**DESIGNING RANDOMIZED COMPARATIVE EXPERIMENTS**

Recall that in an experimental study design, a researcher manipulates something and then measures the effect of that manipulation on some outcome of interest. Randomized comparative experiments involve the *random* assignment of subjects to one treatment or another in the study.

|  |
| --- |
| Comment on Randomization of Treatments |
| *Random* assignment of treatments does not mean the assignment is *haphazard*. Researchers who design randomized comparative experiments typically use some random device (e.g., table of random digits, software, or even coins) to carry out the randomization. The key is that each subject in the study should be equally likely to be assigned to any of the treatment conditions. |

This random assignment of treatments serves two purposes:

1. As discussed earlier, the random assignment helps to protect against hidden biases due to confounding variables.
2. It helps to prevent the researcher from assigning subjects to treatment groups that are favorable to their research hypothesis.

Next, we’ll discuss some terminology that is typically used when discussing randomized experiments.

|  |
| --- |
| Terminology Used in Randomized Comparative Experiments |
| **Factor** – Every randomized experiment involves at least one explanatory variable, which is typically called a *factor*. |
| **Factor Levels** – The values of the explanatory variable are typically referred to as *levels* of a factor. |
| **Treatment** - A *treatment* is a specific experimental condition applied to the subjects. Treatments are defined by the different factor levels (or combinations of levels if an experiment involves more than one factor). |
| **Experimental Unit** – The smallest unit to which a treatment is randomly applied is called an *experimental unit*. In studies involving humans, the individuals on which we experiment are usually called *subjects* or *participants*. |

These definitions may be easier to understand in the context of a few examples, which are introduced next.

| **Example: Effectiveness of a Smoking Cessation Program** |
| --- |
| A recent study (Fu et al., 2012) investigated the effect of a proactive care intervention on smoking abstinence rates. Suppose the study involved 1600 smokers from a VA medical center. Half of them were randomly assigned to the “usual care” group, which receives typical access to tobacco treatment services from their VA facility. This includes tobacco screening and advice for quitting. Pharmacotherapy is also available in the form of nicotine patches, nicotine gum, or bupropion. Participants in this group may also be referred to a smoking cessation clinic which will provide them with counseling and behavioral/cognitive strategies to quit smoking, and they may also contact their state or the national quitline. The other half of the subjects were randomly assigned to the “proactive care” group. These subjects have all of the aforementioned services available to them, as well; in addition, the proactive intervention combines two other components: (1) proactive outreach (mailed invitation materials followed by telephone outreach) and (2) offer of choice of smoking cessation services (telephone care or face-to-face care) involving counselors who have training in motivational interviewing and smoking cessation counseling. In addition to offering care, the mailed materials and the call also provide motivational enhancement to encourage participants to quit smoking and seek tobacco treatment. The experimental design is summarized in the following diagram.  Identify current smokers from the VA’s EMR Health Factors Dataset at the participating site  *Randomize*  Group 1 (800 subjects)  Group 2 (800 subjects)  Baseline Survey  Usual  Care  Proactive Outreach  Compare smoking cessation rates using data from a 12-month  follow-up survey  Questions:   1. Identify the population of interest in this study. Do you feel the sample is representative of the population? Discuss. 2. Identify the response variable. 3. Identify the factor and factor levels in this study. 4. Identify the treatments in this study. 5. Identify the experimental units in this study. |

| **Example: Effectiveness of Zyban and Nicotine Patches** |
| --- |
| Suppose a study was conducted to investigate the effect of both nicotine patches and antidepressants (specifically, Zyban) on smoking cessation. Suppose the study involved 400 men and women who were 18 and older and had smoked 15 cigarettes or more per day for the previous year. The subjects were all highly motivated to quit and in overall good health. The diagram below portrays the design of this experiment, which used placebo alternatives to both the nicotine patch and the Zyban.  Recruit 400 eligible smokers to participate in the study  *Randomize*  Group 4 (100 subjects)  Group 3 (100 subjects)  Group 2 (100 subjects)  Group 1 (100 subjects)  Treatment 4 Placebo Zyban / Placebo Patch  Treatment 3 Placebo Zyban / Nicotine Patch  Treatment 2 Zyban / Placebo Patch  Treatment 1 Zyban / Nicotine Patch  Compare the percentages of people who have relapsed after 12 months  Questions:   1. Identify the population of interest in this study. Do you feel the sample is representative of the population? Discuss. 2. Identify the response variable. 3. Identify the factor(s) and factor levels in this study. 4. Identify the treatments in this study. 5. Identify the experimental units in this study. 6. Why would the researchers conduct the experiment in this way? Specifically, why not carry out one experiment to investigate the effectiveness of Zyban and a separate experiment to study the effectiveness of the nicotine patch on smoking cessation? 7. How many treatments would the study have if it also incorporated whether or not a subject received counseling to discourage smoking? |

**Basic Principles of Experimental Design**

Good experiments are created with basic principles of experimental design in mind. These principles are as follows:

|  |
| --- |
| Principles of Designed Experiments |
| **Control** – All known sources of variation that may affect the response (other than the factor(s) under study) should be identified and controlled by making experimental conditions as similar as possible for all treatment groups. |
| **Randomize** – Some sources of variation will remain unknown. The random assignment of treatments to experimental units reduces bias by equalizing the effects of these unknown and uncontrolled sources of variation. |
| **Replicate** – Each treatment should be randomly and independently applied to a number of experimental units so that the variability in the responses can be estimated. Also, when the experimental units are not a representative sample from the population of interest, the entire experiment should be replicated with the controlled sources of variation at different levels, if possible. |

Questions:

1. Consider the smoking cessation experiments. Describe what the researchers did in these studies to control for other potential sources of variation. Could their designs have been improved in any way? Discuss.
2. How many replications were there for each of the treatments in the first example involving proactive outreach (i.e., how many times was each treatment randomly and independently assigned to an experimental unit)? What about the second example involving Zyban and nicotine patches?
3. How do researchers decide how many replicates are necessary?

| **Example: Physical Activity/Nutrition Intervention in Middle Schools** |
| --- |
| A group of researchers had the ultimate goal of helping middle school students to increase their physical activity level and to eat healthier. They designed both of the following programs:   * A “’Physical Activity Intervention” to increase activity in physical education classes and during leisure periods throughout the school day. * A “Nutrition Intervention” that improved school lunches and offered ideas for healthy home-packed lunches.   Sixty schools in their state agreed to participate (the total number of students involved was therefore around 9,000) . Each of these participating schools was randomly assigned to receive one of the following:   * Neither intervention program * Only the “Physical Activity Intervention” program * Only the “Nutrition Intervention” program * Both intervention programs   The investigators then observed the physical activity level and lunchtime consumption of fat on the students in these schools and then made comparisons across groups.  Questions:   1. Identify the factors and their levels in this experiment. 2. How many treatments exist, and what are they?      1. Identify the experimental units in this experiment. 2. How many replications exist for each treatment (i.e., how many times was each treatment randomly and independently assigned to an experimental unit)? 3. Identify the response variable(s) in this experiment. |

| **Example: Comparing Detergents** |
| --- |
| Suppose a study is conducted to compare the effectiveness of two different detergents on stain removal. Ten pieces of white fabric are stained with grape juice. Half of them are randomly assigned to Detergent 1 and are washed in a single load using an industrial washer. The other half are assigned to Detergent 2 and are also washed in a single load using the same industrial washer (the temperature, wash cycle settings, etc., were held constant). The percent of stain removed is measured on each piece of fabric, and comparisons are made across detergents.   1. Do you see any problems with this experimental design? Discuss. 2. Propose a new, improved study design. Sketch a diagram below to describe this design. 3. Identify the response variable.      1. Identify the factor(s) and factor levels in your improved study design. 2. Identify the treatments in your improved study design. 3. Identify the experimental units in your improved study design. 4. In the context of this design, discuss how you employed the fundamental design concepts of (1) control, (2) randomization, and (3) replication. |

**Types of Experimental Designs**

Most of the examples we have discussed up to this point are examples of a **completely randomized design**.

|  |
| --- |
| Completely Randomized Design |
| In a **completely randomized design**, the treatments are allocated completely by random chance to the experimental units. This design relies solely on randomization to equalize the effects of extraneous variables. |

Sometimes, the completely randomized design is not adequate. Consider the smoking cessation example involving proactive outreach. In the actual study, there were 6,400 smokers available from four different VA medical centers. If the researchers were to have used a completely randomized design, they would have randomly assigned 3,200 subjects to the “proactive outreach” group and the other 3,200 to the “usual care” group.

Question: Can you think of any potential shortcomings to designing the study in this way? Hint: Think of how the four different VA medical center sites may differ. Does the proposed design control for these differences in any way?

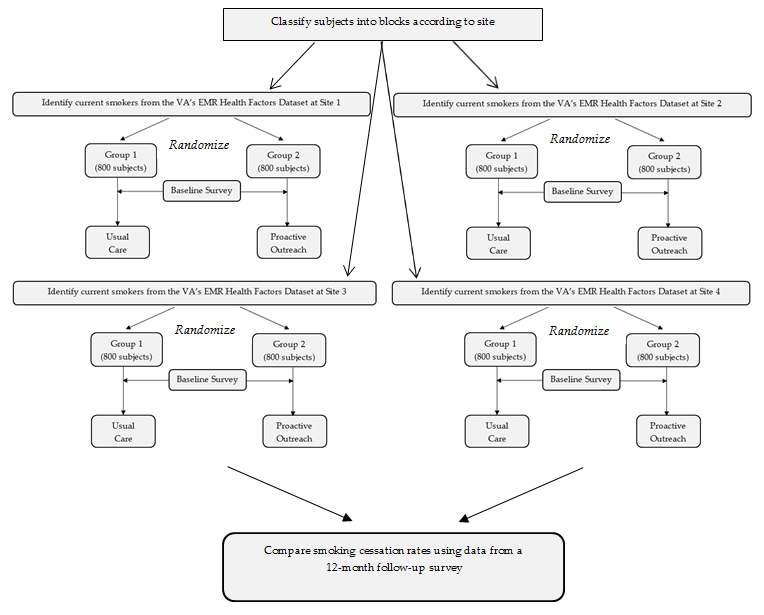
To design the experiment so that the differences between the four sites don’t interfere with our attempts to see differences between the two treatment groups, we may consider using a **randomized block design**.

|  |
| --- |
| Randomized Block Design |
| In a **randomized block design**, the experimental units are first grouped together into homogeneous groups (called blocks), and then treatments are allocated at random to the experimental units within each block. This type of design is used to reduce known sources of variation in the response variable across experimental units and should be used when such sources are known and can be determined for each experimental unit. Note that if each treatment appears once in each block, the design is referred to as a **randomized complete block design**. |

|  |
| --- |
| **Blocking is used to control the factors you can see; randomization helps  balance the ones you cannot see.”** *- Richard L. Scheaffer* |

Note that when a randomized block design is used, the researchers aren’t interested in studying the effects of the blocking factors themselves. Instead, the goal is to reduce variability so that the effects of the treatment factors under study can be seen more easily.

The following diagram shows the actual experimental design used by Fu and colleagues in the 2012 smoking cessation study. Note that, in some sense, four parallel experiments are conducted (one for each site), and the results are combined.



| **Example: Comparing Two Memorization Methods** |
| --- |
| Suppose researchers want to compare two methods for increasing memorization skills. Ten subjects have been recruited for the study, and each will receive only one of the two treatments; furthermore, researchers suspect that a subject’s age may affect their capacity to memorize and want to account for this in the study design.  Questions:   1. What is the downside to simply setting up the study as a completely randomized design (CRD)? 2. Propose an experimental design that is better than a CRD for this situation, and sketch a diagram of this design below. 3. Identify the factor(s) and factor levels in this study. 4. Identify the treatments in this study. 5. Identify the experimental units in this study. 6. In the context of this design, discuss how you might employ the fundamental design concepts of (1) control, (2) randomization, and (3) replication. |

|  |
| --- |
| Matched-Pair Designs |
| A **matched-pair design** is a special case of a block design. In this type of design, treatments are assigned in one of two ways:   * Blocks are created by matching two individuals according to some criteria. Then, treatments are randomly assigned to the individuals within each pair (i.e., block). * The same individual receives each of two treatments, and randomization is used to assign the order of the two treatments. Note that in this case, the individual essentially serves as the block. |

| **Example: Comparing Methods for Teaching Reading** |
| --- |
| Suppose researchers want to investigate two new methods for teaching children to read: visual and phonic. Thirty students who are ready to start learning to read have been recruited to participate in the study. The simplest way to conduct this study is to use a completely randomized design: randomly assign half of the students to the visual method and the other half to the phonic method; then, measure and compare the reading abilities of the children at the end of the year.  Questions:   1. What other factor(s) might affect a child’s reading ability? 2. How might this study design be improved to account for these other factors? Propose at least two alternative designs, and discuss the benefits/drawbacks of each. |

| **Example: Randomized Crossover Trial** |
| --- |
| Background The objective of the study was to determine metabolic and sensory effects of adding fiber to bread. **Subjects**  A total of 83 university students aged between 18–35 years (mean 21 years; SD 2.8), with normal body mass index (mean 22.5 kg/m2; SD 2.7), and normal fasting blood glucose (mean 4.6 mmol/l; SD 0.5) were recruited. Exclusion criteria included being diagnosed with chronic or digestive diseases, food allergies, pregnancy, and taking medications or supplements likely to influence glucose metabolism or gastrointestinal wellbeing. The study was approved by and conducted in accordance with the ethical standards set by the University of Otago Human Ethics Committee.  **Study design**  The experiment consisted of two randomized double-blinded crossover trials of two fiber breads: 1) A prototype product derived from fruit supplied by Anagenix (Petone, New Zealand) and 2) FibreMax™ (New Image International, Auckland, New Zealand). The fruit fiber was predominately an insoluble, non-viscous fiber. Fibre-Max™ contained a mix of soluble and insoluble polysaccharides that formed a gel with water comprising chicory root extract, psyllium, soy fiber, oat bran and pectin. A quantity of 15 g of powder for both products provided approximately 10 g of fiber. Participants were randomly assigned to consume a control bread, and either a fruit fiber or FibreMax™ bread during separate sessions separated by a two week washout. The serving was two slices of bread with 10 g of margarine (Craig’s, Heinz Wattie’s Ltd) accompanied by 250 ml of water. Bread and water were consumed within 15 minutes. A standard protocol for fasting and a recommendation to include carbohydrate in the meals prior to test days was given. Additionally, participants were reminded to avoid alcohol and vigorous physical activity prior to clinic days.  **Results** Consumption of the fruit fiber bread reduced postprandial glycemia by 35% (95% CI 13 to 51; P = 0.004) and cumulative energy intake by 368 kJ (95% CI 163 to 531; P = 0.001). There was little influence on satiety …  FibreMax™ enriched bread reduced glycemia by 43% (95% CI 17 to 61; P = 0.004) without influence on energy intake or satiety.   Question: What is the advantage to using each subject as a block, as was done in this experiment? |

|  |
| --- |
| Repeated Measures Designs |
| A **repeated measures design** is also a special case of a block design. As the name implies, this type of design involves measuring the same subject several times throughout the study. Repeated measures designs typically arise in one of two ways:   * Each subject is assigned a treatment at random, and the response is measured over time. * Each subject receives all treatments at different points in time. |

| **Example: Effects of Alcohol and Marijuana on Driving Ability** |
| --- |
| Suppose researchers want to use a driving simulator to investigate the extent to which drinking alcohol and smoking marijuana impair driving ability. Three treatments will be compared: sober, under alcohol, and under marijuana. There are 30 subjects available for the study.  Questions:   1. Researchers could use a completely randomized design. What would be the downside of doing so? 2. Researchers could also use a repeated measures design (i.e., each subject could serve as a block and receive each of the three treatments). What are the advantages and disadvantages to designing the study in this way? 3. Propose a reasonable design for this experiment. Think of the disadvantage(s) you listed in the previous question. Does your proposed design address your concern(s)? |

**Other Design Considerations**

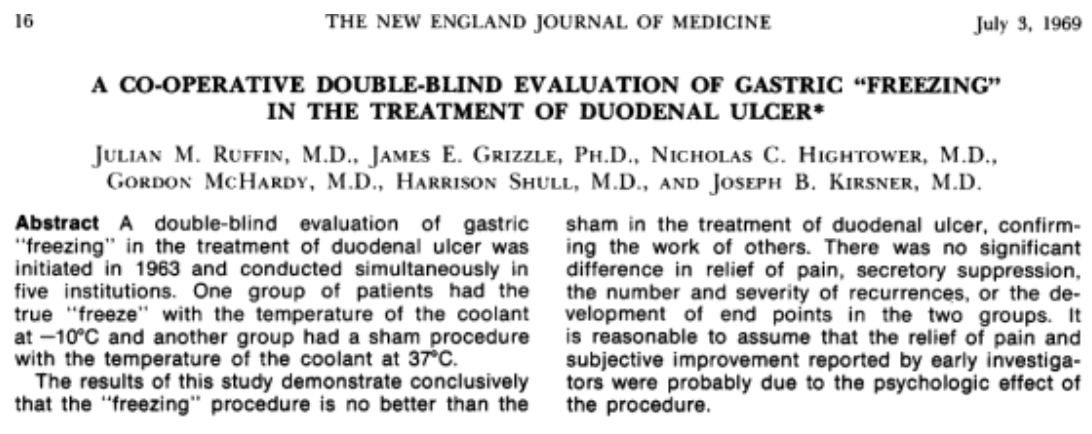
|  |
| --- |
| Control Groups |
| To determine whether a treatment had an effect, a researcher must be able to identify what would have happened to the response variable if that treatment had not been applied. So, researchers use control groups (the subjects in this group are treated identically to all other subjects in the study except that they don’t get the active treatment).  Earlier, the concept of “control” was introduced as one of the principles of designing experiments. In that context, resources control potential sources of variation by keeping them constant. Here, the control treatment is considered as another level of the factor under study so that we can obtain a “baseline” measurement. |

| **Example: Gastric Freezing** |
| --- |
| In 1962, Wangsteen and his colleagues introduced a technique known as gastric freezing for the treatment of duodenal ulcer pain ([link to original paper](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1466227/pdf/annsurg00846-0069.pdf)). In their case study of 31 patients, all reported marked or complete relief of pain, and many clinics started using gastric freezing to treat duodenal ulcer pain. A few years later, other researchers set up a randomized comparative experiment which utilized a control group to investigate the effectiveness of gastric freezing for treating pain associated with duodenal ulcers and published their results in the *New England Journal of Medicine* (Ruffin et. al., 1969). Of 137 patients with duodenal ulcers, 69 were assigned at random to gastric freezing. The other 68 were assigned to the control group (the solution used to treat the subject was held at 37 degrees Celsius). The study resulted in very similar proportions of marked or complete relief of pain between the two groups.   Questions:   1. Discuss the major problem with the design of the original study. 2. What phenomenon could potentially explain these results? |

|  |
| --- |
| Placebos |
| A special kind of control group is often used in experiments involving human subjects. Research has shown that people respond simply to being treated, even if that treatment is with a **placebo** (a “fake” treatment). To separate the placebo effect from the actual effect of the treatment under study, the placebo is used as the control treatment. It is important that subjects not be able to tell whether they received the placebo or the actual treatment. |

|  |
| --- |
| Blinding |
| To avoid bias that can arise if either the subject or someone else involved with the study knows which treatment a subject has received, good experimental designs employ the concept of **blinding**. There are two groups that can potentially bias the results of an experiment:   * Those who could influence the results (e.g., the subjects, treatment administrators, etc.) * Those who evaluate the results (e.g., judges, treating physicians, etc.)   When all of the individuals in only one of these groups are blinded as to which treatment subjects have received, the study is called **single-blind**. When everyone in both groups is blinded, the experiment is **double-blind**. |

Example of a double-blinded experiment:



Question: Why is it important that this experiment be double-blind?

|  |
| --- |
| In summary, the best experiments are usually   * Randomized * Comparative * Double-blind * Placebo-controlled |
| **Example: The Salk Vaccine Trials** | |
| In 1954, the Public Health Service and National Foundation for Infantile Paralysis (NFIP) organized an experiment to test the effectiveness of Dr. Jonas Salk’s polio vaccine. Polio is an infectious viral disease that enters through the mouth and is typically spread by contaminated drinking water or food. Most healthy people infected with the virus experience no more than mild fever or diarrhea; in some cases, however, the virus spreads to the bloodstream and central nervous system, causing various degrees of paralysis (and in extreme cases, death).  In one part of the study, the researchers conducted what was known as the *observed control experiment*. This design called for the vaccination of second-graders at selected schools in selected areas of the country (with the consent of the children’s parents). These vaccinated second-graders formed the treatment group. The first and third-graders at these schools were not given the vaccination, and they formed the control group. This observed control experiment suffered from the following problems:   * Selection bias – The treatment group was self-selecting, since parents had to give their consent. Higher-income parents were more likely to consent; furthermore, the incidence of polio tends to be higher among their children (strangely enough, in communities with poor hygiene, more people tend to have a natural immunity to the virus). So, the selection bias is *against* the vaccine. * Diagnostic bias – Parents and doctors attempting to diagnose illnesses would know the child’s grade level and whether the child got the vaccine. Even when trying to be objective, they might tend to be biased *in favor* of the vaccine (i.e., they are more likely to diagnose polio in an unvaccinated child and less likely in a vaccinated child). * Polio spreads through contact. If the incidence were higher in second-graders, this would bias the study *against* the vaccine. If it were lower for second-graders, this would bias the study *in favor* of the vaccine.   An improved study design, a randomized comparative experiment, was proposed to overcome the deficiencies of the observed control experiment. In this study, the control group and treatment group were both chosen from the same population: children whose parents gave consent (this helped to control for the confounding effect of parental income and provided a valid **control** group). The children were randomly assigned to either the treatment or control groups (i.e., the subject was **randomized**), and children in the control group were given a **placebo** (an injection of salted water). Neither the children nor their doctors knew who was in the treatment or control groups (i.e., the experiment was **double-blind**).  The following table shows the incidence rate per 100,000 for both treatment groups. Do you think there was bias in the observed control experiment?   |  |  |  | | --- | --- | --- | |  | Observed Control Design | Randomized Experiment | | Vaccinated Group | 25 per 100,000 | 28 per 100,000 | | Control Group | 54 per 100,000 | 71 per 100,000 | | |