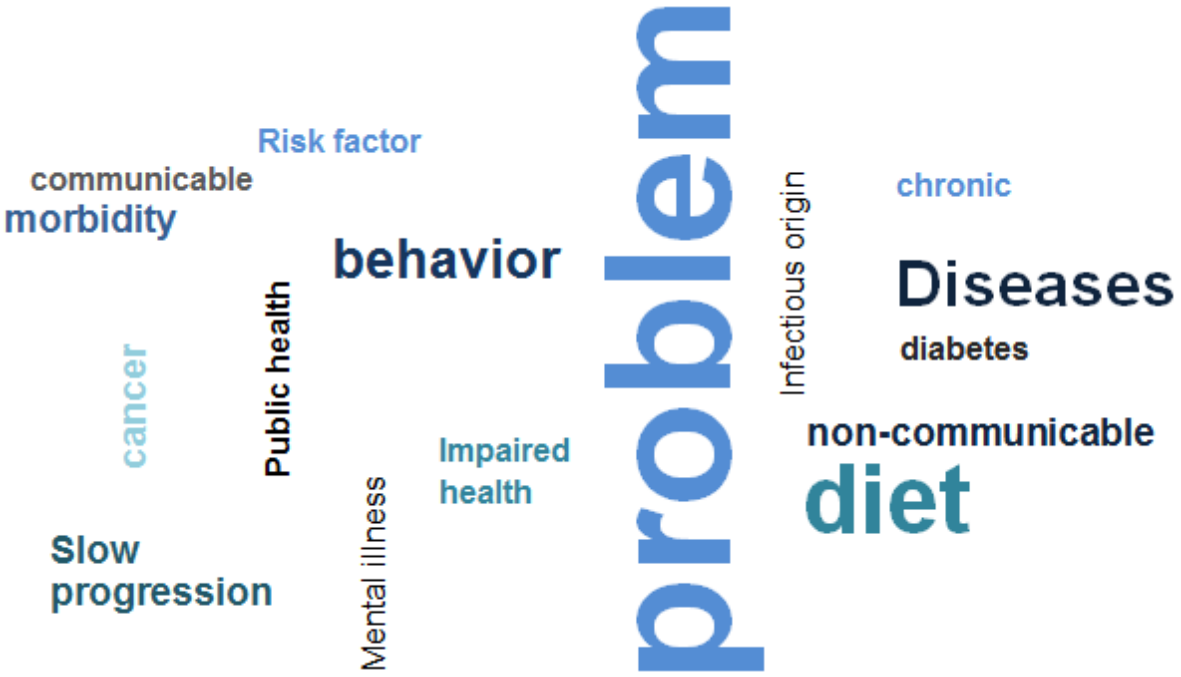
FACILITATOR GUIDE



Descriptive and Analytic Studies

Created: 2013

Descriptive and Analytic Studies. Atlanta, GA: Centers for Disease Control and Prevention (CDC), 2013.

Table of Contents

Descriptive and Analytic Studies

LEARNING OBJECTIVES

At the end of the training, participants will be able to:

- Identify the following for an NCD problem:
  - Type of study to conduct
  - Sampling methods to use
  - Measure of association to calculate for a particular study
- Interpret the results of descriptive and analytic studies.

ESTIMATED COMPLETION TIME

- 5 hours, 30 minutes (4 hours for interactive lecture; 1 hour, 30 minutes for Skill Assessment).

TRAINING TECHNIQUES

- Present content and examples by using lectures and group discussion. Skill assessment will be in small groups.

PREREQUISITES

- *Introduction to NCD Epidemiology*
- *NCD Burden of Disease*
- *NCD Data Sources*
- *Surveillance in Public Health*

MATERIALS AND EQUIPMENT

For the Facilitator:

- PowerPoint file for presentation
- Computer with projector
- Flip chart or chalk board

For the Participant:

- Participant Guide

REFERENCES AND RESOURCES

- Aschengrau A, Seage GR. Essentials of Epidemiology in Public Health, 2nd edition. Sudbury, Massachusetts: Jones and Bartlett Publishers; 2008.
- Aschengrau A, Seage GR. Essentials of Epidemiology in Public Health, 2nd edition. Case-control Studies PowerPoint™ Slides: http://publichealth.jbpub.com/aschengrau/ppts/case-control%20studies.ppt accessed on February 22, 2011

- Gordis, L. Epidemiology, 2nd edition. Philadelphia, PA: W.B. Saunders Company; 2000.
- Herold JM and Peavy JV. Surveys and Sampling. Field Epidemiology, 2nd ed. Ed. Gregg M. New York: Oxford University Press, 2002.
- Oleckno WA. Essential epidemiology: principles and applications. Prospect Heights, IL 2002;108.
- Remington RP, Brownson RC, Wegner MV, ed. Chronic Disease Epidemiology and Control. 3rd ed. Washington DC: American Public Health Association; 2010
- Rothman K.J., Greenland S. Modern Epidemiology, Second edition, Philadelphia, PA, 1998.
- Stehr-Green, J and Stehr-Green P, Survey Design Part 1: Sampling. North Carolina Center for Public Health Preparedness Training Website. Accessed on January 31, 2011 at http://cphp.sph.unc.edu/training/HEP_SDP1/certificate.php
- Stöckl H, Watts C, Kilonzo Mbwambo JK. Physical violence by a partner during pregnancy in Tanzania: prevalence and risk factors. Reprod Health Matters. 2010 Nov;18(36):171-80.
- Stern, F, Halpern, W, Hornung, R, Ringenburg, V and McCammon, C. Heart Disease Mortality Among Bridge and Tunnel Officers Exposed to Carbon Monoxide. American Journal of Epidemiology. 1988;128:1276-1288.
- Unwin N, James P, McLarty D, Machybia H, Nkulila P, Tamin B, Nguluma M, McNally R. Rural to urban migration and changes in cardiovascular risk factors in Tanzania: a prospective cohort study. BMC Public Health. 2010 May 24;10:272.

PREPARATION CHECKLIST

The following are action items to be completed by the facilitator prior to training:

___ Review slides
___ Prepare a blank table for slide 69
___ Print copies of the abstracts at the back of this facilitator guide to distribute after participants complete the skill assessment *(optional)*

FONT GLOSSARY

The following fonts are used in this guide:

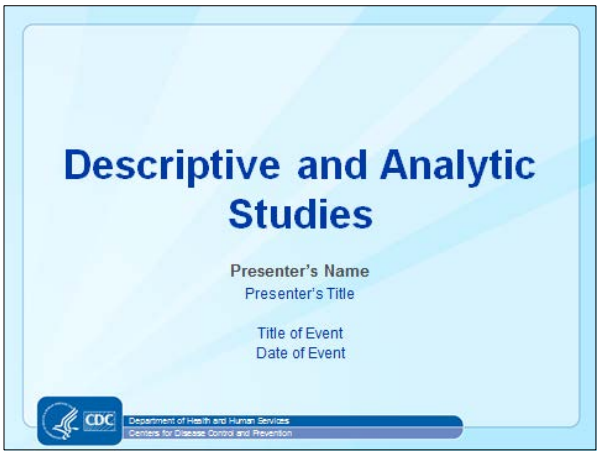
Font Type	Font Meaning
Plain	Script
Bold	**Instructions**
Italics	*Answers*

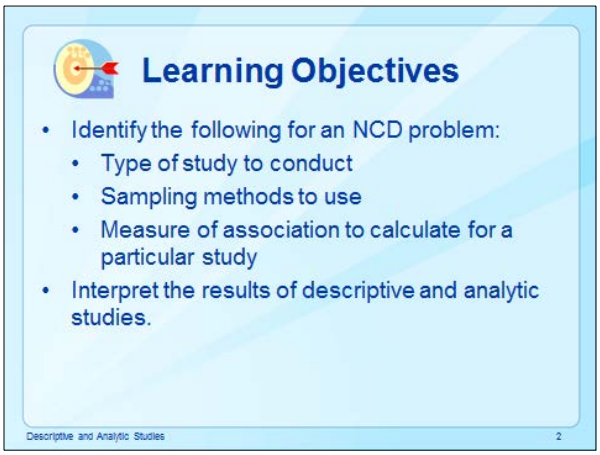
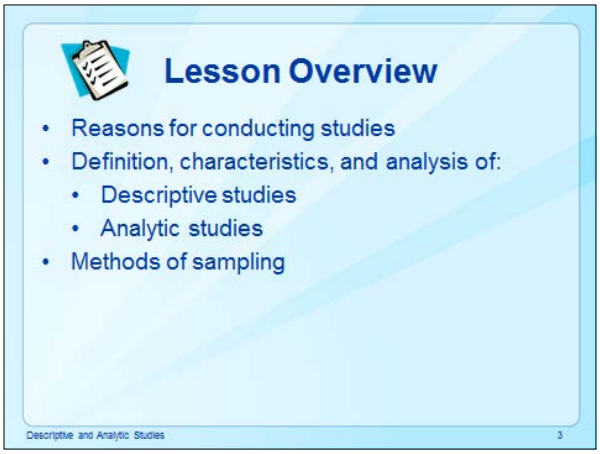
ICON GLOSSARY

The following icons may be used in this guide:

Image Type	Image Meaning
 Activity Icon	Small group exercise.
 Flip Chart Icon	Write responses during facilitator-led discussions or debriefs.
 Question Icon	Question for facilitator to ask participants.
 Tip Icon	Supplemental information discussion.
 Stop Icon	Do not start the assignment until your facilitator tells you to begin.

MODULE CONTENT

Duration/ Slide Number	What to Do/What to Say
5 minutes **Slide 1**	
Question	• **Welcome participants to this lesson.** • **Introduce yourself if you are a new facilitator.** • **Ask participants about their experience conducting descriptive and analytic studies.** • **Explain that this lesson will provide them with an overview of descriptive and analytic studies used in public health.** • **Tell participants that this lesson will take approximately 4.5 hours to complete.** • **Explain that during the lesson they will complete 4 practice exercises; at the end of the lesson, they will complete a skill assessment with a small group.**

Duration/ Slide Number	What to Do/What to Say
1 minute Slide 2	 • **Read the slide.**
1 minute Slide 3	 • **Read the slide.**

Duration/ Slide Number	What to Do/What to Say

5 minutes

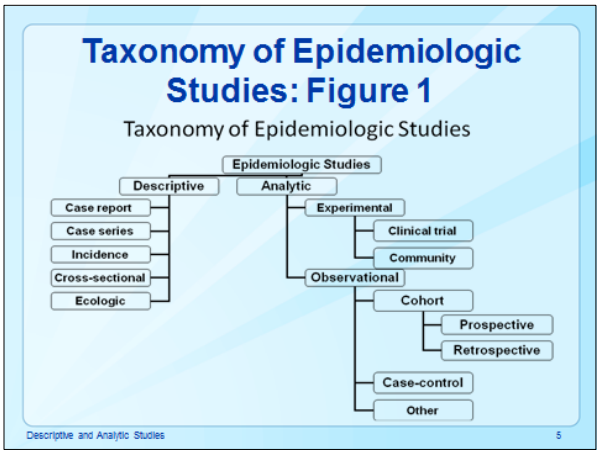
Slide 4

Question

- **Ask the question on the slide.**
- **Solicit responses and review the answers on the slide.**

5 minutes

Slide 5



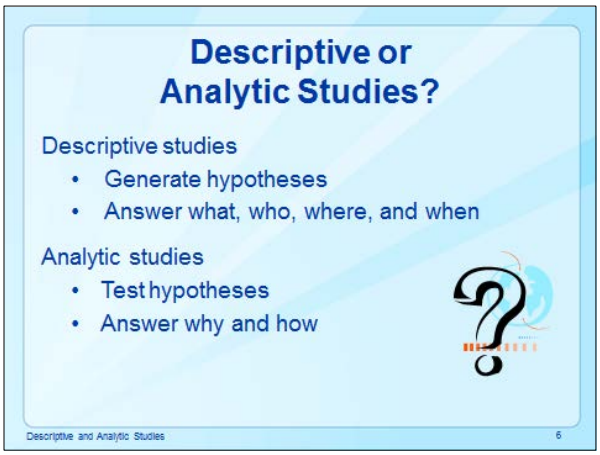
Question

- **Ask participants to turn to the appropriate page in their participant workbook to review the taxonomy of epidemiologic studies.**
- **Ask for a show of hands of who has had experience conducting these types of studies (cross-sectional, cohort, case-control).**
- **Tell participants you will begin the discussion on descriptive studies, in particular, cross-sectional.**

Duration/ Slide Number	What to Do/What to Say

3 minutes

Slide 6



Question

- **Ask:** Why would you conduct a descriptive study? What types of questions do descriptive studies ask?
- **CLICK SLIDE IN POWERPOINT to show answers.**
- **Explain that surveillance data can be used for descriptive studies.**

Question

- **Ask:** Why would you conduct an analytic study? What types of questions do analytic studies ask?
- **CLICK SLIDE IN POWERPOINT to show answers.**

1 minute

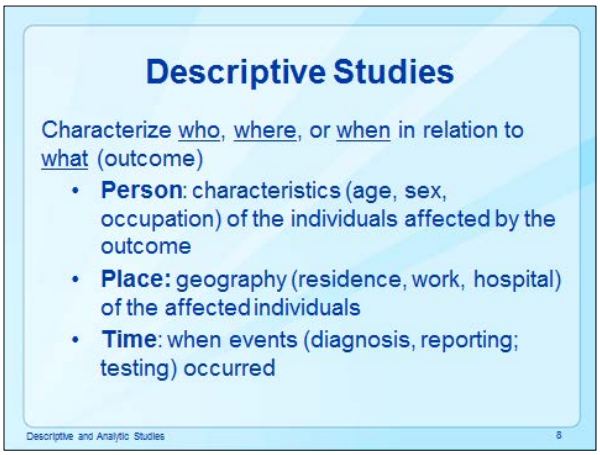
Slide 7

- **Tell participants that you will spend time discussing descriptive studies in more detail.**

Duration/ Slide Number	What to Do/What to Say

2 minutes

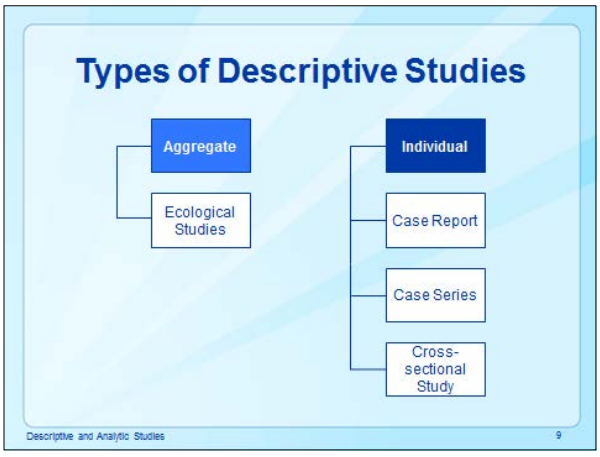
Slide 8



- **Read the slide.**
- **Explain that once the data have been identified, or collected, an analysis of the data should be done by summarizing person, place, and time into frequencies, proportions, and rates.**

5 minutes

Slide 9



- **Read the slide.**
- **Explain that descriptive studies can be categorized by the type of data used in the study. Data at a group-level, such as a city, district, or country, are referred to as aggregate-level data. An example of aggregate-level data would be the population growth rate of a country. Ecological studies are aggregate descriptive studies.**
- **Explain that the types of descriptive studies that use individual-level data include case reports, case series,**

Duration/ Slide Number	What to Do/What to Say

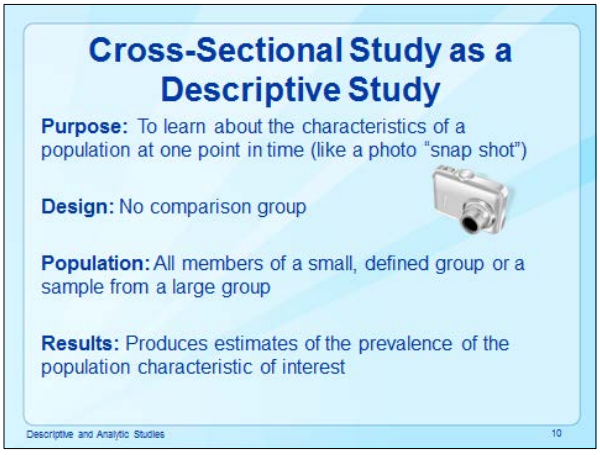
and cross-sectional studies.

Question

- **Ask:** Who can define case report? Case series? Cross-sectional study?
- **Suggested answers:** *A case report is simply a summary of one individual's experience with a particular outcome. Sometimes a single patient's illness or circumstances is so unusual that it warrants publication.*
- *A case series is a summary of a series of individuals with the same outcome. For example, in 1971 two physicians reported a case series of young women with clear cell adenocarcinoma of the vagina. This type of cancer was an exceedingly rare tumor that occurred predominantly in older women. This case series led to an analytic study that documented the association between this cancer and having been exposed to DES in utero (i.e., "DES babies.")*
- *A cross-sectional study is an example of a descriptive epidemiologic study that we will discuss on the next slide.*

3 minutes

Slide 10

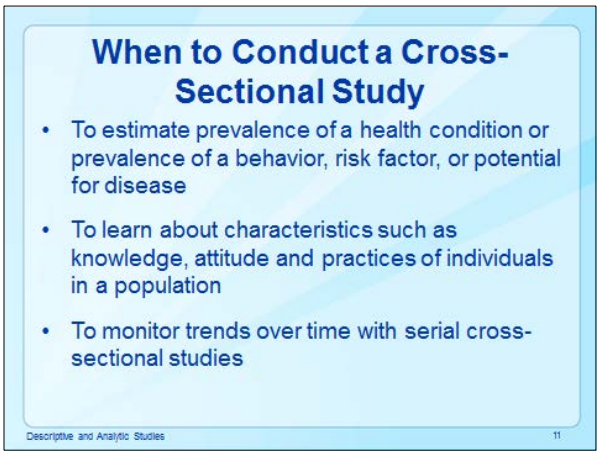


- **Explain that a cross-sectional study can be descriptive or analytic, depending on the purpose of the study and how it is designed. Here we discuss a descriptive study.**
- **Read the slide.**

Duration/ Slide Number	What to Do/What to Say

2 minutes

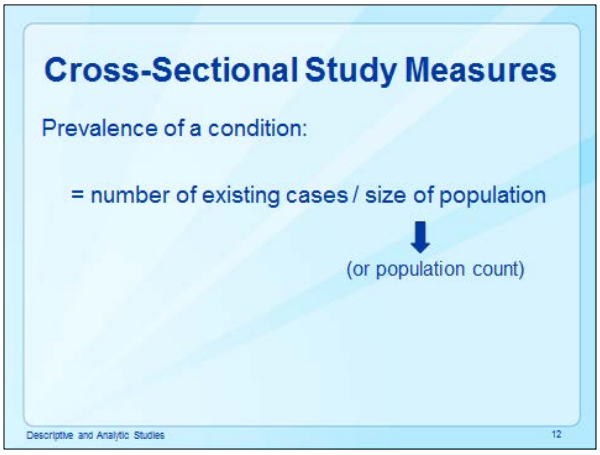
Slide 11



- **Read the slide.**
- **Explain that cross-sectional studies are especially suited for studying chronic conditions and their risk factors (e.g., use of a smoking). It is also well suited to study other behaviors that can lead to acute conditions, such as use of seat belts.**
- **Explain that a cross-sectional study uses a representative sample of the population. We have several ways of sampling a population; these are described later on in this module.**

2 minutes

Slide 12



Question

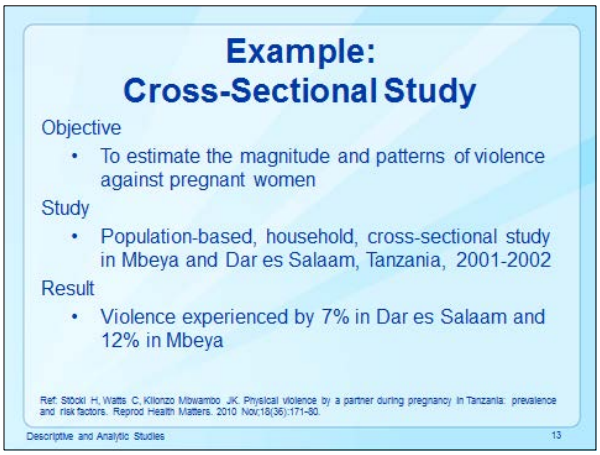
- **Explain that cross-sectional study measures prevalence.**
- **Ask:** Does anyone remember how prevalence is calculated?
- **CLICK to show the answer.**

Duration/ Slide Number	What to Do/What to Say

- **Suggested answer:** *Prevalence is the number of persons with the existing condition (cases) divided by the size of population or population count.*

3 minutes

Slide 13



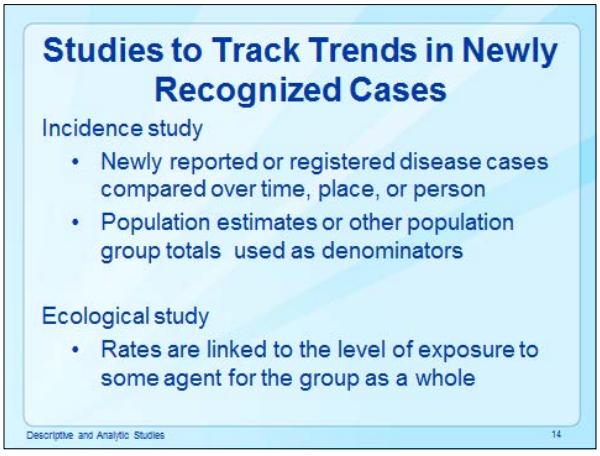
- **Read the slide.** *A case of violence can be very brief and would seldom be witnessed at the time of a visit to a person to obtain these data. Therefore, the investigator needs to specify or explain that this means violence at some defined time before the interview (or other determination of a violent act). Basically the goal is to obtain permanent or habitual / recurring exposures in cross sectional studies.*
- **Explain that these are the results of a real study.**

Question

- **Ask:** Who can provide another example of a cross-sectional descriptive study?
- **Possible answer:** *WHO STEPS survey.*

1 minute

Slide 14

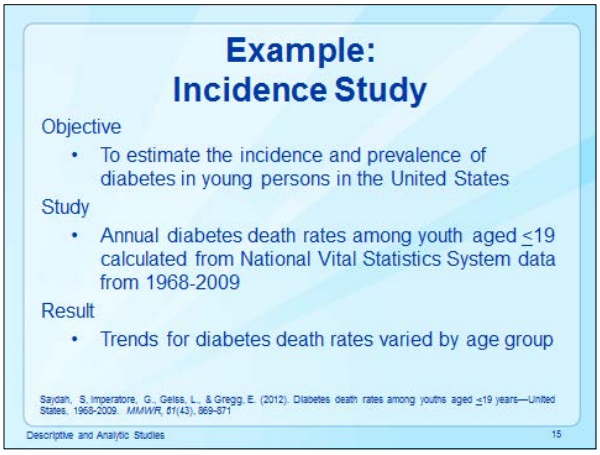


Duration/ Slide Number	What to Do/What to Say

- **Read the slide.**

1 minute

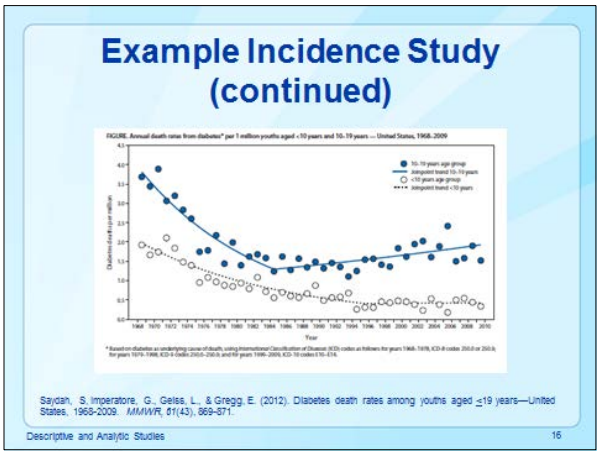
Slide 15



- **Read the slide.** *Among youth, deaths from diabetes are caused most often from direct acute complications of diabetes, such as ketoacidosis or hypoglycemia. For the most part, these causes of deaths are preventable. In these cases diabetes is listed as the underlying cause. Thus, investigators defined these deaths by using ICD codes that were relevant for different time periods to identify diabetes as the underlying cause of death.*

2 minutes

Slide 16

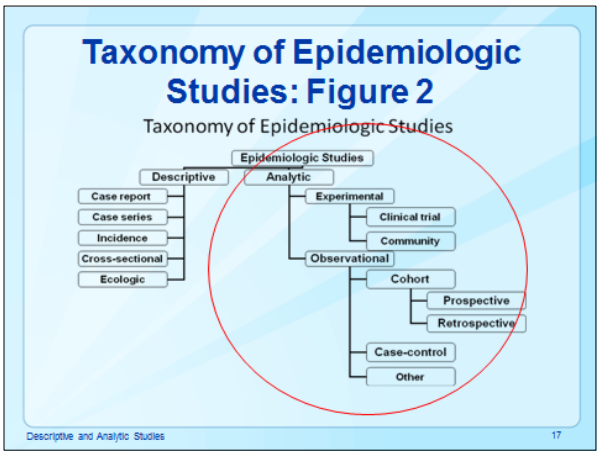


- **This slide reveals that there were different patterns of decrease in diabetes death rates among age groups. The percentage decrease was greater among youth <10 years than among youths 10-19 years.**
- **Explain that these are the results of a real study.**

Duration/ Slide Number	What to Do/What to Say

3 minutes

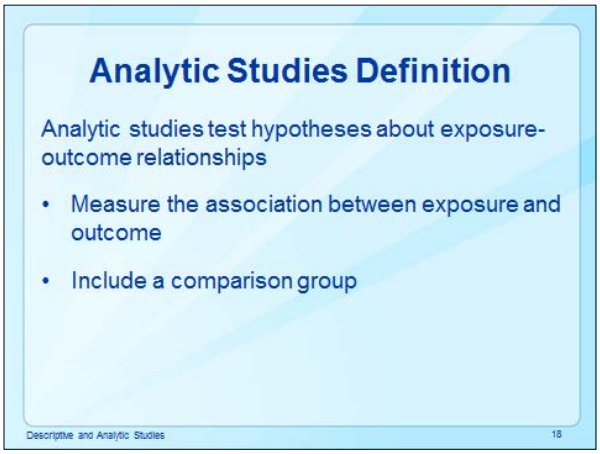
Slide 17



- **CLICK slide in PowerPoint until red circle appears.**
- **Tell participants that you will now discuss the right side of this flowchart: Analytic Studies.**

2 minutes

Slide 18

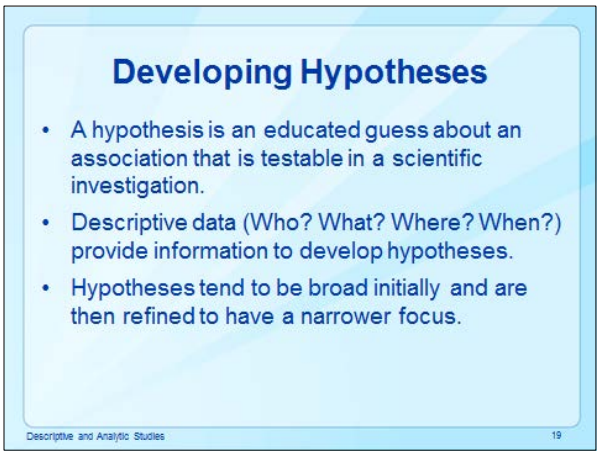


- **Read the slide.**
- **Explain that the key difference between analytic and descriptive studies is that specific exposures are determined that are not available otherwise. Descriptive studies are limited to demographic characteristics and some regular environmental measurements (e.g., mean particulate air pollution levels.) Therefore, analytic studies can measure the association between exposure and outcome, addressing the "why" question about the cause of disease.**

Duration/ Slide Number	What to Do/What to Say

3 minutes

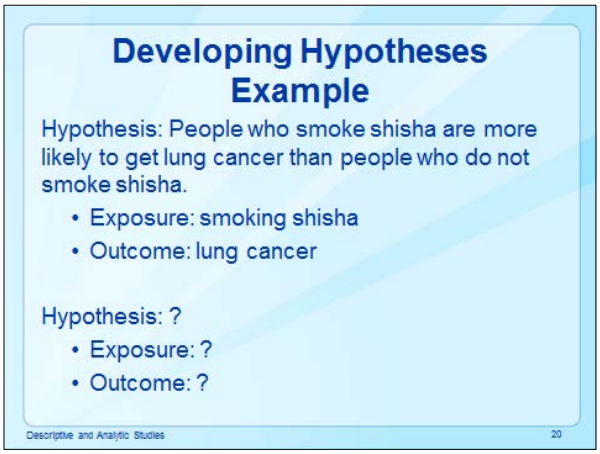
Slide 19



- **Say:** A hypothesis is an educated guess about an association between an exposure and an outcome that can be tested in a scientific investigation.
- **Say:** To develop a hypothesis, we need some information on which to base our educated guesses. Descriptive data provide information that can be used to develop hypotheses by using available data to suggest exposures that can be measured more exactly in analytic studies. These descriptive data include answers to the questions: Who? What? Where? Or When?
- **Explain that hypotheses tend to be broad initially, when less is known about what is causing a disease or health event. They are then refined to have a narrower focus.**
- **Say:** Let's go through some examples to make this clearer.

3 minutes

Slide 20



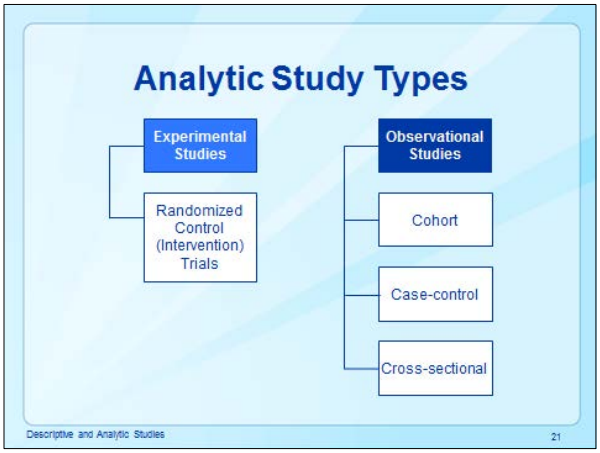
Duration/ Slide Number	What to Do/What to Say

Question

- **Read the example of a hypothesis.**
- **Ask participants to provide you with the exposure and outcome.**
- **CLICK slide in PowerPoint until the answers appear.**
- **Ask participants for another example of a hypothesis. Ask other participants for the exposure and outcome for that hypothesis.**

3 minutes

Slide 21

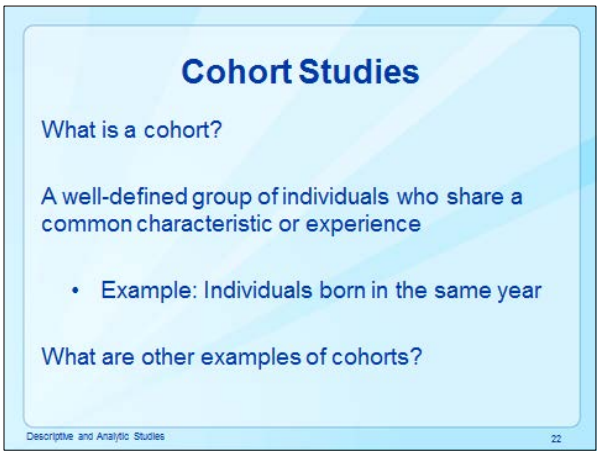


- **Explain that there are two types of analytic studies: experimental and observational, but for NCDs, the focus is on observational.**
- **Say:** In experimental studies, investigators assign study participants to an exposure of interest either blindly or knowingly (e.g., a health education campaign). In observational studies, investigators attempt to quantify exposures that study participants chose themselves (e.g., smoking) or had happen to them (e.g., an exposure to an unintended release of toxic fumes).
- **Read the three types of observational studies.**
- **Tell participants that they will learn about cohort, case-control and more about cross-sectional studies in this module.**

Duration/ Slide Number	What to Do/What to Say

3 minutes

Slide 22

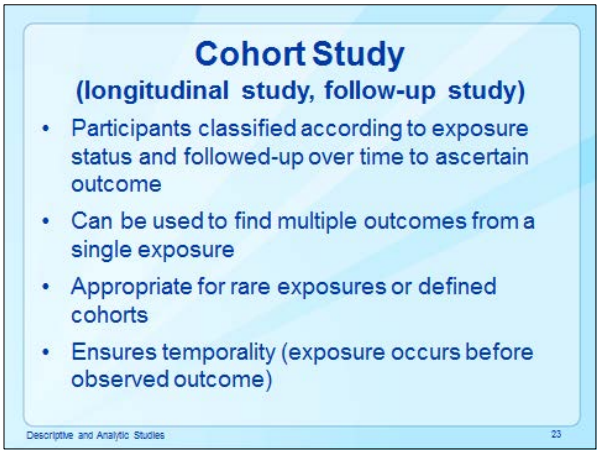


Question

- **Ask:** What is a cohort**?**
- **CLICK to show answer.**
- **Read the slide.**
- **CLICK to show question.**
- **Ask:** What are other examples of cohorts?
- **Suggested answers***: include people who work in the same place, children in the same grade at school, members of one tribe, women aged 65 or older.*

3 minutes

Slide 23



- **Read the slide.**
- **Explain that cohort studies are also called longitudinal studies or follow-up studies. They are used to ensure that the exposure being observed occurred before the outcome. Among observational studies, a cohort study**

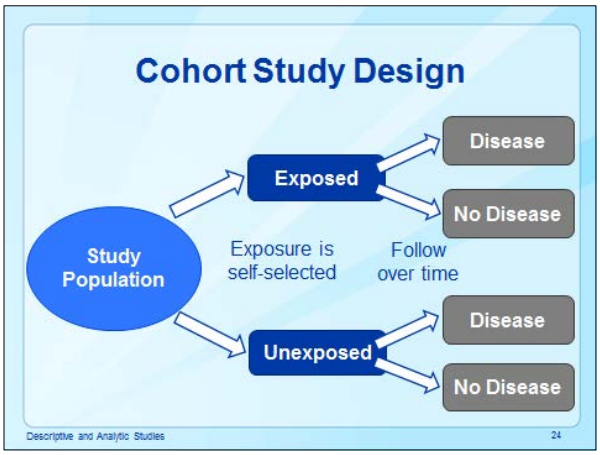
Duration/ Slide Number	What to Do/What to Say

produces the strongest evidence of a causal relationship between exposure and outcome. Even though it may not prove causality, a cohort study may provide sufficient evidence to initiate interventions. (Review Bradford Hill regarding causation in epidemiology—Hill, Austin Bradford (1965). "The Environment and Disease: Association or Causation?". *Proceedings of the Royal Society of Medicine* 58 (5): 295–300. PMC 1898525. PMID 14283879.)

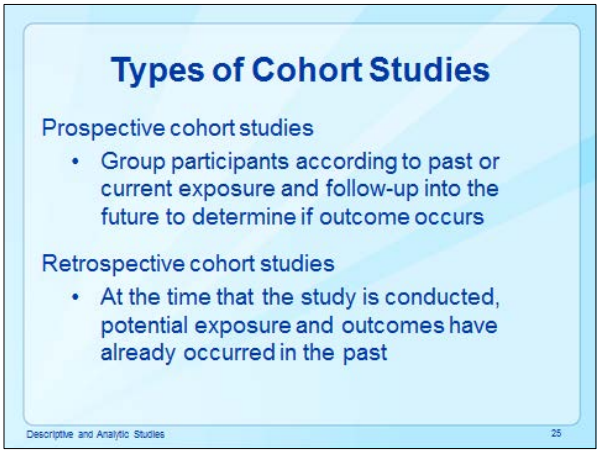
- **Explain that a cohort study is particularly good for studying the effects of rare exposures (e.g., the effects of a release of radioactive material from a nuclear facility on the people in the facility and in the facility vicinity).**

3 minutes

Slide 24



- **Say:** Here is a diagram that helps illustrate how a cohort study is designed. You start by identifying your study population (or cohort) and then determine who is exposed and who is not exposed. Exposure status is self-selected or by circumstance, as opposed to a randomized controlled trial, where exposure is assigned by the investigator. Then, we compare the disease experience between persons who were exposed and persons who were not exposed by following the groups over time to determine if they develop the outcome.
- **Explain that it is important for cohort members to not have the outcome at the beginning of the study because they should be "at risk" for the outcome. Participants cannot be "at risk" for the outcome if they already have the outcome. Any potential participant who already has the outcome should be excluded from the study. (However, this can be a particular problem in diseases such as cancer or other chronic diseases that are slowly**

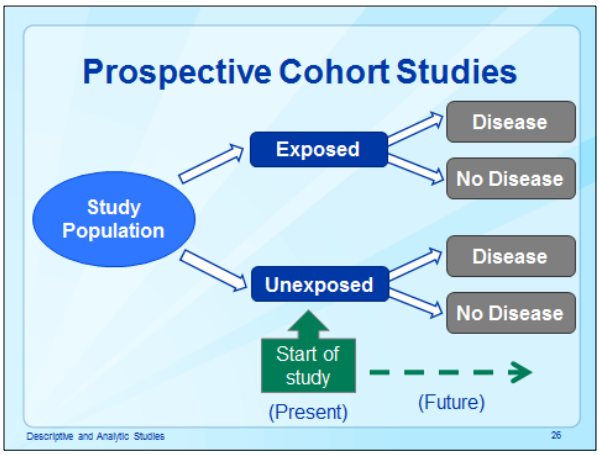
Duration/ Slide Number	What to Do/What to Say
	progressive, resulting in the person having undetectable disease for many years. These persons would not be excluded from the cohort when the disease is finally detected.)
3 minutes **Slide 25**	

- **Explain that there are two types of cohort studies, prospective and retrospective cohort studies, which are based on the timing of the investigator in observing the process of exposure and disease.**
- **Say:** In a prospective cohort study, the investigator groups participants based on the past or current exposure status and then follows them into the future to see whether or not they experience the outcome.
- **Say:** In a retrospective cohort study, the investigator examines exposures and outcomes that have already occurred in the past. Retrospective cohort studies are sometimes called historical studies.

Duration/ Slide Number	What to Do/What to Say

5 minutes

Slide 26



- **Say:** As the name implies, prospective studies look forward in time. At the start of the study the investigator categorizes the members of the study population as exposed or unexposed based on their past or current exposure status.

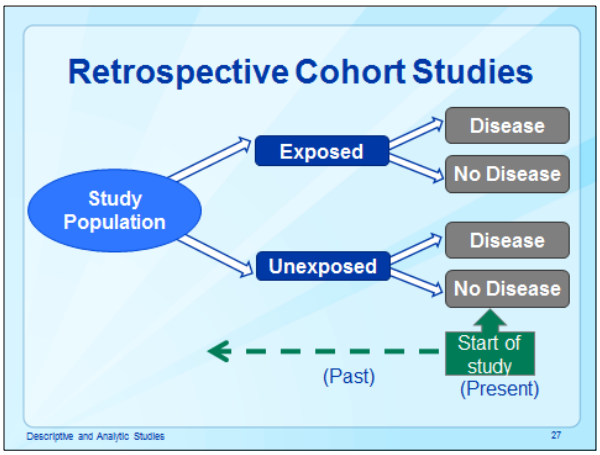
Question

- **Ask:** Can anyone describe an example of a prospective study?
- **Provide the following example***: In 1948 in the U.S. city of Framingham, investigators enrolled over 5,000 men and women, administered questionnaires and examined them to document a wide variety of behaviors and other possible risk factors*
- **CLICK SLIDE IN POWERPOINT to make black box and "(Present)" appear.**
- **Say:** Then the investigator follows the study subjects into the future to assess whether or not they have the outcome.
- **Say:** In Framingham, these men and women were examined every two years to document the occurrence of any type of cardiovascular disease.
- **CLICK SLIDE IN POWERPOINT to make arrow and "(Future)" appear.**

Duration/ Slide Number	What to Do/What to Say

5 minutes

Slide 27



- **Say:** In a retrospective cohort study, the investigator starts the study after both the exposure and outcome have occurred. Investigators are often able to use existing records (e.g., occupational health records) to select a cohort, determine each participant's exposure status, and classify outcomes.
- **CLICK SLIDE IN POWERPOINT to make black box and "(Present)" appear.**
- **Say:** The investigator looks back in time to see who was exposed and unexposed and subsequently, who developed the outcome of interest.
- **CLICK to make arrow and "(Past)" appear.**

Question

- **Ask:** Can anyone provide an example of a retrospective cohort study:
- **Provide the following example, if needed:** Exposure to high levels of carbon monoxide has been shown to increase the risk of cardiovascular disease. Public health officials in New York City wanted to know whether workers with long-term low-level exposure to carbon monoxide were also at increased risk of developing cardiovascular disease. They decided to conduct a retrospective cohort study, looking at causes of death among traffic control officers on bridges and in tunnels compared with the general New York City population. This investigation documented an increased rate of mortality from atherosclerotic heart disease, particularly among the tunnel workers. (Ref: Stern, et al.)
- **Explain that when you want to determine whether a cohort study is prospective or retrospective, the easiest**

Duration/ Slide Number	What to Do/What to Say

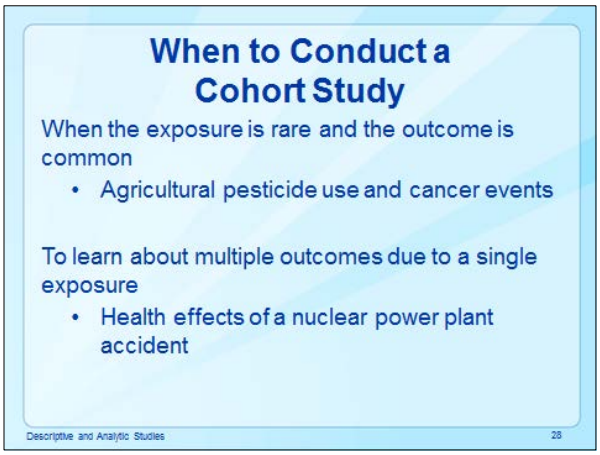
way to think about it is to ask yourself "Where in the sequence of events was the investigator in relationship to the outcome of interest?"

- **If the outcomes took place after the investigator started the study, then it is a <u>prospective</u> cohort study.**
- **If the outcomes had already occurred when the investigator started the study, then it is a <u>retrospective</u> cohort study.**

- **In some cases, the investigator starts a study and some people have already had the outcome, and some people will be followed to see if they develop the outcome in the future. These types of cohort studies are called <u>ambi-directional</u> cohort studies.**

2 minutes

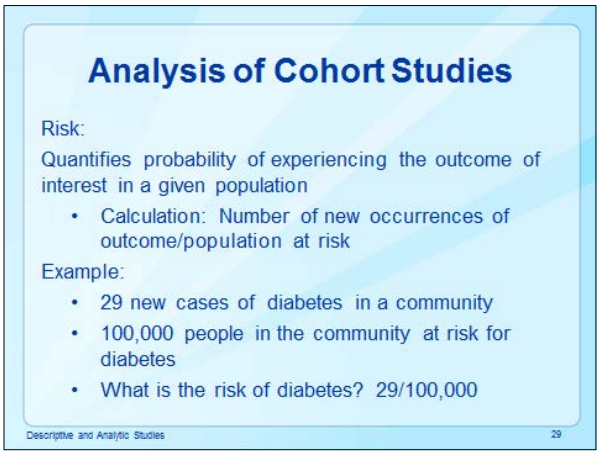
Slide 28



- **Read the slide.**

2 minutes

Slide 29



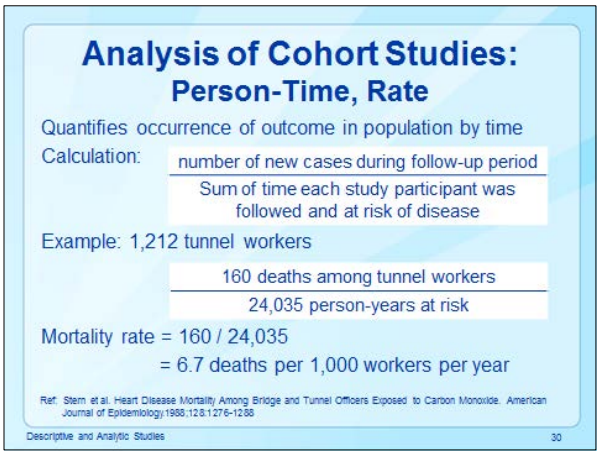
Duration/ Slide Number	What to Do/What to Say

- **Explain that cohort studies evaluate a population's risk of disease. Risk is calculated by dividing the number of new occurrences of the disease by the population at risk when the observation period begins.**
- **Read the example.**
- **Ask:** What is the risk of diabetes?
- **CLICK SLIDE IN POWERPOINT to show the answer.**

Question

3 minutes

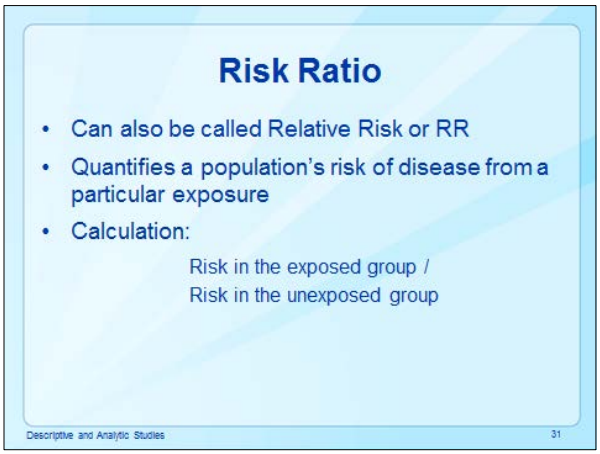
Slide 30



- **Read the slide.**

2 minutes

Slide 31



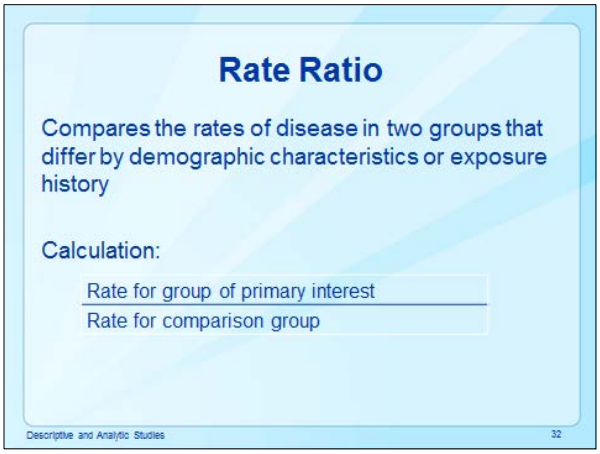
- **Explain that two risks can be compared to provide a risk ratio (RR), also known as a relative risk. This measure is a ratio that compares the risk of disease in a group with a particular exposure to an unexposed group.**

Duration/ Slide Number	What to Do/What to Say

- **Read how risk ratio is calculated.**

2 minutes

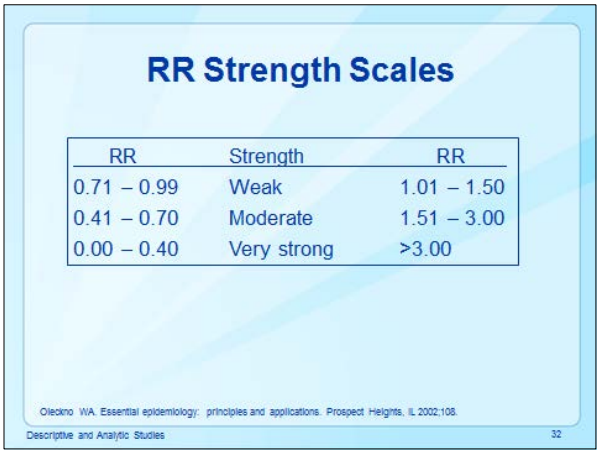
Slide 32



- **Explain that rate ratio is another tool that is helpful for comparing rates between groups.**
- **Read the slide.**

3 minutes

Slide 33



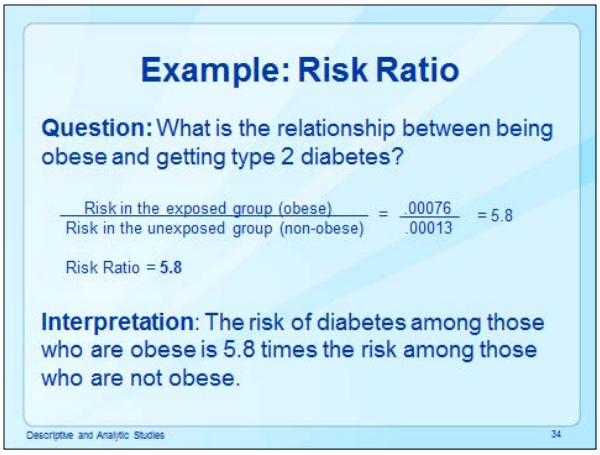
- **Say:** Once you calculate a risk ratio, rate ratio, or any measure of association, it is important to understand what this value means. The scale on this slide is one of many standard scales used to assess the strength of a risk ratio. This scale shows the strength of association between the exposure and outcome for a cohort study.
- Keep in mind that the confidence interval attached to your measure of association will also impact your interpretation of the data.

Duration/ Slide Number	What to Do/What to Say

- **Read the slide.**
- **Explain that you can use similar guidelines for odds ratios and other ratio measures of effect.**

3 minutes

Slide 34



- **Explain that is a fictional example.**
- **Read the question on the slide.**
- **Explain that to answer the question, they would calculate the risk ratio, which compares the risk of diabetes among those who are obese with the risk among those who are not obese.**
- **Show the calculation for risk ratio, but explain that this module does not teach them *how* to do the calculation – only to interpret the result.**
- **Read the interpretation.**

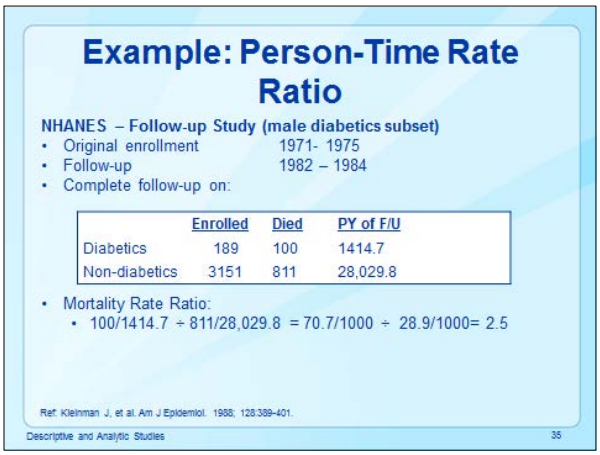
Question

- **Ask:** Is 5.8 a weak, moderate, or very strong association?
- **Answer:** *Very strong*.

Duration/ Slide Number	What to Do/What to Say

5 minutes

Slide 35



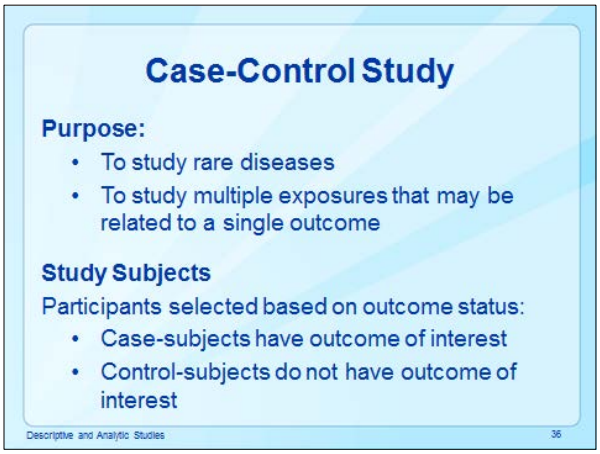
- Say: In this example of person-time rate ratio, let us assume that the question that needed to be answered was: How does the mortality rate for diabetics in the study compare with the mortality rates for non-diabetics in the study?
- **Ask for a volunteer to explain the example on the slide or person-time rate ratio.**

Question

- **Ask:** How would you interpret the mortality rate ratio of 2.5?
- **Answer:** *The mortality rate of diabetics was 2.5 times as high as the mortality rate of non-diabetics.*
- **Ask:** Is 2.5 a weak, moderate, or very strong association?
- **Answer:** *moderate*.

2 minutes

Slide 36



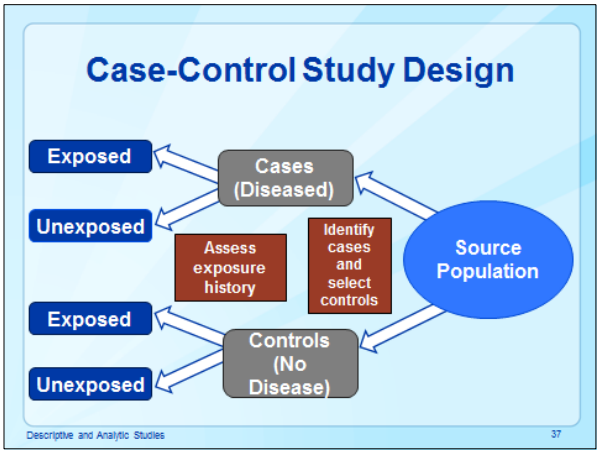
- **Tell participants you will now discuss another type of observational study: case-control.**

Duration/ Slide Number	What to Do/What to Say

- **Read the slide.**

3 minutes

Slide 37

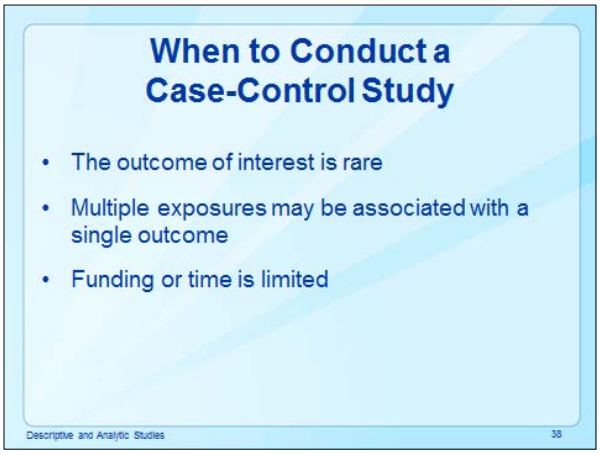


- **Say:** Here is a diagram that helps illustrate how a case-control study is designed. You start by identifying your cases from the source population and then selecting appropriate controls from the same source population. Next, you collect data on exposure history from all of the cases and controls. Then, we compare the exposure experience between cases and controls to see if there is a difference.
- **Explain that selecting cases and controls for a study is not as simple as finding people with and without the outcome of interest. Cases may be difficult to find if the outcome is rare.**
- **Explain that potential sources of cases include hospitals, clinics, and registries. The selection of appropriate controls requires an approach to minimize bias in study results. Selection of controls is beyond the scope of this presentation but references are listed at the end of the presentation for those who are interested.**

Duration/ Slide Number	What to Do/What to Say

1 minute

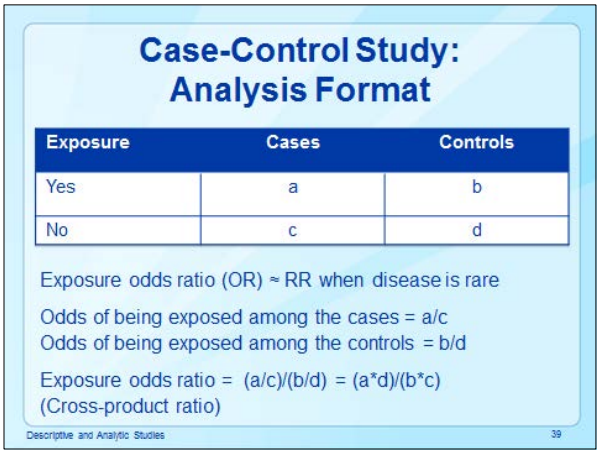
Slide 38



- **Read the slide.**

5 minutes

Slide 39



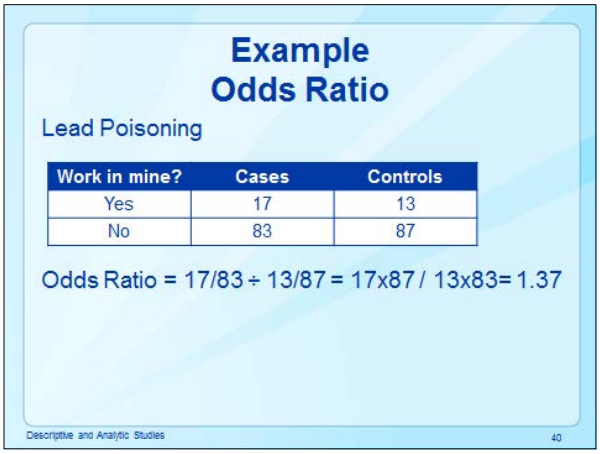
- **Say:** The exposure odds ratio (often referred to as simply, the odds ratio or OR) approximates the risk ratio when the disease is rare (less than 5% in the source population).
- In case-control studies, the low totals for the exposed and unexposed groups simply reflect how many cases and controls were enrolled in the study; they are not the true denominators needed to calculate risk of disease among the exposed and unexposed groups. In other words, we cannot calculate risk of disease.
- If we cannot calculate risks, we cannot calculate risk ratios.
- However, the controls are intended to represent the population-at-risk, but are only a sample of that population. We can calculate an odds ratio. And if the disease is rare, the OR provides a good approximation of the RR.

Duration/ Slide Number	What to Do/What to Say

- **Read on the slide how to calculate OR.**

5 minutes

Slide 40



- **Read the question on the slide.**
- **Explain that to answer the question, they would calculate the odds ratio.**
- **Show the calculation, but explain that this module does not teach them *how* to do the calculation – only to interpret the result.**

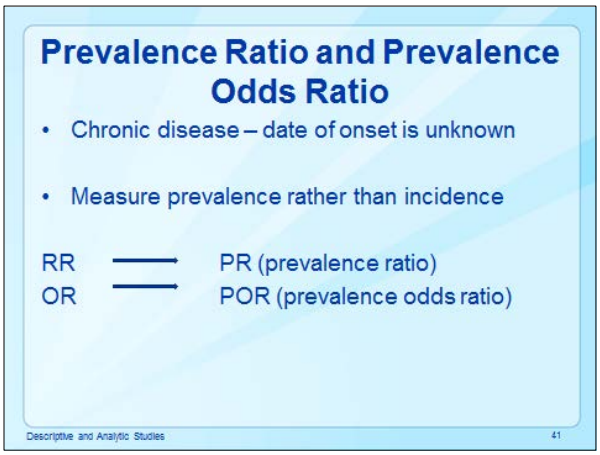
Question

- **Ask:** How would interpret the OR?
- **Answer:** *Mine workers had 1.4 times the odds of developing lead poisoning than did people who did not work in a mine.*
- **Say:** We can extrapolate and say that these data are consistent with the finding that mine workers may be at slightly increased risk of lead poisoning compared to non-miners. We say “slightly” because: 1) 1.4 represents a weak association, and 2) we have not calculated p-value or confidence intervals to determine whether this slight increase could simply be due to chance.

Duration/ Slide Number	What to Do/What to Say

2 minutes

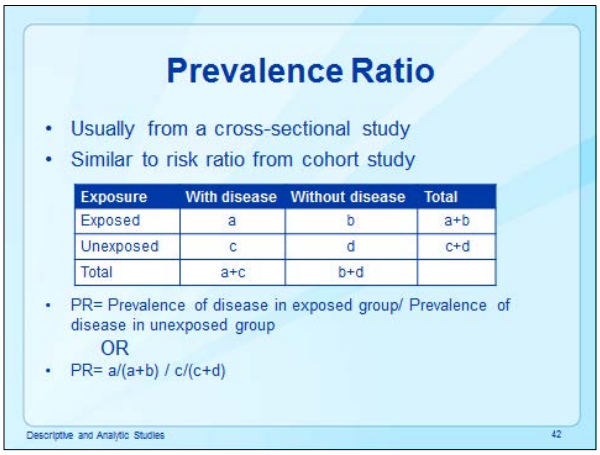
Slide 41



- **Say:** For many chronic diseases, date of onset is unknown and burden of disease is important. Therefore, you are more likely to measure prevalence rather than incidence. If you measure the outcome in terms of prevalence, the corresponding measures of association are the **prevalence ratio** and **prevalence odds ratio**.

3 minutes

Slide 42

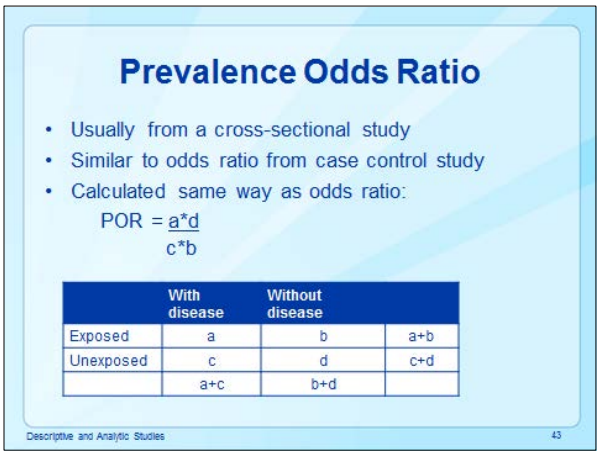


- **Explain that the measure of association used for a chronic disease cross-sectional study is prevalence ratio.**
- **Mention that a Prevalence Ratio could also arise from a retrospective cohort study when the onset is not known. Also remember that for many chronic diseases the onset of the problem is silent and may come many years before any overt disease.**
- **Read the remainder of the slide.**

Duration/ Slide Number	What to Do/What to Say

3 minutes

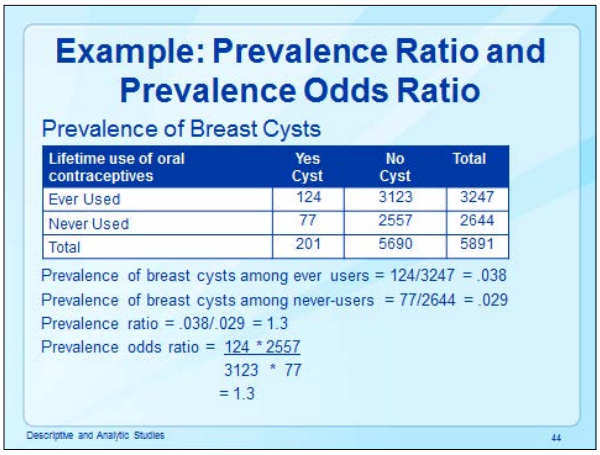
Slide 43



- **Read the slide. Prevalence odds ratios can also be used in case control studies if the cases are prevalent cases rather than new cases. This can often occur with chronic conditions because the onset or another marker for the beginning of the disease is not possible to determine.**
- **Tell participants that you will now show an example of PR and POR.**

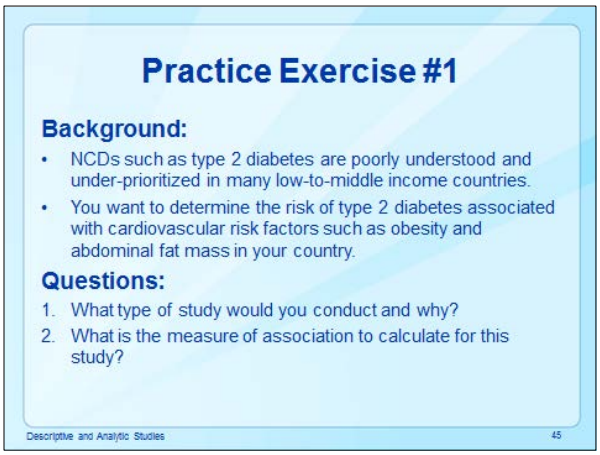
5 minutes

Slide 44



- **Say:** This is a hypothetical example of a cross-sectional study where investigators wanted to assess whether the prevalence of breast cysts was associated with having ever used oral contraceptives.
- The results are shown in this table.
- **Ask:** Who can interpret the PR of 1.3?
- ***Possible answer:*** *The probability of having breast cysts is*

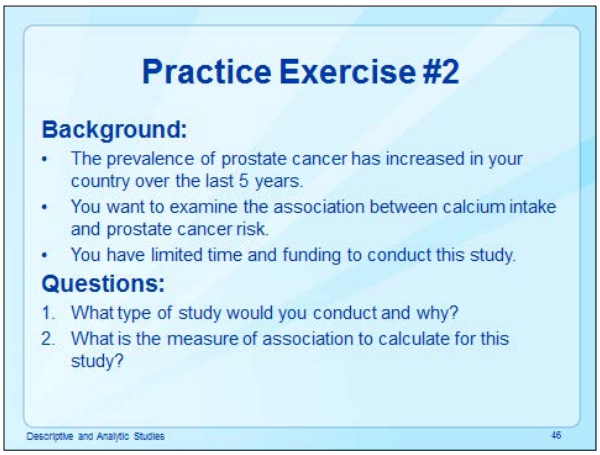
Question

Duration/ Slide Number	What to Do/What to Say
Question	*1.3 times as high in ever users of oral contraceptives as in never users of oral contraceptives.* • **Ask:** Who can interpret the POR of 1.3? • ***Possible answer:*** *The odds of getting breast cysts is 1.3 times as high in ever users of oral contraceptives than with never users of oral contraceptives.*
Question	• **Ask:** Why do you think the PR and POR are very similar? • ***Answer:*** *Because overall prevalence is very low.* • **Ask:** What do you think about testing >5,000 people to get this result? • ***Answer:*** *A case control study using medical records to identify new breast cyst cases or prevalent breast cyst cases and unaffected controls would require 10 times fewer persons. Do not do massive studies when a much more efficient design gets the same answer. Use cross sectional design for common conditions like hypertension.* • **Tell participants that they will now practice what they have learned about descriptive and analytic studies.**
10 minute **Slide 45**	
Activity	• **Ask participants to work in small groups or with another colleague to complete Practice Exercise #1.** • **After 5 minutes, ask for one group to provide the answer to question 1 and another group to answer question 2.** • ***Possible answer to question 1:*** *Because you want to learn about multiple risk factors for diabetes, you would conduct a case control study.* • ***Possible answer to question 2:*** *Odds ratio*

Duration/ Slide Number	What to Do/What to Say

10 minutes

Slide 46

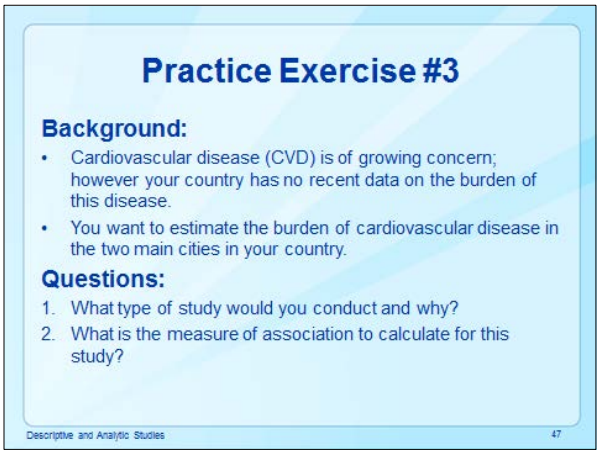


Activity

- **Ask participants to work in small groups or with another colleague to complete Practice Exercise #2.**
- **After 5 minutes, ask for one group to provide the answer to question 1 and another group to answer question 2.**
- ***Possible answer to question 1:*** *You could conduct a case control study because funding and time is limited.*
- ***Possible answer to question 2:*** *Since the controls are only a sample of the population at risk, you would calculate odds ratio.*

10 minutes

Slide 47



Activity

- **Ask participants to work in small groups or with another colleague to complete Practice Exercise #3.**
- **After 5 minutes, ask for one group to provide the answer to question 1 and another group to answer question 2.**
- ***Possible answer to question 1:*** *You could conduct a cross-*

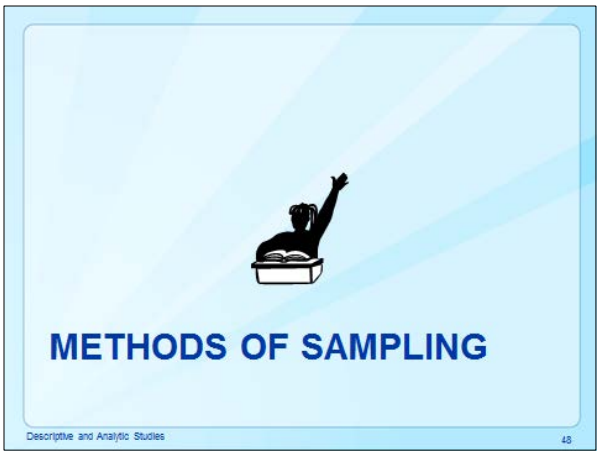
Duration/ Slide Number	What to Do/What to Say

sectional study to estimate the prevalence of CVD.

- ***Possible answer to question 2:*** *PR and POR*

1 minute

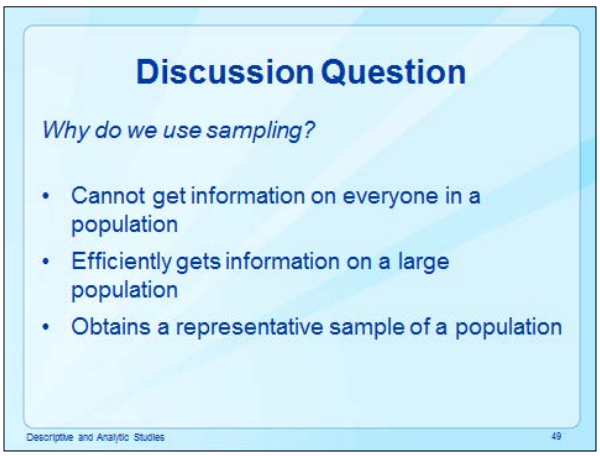
Slide 48



- **Explain that you will next discuss methods of sampling.**

3 minutes

Slide 49



Question

- **Say:** Sampling is a procedure you use to select a specified number of persons in a population for a study.
- **Ask:** Can anyone tell me why we use sampling?
- **CLICK SLIDE IN POWERPOINT to show possible answers.**
- **Possible Answers:** *In most cases we can't get information on every person in the population we want to study, and often it would be prohibitively expensive to try to collect data on every person. Sampling, when done correctly, is an efficient way to get information on a population. A sample should accurately reflect distribution of relevant variables in a*

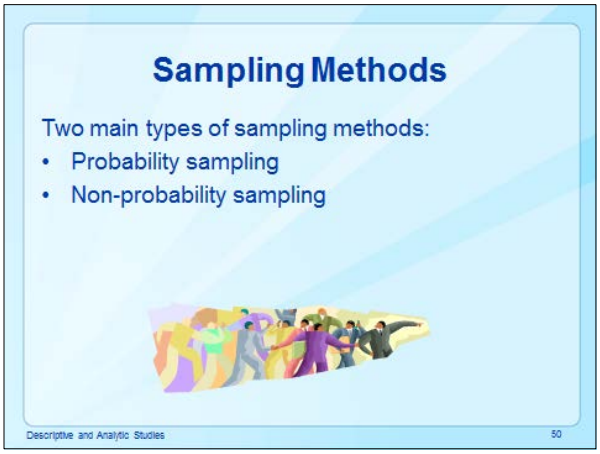
Duration/ Slide Number	What to Do/What to Say

population according to person, place, and time. Representativeness is essential in order to generalize the results of the study to the larger population.

Non-probability sampling is NOT representative of the population. However, this type of sampling can have some practical uses in analytic studies (e.g. as a study group for a cross sectional, retrospective cohort, or cohort study or as a control group for a case control study.

3 minutes

Slide 50

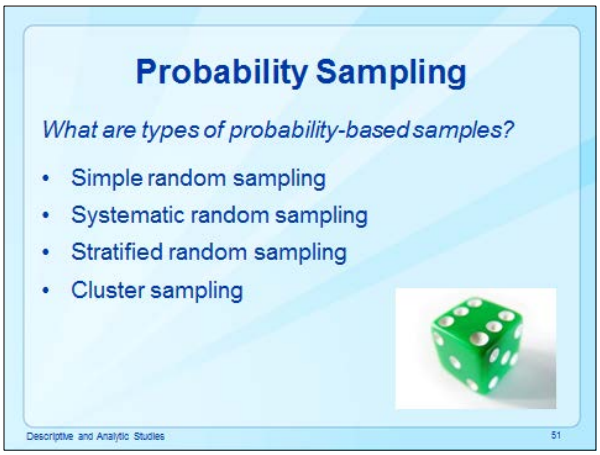


- **Say:** With probability sampling methods, each person has a known (non-zero) chance of being chosen for the sample. Sampling methods using random sampling require that each member of the population has an equal chance of being selected, but in more complex sampling schemes the probability of being selected may not be equal for certain people.
- **Say:** However, there are some situations where a probability sample is not possible because some people have a zero chance of being selected (such as persons without a telephone in a survey that uses random-digit dialing). In non-probability sampling, the chance of being selected is not known and is not representative of the population.
- **Explain that probability sampling is more complex, more time-consuming, and usually more expensive than non-probability sampling. But, with probability sampling, the sample is less likely to be biased and is more likely to provide results that can be generalized to the target population.**

Duration/ Slide Number	What to Do/What to Say

2 minutes

Slide 51

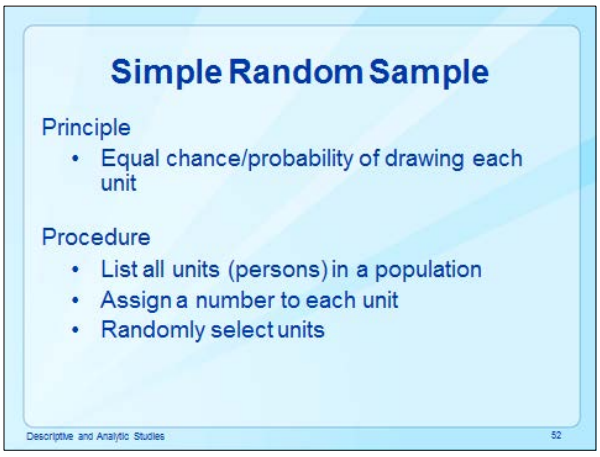


Question

- **Ask:** Can anyone name one or more types of probability-based samples?
- **CLICK SLIDE IN POWERPOINT to show possible answers.**
- **Read the slide.**

3 minutes

Slide 52



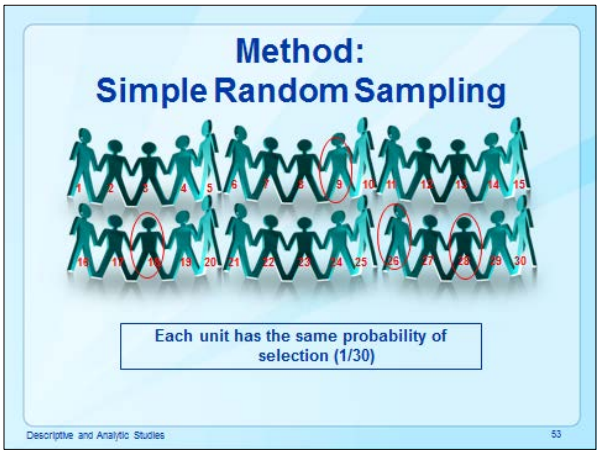
- **Say:** The first method is simple random sampling. Simple random sampling gives every member of the population an equal chance of being selected for the sample.
- **Explain that to take a simple random sample, you**
  - **Determine how many members of the target population you want to include in the sample (your sample size).**
  - **List all of the units (usually persons, but can be households or other types of units) of the population. This is called the sampling frame.**

Duration/ Slide Number	What to Do/What to Say

- **Assign a number to each unit (using consecutive numbers).**
- **Select units randomly using some random process to select the assigned numbers (e.g., a random number generator or a random number table).**

2 minutes

Slide 53

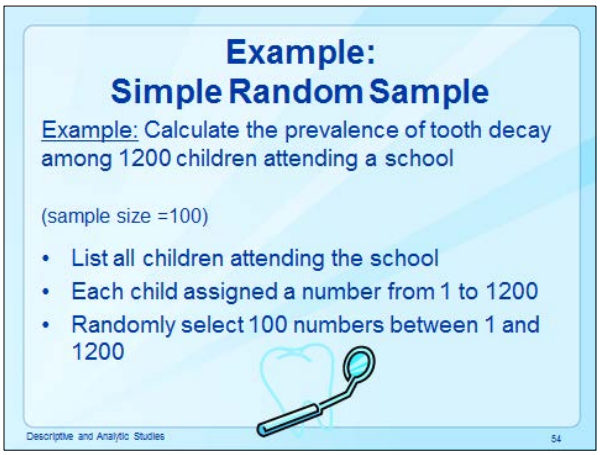


- **Explain that the slide is a sampling frame of 30 units. The red ovals represent 4 units that have been randomly selected.**
- **Read the slide.**
- **Say:** Simple random sampling is easy to conduct when the sampling frame is small. However, it can be expensive and difficult to conduct when the sampling frame is large. For example, if the sampling frame is a large city like (*insert your city)* or (*insert your city)*, it is almost impossible to list and assign each person a number. It is also expensive and logistically difficult if the sampling frame is dispersed over a wide geographic area (e.g., an entire country).

Duration/ Slide Number	What to Do/What to Say

3 minute

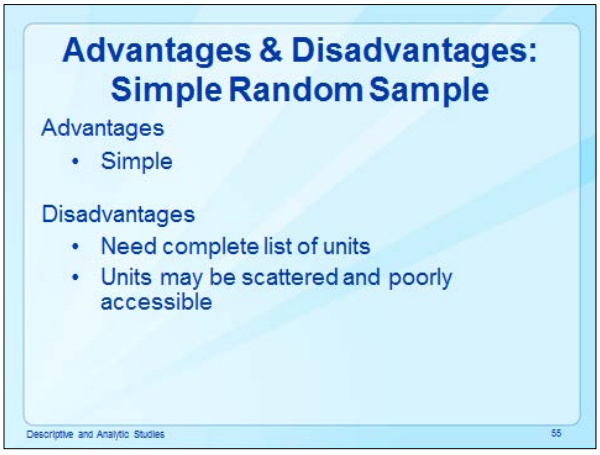
Slide 54



- **Say:** Here is an example of how to select a simple random sample. In this example, we are conducting a study to calculate the prevalence of tooth decay among schoolchildren at a particular school with 1200 students. We calculate a sample size of 100 students.
  - o First, we must make a list of all 1200 students attending the school.
  - o Then we assign a unique number from 1 to 1200 to each child.
  - o Then, we randomly select 100 students by choosing 100 numbers at random between 1 and 1200. These 100 selected students are the random sample. The prevalence of tooth decay among this group should approximate the prevalence of tooth decay in all 1200 students, without having to survey the entire school.

2 minutes

Slide 55

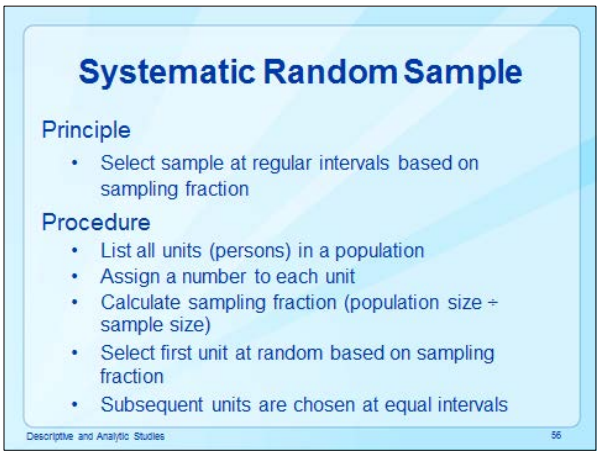


Duration/ Slide Number	What to Do/What to Say

- **Say**: Simple random sampling does not require any additional information on the population other than a complete list of the members and their contact information. For example, you may use a list of your city's registered voters as your population and select individuals from this population.
- **Say**: However, while simple random sampling is easy to apply to small populations, it can be difficult to use with large populations because all members of the population must be identified and enumerated prior to sampling. In addition, since the sample may be widely dispersed geographically speaking, a simple random sample can also result in a more expensive survey if the survey involves personal interviews.

3 minutes

Slide 56

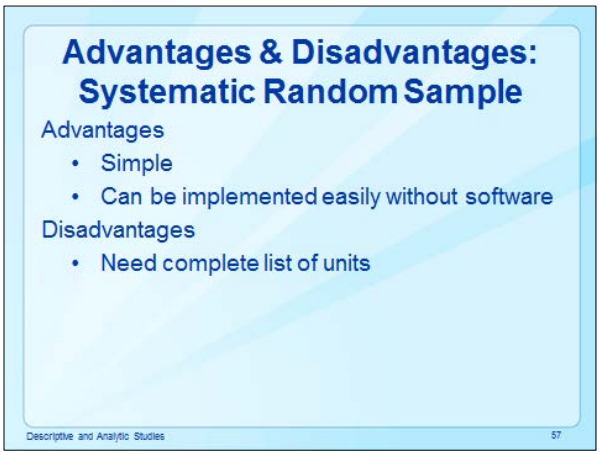


- **Explain that systematic sampling is similar to simple random sampling in that it also gives every member of the population an equal chance of being selected. But rather than randomly selecting all of the participants individually, a selection interval is determined, a starting point is randomly selected, and participants are selected based on the selection interval.**
- **Read the second bullet.**

Duration/ Slide Number	What to Do/What to Say

2 minutes

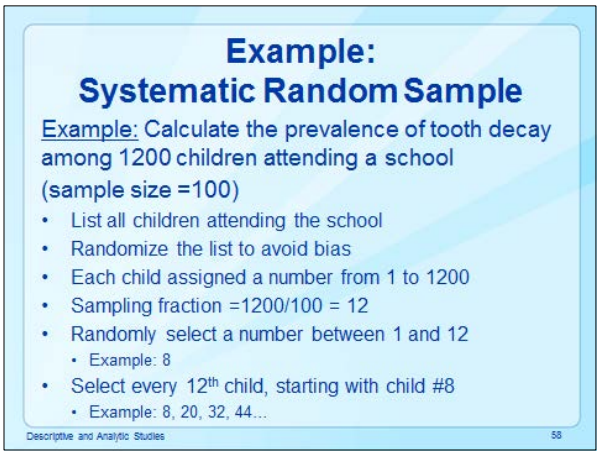
Slide 57



- **Read the slide.**

3 minutes

Slide 58

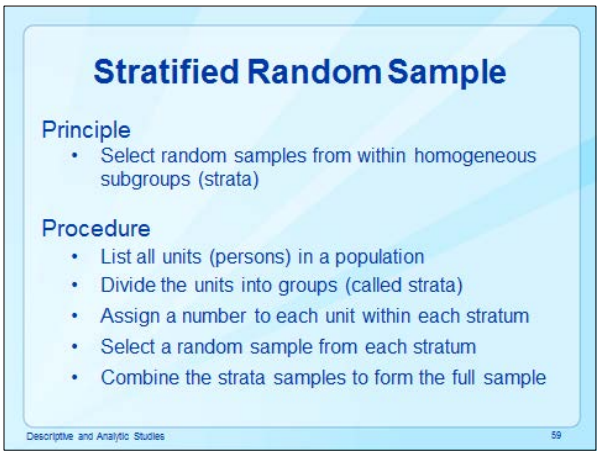


- **Read the example and point out that systematic random sampling is a way to select only a *part of* a population in order to get a representative set of people.**

Duration/ Slide Number	What to Do/What to Say

3 minutes

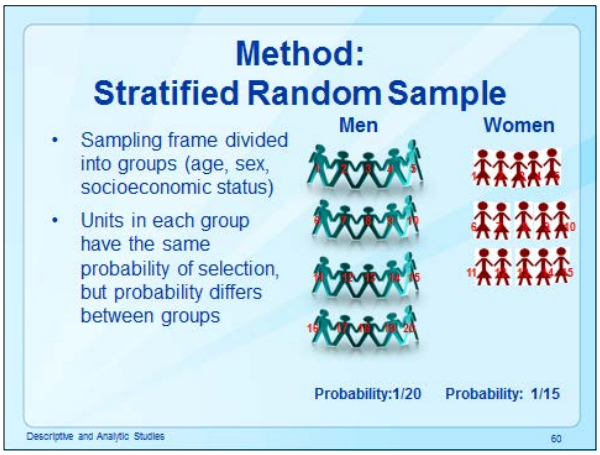
Slide 59



- **Explain that in stratified random sampling, random samples are selected from within homogenous subgroups called strata.**
- **Say:** To take a stratified random sample, you need to list all units in the population, then the population is divided into subgroups (strata) based on at least one common characteristic. For example, you may choose to stratify your sample by race, so that you get equal representation of different races in your sample. A number is assigned to each unit within each stratum, and then a random sample is selected within each stratum. Finally you combine the random samples within each stratum to form the full sample. Each member of a particular stratum has an equal chance of being selected; however, the probability of selection may differ between strata.

3 minutes

Slide 60

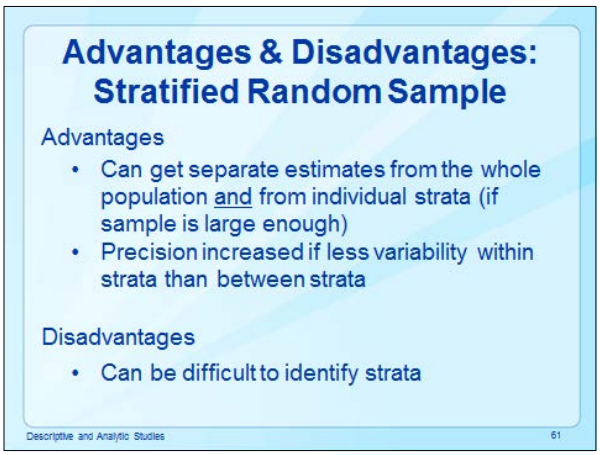


Duration/ Slide Number	What to Do/What to Say

- **Read the first bullet.**
- **Explain that when data are stratified, the selection process is similar to simple random sampling: all units are listed, numbered, and then randomly selected.**
- **Read the second bullet.**
- **Say:** An advantage of stratified random sampling is that investigators can ensure that small subgroups are included in the study. For example, in a survey of truck drivers, simple random sampling could miss the small number of women drivers. By stratifying the sampling frame into men and women, investigators ensure they include women drivers in their survey.

3 minutes

Slide 61



- **Say:** Stratified sampling is a commonly used probability method.
- **Explain that each stratum of the population is assured to be represented in the selected sample. As a result, stratified sampling is often used when one or more of the strata in a population are small relative to the other strata and could, by chance, be largely missed with simple random or systematic sampling.**
- **Explain that with stratified random sampling, separate estimates can be determined for each stratum (a real advantage in many settings) and the overall estimate for the population can be determined based on the estimates for each stratum. Because members within each stratum are more alike (at least on the stratifying characteristic), the estimate of the parameter of interest will be more precise compared to a simple random sample if there is**

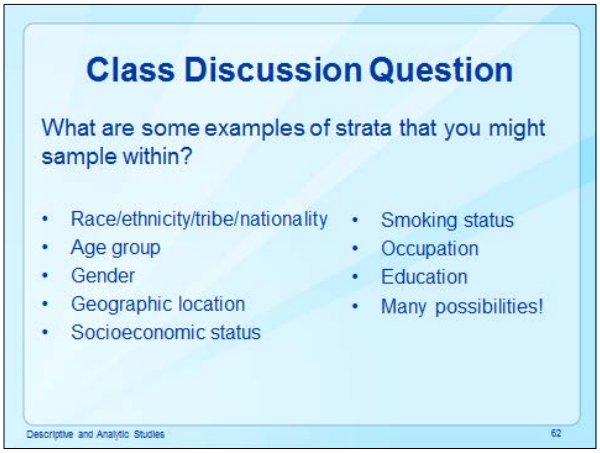
Duration/ Slide Number	What to Do/What to Say

less variability within the strata than between the strata.

- **Say:** A disadvantage to using stratified random sampling is that in some cases the strata can be difficult to identify – such as strata based on race or ethnicity, where multi-racial or ethnic persons may fit in more than one category.

3 minutes

Slide 62



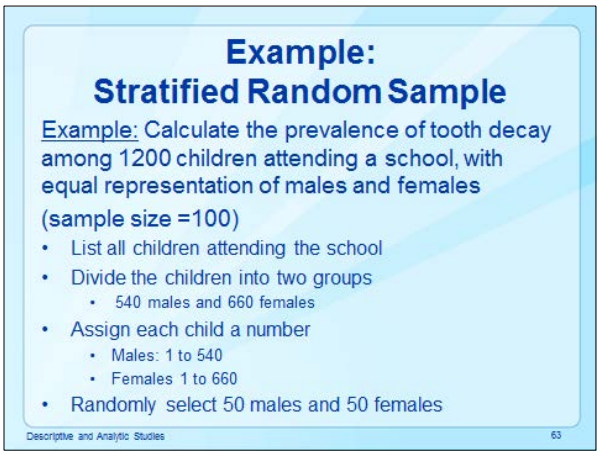
Question

- **Ask:** Can you think of some examples of strata that you might sample within?
- **CLICK SLIDE IN POWERPOINT to show possible answers:**
  - Race/ethnicity/tribe/nationality (e.g., Palestinian vs. Jordanian)
  - Age group
  - Gender
  - Geographic location
  - Socioeconomic status
  - Smoking status
  - Occupation
  - Education
- **Say:** The point of this exercise is to illustrate that there are many possibilities for factors which you may choose to stratify on, depending on what you want to study.

Duration/ Slide Number	What to Do/What to Say

2 minutes

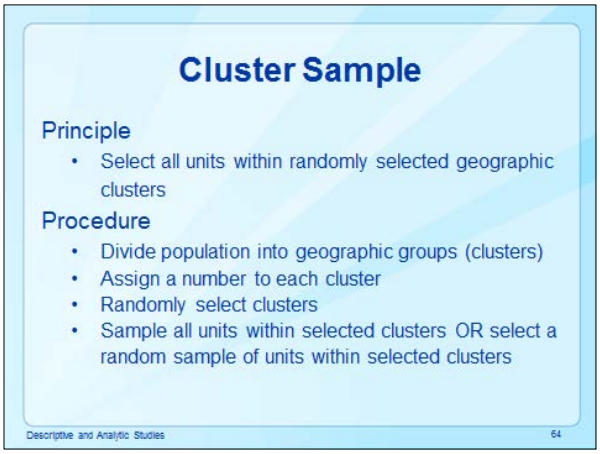
Slide 63



- **Say:** Now let's look at how to take a stratified random sample using the tooth decay study example. In this case we will divide the students into strata based on gender. We divide the students into two groups, males and females, which have 540 and 660 students respectively. We number the students in each group from 1 to 540 for the males and 1 to 660 for the females, and then randomly select 50 students from each group to get a total sample of 100 students.
- **Explain that with this sampling method, we can now estimate the prevalence of tooth decay separately for males and females, as well as for the total population of school children.**
- **Explain that when combined, you need to adjust because the sampling fraction is 50/540 for boys and 50/660 for girls.**

3 minutes

Slide 64

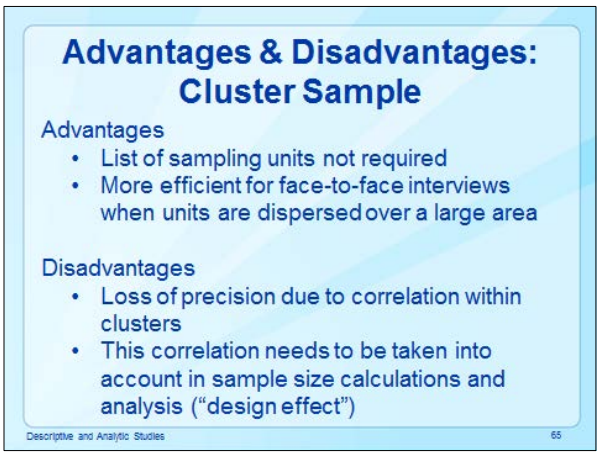


Duration/ Slide Number	What to Do/What to Say

- **Say:** In cluster sampling, the population to be studied is divided into natural, geographically distinct groups or "clusters" (such as schools, villages, or camps).
- Each cluster is assigned a number and then a sample of clusters is then randomly selected.
- Once the clusters are selected, all units within the selected clusters are included in the sample (called one-stage cluster sampling), OR, you can take a random sample of units within the selected clusters (called two-stage cluster sampling).
- This is the method used for WHO 30x7 immunization surveys and for most rapid epidemiologic assessment surveys after disasters.

2 minutes

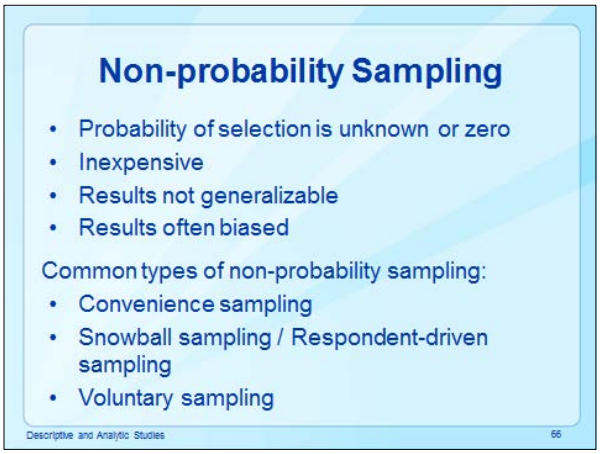
Slide 65



- **Read this slide.**

3 minutes

Slide 66

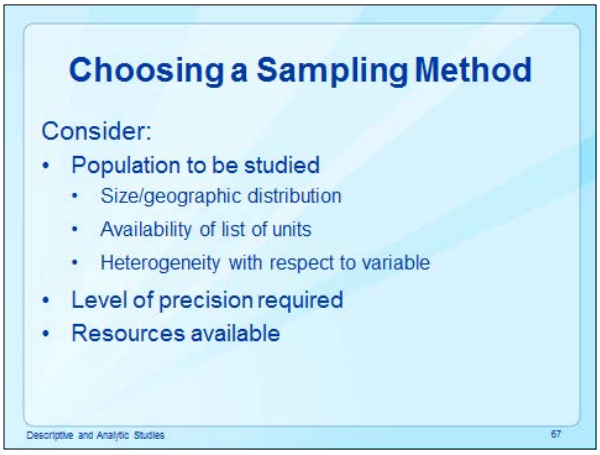


Duration/ Slide Number	What to Do/What to Say

- **Explain that non-probability sampling includes any type of sampling where the probability of selection for each person is unknown and may be zero depending on the method of sampling. While non-probability sampling is relatively easy and inexpensive, non-probability sampling techniques yield non-representative samples, and these techniques cannot be used to infer from the sample to the general population. Furthermore, the results are often biased due to the unequal selection probabilities of different types of people.**
- **Read the examples on the slide.**
- **Explain that "snowball sampling" is commonly used in HIV research of populations that are difficult to locate (e.g., homeless, IV drug users)**

2 minutes

Slide 67

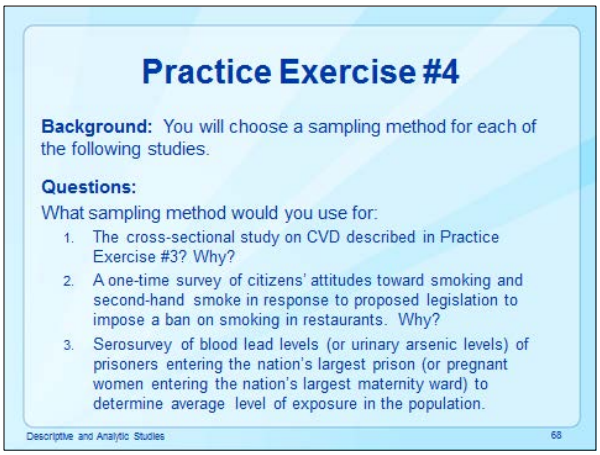


- **Read this slide.**

Duration/ Slide Number	What to Do/What to Say

15 minutes

Slide 68

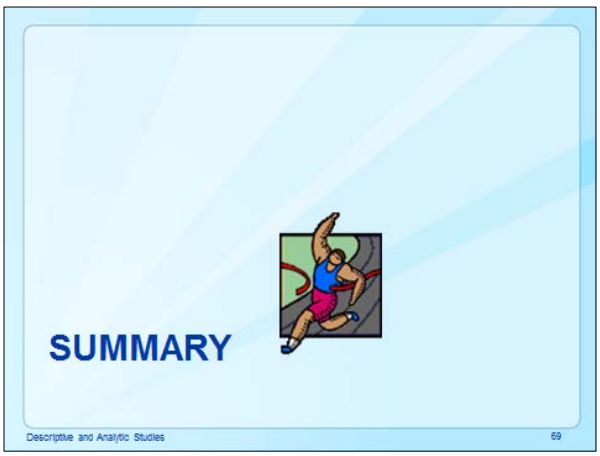


Activity

- **Ask participants to work in small groups or with another colleague to complete Practice Exercise #4.**
- **After 10 minutes, ask some of the pairs or groups to provide you with the answers.**
- ***Possible answers:***
  1. *Stratified random sampling for the cross sectional study on CVD to get separate estimates for the two cities (i.e., strata) in the study.*
  2. *Cluster sampling because it is quick.*
- *Systematic sampling, such as every 10th prisoner. Easiest to administer, does not need any additional information. Could do random sampling, but would need to assign random numbers to each person. Could do stratified sampling, e.g., by residential area, but would need home address of each person.*

1 minute

Slide 69

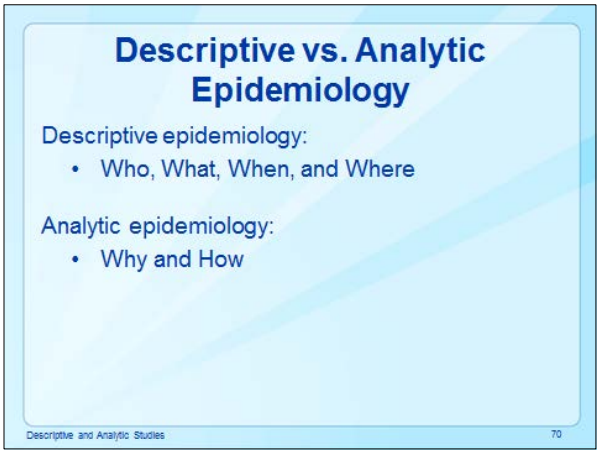


Duration/ Slide Number	What to Do/What to Say

- **Tell participants that you have reached the end of the lesson.**
- **Tell participants that you will now summarize the key points in the lesson.**

2 minutes

Slide 70

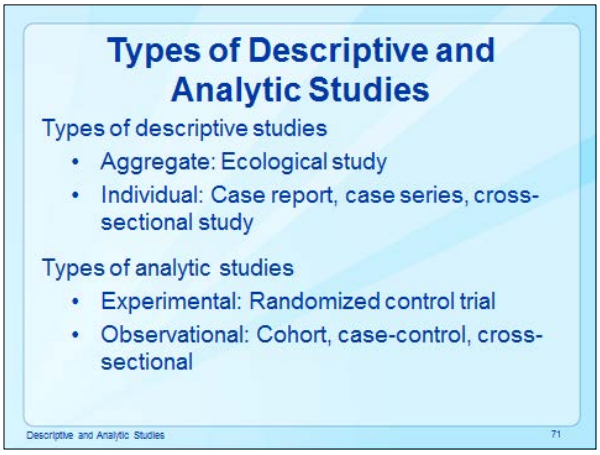


Question

- **Ask**: What are the main differences between descriptive and analytic studies?
- **CLICK slide in PowerPoint and read the bullets**.

2 minutes

Slide 71



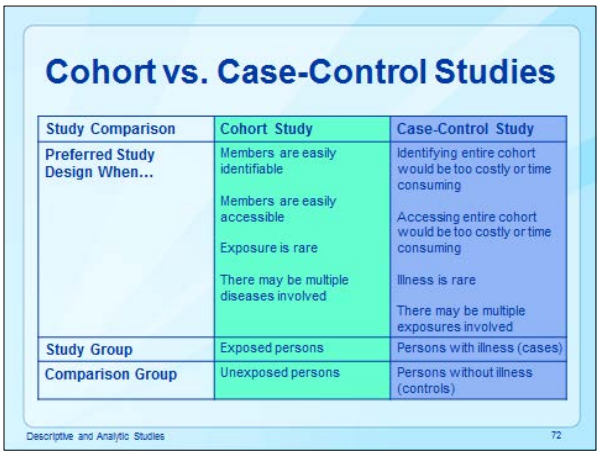
Question

- **Ask**: What are the types of descriptive studies?
- **CLICK slide in PowerPoint and read the bullets**.
- **Ask**: What are the types of analytic studies?
- **CLICK slide in PowerPoint and read the bullets**

Duration/ Slide Number	What to Do/What to Say

3 minutes

Slide 72

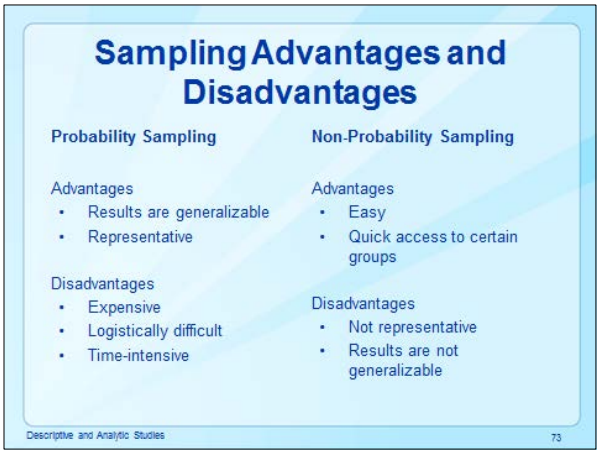


Flip Chart

- **Ask for volunteers to fill in the above table that you should have prepared on a flip chart.**
- **Review the answers and then CLICK on the slide to compare.**

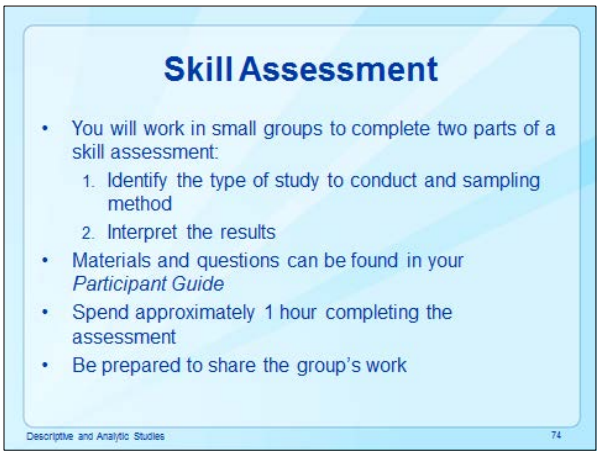
3 minutes

Slide 73



Question

- **Ask participants the main advantages and disadvantages of each sampling method.**

Duration/ Slide Number	What to Do/What to Say
1 hour (plus 20 minutes for review of responses) **Slide 74**	 • **Tell participants that they will now practice the skills they learned using information about actual studies that were conducted.**

Activity

Instructions:

1. You will work in small groups to complete a three-part assessment.
2. Select a member of your group to record your responses.
3. Groups will have approximately 1 hour to complete the assessment.
4. At the end of the assessment, one member from the group will share your summary with the class. (20 minutes)

Part 1. Identify study to conduct and sampling method to use (30 minutes)
Read the following three issues of public health concern and answer the questions that follow.

1. Cardiovascular disease is an increasing problem in Tanzania. The last study to examine risk factors of cardiovascular disease in Tanzania was in 1987. You want to know about the current status of risk factors for cardiovascular disease and would like to examine how the risk factors have changed since the last study in 1987.

   a. What type of study would you conduct and why?

      *Possible answer: For studying cardiovascular disease and the associated risk factors, the most appropriate study would be a **cross-sectional study**. A cross-sectional study is necessary when the objective is to learn about the burden of a health condition in a population. Furthermore, serial cross-sectional studies allow for monitoring changes in risk factors over time, so the results from the 1987 study can be compared to the results of a new cross-sectional study.*

   b. What sampling method would you use?

      *Possible answer: **Simple random sampling** can be used to select a few communities in Tanzania and a small sampling frame (e.g., 200).*

      Njelekela et al. conducted a cross-sectional study in Tanzania in 2000 and the results were published in the following paper:

      Njelekela M, Negishi H, Nara Y, Tomohiro M, Kuga S, Noguchi T, Kanda T, Yamori M, Mashalla Y, Jian Liu L, Mtabaji J, Ikeda K, Yamori Y. Cardiovascular risk factors in Tanzania: a revisit. Acta Trop. 2001 Jun 22;79(3):231-9.

2. In Thailand, breast cancer incidence is increasing, but little is known about the primary risk factors for breast cancer among Thai women. Current understanding of breast

cancer risk factors are from studies in high-income countries. It is unclear if identified risk factors in high-income countries are the same in Thailand. You want to learn about the risk factors for breast cancer in Thailand females.

a. What type of study would you conduct and why?

*Possible answer: A **case-control study** is appropriate in this situation as it is less expensive and requires less time to identify females with the outcome (breast cancer) and compare their exposures to individuals without the outcome (no breast cancer). Furthermore, a case-control study permits investigation of multiple exposures.*

b. How would you use **simple random sampling** to select 860 controls?

Possible answer: Assign number to each woman without breast cancer (from 1 to 47,271) in the population. Use random number generator to randomly select 860 of the women.

c. How would you use **systematic random sampling** to select 860 controls?

Possible answer: Randomize the list of 47,271 women without breast cancer to avoid biases. Assign number to each woman without breast cancer (from 1 to 47,271). Calculate sampling fraction as 47,271 / 860 = 54.966 $\approx$ 55. Select random number between 1 and 54, e.g., 13. Select women no. 13, 13+55=68, 68+55=123, 123+55=178, etc., up to woman no. 13+(55x859)=47,258.

d. How would you use **stratified random sampling** to select 860 controls?

Possible answer: Determine the age distribution of the women with breast cancer by single year (as was done in the actual study) or perhaps by 5-year age group. Stratify the list of women without breast cancer by the same age grouping. For each age group, multiply the percentage from the case frequency distribution times 860 to determine the number of controls needed from each age group stratum, e.g., if 3 of the 43 women were in the 25-29 year age group, you would need (3/43)x860=60 controls from the 25-29 year age group. Randomly select the required number of controls from each age group stratum. The controls would be said to be frequency-matched by age.

Jordan et al. conducted a case-control study on this topic in Thailand and the results were published in the following paper:

Jordan S, Lim L, Vilainerun D, Banks E, Sripaiboonkij N, Seubsman SA, Sleigh A, Bain C. Breast cancer in the Thai Cohort Study: an exploratory case-control analysis. Breast. 2009 Oct;18(5):299-303. Epub 2009 Oct 2.

3. Non-smoking women in China are disproportionally exposed to environmental tobacco smoke exposure in their homes and at work. You want to know if non-smoking Chinese women exposed to environmental tobacco smoke have increased mortality compared to non-smoking Chinese women who live and work in smoke-free

environments. You'd also like to know if environmental tobacco smoke exposure has an impact on the risk of cardiovascular and cancer deaths among Chinese women.

a. What type of study would you conduct and why?

*Possible answer: A **prospective cohort study** is the best way to assess multiple outcomes (different kinds of death) from a particular exposure (environmental tobacco smoke exposure). And because we will need to interview the women about their past smoke exposure (as children and adults), this study must be prospective to have the most accurate exposure data.*

Since we know that smoking has negative health outcomes, we cannot assign the exposure; we must use an observational method to study this problem. Also, as the outcome is not rare (all-cause mortality, and mortality due to cancer or CVD), a case-control study is not appropriate here either.

Wen et al. conducted a prospective cohort study to address these concerns, and the results were published in the following paper:

Wen W, Shu XO, Gao YT, Yang G, Li Q, Li H, Zheng W. Environmental tobacco smoke and mortality in Chinese women who have never smoked: prospective cohort study. BMJ. 2006 Aug 19;333(7564):376. Epub 2006 Jul 12.

Part 2. Interpreting the results (30 minutes)

Interpret the results for each study in 1-2 concise sentences. If applicable, explain whether the exposure was associated with an increased or decreased chance of experiencing the outcome.

1. You found that the prevalence of hypertension was 41.1% in men and 38.7% in women.

   Possible answer: At the time the survey was conducted, 41.1% of all men and 38.7% of all women in the study population had existing hypertension. These are not NEW cases of hypertension, since cross-sectional studies can only give you the counts and proportions of existing disease at a specific point in time.

2. You found that women with older siblings had a statistically significant odds ratio of 0.3 for breast cancer.

   Possible answer: Among female students at STOU in Thailand, those with older siblings had 0.3 times the odds of developing breast cancer than did those who did not have older siblings. This means that the exposure (having older siblings) was associated with a decreased chance of developing breast cancer.

3. You found that non-smoking women whose husbands smoked had a statistically significant risk ratio of 1.19 for mortality due to cancer.

Possible answer: Among non-smoking women in the Shanghai region of China, those whose husbands smoked had 1.19 times the risk of mortality due to cancer than those whose husbands did not smoke. This means that living with a husband who smokes is associated with an increased chance of death due to cancer.

Cardiovascular risk factors in Tanzania: a revisit.

Njelekela M, Negishi H, Nara Y, Tomohiro M, Kuga S, Noguchi T, Kanda T, Yamori M, Mashalla Y, Jian Liu L, Mtabaji J, Ikeda K, Yamori Y.

Acta Trop. 2001 Jun 22;79(3):231-9.

Abstract

In this assessment of cardiovascular risk factors, we examined the prevalence of selected risk factors according to the World Health Organisation (WHO) CARDIAC Study protocol and compared them with a similar study conducted more than a decade ago. The survey was carried out in Dar es Salaam (D, urban), Handeni (H, rural) and Monduli (Mo, semi-nomadic area). Subjects aged 47-57 were recruited randomly for blood pressure and anthropometrical measurements, 24 h urine collection and blood sampling. A structured questionnaire was used to obtain dietary information. The 1998 survey studied 446 subjects, while the 1987 survey included 496 men and women. The measured weight, body mass index (BMI) and prevalence of obesity (BMI > or = 30 kg/m(2)) increased significantly among women in the 1998 survey in rural Handeni and urban Dar. The overall prevalence of obesity was higher for women in the most recent survey (22.8%, P < 0.0001). Diastolic blood pressure (DBP) was higher in the most recent survey for women in Handeni. The overall prevalence of hypertension (blood pressure > 160/95 mmHg, or antihypertensive drug use), rose to 41.1% in 1998, (P < 0.001) for men and to 38.7% (P < 0.05) for women. The mean total serum cholesterol and prevalence of hypercholesterolaemia increased significantly in the most recent survey in the three studied areas. The overall prevalence of hypercholestrolaemia (serum cholesterol > 5.2 mmol/l) was higher in the 1998 survey for both men (21.8%, P < 0.0001) and women (54.0%, P < 0.0001). The mean HDL cholesterol increased significantly in the most recent survey, with a significant reduction in the mean atherogenic index, though these were still at higher levels (men 5.8, P < 0.0001; women 5.1, P < 0.0001 vs. 1987). A strong positive correlation was observed between blood pressure (SBP and DBP) and body mass index, total serum cholesterol and sodium to potassium ratio. These data suggest that for the past decade there has been an increase in the mean levels and prevalence of selected cardiovascular risk factors in Tanzania.

Breast cancer in the Thai Cohort Study: an exploratory case-control analysis.

Jordan S, Lim L, Vilainerun D, Banks E, Sripaiboonkij N, Seubsman SA, Sleigh A, Bain C.
Source

Breast. 2009 Oct;18(5):299-303. Epub 2009 Oct 2.

Abstract

Breast cancer incidence may be increasing in Thailand but very little research has assessed core breast cancer risk factors in this country. We used baseline questionnaire data from a national cohort study of Thai Open University students in an exploratory case-control study of breast cancer. The study included 43 female cases and 860 age-matched controls selected from the remaining 47,271 female cohort participants. Odds ratios and 95% confidence intervals were calculated using conditional logistic regression. The women were predominantly premenopausal. Taller women had an increased risk of breast cancer (OR=2.3, 95% CI 1.1-4.8, for height >or=160cm vs <or=154cm) as did women with non-insulin dependent diabetes mellitus (OR=8.4, 95% CI 1.7-41). Women with older siblings had a reduced risk of breast cancer compared to those firstborn (OR=0.3, 95% CI 0.2-0.7). Although limited by small case numbers, our findings suggest substantial increases in breast cancer rates in Thailand could be expected in the future.

Environmental tobacco smoke and mortality in Chinese women who have never smoked: prospective cohort study.

Wen W, Shu XO, Gao YT, Yang G, Li Q, Li H, Zheng W.

BMJ. 2006 Aug 19;333(7564):376. Epub 2006 Jul 12.

Abstract

OBJECTIVE: To evaluate the association of environmental exposure to **tobacco** smoke from husbands and from work, as well as from family members in early life, with all cause mortality and mortality due to cancer or cardiovascular disease in Chinese women.

DESIGN: Ongoing prospective **cohort** study in Shanghai, **China**.

PARTICIPANTS: Of 72,829 women who had never smoked, 65,180 women provided information on smoking by their husbands, and 66,520 women provided information on exposure to **tobacco** smoke at work and in early life from family members.

MAIN OUTCOME MEASURES: All cause mortality and cause specific mortality with the main focus on cancer and cardiovascular disease. Cumulative mortality according to exposure status, and hazard ratios.

RESULTS: Exposure to **tobacco** smoke from husbands (mainly current exposure) was significantly associated with increased all cause mortality (hazard ratio 1.15, 95% confidence interval 1.01 to 1.31) and with increased mortality due to cardiovascular disease (1.37, 1.06 to 1.78). Exposure to **tobacco** smoke at work was associated with increased mortality due to cancer (1.19, 0.94 to 1.50), especially lung cancer (1.79, 1.09 to 2.93). Exposure in early life was associated with increased mortality due to cardiovascular disease (1.26, 0.94 to 1.69).

CONCLUSIONS: In Chinese women, exposure to environmental **tobacco** smoke is related to moderately increased risk of all cause mortality and mortality due to lung cancer and cardiovascular disease.