L A B

2

Experimentation



to test the hypothesis.

This week's lab covers an introduction to experimental design. Experimentation is the heart of science. In fact, modern science considers only evidence that is based on direct empirical observation during rigorously controlled experiments (those that can be repeated by others). Even though health professionals must understand and prepare scientific reports, the scientific methodology can seem daunting and even a bit boring. We hope this lab will convince you that the challenge is more interesting and understandable than many scientists project in overly sophisticated books and periodicals. There is an art to the scientific process. After all, most decisions in life do not have solid scientific explanations telling us exactly what is best to do.

We start with a discussion of the hypothesis from an insider's view. Most likely this is not a view that was presented in your previous science classes where you memorized the steps of the scientific method. You will learn some of the inherent bias and error "built into" the scientific method. It is a bit of a philosophy lesson. After a consideration of the hypothesis, you will learn some of the particular tools of experimentation including controls, blind and double-blind studies, and placebo effects. In addition, we offer a recommendation about when to stop experimentation and when to apply known health knowledge (The 85% Rule). All of this is just a beginning consideration of how science is used and misused in the health professions.

Exercise #1	The Hypothesis
Exercise #2	Experimental Design
Exercise #3	The 85% Rule
Exercise #4	Placebo Effects in Sport and Medicine



Should we try to prove the hypothesis correct, or try to prove it wrong?

Exercise #1 The Hypothesis

The fundamental goal of science is to understand nature through observation and experimentation. The scientific method begins with observation. There is something in the world that we want to understand. Observation is followed by offering basic questions and possible explanations for what we are observing. "Do people with more good health behaviors live longer?" That question is the beginning of a *hypothesis*. Although the hypothesis starts as a question, it is written as a declarative statement "People with more good health behaviors live longer." An experiment must then be designed to test whether the hypothesis is true or not. Often the hypothesis is stated in a very narrow and specific way because it is only a part of a much bigger explanation.

It turns out that the very specific way that you state a hypothesis leads to a particular experimental design, and also determines the kind of possible error that you might make in your study. Today, science generally agrees that all experiments should be designed to test what is called the *Null Hypothesis*. The simplest statement of a null hypothesis is: **Sample A = Sample B**

When we like an idea, we have a natural bias to try to prove it correct. Statisticians found that a better way is to test ideas by trying to prove them wrong. People are naturally very good at finding flaws in almost any idea. If we can't prove the idea wrong, then we feel more confident that the idea is probably correct. And we've avoided the natural bias of trying to prove our ideas correct. However, no matter how good science becomes at reducing mistakes and bias, there always remains some inherent error in its methods. Scientists accept that research errors do happen. Therefore, error is one of the first topics taught when scientists are trained to use the scientific method. There are two kinds of error that can be made when a researcher tests a null hypothesis.

Table 2.1	. Error When	Using the	Scientific	Method.
-----------	--------------	-----------	------------	---------

Two Possible Errors When Testing the Null Hypothesis				
	The null hypothesis is actually true			
Type I Error	(Sample A = Sample B), but the experimenter			
	mistakenly concludes it is false.			
	The null hypothesis is actually false			
Type II Error	(Sample A \neq Sample B), but the experimenter			
	mistakenly concludes it is true.			

- **? Question** 1. Is the hypothesis written as a question?
 - 2. State the null hypothesis in your own words.
 - 3. Is the correct approach of science to try to prove the hypothesis right, or to try to prove it wrong?
 - 4. The way that you state the null hypothesis determines the kind of ______ that you might make in your conclusions.
 - 5. Define Type I error in your own words.
 - 6. Define Type II error in your own words.



You have learned that science can make two kinds of mistakes—Type I and Type II. The next question is, "Can we avoid each of these errors?" It turns out that you can reduce Type I error to as low a risk as you choose. For example, if you want to reduce possible Type I error when using the Standard Error Test (a statistical test that you will use in the Data Analysis Lab), then you can require that the two sample means must differ by 3 S.E. units in order to conclude that they are different. (You will use a threshold of only 2 S.E. units in this class.) If you want to reduce Type I error even more, you could require 4 S.E. units as the threshold. This means that in order to avoid Type I error, I am willing to require more evidence, and then even stronger evidence. I could set the bar so high that no one can get enough evidence to



That theory is worthless. It isn't even wrong! - Wolfgang Pauli (1900-1958) a pioneer of quantum mechanics

prove the hypothesis wrong. I am now opening myself to Type II error. The Null hypothesis is actually wrong, but I'm forcing it to be considered correct.

The dilemma of scientific testing is that when you reduce the chance of a Type I error, the chance of a Type II error automatically increases. This is because the factors creating these two kinds of error are linked. When either one of them is reduced, the other one tends to increase.



When scientists reduce Type I error, they must make adjustments in their experimental design (such as greatly increasing the sample size) to prevent Type II error from increasing out of their control. However, increasing the sample size will take more time and cost more money. Most investigators won't do it. So, if we are somewhat unwilling or unable to avoid error, how do we avoid a catastrophe when error does happen? One way of avoiding a catastrophe of Type I and Type II errors lies in the way that the null hypothesis is written. And there are two very important rules for stating the null hypothesis.

- **Rule #1:** The null hypothesis is written in the equality form. (Heart rates on a test day are the same as heart rates on a normal day.)
- **Rule #2:** The null hypothesis should be stated in the "safest and most conservative way" to minimize danger. It must be stated in such a way that a Type II error won't result in someone dying or something equally serious.

The main reason that scientists use the equality form for stating the hypothesis is because this sets up a situation in which they can attempt to prove the hypothesis wrong. We excel at thinking up experiments designed to try to prove an idea wrong. It is much more difficult to devise ways of proving that an idea is correct, especially when not much is known about the phenomenon. So, scientists deliberately set the Type I error to a very high threshold (low risk), and then design experiments to try to prove the



We love to prove an idea wrong.

hypothesis wrong. However, when they set the Type I error to a low risk, this increases the chance of a Type II error. Remember, Type II error means that the null hypothesis (Sample A = Sample B) actually is false, but the experimenter mistakenly concludes it is true. Because science uses the approach of reducing Type I error to a minimum, it is automatically vulnerable to Type II error. Therefore, Rule #2 must be followed (safe and conservative hypothesis) to avoid the "catastrophe of error."

- **? Question** 1. What are the two important rules for stating the null hypothesis?
 - 2. What is the value of stating the hypothesis in an equality form? (Sample A = Sample B)
 - 3. Does this help to reduce Type I error?
 - 4. Why do scientists try to prove ideas wrong, rather than trying to prove them correct?
 - 5. Does this help to reduce Type I error?
 - 6. Can scientists set the Type I error to as low a risk as they want?
 - 7. What happens as science reduces Type I error?
 - 8. Why is it usually a challenge to greatly increase the sample size?

9. What kind of error is science most vulnerable to?

```
Type I or Type II
```

10. What is it about science that naturally leads to Type II error?

What is a Safe Null Hypothesis?

During the early years of World War II, London was being battered by daily bombing attacks. Many people had been killed, and the city was in chaos. It was a bleak time for the British people. Then a truly amazing man, the Prime Minister, took control of the situation. His name was Winston Churchill, and he inspired his people to persevere.

One of Churchill's brilliant inspirations occurred when he ordered a group of the best British scientists and mathematicians to meet with him. This stern and pugnacious man walked into the room of experts and stared at each of them without saying a word. After what must have seemed like an eternity, he bellowed, "We will not cower to the scoundrels! London must go on!" Then he proceeded to give these men and women an incredible challenge. They were to determine which areas of London would be safe from the bombing—safe for work, safe for transportation, safe for schools and the nation's children.

The experts were shocked. They had been ordered to assume an enormous responsibility. What if they were wrong? People would die! This was their state of mind as they worked feverishly to develop a method of testing various hypotheses so they could identify which areas would be safe from bombing. Soon they came to realize that there were two very different kinds of null hypotheses – the safe and the risky.

Scientists test a hypothesis by comparing data collected from two samples or by comparing a sample to an expected value. Remember, the simplest statement of the null hypothesis is Sample A = Sample B. But, how we define Sample A and B becomes crucial! In the bombing of London, the safest null hypothesis was:

> "Zone A, an area they wondered if safe, is the same as known bombing zones (Zone B)." (Zone A is a bomb area.)



We will not cower to the scoundrels!

-Winston Churchill June, 1940

What if the hypothesis is false, but you think it's true?

With a "Safe" hypothesis, Nobody dies! Why is this the safest way to state the null hypothesis? Remember, science tries to prove the hypothesis wrong. Science is at its best and most certain when proving a hypothesis wrong. Statistical comparisons were designed to test this hypothesis. The statistical threshold for making a Type I error was set at an extremely low risk. If England's scientists were successful at disproving this hypothesis ("Zone A is a bomb area"), then they could be reasonably certain that they had actually identified a place where bombs would not fall—a safe area for England's people.

Most important of all, the scientists were able to protect themselves against the danger of making Type II error. Remember, science is most vulnerable to Type II error. Type II error in a safely stated null hypothesis would have led them to mistakenly conclude that their hypothesis was true (that Zone A was a bomb area), and they would warn people away. It would be a mistake. But, no one would die because of it!

The Risky Null Hypothesis

"Zone A, an area they wondered if safe, is the same as known unbombed zones (Zone B)." (Zone A is a safe area.)

This is the risky way to state the null hypothesis. It leads to catastrophe if a Type II error is made. A Type II error for this hypothesis would mistakenly conclude that Zone A is a safe area when it actually is a bomb area. People would die because of this mistake. It is a dramatic example that demonstrates why science uses only the "safe and conservative" way of investigating problems. A safe and conservative null hypothesis minimizes the danger coming from a Type II error. The "safe and conservative" null hypothesis became established in science. However, it is important to mention that not all researchers follow this guideline. Some even seem ignorant of it. By protecting against a Type II error, science makes it harder to prove their favorite idea correct. People don't like that! And there are bitter disagreements about what is a safe hypothesis.

1. Explain the importance of stating the null hypothesis in the "safest and most conservative way" to minimize danger.



But, what if it's false, and you think it's true?

- People die!

?Question

2. Using the same approach as science, decide which hypothesis is the "safest and most conservative" null hypothesis. (Circle your choice.)

Hypothesis A: Defendant is assumed to be innocent, and the trial tries to prove him guilty.

Hypothesis B: Defendant is assumed to be guilty, and the trial tries to prove him innocent.

3. If you made a Type II error using Hypothesis **A**, what would be the result?

Guilty man goes free or Innocent man goes to jail

4. If you made a Type II error using Hypothesis **B**, what would be the result?

Guilty man goes free or Innocent man goes to jail

5. Any hypothesis must be considered in a context of possible bias and statistical error. For example, research into global warming can be approached from two directions.

Hypothesis A: Area in question is an example of global warming.

Hypothesis B: Area in question is an example of non-global warming.

This is another example of a situation where we need to know which hypothesis would be the most dangerous. What is your opinion?

6. Another example of two different hypotheses involves the testing of drugs and treatments by the FDA.

Hypothesis A: The test drug (or treatment) is safe.

Hypothesis B: The test drug (or treatment) is unsafe.

The FDA requires the approach illustrated by Hypothesis **B** to be the acceptable method. Why have they chosen that approach?

 Proponents of alternative medicine and nutritional supplements generally use the research approach represented by Hypothesis A below. They argue that this is the best way to be "open to" new treatments that are alternatives to traditional medicine.

Hypothesis A: The test drug (or treatment) is effective.

Hypothesis B: The test drug (or treatment) is ineffective.

What is the difference between this approach and the one required by the FDA? What are your opinions?

"It is easy to obtain confirmations, or verifications, for nearly every theory – if we look for confirmations."

Karl Popper (1902-1994),
Philosopher, exponent of scientific falsification

Exercise #2 Experimental Design

In science, we try to make sure that the work done during the day gets us a little closer to understanding what is really going on in the world. *Experimental design* is the planning of an experiment so that we get a yes or a no answer to the question we are asking.

The Experimental Control

The scientific method begins with a possible explanation of something we are observing. That explanation is a called the *hypothesis*. It generates the questions we are trying to answer in an experiment. For example, maybe adding more water to a lawn will make it grow faster. We could design an experiment to test that hypothesis. In the *experimental design* we might decide to water the lawn twice as much as normal rainfall. Let's suppose that this experiment is done for a month, and we observe that the lawn actually does grow quite a bit faster. Can we safely conclude that the hypothesis has been tested?

What if someone were to point out that the weather was warmer than normal during our experiment, and that is why our experimental lawn grew faster than normal? Or suppose that someone else said that our lawn really didn't grow any more than her lawn which was watered only by rainfall during the same month?

> Because we can't answer these questions, our "experimental design" has failed. Why?

We did not control our experiment! An *experimental control* is a duplicate procedure that is set up exactly like the experiment except that the factor being tested (more water) is left out. So, in our experiment we should have monitored a nearby lawn that received only normal rainfall during the same warm month. Then we could compare the growth rate of that "control lawn" to our "experimental lawn" which was receiving extra watering. The results from observing the control lawn would have answered both questions from our critics.



If you fail to include the experimental control, your critics won't fail to notice.



All experiments need a control!



The blind study does not let the subjects know which treatment they are getting.

What are blind and double-blind studies?

As we said previously, there are disagreements about which methods are required to be "real" science. Although there is agreement on most guidelines, not everyone follows them with equal fervor. Even if they did, bias would still remain because it seems to be a universal human trait. The most rigorous and objective experimentation and analysis are based on the "*principles of negation*" – constantly try to prove your ideas wrong. Testing of popular health remedies does not follow this guideline, but health professionals usually do. One of the research tools spawned from the principles of negation is the blind experiment.

The *simple blind study* does not let the subjects know which treatment they are getting (experiment or control). The purpose of this design is to reduce placebo effects. In *double-blind experiments*, neither the researchers nor subjects know who is in the control or test groups. The purpose of this design is to reduce placebo effects and experimenter bias. A neutral thirdparty is the only one aware of the real grouping. All subjects are randomly assigned to be in either the test treatment or control group, and there can be an independent committee to supervise the study as a way of detecting any unexpected bias that may develop over time during the investigation. Sometimes an experimental treatment is so notably effective that everyone involved becomes aware that something is happening. Studies have been suspended in cases where tested drugs were so effective that it was unethical to withhold treatment to the control subjects and general public. Of course, there are disagreements in the "grey areas" of research findings, which is why the general principle of negation is valuable – "I should always try to disprove my hopeful drug."

?Question 1. What is the principle of negation?

- 2. Describe the control in a properly designed experiment.
- 3. Define simple blind study.
- 4. Define double-blind study.

Exercise #3 The 85% Rule

How many causes of an important health problem do you have to discover? When do you know enough to recommend treatments for an important health issue? Must you know everything that contributes to physical performance before saying anything really definite and helpful? We are swamped with a flood of new explanations that seem to come from everywhere – and so much competition of opinions. Do you say to yourself, "I thought there was a basic truth, but I don't know anymore"?



85% is enough. Get on with it!

We offer you the 85% Rule. This does not mean that you wait until you are 85% certain of something – we actually want you to be 95% certain of something. You'll learn about that in the Data Analysis Lab. Rather, we want you to focus on the central causes of 85% of a health problem. Determine what contributions to 85% of nutrition or performance improvements. In other words, when you have discovered the important variables that explain 85% of a particular situation, you can stop! If you don't stop, you will go on and on and lose sight of the important factors that would help patients or clients. Let us show you two situations where the 85% Rule can be used, and you will learn more examples throughout this semester's labs.

Case #1

What are you going to say when people come to you wanting to lose weight? Break into lab groups and take 5 minutes to come up with five things you would say to your patient (client).



Group Discussion

1.

2.

3.

4.

5.

Quickly check with other lab groups to see how many truly different suggestions were made by the entire class. How many? _____

Case #1 is on the short end of the 85% Rule. Physiologists have shown that only two variables are needed to explain body weight imbalance. Those two are energy input (Calories in food) and energy output (Calories used in metabolism). Of course, there are many individual factors affecting energy input and output, and each could be a tool for helping people. But in this case we have all that we need to begin our therapies. Consider what happens when we read one of many news articles where people don't stop with the already known explanations for weight imbalance.

News Flash!

"The Alabama group puts forth these ten explanations for obesity."

- 1. Sleep debt
- 2. Pollution
- 3. Air conditioning
- 4. Decreased smoking
- 5. Medicine
- 6. Age and Ethnicity
- 7. Older moms
- 8. Ancestors' environment
- 9. Obesity linked to fertility
- 10. Unions of obese spouses

This news report has focused your attention on ten new explanations. Instead of doing that, stay focused on energy input and energy output, and use any "tricks" that will help you to help your patient. These tricks are many, and some even come from researchers who delve beyond the 85% Rule. None of this discussion is to disparage the findings of those who do not stop with the 85% Rule. But, this rule will keep you from getting lost in the forest of competing explanations.

Case #2

Most of the time you will discover that a handful of variables will explain 85% of the problem you face. However, this next example is an exception. It requires more than 15 causes to explain 85% of the deaths in 2011. (See Table 2.2.) The causes of death are listed in decreasing order of impact starting with Heart Failure (24.3% of deaths) and ending with Infections caused by aspirating food or liquids (0.6% of deaths).



How many factors control weight imbalance?

Cause of Death	%	Cumulative %
Heart Failure	24.3	24.3
Cancer	23.3	47.6
Lower Respiratory	5.6	53.2
Cardiovascular	5.3	58.5
Accidents	5.0	63.5
Alzheimer's	3.4	66.9
Diabetes	2.8	69.7
Kidney Failure	2.1	71.8
Flu	2.0	73.8
Suicide	1.5	75.3
Blood Poisoning	1.4	76.7
Liver Failure	1.3	78.0
BP Kidney Failure	1.0	79.0
Parkinson's	0.9	79.9
Aspiration Infection	0.6	80.5

Table 2.2. The Leading Causes of Death Listed in Cumulative Order.(Government Statistics, 2011)



How many causes of death?

There is another variation of the 85% Rule to help in this case. Notice that the right-hand column of numbers is called Cumulative %. As you go down the list, all of the previous percent of deaths are being added together. These cumulative percents tell us how many more deaths are accounted for as we include more causes. By the time we get to #15 (Aspiration Infection) 80.5% of deaths have been accounted for.

The impact of each cause of death quickly decreases as we go down the list. It drops sharply at the third cause (Lower Respiratory), and another lessening happens with the sixth cause (Alzheimer's). Although the strict use of the 85% Rule doesn't help much in this case, we can see that nearly 50% of the death problem can be explained by causes #1 and #2. And, unless you want to chase after the entire list and beyond to get to 85%, you could stop with Cancer or perhaps Accidents. That would be using similar thinking to the 85% Rule. The focus is on the primary causes of a particular health problem. Your job now is to graph the Cumulative % of Deaths as More Causes of Death are included. Do that in Figure 2.1.





? Question

- 1. State the 85% Rule in your own words.
- 2. What is the advantage of using the 85% Rule?
- 3. What would be your greatest objection to using the 85% Rule?

Exercise #4 Placebo Effects in Sport and Medicine

The placebo is a special control treatment sometimes used in human research. Experiments in sport and medicine are usually designed to test whether a treatment will improve a health or performance problem. In setting up the experimental design there is a factor that we think may be affecting things. For example, the Coronary Drug Project (discussed next) tested whether niacin, aspirin, and a couple of other drugs benefit people with coronary heart disease. Those factors are called the *experimental treatments* in a scientific study. But in order to be more sure of the experimental treatment, we need to know what would happen without it. That situation (without the experimental factor) is called the *control treatment*. You must control your experiment. The control treatment in the Coronary Drug Project included people exactly like the experimental groups except they did not receive an experimental drug.

Placebos are a kind of control treatment sometimes used in human research. The placebo is set up exactly like a control, except the subject being tested is told that he is getting the experimental treatment. The person is then given an inert substance or a known ineffective therapy. If the person responds in any way different from "no effect", then that effect on the subject is called the **placebo effect**. It doesn't always happen, but when it does, what the heck is going on? And is it important in the health professions?



The placebo treatment showed a 50% reduction in mortality.



Is there any clue revealed by who responds to the placebo treatment?

The Coronary Drug Project (started in 1965) is an example of a carefully designed study that tested the effects of various drugs on heart disease. It also revealed the importance of a placebo effect. The placebo group received a sugar pill (lactose) instead of any drug treatment. All groups were compared to each other and to a large control group of people who were not part of the project and received no treatments at all. Some of the treatments (niacin and aspirin) showed benefits of about a 10% lower mortality than the placebo group. But the placebo group surprised researchers with its own benefit of 50% the mortality of the control group.

We can explain the beneficial effects of certain drugs, but why should placebo treatment have benefit? This is the central question to all placebo effects. And three possible explanations are presented in the CDP's report:

- The placebo has a genuine physiological effect, or
- People who were willing to fully participate in a study (not drop out) are basically healthier people than those not willing to participate, or
- People who were willing to participate in the study were more diligent in other aspects of their lives, and those habits led to better health.

Research has shown that certain areas of the brain become more active during the placebo conditions. But this does not prove that these brain areas are the cause of the effect - only that the brain is responding. Some brain areas produce more dopamine which could be part of the placebo effect. Regardless of which theory is correct, the "good" effects of the placebo are typically reversed when the subject is told that the treatment is a placebo with no value. Placebos can cause intoxication, allergic reaction, greater weightlifting ability, more endurance and speed, and many other interesting changes. Size and color of pills, "branding", high price, and even whether it is a capsule drug can matter. And it helps to see someone else who is already benefiting.

Alzheimer's patients show no placebo response. Children show more response than adults. And some people just seem more susceptible than others. The placebo effect for strength-building is exaggerated when someone very fit recommends the placebo treatment. Men are very responsive when it involves body strength. It works better on those who think it will work. Maybe the person believes, and the belief helps. Or maybe he believes and starts doing something different that is actually the beneficial factor. In some clinical conditions such as irritable bowel syndrome, pain and anxiety, high blood pressure, and other disorders there can be quite long lasting benefits – sometimes weeks, months, or even a year or two. About 30% of people show some placebo response, although there is 0% in some conditions (infection) and as high as 80% in targeted lab studies that don't have particular clinical application.



Men are very responsive to the placebo effect when it involves body strength.

So what can we conclude about placebos? The effects are significant but small in magnitude. Changes in physical strength are usually in the range of 2-5% above normal, and workout performance or event endurance rarely exceeds 10% benefit. There is evidence and strong opinion by some observers that the effect is too unreliable compared to "real" treatments. On the other hand, the placebo effect is so consistent that government regulation requires all drug testing to show results beyond normal control conditions and also beyond any previously demonstrated placebo effects. That is a standard indeed!

Some people argue that the clinical use of placebo treatment is very limited. After all, patients can get better on their own, and maybe that's all that is happening. But when you consider that opiates don't relieve pain 30-50% of the time, and some patients can't tolerate them anyway, what can you do? Placebos do work for pain relief on 20-40% of patients, so that is something. And doctors can use placebos to see if their patients are malingering. But the World Medical Association states that placebos should be used in only two situations:

- When a placebo is necessary for the testing of important medical treatments, or
- When minor conditions are being investigated, and patients who receive the placebo are not subject to risk.



Placebos are effective pain relievers for 20-40% of patients.

What about the marketing and sales of health treatments, weight loss products, sports and training supplements, etc.? Most of these deliver results in the same range as the placebo effect. How will you know if what you recommend to your patients and clients is beyond placebo? There are billions of marketing dollars spent on placebo treatments. What are you going to do? **?Question** 1. What is the key difference between a normal experimental "control" and the placebo treatment?

- 2. How did the drug niacin compare with the placebo in the Coronary Drug Project?
- 3. How much better was the placebo group than the control group?
- 4. What were the three explanations for the placebo effect reported in the Coronary Drug Project?

- 5. One group of people show no placebo response, and this suggests that the thinking brain is involved. Which group is that?
- 6. The current government standard for judging the effectiveness of a drug or treatment is that its benefits must be greater than
- 7. Both niacin and caffeine have noticeable effects on all people. Niacin makes you feel "warm in the face", and caffeine "peps you up". Researchers have proven that if either or both are added to a supplement, that supplement will have a doubling of the placebo effect. Discuss how this can influence the health and exercise industries.

8. How could you effectively use the placebo treatment in your profession? Be specific.

9. What health and exercise products use "branding" and "buff" pictures to promote their use? Why would these promoters spend large amounts of money to get endorsements of their products?

10. Is there any health or exercise product that can't be explained by the placebo effect? List them.

11. How could you counter the marketing of placebo products and treatments targeted at patients and clients of your profession? Be specific.



Group Discussion

Take 10 minutes to share your answers to questions 7-11 with other lab groups. List the five best ideas of the class.

1.

2.

3.

4.

5.