

$$\bar{p} = \frac{P_1 + P_2}{2} = \frac{0.45 + 0.1}{2} = 0.275$$

Equation 13.6

With this, we turn to Altman's nomogram, and draw a line connecting a power of 0.8 to a standardized difference of 0.78. We look to see where this line intersects that representing a 0.05 significance level and see that 54 patients will be required in each arm of the study. In the case of a continuous outcome variable, we follow an identical procedure, except when calculating the standardized difference.

In summary, when designing and analyzing clinical trials there are four major parameters of interest:

1. The significance level of the study
2. The power of the study
3. The effect size one can detect
4. The sample size

For a fixed significance level and power, as we increase the sample size we can detect smaller and smaller differences in the outcomes between the treatment and control groups. It is important to place the effect size we wish to determine in the proper clinical context. There is an important difference between the statistically significant difference we can detect given the parameters above and what may be clinically meaningful. The smallest difference that is clinically meaningful is sometimes called the minimum clinically important difference (MCD). It is important not to waste time and money enrolling patients in order to detect differences that are substantially less than the MCD, since these differences are of no clinical importance.

We have seen that in order to calculate a sample size, we need to estimate the difference that might exist between the treatment and control groups and the expected variance in the data. In practice, the difference may be larger than predicted or the spread in the data may be smaller than expected. This, of course, is the dream of every investigator. In this case, the

## *Clinical Trials*

number of patients required to reach statistical significance will be smaller than initially predicted.[13] However, it is also possible that the differences between the treatment and control groups will be smaller than predicted or the spread in the data will be larger; in order to reach significance, more patients will be required than originally estimated.[13]

Frequently, special committees called Data Safety and Monitoring Boards (DSMB) are convened in order to monitor interim data in clinical trials to ensure the safety of participating subjects.[15] The federal government requires that a DSMB be used to monitor all phase III clinical trials. Typically, a DSMB is composed of scientists and clinicians who are knowledgeable about the field of interest, but are not otherwise involved in the study. The role of the DSMB is to analyze all adverse events reported in a trial and to perform interim analysis of clinical outcomes to determine whether a trial should be stopped early.

If an interim analysis indicates that a new treatment is substantially better than the standard of care, then the study may be stopped early and the new treatment offered to both the study group and the control group.[11] Conversely, if an interim analysis reveals substantially increased risk associated with a new treatment, then a study may also be stopped early to prevent additional harm. Making decisions to stop a trial early are frequently difficult, because they force researchers to make judgments about when interim results cross the boundary from suggestive to conclusive. Making decisions to stop a trial or proceed with a trial following an interim analysis can also raise serious ethical questions.

For example, a recent trial was carried out to assess the efficacy of a new treatment for severe sepsis (blood stream infection).[15] Researchers designed a trial to compare a new drug, tifacogin, to that of a placebo; the primary outcome was mortality. The trial was designed with a sample size of approximately 1500 subjects. An interim analysis conducted after 722 patients had been accrued showed a mortality rate of 38.9% in the placebo group and 29.1% in the tifacogin group. This difference was significant at the 0.006 level. The DSMB was forced to decide whether to stop the trial early, given the apparently lower mortality rate associated with the new drug. Ultimately, the board decided that the results were not sufficiently compelling and the trial was continued. Neither the researchers nor the prospective trial participants were informed of the interim results.

Was this decision scientifically sound? Was it ethical? The interim analysis, although not definitive, suggested that the new drug reduced mortality by about 25% compared to placebo. If the researchers had been informed of the interim results, it

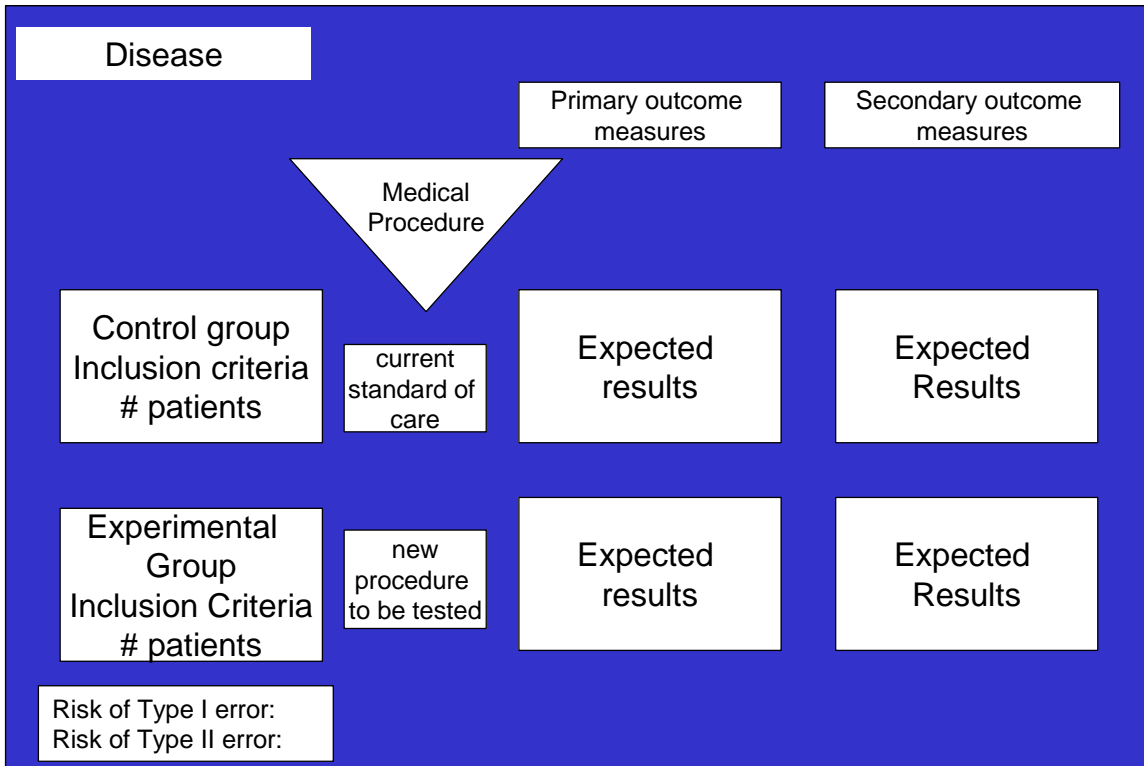
might have made them more hesitant to enroll their patients in the trial and to follow the study protocol. In the end, recruitment was continued to the final sample size. When all the data were analyzed, it was determined that mortality in the tifacogin group was 34.2% compared to 33.9% in the placebo group—a difference that was neither clinically nor statistically significant.

To address these ethical concerns regarding decisions made by DSMBs, it has been recommended that informed consent documents be modified to indicate if a trial will be monitored by a DSMB. If so, during informed consent, the role of the DSMB should be explained, noting that the DSMB may make recommendations to continue a trial even in the face of evidence suggesting effectiveness of one of the treatments. The informed consent process should make clear that, in the interests of maintaining the scientific integrity of the study, interim results will not necessarily be made available to patients enrolled in the study.[15]

**Bioengineering and Global Health Project**

**Project Task 10: Design a clinical trial to test your new technology.**

In this task, you will design a clinical trial to test the new technology that you have developed. As part of this design, you must choose the subjects to be tested including the control group (receives standard of care) and the experimental group (receives new technology). You must specify the primary and secondary outcomes that you will monitor in the trial. You must calculate a statistically justified sample size for the trial. Complete the table below and provide a one page summary of the trial design and sample size justification.



**Chapter 13 Homework**

1. Complete the Human Participant Protections Education for Research Teams course which can be found at: <http://cme.cancer.gov/clinicaltrials/learning/humanparticipant-protections.asp>. This course presents information about the rights and welfare of human participants in research. The tutorial is designed for those conducting research involving human participants. Complete the exercises at the end of each of the six content areas. Print a certificate of completion upon completing the course.
2. You are a researcher working on a vaccine for malaria, which is caused by a parasitic protist. You are interested in performing clinical trials for your vaccine.
  - a. During Phase I testing, you must recruit about 100 healthy volunteers. What are you trying to determine through this testing?
  - b. During Phase II testing, what should your malaria vaccine be tested against?
  - c. Phase III testing involves a double-blind study. What does this mean and why is it important?

**3.** You are designing a study to test a new implantable artificial kidney for patients with end-stage renal disease. You divide your patients into two groups: one will receive an implanted artificial kidney, the other will receive tri-weekly hemodialysis (standard of care in end stage renal disease). Your primary endpoint is mortality for all causes at one year, a secondary endpoint is patient quality of life at one year, which will be assessed via questionnaire.

- a. What is a type I error? What are the possible consequences of making a type I error in this study? What is a type II error?
- b. What are the possible consequences of making a type II error in this study?
- c. Define the p-value.
- d. Define power (if you use Greek letters in your definition, you must define these as well).
- e. Why would blinding be difficult in this study?
- f. Which of your endpoints is more likely to be affected by a lack of blinding?
- g. Assuming you expect 30% mortality at one year for the control group and 20% for the treatment group, what sample size would be required to achieve 80% power? What p-value should you use? Justify this value.

**4.** You are designing a clinical trial to compare the performance of a new thrombolytic agent to dissolve blood clots associated with acute myocardial infarction. You design a trial to compare the new agent to the standard of care, streptokinase. You choose the 30 day mortality as your primary endpoint. There will be some statistical uncertainty associated with the measured mortality rate in the treatment and control groups. Your goal in selecting the sample size for the trial is that this uncertainty be significantly less than the difference in the mortality rate between control & treatment group. We must set acceptable levels for the risks of type I and II error.

- a. Define type I error and type II error.
- b. Suppose you expect a mortality rate in the group treated with the new drug stent of 5%, while the expected restenosis rate in the group treated with the current stent is 7%. You calculate a standardized difference of 0.19. If you can tolerate a 20% risk of type II error and a 5% risk of type I error, how many patients are needed in the trial? Use the following figure to indicate how you calculated your answer.
- c. If the mortality rate for the new drug was expected to be 1%, would the required sample size increase or decrease?
- d. List one secondary outcome you would want to monitor in this trial.

**5.** Consider the Acute Respiratory Distress Syndrome Network trial of low versus traditional tidal volume ventilation in patients with acute lung injury and acute respiratory distress syndrome published in 1996. Mortality rates in the low and traditional volume groups were 31% and 40%, respectively, corresponding to a reduction of 9% in the low volume group. What sample size would be required to detect this difference with 90% power using a cutoff for statistical significance of 0.05?

## References

- [1] Rimm EB, Stampfer MJ, Ascherio A, Giovannucci E, Colditz GA, Willett WC. Vitamin E consumption and the risk of coronary heart disease in men. *The New England journal of medicine*. 1993 May 20;328(20):1450-6.
- [2] Stampfer MJ, Hennekens CH, Manson JE, Colditz GA, Rosner B, Willett WC. Vitamin E consumption and the risk of coronary disease in women. *The New England journal of medicine*. 1993 May 20;328(20):1444-9.
- [3] Stephens NG, Parsons A, Schofield PM, Kelly F, Cheeseman K, Mitchinson MJ. Randomised controlled trial of vitamin E in patients with coronary disease: Cambridge Heart Antioxidant Study (CHAOS). *Lancet*. 1996 Mar 23;347(9004):781-6.
- [4] Yusuf S, Dagenais G, Pogue J, Bosch J, Sleight P. Vitamin E supplementation and cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. *The New England journal of medicine*. 2000 Jan 20;342(3):154-60.
- [5] Lonn E, Bosch J, Yusuf S, Sheridan P, Pogue J, Arnold JM, et al. Effects of long-term vitamin E supplementation on cardiovascular events and cancer: a randomized controlled trial. *Jama*. 2005 Mar 16;293(11):1338-47.
- [6] Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *Jama*. 2005 Jul 13;294(2):218-28.
- [7] Shoemaker AL, College C. What's Normal?-Temperature, Gender, and Heart Rate. *Journal of Statistics Education*. 1996;4(2).
- [8] Neely JG, Stewart MG, Hartman JM, Forsen JW, Jr., Wallace MS. Tutorials in clinical research, part VI: descriptive statistics. *The Laryngoscope*. 2002 Jul;112(7 Pt 1):1249-55.
- [9] Bernson M. Tutorial: An Introduction to Clinical Trials Part I of II: Purposes and Phases. *The Next Generation*. 2005;1(4).
- [10] U.S. National Institutes of Health. *ClinicalTrials.gov*. 2007 [cited 2007 June 7]; Available from: <http://www.clinicaltrials.gov/>
- [11] The Patient Education Institute. *Clinical Trials*. MedlinePlus 2005 January 25 [cited 2007 June 7]; Available from: <http://www.nlm.nih.gov/medlineplus/tutorial.html>
- [12] Bernson M. Tutorial: An Introduction to Clinical Trials Part II of II: Statistics and Experimental Design. *The Next Generation*. 2005;1(5).
- [13] Neely JG, Karni RJ, Engel SH, Fraley PL, Nussenbaum B, Paniello RC. Practical guides to understanding sample size and minimal clinically important difference (MCID). *Otolaryngol Head Neck Surg*. 2007 Jan;136(1):14-8.
- [14] Ogundipe L. Sample size determination in clinical research: 2. *Hosp Med*. 2000 Nov;61(11):797-8.
- [15] Slutsky AS, Lavery JV. Data safety and monitoring boards. *The New England journal of medicine*. 2004 Mar 11;350(11):1143-7.