Slide 1: Introduction

Hello, and welcome to module two: Fundamentals of Epidemiology. This presentation will focus on experimental studies. My name is Jeffery Bethel, assistant professor of epidemiology, East Carolina University, Brody School of Medicine.

Slide 2: Acknowledgements

APTR wishes to acknowledge the following individual that developed this module:

- Jeffery Bethel, PhD
  Department of Public Health
  Brody School of Medicine at East Carolina University

This education module is made possible through the Centers for Disease Control and Prevention (CDC) and the Association for Prevention Teaching and Research (APTR) Cooperative Agreement, No. 5U50CD300860. The module represents the opinions of the author(s) and does not necessarily represent the views of the Centers for Disease Control and Prevention or the Association for Prevention Teaching and Research.

Slide 3: Presentation Objectives

There are four objectives in this presentation. They are: Recognize the use of experimental studies as an epidemiologic study design, distinguish between types of experimental studies, describe key features of conducting experimental studies, and to recognize special considerations of experimental studies.
Slide 4: Analytic Study Designs

In this presentation I will cover experimental studies, which are different than observational studies in that the researcher allocates treatment or exposure to study participants. Observational studies are natural studies and are covered in another presentation.

Slide 5: Experimental Studies

The goal of public health as well as clinical medicine is to essentially modify natural history of disease to decrease morbidity and mortality. The question then becomes, how do we know which preventive and therapeutic measures are most effective to achieve this goal? To do that researchers carry out formal studies. Specifically experimental studies for the purposes of this presentation to determine the value of various measures.

Slide 6: Study Design Hierarchy

As you see in this diagram, experimental studies sit at the top of the hierarchy of epidemiologic study design and produce the most valid results but come at a very high cost.
Slide 7: Experimental Studies

The reason experimental studies are at the top of the hierarchy is because they most closely resemble controlled laboratory experiments. In the controlled laboratory experiments, investigators regulate all the important aspects of the experimental conditions and allow the subjects to differ only for the purposes of testing the hypothesis. For example, if the investigator was testing the toxicity of a chemical in mice, he or she would include genetically similar or identical mice, assign mice to a test and control group, maintain the same physical environment for the mice, give them the same diet, and maintain the same daily schedule for the mice. That is so the investigator is in control of all aspects of the study. Experimental studies, for the purpose of epidemiologic study designs, are the gold standards of epidemiologic research. Their high status and validity enables them to pick up small and modest effects.

Slide 8: Scurvy

One of the first experimental studies of note involves scurvy. James Lind originally identified the symptoms of scurvy among sailors at sea after as little as a month. The symptoms included spongy and bleeding gums, bleeding under the skin and extreme weakness. So Lind conducted an experimental study onboard a British ship in the mid-1700s in which he created six groups of two ill sailors each. He gave each group a different diet. He found that the group eating oranges, limes, and lemons, were fit for duty within six days. As a result of this experimental study, the British navy required limes and lime juice to be included in the diet of its sailors.
Module 2: Fundamentals of Epidemiology – Experimental Studies

Transcript

**Slide 9: Potential Uses**

More contemporary uses of experimental studies include evaluating new drugs and other treatments for diseases, evaluating new medical and health care technology, evaluating new screening programs or techniques, or evaluating new ways of organizing or delivering health services. You could compare home versus hospital care following a myocardial infarction.

**Slide 10: Preventive V. Therapeutic**

Let’s get into some study design issues regarding experimental studies. Specifically, the various ways to classify them. You can classify experimental studies based on their purpose as either preventive or therapeutic. In preventive studies we would ask the question: Does a prophylactic agent given to healthy or high-risk individuals prevent a disease? Whereas in a therapeutic study we would ask the question: Does the treatment given to diseased individuals reduce the risk of recurrence, improve survival or quality of life? Again, these types of studies are classified by their purpose; preventive or therapeutic.
We can also classify experimental studies by which unit the treatment is assigned. Treatment can be allocated to the individual or the community. In an individual based experimental study we could ask the question: Do women with stage I breast cancer given a lumpectomy alone, survive as long without recurrence of disease, as women given a lumpectomy plus radiation? This is at the individual level and in this case, the female. In a community based experimental study we could ask the question: Does fluoride in the water supply decrease the frequency of dental caries in a community compared to a similar community without such water treatment? Again, the unit to which the treatment is assigned is at the individual or community level.

Here’s the overall study design of an experimental study. In this case a preventive study as noted by the outcome, disease or no disease. You’ll notice here that the key to an experimental study is random assignment of the study participants to treatment groups. In this case the current treatment or the new treatment. Once groups are assigned and treatment has been given, investigators follow these groups of people to monitor for the outcome of interest. In this case, a preventive study looking for disease or no disease.

Here is the same overall design of experimental study, but for a therapeutic study. Again, the key is a random assignment of study participants to a current treatment or new treatment and then examining or following for the outcome of interest. In a therapeutic study were looking for an improvement or no improvement.
Let’s take the figure from the previous slides a little further and get into more specific aspects of experimental studies. The overall conduct of an experimental study can be thought of in these terms: First, form hypothesis. Second, recruit participants based on specific criteria and obtain their informed consent. Third, randomly allocate eligible and willing subjects to receive one of the two or more interventions being compared. Next monitor the study groups for the outcome of your study. Again, that’s in the context of a preventive study or a therapeutic study. Finally, compare the rates of the outcome between the various groups.

Let’s go through each of those steps one at a time. In an experimental study, a hypothesis is initially formed. So based on the research question this is what the investigators think will happen by the end of the study. For example, a hypothesis in a study could be: Women with stage 1 breast cancer given a lumpectomy alone will survive as long without recurrence of disease as women given a lumpectomy plus radiation. Or, in a community based study: Water supply with fluoride will decrease the frequency of dental caries in a community compared to a similar community without water treated with fluoride. Again, this is what the investigators think will happen by the end of the study.
Slide 16: Inclusion Criteria for Participants

Once we form a hypothesis we need to identify participants. That is, who will be in our study? To do this, we need to develop specific criteria to use to recruit participants into the study. These criteria must be defined specifically before the study begins. This is done for two reasons: 1, to remove subjectivity and 2, to allow the study to be reproducible. That is, other researchers can use our criteria to reproduce the study.

Slide 17: Women’s Health Study

So here are the inclusion criteria for the Women’s Health Study, which was a randomized trial of low dose aspirin in the primary prevention of cardiovascular disease in women. The criteria that the investigators developed were: participants had to be at least 45 years of age, female, no history of coronary heart disease, cerebrovascular disease, cancer, other major chronic illness. Participants cannot have had a history of side effects to any of the study medications. They weren’t taking any of the following medications more than once per week: aspirin, NSAIDs, supplements of vitamin A, E, or beta-carotene. And they were not taking anticoagulants or corticosteroids.

Slide 18: Sample Size

Once we determine who is eligible to participate in our study, we need to determine how many participants are needed. And there are various programs and tables that will enable researchers to calculate a sample size based on various parameters.
Slide 19: Type I & II Errors

We can think about these parameters based on the following table. Let’s cover two concepts before we get too far into this table. One is the idea of the sample versus the population. Given most if not all studies are based on a sample rather than the entire population, we can compare our results or our conclusions from our sample to what the truth is in the population. Next is the idea of the null hypothesis versus the alternate hypothesis. Null denoted by $H_0$ and the alternate denoted by $H_1$.

Now the null hypothesis is that there is no difference in effect between the test and control groups, or the various groups being compared. The alternative hypothesis is that there is a difference in the effect, in this case between the experimental test group and the control group. Ok, now let’s get into this table a little bit further. So there are four different possibilities when comparing the conclusions from a sample to the truth in the population. You’ll see here that we could conclude from the sample that there is no difference in effect between the two groups. That is, we fail to reject the null hypothesis. Whereas the truth in the population is also that there is no difference, so in this case we made a correct decision based on our sample. Now let’s move to the lower right corner. In this instance we conclude from the sample that we reject the null and there is a difference between the test and control groups which is in agreement with the truth in the population, that there is a difference in the effect between the two groups. This is the correct decision. This is also thought of as probability $1-\beta$. Now let’s look at the diagonal squares where the conclusions from the samples do not agree with the truth in the population. Looking at the lower left, we can conclude from the sample that there is a difference in the effect between the two groups. Whereas the truth in the population is that there is no difference. When this occurs we’ve committed what is called a type I error. You can think of this as a false positive. And the probability of this occurring is denoted by $\alpha$. Now let’s go up to the top right hand corner. The conclusion from our sample is that there is no difference in the effect between the two groups. Whereas, the truth in the population is that there is a difference. When this occurs we’ve committed what’s called a type II error. You can think of that as a false negative and the probability of this occurring is $\beta$. Now how does this relate to sample size? We can reduce the occurrence of type I and type II error by increasing sample size.
Slide 20: Sample Size Required Parameters

Here are all the required parameters we need to calculate a sample size. First we need the estimate of the difference in the effect to be detected. That is, what we hypothesize will be the difference in the outcome between the test and control group. Next we need an estimate of the occurrence of our outcome in either the test or control group. Next we need to set a level of significance, \( \alpha \) to be set at 0.05. We also need to set the level of power desired, that’s 1-\( \beta \), typically set at 0.8. Next we need to determine whether we are testing a 1-sided or 2-sided test. A 2-sided test will have double the sample size of a 1-sided test.

Slide 21: Participant Allocation

Now that we have determined how many participants we need for our study, we need to allocate participants to test and control groups. So lets take a step back here for a second. In all experiments we need to have a control group. Ideally we want to compare the outcome among individual in the treatment group or exposed group to what the outcome would have been in that same individual had he or she not been exposed or not received the treatment. By the way, this comparison is counterfactual and not possible. However, we can compare the outcome among the exposed or treatment group to the outcome in the substitute population. This is where the control group comes in. And finally the validity of our inference of our findings depends upon finding a valid substitute or comparison population.
Slide 22: Participant Allocation

So in experimental studies we randomly assign participants to one of the intervention groups, that is, the test or control. This is how we allocate participants to our various intervention groups. Now randomization essentially is a situation in which the next assignment of a participant to a test or control group is unpredictable. There are various ways to do this. You can toss a coin, you can use a random number table, opaque envelopes, or a computer can make the assignment.

Slide 23: Randomization

So why is research randomized? Two reasons: The randomization reduces the selection bias in the allocation of treatment. That is, researchers are not explicitly determining whether participant 1 goes to treatment group and 2 goes to control group. Randomization takes care of that. Also, each participant has an equal chance of being in the test or control group. This removes the roll of the investigator in allocating treatment. The secondary purpose of experimental studies is that if you are randomizing a large enough sample size, randomization will most likely produce treatment and control groups with similar baseline characteristics. And it also controls for known and unknown factors. Risk factors, behaviors of the participants in your study.
Now lets take a look at an example in which the secondary purpose of randomization was achieved. In this example, nearly 20,000 participants were randomly allocated to a test and control group. After allocation the baseline characteristics of the two groups can be compared. As you see here in this table, the test and control group had a similar distribution of male participants, white participants, current smokers, and patients with a history of hypertension, stable angina, and high cholesterol. That is, these groups are comparable.

Whereas, here is an example of where a relatively small sample size of 500 was randomly allocated to test and control groups. In this instance, randomization produced two groups that were not equal in terms of the distribution of white participants. You see in the test group 48% were white and in the control group 38% were white. So again with a relatively large sample size randomization will most likely produce two groups that are similar in terms of baseline characteristics and unknown factors.

Once participants have been allocated to the various treatment groups, it is time to administer the treatments and collect baseline data. In terms of treatment, researchers must keep track of the treatment group to which each participant has been assigned. And therefore keep track of which therapy he or she is receiving. In terms of baseline data researchers can collect demographic and other risk factor data to then assess whether randomization has achieved its secondary purpose of comparability between treatment and control groups.
Module 2: Fundamentals of Epidemiology – Experimental Studies

Transcript

Slide 27: Data Collection

Once treatments have been assigned and given, it’s time to collect our data. Specifically measuring the outcome under study. Just like eligibility criteria need to be explicitly defined before a study begins, researchers must define the outcome of interest beforehand as well. Measurement of the outcome or disease can be in the context of a preventive or therapeutic study. This must be done uniformly for all participants. Outcomes can be measured through various means including a physical exam, self report, medical records, a combination of two or more of these. Could be a self report followed by a medical review.

Slide 28: Women’s Health Study

Let’s get back to that Women’s Health Study we talked about earlier. In that study three outcomes were being investigated. One, myocardial infarction, and that was defined as symptoms meeting WHO criteria. And also abnormal levels of cardiac enzymes or diagnostic electrocardiograms. Another outcome under investigation was stroke. That was defined as new neurologic deficit of sudden onset that persisted for at least 24 hours. The third outcome was death from cardiovascular disease. The researchers defined and measured this outcome by examining autopsy reports death certificates, medical records and other information obtained from the next of kin or other family members. Again the key here is that these outcomes were defined before the study began and uniformly applied to all study participants.
Another issue regarding data collection is the masking or blinding of participants as well as researchers. The purpose of blinding is to prevent the conscious and subconscious bias in research. Researchers invest a lot of time and energy in designing and implementing their study and they may be subconsciously swayed to classify a participant as diseased when he or she may not truly be diseased or have the outcome under study. In experimental studies a placebo is typically used to mask the participants as to their treatment. So we can think of experimental studies as single blind. That is the participants don’t know what treatment they are receiving, again with the use of a placebo. Or a double blind where the participants are blinded and the observers or the people collecting the data do not know the treatment status of the participant. For example, when assessing or measuring the outcome of a given participant, the people collecting the data do not know if the participant is in the treatment or control group. Again, this reduces the conscious and subconscious bias in research.

So far we’ve been talking about experimental studies as parallel designed, classified by the method of treatment administration. That is participants in each group simultaneously receive one study treatment. These groups consist of different participants. However the crossover design is also possible and it involves a planned reversal of the test and control groups. So in this design each participant serves as his or her own control.
**Slide 31: Crossover Design**

Here is a figure depicting the crossover design. You’ll see that the outcome of interest is measured in each group, and then, typically after a washout period, the groups switch treatment and the outcome of interest is measured again. So there are a couple considerations when using this method. One, the washout period may be needed. That is the time between when group 1 receives a new treatment and then switches to the current treatment. The order of the treatments may matter and may have an effect on the outcome under study. Lastly, this study design may not be feasible if the treatment is surgical or if the treatment cures the disease or treatment under study.

**Slide 32: Simple v. Factorial**

Another design option is based on the number of treatments being tested. In a simple design, each group receives one treatment, one drug. In a factorial design the same study population is used to compare 2 or more treatments. It’s similar to a three armed study but with fewer participants. There are a couple of assumptions when using a factorial design. Anticipated outcomes are different. And secondly the modes of action are independent that is there is no interaction.

**Slide 33: Factorial**

So here is a generic way to look at a factorial design. So you have four groups of study participants. In this instance we’re comparing drug A to drug B. One group receives both drug A and drug B. One group receives A only. One group receives B only. And one group receives neither A nor B. Then we can compare the efficacy of drug A and drug B.
Slide 34: Factorial

Here is a more specific example of factorial design in which the four groups of study participants are receiving two different treatments. One group is receiving aspirin and beta-carotene. One group is receiving aspirin only. The third group is receiving beta-carotene only. And the fourth group is receiving neither aspirin nor beta-carotene. And again the efficacy of aspirin and beta-carotene can be assessed independently.

Slide 35: Noncompliance

Another design issue that must be addressed is the issue of noncompliance. There are two forms of noncompliance including overt noncompliance, which occurs when a participant notifies the investigator that he/she is dropping out of the study. We consider them loss to follow up or drop outs from the study. Covert noncompliance occurs when a participant can stop taking their assigned treatment or take the assigned treatment on a modified schedule without telling the investigators. The investigators do not know this is occurring. So investigators need to build compliance checks into their study. For example they could test the urine of participants for the metabolite of the treatment, or they could count pills. Some measure to make sure that their participants are adhering to the schedule and taking the assigned treatment. However, this is easier said than done. Between the two, covert noncompliance can have more serious effect on the validity of the results of the experimental study.
Once data collection is complete, (the outcome has been measured in the different treatment groups) it’s time to analyze our data and express our results. There are many ways to do this in experimental studies. One way is to report efficacy of the treatment. That is, the reduction in risk of the disease, the risk of death or the risk of complications. Again it goes back to the purpose of the study, be it preventive or therapeutic. Now here’s an example involving vaccines. We could express efficacy of the vaccines as the rate of our outcome in the placebo group minus the rate of our outcome in the vaccine group, all divided by the rate of the outcome in the placebo group.

Other ways of expressing results include calculating a relative risk. That is the incidence of the outcome in our test group divided by the incidence of the outcome in our control group. We can also calculate a Kaplan-Meier plot or a hazard ratio. We could calculate the NNT, which is the number of patients needed to be treated to prevent 1 adverse event, typically used in clinical trials. We can calculate NNH or number needed to harm which indicated the number of patients treated to cause harm in one patient who would not otherwise have been harmed.
Ok, now that we’ve expressed our results we need to think about what these results mean. That is, how valid are these results. We can think about validity in two ways: Internal validity and external validity. Regarding internal validity, that is, the extent to which the study groups are comparable. Or comparability, that is how similar are the test and control groups. The comparability of our test groups is reflected by how participants were selected or how treatment was allocated. And again, in the case of experimental studies the key here is randomization. Participants were randomly allocated to test and control groups. The other form of validity is external validity. That is, the extent to which the results of the study can be applied to people not in it. We can think of external validity as generalizability or representativeness. So how representative are our study participants of people in the target or reference population.

Here is a figure expressing internal and external validity from the previous slide. External validity is how representative is the study population of the reference population. With the example of the Women’s Health Study, the study population included female health professionals over 45 years of age in the U.S. and other various eligibility criteria. So the question is, how representative was that study population of all women or of all women over 45 in the US? Internal validity again, is comparability. How comparable were the participants in the test and control group or the current treatment and new treatment groups.
Slide 40: Validity of Results

We can assess the validity of our results by assessing a few items. These items effecting internal validity include loss to follow-up, and lack of randomization. Regarding loss to follow-up, since prospective studies follow participants over time we talked about the issue of drop-outs and loss to follow-up. A high degree of loss to follow-up will effect the internal validity. Particularly when you have a higher rate of loss to follow-up in the test rather than the control group. Or a higher rate of loss to follow-up in one group than the other. The question then becomes how comparable are those groups when one group has a much higher rate of loss to follow-up. Lack of randomization can also affect internal validity. However, as we talked about with experimental studies, the key is randomization. Items effecting external validity or representativeness of the study population include also loss to follow-up, a low response rate and narrow inclusion criteria. Thinking about loss to follow-up, if a study has a high loss to follow-up rate then you could ask the question, how representative are the people who remained in the study to the reference population? Also a low response rate, of all the people invited to participate in the study, if you have a low response rate, again, is this group who agreed to participate different or representative of the people who declined to participate? Lastly, narrow inclusion criteria. Long or narrow inclusion criteria further limit or restrict this group. So again the question is how representative is this group given this narrow inclusion criteria of the people not in the study?

Slide 41: Ethical Considerations

Looking back at the experimental design we also need to address ethical considerations involving the study. First regarding randomization. In randomization each participant has an equal chance of being assigned to a test or control group. Therefore there must be a genuine uncertainty about which treatment arm is better. If there is not a genuine uncertainty about which treatment arm is better you would be depriving one half of your study population of a superior or better treatment. Next, when to stop the study. If, as participants are being followed for the outcome under investigation, investigators determine there are harmful effects of one treatment arm, the study may need to be stopped. So that is one path is either being unduly harmed by this treatment, or they are being deprived of a beneficial effect of the other treatment.
Typically an outside board monitors for these effects. Lastly, informed consent.
Some trials enroll participants immediately after diagnosis when obtaining informed consent may be problematic. The effect of this new diagnosis may affect their informed consent into the study.

**Slide 42: Disadvantages**

There are several disadvantages of conducting experimental studies. One is that they are expensive and time-consuming. Given the large sample size needed to determine effects between treatment groups and the long amount of time these participants may need to be followed, that leads to a high cost. Next, ethical concerns may arise regarding when to stop the study, how to gain informed consent, and other issues we just talked about in the previous slide. Next, a large number of participants may be required. So based on our sample size calculation, to detect a small difference we may need a large group of people, to detect that small difference in our outcome between the test and control groups. Next, if we have narrow inclusion criteria, we may be limiting our generalizability of our results. That is the list of criteria we establish was so narrow that the people who were eligible to participate in our study are not representative of the target population. Next compliance may be an issue. Investigators need to build in compliance checks among their participants to make sure they are taking the assigned treatment at the right time. Lastly, influence of sponsorship should be monitored.
However there are advantages to experimental studies. First, randomization tends to balance risk factors across study groups, which we talked about several times. Next, blinding participants reduces bias in the ascertainment of outcomes. We talked about single-blind in which participants are blinded as to which test or control group they’re in. And double-blind in which the participants and the researchers are blinded. The next advantage of experimental studies is that they are prospective. They are ensuring that the test or treatment allocation precedes the outcome of interest. Next, experimental studies can eliminate bias by comparing two identical or nearly identical similar groups. Again, that is achieved through randomization. Finally, detailed information can be collected at baseline and throughout the study for making comparisons between those at baseline. And you can compare baseline information between those who stay in the study and those who are lost to follow-up to see if they are similar.

In summary, experimental studies top the epidemiologic study design hierarchy in terms of validity and that is because they most closely resemble laboratory-controlled experiments. In experimental studies, investigators assign treatment to participants and that’s through randomization. Randomization reduces selection bias in treatment allocation. We also talked about how data collection must be conducted systematically. That is, inclusion/exclusion criteria must be defined before the study begins and applied uniformly. The assessment or ascertainment of outcome must be done systematically and uniformly between test and control groups. Finally, noncompliance and drop-outs must be minimized to increase the validity of our results. Thank you that concludes Module 2 Experimental Studies.
Collaborating Institutions

- Center for Public Health Continuing Education
  University at Albany School of Public Health
- Department of Community & Family Medicine
  Duke University School of Medicine

Advisory Committee

- Mike Barry, CAE
- Lorrie Basnight, MD
- Nancy Bennett, MD, MS
- Ruth Gaare Bernheim, JD, MPH
- Amber Berrian, MPH
- James Cawley, MPH, PA-C
- Jack Dillenberg, DDS, MPH
- Kristine Gebbie, RN, DrPH
- Susan Jari, MD, MPH, FACP
- Denise Koo, MD, MPH
- Suzanne Lasorina, MD, MPH
- Rika Mawson, MD, MPH
- Dan Mawson, MD
- Steve McCurdy, MD, MPH
- Susan M. Meyer, PhD
- Safar Riaz, MD, MBBS
- Nawzad Shawir, MBBS

APTR

- Sharon Hull, MD, MPH
  President
- Allison L. Lewis
  Executive Director
- O. Kent Nordvig, MEd
  Project Representative