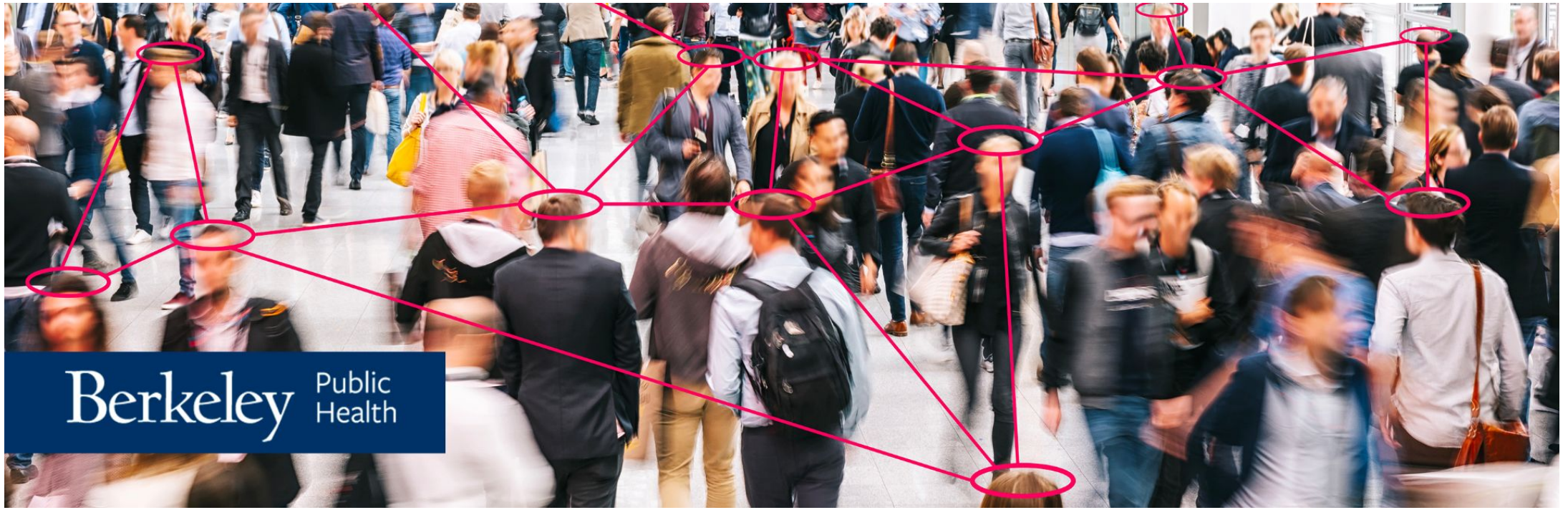


Epidemiologic Methods II

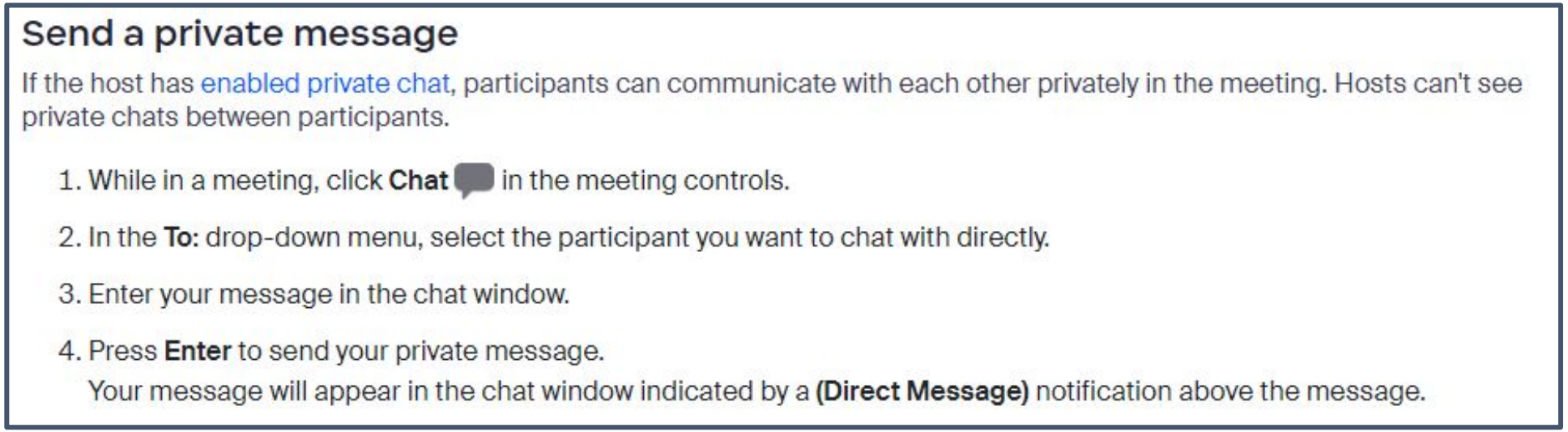
PHW250B

Week 7: Case-Control Studies

October 8, 2025

Recording sessions

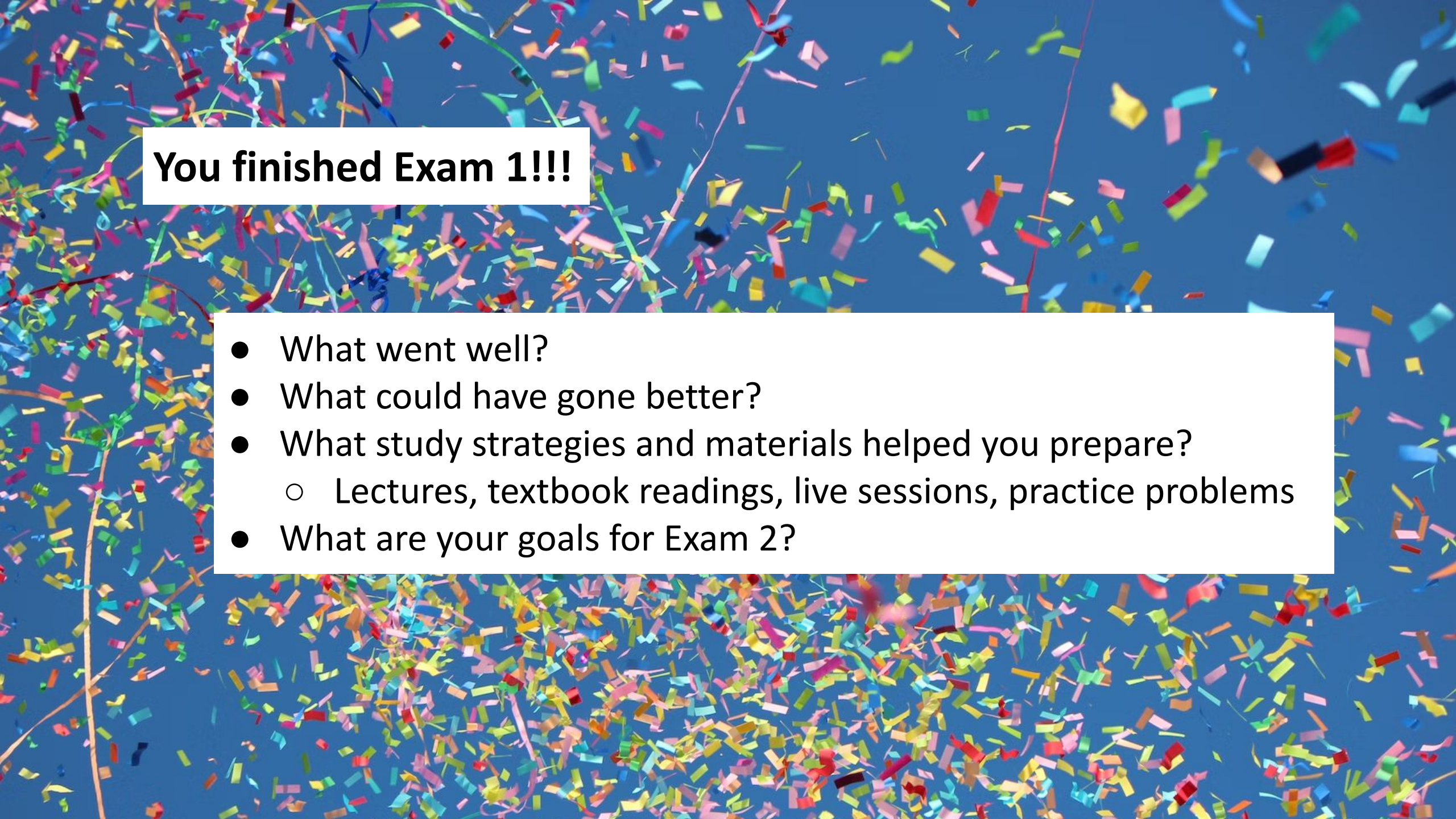
- All live sessions will be recorded and available on bCourses.
- If you want remain anonymous during the session, please send a private message to me on Zoom.



Send a private message

If the host has enabled private chat, participants can communicate with each other privately in the meeting. Hosts can't see private chats between participants.

1. While in a meeting, click **Chat** in the meeting controls.
2. In the **To:** drop-down menu, select the participant you want to chat with directly.
3. Enter your message in the chat window.
4. Press **Enter** to send your private message.
   Your message will appear in the chat window indicated by a **(Direct Message)** notification above the message.



You finished Exam 1!!!

- What went well?
- What could have gone better?
- What study strategies and materials helped you prepare?
  - Lectures, textbook readings, live sessions, practice problems
- What are your goals for Exam 2?

Agenda

- **Ed Discussion questions**
- **Practice problems**
  - 3 on principles of case control studies
  - 2-3 on types of case-control studies
- **Q & A**

Ed Discussion

Question 3.3:

When designing your study, you excluded pregnancies in 2020 because of the change in rate of preterm birth during the COVID-19 lockdown. What assumption for the calculation of *incidence density* would the inclusion of the 2020 pregnancies violate?

Common confusion:
“Constant risk over the interval period” vs. “No secular trends”

In This Example:

- COVID-19 lockdown caused a **population-level change** in preterm birth rates.
- Including 2020 pregnancies would violate **no secular trends**, because the background rate was no longer stable.

Ed Discussion

CONCEPT	APPLIES TO	ASSUMPTION	MEANING
CONSTANT RISK	Cumulative incidence	Risk (probability) is stable across a fixed follow-up period	Closed cohort over a defined time
NO SECULAR TRENDS	Incidence density (rate)	Population rate of disease is stable over *calendar time*	Open cohort; rate not changing due to external events

The logic about changing rates was correct — it's just that the assumption has a different name for incidence density.

Agenda

- **Ed Discussion questions**
- **Practice problems**
  - 3 on principles of case control studies
  - 2-3 on types of case-control studies
- **Q & A**

Week 7, Tab 1 Problems. #1

Define or explain the following:

- Study base principle

- Deconfounding principle

- Comparable accuracy principle

Week 7, Tab 1 Problems. #1

Define or explain the following:

- Study base principle

Cases and controls should be “representative of the same base experience” in order to minimize selection bias

Selection bias: Systematic error in the recruitment or retention of our study participants.

Example adapted from Szklo: You are study a rare disease where cases are identified in a major referral hospital and controls are patients with other conditions identified in the same hospital. Patients of the rare disease come to the hospital via referral and may be more affluent because they have the means to travel to this hospital. As result, exposures related to higher socioeconomic status may differ between cases and controls.

Week 7, Tab 1 Problems. #1

Define or explain the following:

- Deconfounding principle

The measures of association should not be distorted by confounding, unmeasured confounders should be minimized in the design (e.g. stratification, matching)

Example: You are studying a developmental disability that is more common among infants with birthing parents over the age of 40 years. You are interested in hypertension as an exposure, which is also associated with age. In order to reduce differences between cases and controls, you decide to match on age.

Week 7, Tab 1 Problems. #1

Define or explain the following:

- Comparable accuracy principle

The degree of accuracy in measuring the exposure of interest for the cases should be equivalent to the degree of accuracy for the controls

Example of lacking comparable accuracy: The exposure is the BRCA gene for breast cancer. Researchers assess whether the cases have the BRCA gene via genetic testing, while controls are surveyed about family history of breast cancer.

Quick check:

What type of bias occurs if our cases and controls come from *different* source populations?

What is one design strategy we can use to reduce confounding in a case-control study?

If exposure is measured more accurately among cases than controls, what type of bias might result?

Quick check:

What type of bias occurs if our cases and controls come from *different* source populations?

Selection bias

What is one design strategy we can use to reduce confounding in a case-control study?

If exposure is measured more accurately among cases than controls, what type of bias might result?

Quick check:

What type of bias occurs if our cases and controls come from *different* source populations?

Selection bias

What is one design strategy we can use to reduce confounding in a case-control study?

Matching or stratification

If exposure is measured more accurately among cases than controls, what type of bias might result?

Quick check:

What type of bias occurs if our cases and controls come from *different* source populations?

Selection bias

What is one design strategy we can use to reduce confounding in a case-control study?

Matching or stratification

If exposure is measured more accurately among cases than controls, what type of bias might result?

Information bias - specifically differential misclassification

Week 7, Tab 1 Problems. #2

Read the abstract and methods of the following article to answer problem 2.
Coleman-Phox K, Odouli R, Li DK. Use of a fan during sleep and the risk of sudden infant death syndrome. Arch Pediatr Adolesc Med. 2008;162(10):963-8.

Objective To examine the relation between room ventilation during sleep and risk of sudden infant death syndrome (SIDS).
Design Population-based case-control study.
Setting Eleven California counties.
Participants Mothers of 185 infants with a confirmed SIDS diagnosis and 312 randomly selected infants matched on county of residence, maternal race/ethnicity, and age.
Intervention Fan use and open window during sleep.
Main Outcome Measure Risk of SIDS.

a. What was the exposure and outcome of interest in this study?

Exposure = fan use during sleep
Outcome = sudden infant death syndrome (SIDS)

Week 7, Tab 1 Problems. #2

Read the abstract and methods of the following article to answer problem 2.
Coleman-Phox K, Odouli R, Li DK. Use of a fan during sleep and the risk of sudden infant death syndrome. Arch Pediatr Adolesc Med. 2008;162(10):963-8.

b. What type of study did the authors use? Why did the authors choose this type of study design to answer this question?

- This is a case-control study.
- The authors probably chose this study design because SIDS is a relatively uncommon outcome (0.53 deaths/1,000 live births). It would have been time-consuming and expensive to enroll and follow a prospective cohort of infants that was large enough to eventually include enough cases of SIDS for the analysis.
  - Thus, a case control study is advantageous because the investigators can start by identifying cases of SIDS rather than waiting for them to occur.

Week 7, Tab 1 Problems. #2

Read the abstract and methods of the following article to answer problem 2.
Coleman-Phox K, Odouli R, Li DK. Use of a fan during sleep and the risk of sudden infant death syndrome. Arch Pediatr Adolesc Med. 2008;162(10):963-8.

Methods: The participating counties were selected on the basis of their proximity to the two investigation centers, located in the San Francisco Bay Area in Northern California and Los Angeles County in Southern California…SIDS cases were identified from all infant deaths reported to the California Department of Health Services and to the Los Angeles County coroner's office with a diagnosis of SIDS or presumed SIDS during the study period.

c. Was this a primary or secondary study base? What are the advantages and disadvantages of this type of study base?

This is a **primary study base.**

- The authors identified the population first (11 California counties) between May 1, 1997 and April 30, 2000; and then identified cases based on that population. This means that controls are selected from this population also.

Week 7, Tab 1 Problems. #2

Read the abstract and methods of the following article to answer problem 2.
Coleman-Phox K, Odouli R, Li DK. Use of a fan during sleep and the risk of sudden infant death syndrome. Arch Pediatr Adolesc Med. 2008;162(10):963-8.

d. What are the advantages and disadvantages of this type of study base (primary)?

- **Advantages:**
  - We can be more confident that the controls come from the population that gave rise to the cases (i.e. because they are all the other births in that county that would have become cases had they died of SIDS).
  - We can be more confident that exposure in controls estimates the distribution of exposure in the general population.

- **Disadvantages**:
  - We must consider whether all cases have been identified,
  - Controls: controls from this population may be less motivated to participate, it may be expensive to track down and interview controls from such a wide population base, and differential misclassification of exposure via recall bias may be more of a concern. (NOTE: This is an issue that will come up with secondary study bases as well)

Week 7, Tab 1 Problems. #4

Streptococcus pneumoniae is a potentially serious infection that disproportionately affects young children. This infection is more severe in young children than in otherwise healthy adults and can prove deadly. *Streptococcus pneumoniae* is a common complication of influenza infection in young children and could potentially be prevented by influenza vaccination. There is, therefore, a need for research on the possible benefit of influenza vaccination with respect to prevention of *Streptococcus pneumoniae* in the general population of children aged 2-5.

You are an epidemiologist for Kaiser Permanente, CA, with access to administrative and medical records for a cohort of 10,000 children age 2-5 years living in the state of California beginning in the year 2007 through the present. Though the database is comprehensive for all services obtained within the Kaiser system, some children receive flu vaccines through school programs or other public health services (necessitating record abstraction and potentially in-person interviews) making it too expensive to assess exposure for the entire cohort. With flu season fast approaching, you want to conduct an efficient study to assess the hypothesis that influenza vaccination in young children might prevent streptococcus pneumonia infection.

Week 7, Tab 1 Problems. #4

Design a density case-control study that tests your hypothesis, and answer questions related to your design in the spaces provided below.

a) Identify the exposure and outcome of your study. *(2 points)*

b) Describe your source population and define the type of study base. *(2 points)*

Week 7, Tab 1 Problems. #4

Design a density case-control study that tests your hypothesis, and answer questions related to your design in the spaces provided below.

a) Identify the exposure and outcome of your study. *(2 points)*

Exposure: Influenza vaccination
Outcome: *Streptococcus pneumoniae* infection

b) Describe your source population and define the type of study base. *(2 points)*

Week 7, Tab 1 Problems. #4

Design a density case-control study that tests your hypothesis, and answer questions related to your design in the spaces provided below.

a) Identify the exposure and outcome of your study. *(2 points)*

Exposure: Influenza vaccination
Outcome: *Streptococcus pneumoniae* infection

b) Describe your source population and define the type of study base. *(2 points)*

The source population is the 2007-2010 Kaiser population of children aged 2-5. This represents a Primary Study Base. Explanation: Kaiser is an unusual entity-- we would expect students to be familiar with this, but in northern California, their level of population coverage is very high across income levels. Thus, some may consider this a primary base because it can be thought of as representative of the population.

This could also be thought of as a secondary study base because the population is comprised only of Kaiser patients, so controls are individuals who would have shown up to Kaiser had they had *Streptococcus pneumoniae* infection.

Week 7, Tab 1 Problems. #4

Design a density case-control study that tests your hypothesis, and answer questions related to your design in the spaces provided below.

c) Name the sampling strategy employed in this type of study design, and in 4-5 sentences, describe how this sampling strategy would be used in this specific study. *(3 points)*

Week 7, Tab 1 Problems. #4

Design a density case-control study that tests your hypothesis, and answer questions related to your design in the spaces provided below.

c) Name the sampling strategy employed in this type of study design, and in 4-5 sentences, describe how this sampling strategy would be used in this specific study. *(3 points)*

Risk-Set sampling.
Cases are all source population/cohort members with a validated diagnosis of pneumonia between 2007 and the present. Controls are randomly selected from the cohort population at risk for pneumonia (anyone who does not have pneumonia) at the time of pneumonia onset/ date of pneumonia diagnosis of the matched case. Controls can be matched to more than one case, and controls can become cases at any point.

Week 7, Tab 1 Problems. #4

Design a density case-control study that tests your hypothesis, and answer questions related to your design in the spaces provided below.

d) State the appropriate measure(s) of association that you would calculate in this type of study.

e) What other relative measure does this measure of association approximate?

f) When reporting your findings from this density case-control study, do you think you would be justified in generalizing your results to the entire population of the state of California? Why/why not?

Week 7, Tab 1 Problems. #4

Design a density case-control study that tests your hypothesis, and answer questions related to your design in the spaces provided below.

d) State the appropriate measure(s) of association that you would calculate in this type of study.

Odds ratio (OR).

e) What other relative measure does this measure of association approximate?

f) When reporting your findings from this density case-control study, do you think you would be justified in generalizing your results to the entire population of the state of California? Why/why not?

Week 7, Tab 1 Problems. #4

Design a density case-control study that tests your hypothesis, and answer questions related to your design in the spaces provided below.

d) State the appropriate measure(s) of association that you would calculate in this type of study.

Odds ratio (OR).

e) What other relative measure does this measure of association approximate?

Incidence Density Ratio.

f) When reporting your findings from this density case-control study, do you think you would be justified in generalizing your results to the entire population of the state of California? Why/why not?

Week 7, Tab 1 Problems. #4

Design a density case-control study that tests your hypothesis, and answer questions related to your design in the spaces provided below.

d) State the appropriate measure(s) of association that you would calculate in this type of study.

Odds ratio (OR).

e) What other relative measure does this measure of association approximate?

Incidence Density Ratio.

f) When reporting your findings from this density case-control study, do you think you would be justified in generalizing your results to the entire population of the state of California? Why/why not?

No. People who have HMO health insurance are fundamentally different from people in CA who do not (SES, employment, etc…)

Agenda

- **Ed Discussion questions**
- **Practice problems**
  - 3 on principles of case control studies
  - 2-3 on types of case-control studies
- **Q & A**

Overview on Design Types

Design Type	Timing of Control Selection	Can a Case Also Serve as a Control?	Matching Variable	Rare Disease Assumption Needed for OR ≈ IDR/RR?
Cumulative Case-Control	At the end of follow-up	No	None (often unmatched)	Yes
Density (Risk-Set) Case-Control	At the time each case occurs	Yes	Time at risk	No
Case-Cohort	At baseline (start of follow-up)	Yes	None (subcohort sampled)	No

Week 7, Additional Problems. #1

A study to examine the relationship of inflammatory markers (such as interleukin-6 and C-reactive protein) to incident dementia was conducted within the Rotterdam Study cohort (n=6713). A random sample of the total cohort at baseline (n=727) and the 188 individuals who developed dementia on follow-up were compared. Serum inflammatory markers were measured in cases and in the random sample.

a. Which type of study have the authors conducted? Be as specific as possible.

b. What measure of association can we calculate when using the type of study you identified in part a?

Week 7, Additional Problems. #1

A study to examine the relationship of inflammatory markers (such as interleukin-6 and C-reactive protein) to incident dementia was conducted within the Rotterdam Study cohort (n=6713). A random sample of the total cohort at baseline (n=727) and the 188 individuals who developed dementia on follow-up were compared. Serum inflammatory markers were measured in cases and in the random sample.

a. Which type of study have the authors conducted? Be as specific as possible.

Case-cohort study. This is a case-control study within a defined cohort, in which the control group was a random sample of the total cohort at baseline.

b. What measure of association can we calculate when using the type of study you identified in part a?

Week 7, Additional Problems. #1

A study to examine the relationship of inflammatory markers (such as interleukin-6 and C-reactive protein) to incident dementia was conducted within the Rotterdam Study cohort (n=6713). A random sample of the total cohort at baseline (n=727) and the 188 individuals who developed dementia on follow-up were compared. Serum inflammatory markers were measured in cases and in the random sample.

a. Which type of study have the authors conducted? Be as specific as possible.

Case-cohort study. This is a case-control study within a defined cohort, in which the control group was a random sample of the total cohort at baseline.

b. What measure of association can we calculate when using the type of study you identified in part a?

You calculate an OR. An OR is the only measure of association you can calculate with this study design (that is commonly used). The OR in this situation approximates a risk ratio if control sampling is independent of exposure status so that the exposure experience in the controls represents the exposure experience of the at-risk population at baseline.

Week 7, Additional Problems. #1

A study to examine the relationship of inflammatory markers (such as interleukin-6 and C-reactive protein) to incident dementia was conducted within the Rotterdam Study cohort (n=6713). A random sample of the total cohort at baseline (n=727) and the 188 individuals who developed dementia on follow-up were compared. Serum inflammatory markers were measured in cases and in the random sample.

c. If the authors wished to study the relationship of inflammatory markers to stroke, could they use the same control group? Why or why not?

Week 7, Additional Problems. #1

A study to examine the relationship of inflammatory markers (such as interleukin-6 and C-reactive protein) to incident dementia was conducted within the Rotterdam Study cohort (n=6713). A random sample of the total cohort at baseline (n=727) and the 188 individuals who developed dementia on follow-up were compared. Serum inflammatory markers were measured in cases and in the random sample.

c. If the authors wished to study the relationship of inflammatory markers to stroke, could they use the same control group? Why or why not?

Yes, the same control group could be used for different outcomes as long as they were all at risk for stroke as well as dementia. Taking a random sample from the cohort approximates the exposure experience of the at-risk population that gave rise to the cases. They would, however, need to use a different case group.

Week 7, Additional Problems. #2

For each of the following statements about control selection in case-control studies, circle either TRUE or FALSE:

Controls in a case-cohort study should be selected to reflect the exposure distribution in the underlying cohort at baseline. TRUE FALSE

Cases in a cumulative case-control study can also serve as controls. TRUE FALSE

In a density case-control study, controls are matched to cases on exposure status. TRUE FALSE

In a cumulative case-control study, the rare disease assumption is not necessary for the OR to estimate the IDR. TRUE FALSE

Week 7, Additional Problems. #2

For each of the following statements about control selection in case-control studies, circle either TRUE or FALSE:

Controls in a case-cohort study should be selected to reflect the exposure distribution in the underlying cohort at baseline. TRUE FALSE

Cases in a cumulative case-control study can also serve as controls. TRUE FALSE

In a density case-control study, controls are matched to cases on exposure status. TRUE FALSE

In a cumulative case-control study, the rare disease assumption is not necessary for the OR to estimate the IDR. TRUE FALSE

Week 7, Additional Problems. #2

For each of the following statements about control selection in case-control studies, circle either TRUE or FALSE:

Controls in a case-cohort study should be selected to reflect the exposure distribution in the underlying cohort at baseline. TRUE FALSE

Cases in a cumulative case-control study can also serve as controls. TRUE FALSE
This is true for case-cohort and density-sampled case controls, not for cumulative case-controls

In a density case-control study, controls are matched to cases on exposure status. TRUE FALSE

In a cumulative case-control study, the rare disease assumption is not necessary for the OR to estimate the IDR. TRUE FALSE

Week 7, Additional Problems. #2

For each of the following statements about control selection in case-control studies, circle either TRUE or FALSE:

Controls in a case-cohort study should be selected to reflect the exposure distribution in the underlying cohort at baseline. TRUE FALSE

Cases in a cumulative case-control study can also serve as controls. TRUE FALSE
This is true for case-cohort and density-sampled case controls, not for cumulative case-controls

In a density case-control study, controls are matched to cases on exposure status. TRUE FALSE
They are matched on time at risk. If we match on exposure status, we have nothing left to analyze and we have designed a useless study.

In a cumulative case-control study, the rare disease assumption is not necessary for the OR to estimate the IDR. TRUE FALSE

Week 7, Additional Problems. #2

For each of the following statements about control selection in case-control studies, circle either TRUE or FALSE:

Controls in a case-cohort study should be selected to reflect the exposure distribution in the underlying cohort at baseline. TRUE FALSE

Cases in a cumulative case-control study can also serve as controls. TRUE FALSE
This is true for case-cohort and density-sampled case controls, not for cumulative case-controls

In a density case-control study, controls are matched to cases on exposure status. TRUE FALSE
They are matched on time at risk. If we match on exposure status, we have nothing left to analyze and we have designed a useless study.

In a cumulative case-control study, the rare disease assumption is not necessary for the OR to estimate the IDR. TRUE FALSE
In cumulative case-control, we DO need the rare disease assumption to assume the OR estimates the IDR. In a density-sampled case-control, we do not need the rare disease assumption

Week 7, Tab 1 Problems. #3

Numerous studies have explored whether long-term cellular phone use increases the risk of developing brain tumors. Many of these studies were conducted in Sweden and Denmark, countries that have population censuses as well as cancer registries that are thought to capture all cancer cases in the population. In addition, some studies have received contact information of cell phone users from cell phone companies, which allowed them to follow cell phone subscribers over time. Both case control studies and cohort studies have been used to explore this research question.

a. You are an epidemiologist who is asked to design a case-control study to answer this question in Sweden. Would you choose a primary or secondary study base? Why? (2 points)

Week 7, Tab 1 Problems. #3

Numerous studies have explored whether long-term cellular phone use increases the risk of developing brain tumors. Many of these studies were conducted in Sweden and Denmark, countries that have population censuses as well as cancer registries that are thought to capture all cancer cases in the population. In addition, some studies have received contact information of cell phone users from cell phone companies, which allowed them to follow cell phone subscribers over time. Both case control studies and cohort studies have been used to explore this research question.

a. You are an epidemiologist who is asked to design a case-control study to answer this question in Sweden. Would you choose a primary or secondary study base? Why? (2 points)

It would be preferable to use a primary study base, and because high quality cancer and population registries are available, it would be possible to define the primary study base as the whole population in Sweden.

Week 7, Tab 1 Problems. #3

Numerous studies have explored whether long-term cellular phone use increases the risk of developing brain tumors. Many of these studies were conducted in Sweden and Denmark, countries that have population censuses as well as cancer registries that are thought to capture all cancer cases in the population. In addition, some studies have received contact information of cell phone users from cell phone companies, which allowed them to follow cell phone subscribers over time. Both case control studies and cohort studies have been used to explore this research question.

b. How would you select cases, given the information provided above? How would you select controls? Discuss the extent to which your control selection method controls for time as a confounder in the design phase. (3 points)

Week 7, Tab 1 Problems. #3

Numerous studies have explored whether long-term cellular phone use increases the risk of developing brain tumors. Many of these studies were conducted in Sweden and Denmark, countries that have population censuses as well as cancer registries that are thought to capture all cancer cases in the population. In addition, some studies have received contact information of cell phone users from cell phone companies, which allowed them to follow cell phone subscribers over time. Both case control studies and cohort studies have been used to explore this research question.

b. How would you select cases, given the information provided above? How would you select controls? Discuss the extent to which your control selection method controls for time as a confounder in the design phase. (3 points)

First you would need to define the period of time of interest for your study and define criteria for someone to be a case. Then you could select all of the cases meeting the criteria from the cancer registry during the time period of interest. To select controls, you could draw a sample from the population census. This could be done in a number of different ways:

Density case-control: draw control at each time point when there was a new a case; doing so controls for time in the design phase.

Case-cohort: choose a random sample of the population at risk from the population registry at baseline. You would have to cross-check with the cancer registry to make sure that people in this population database do not have brain tumors at that time. This type of control sampling does not control for time in the design phase and would require adjustment for time in the analysis phase.

Week 7, Tab 1 Problems. #3

Numerous studies have explored whether long-term cellular phone use increases the risk of developing brain tumors. Many of these studies were conducted in Sweden and Denmark, countries that have population censuses as well as cancer registries that are thought to capture all cancer cases in the population. In addition, some studies have received contact information of cell phone users from cell phone companies, which allowed them to follow cell phone subscribers over time. Both case control studies and cohort studies have been used to explore this research question.

c. If the cell phone records do not directly allow you to assess exposure status, what is another way of gathering exposure information for cases and control? Discuss any potential disadvantages of the exposure assessment method you suggested.

Week 7, Tab 1 Problems. #3

Numerous studies have explored whether long-term cellular phone use increases the risk of developing brain tumors. Many of these studies were conducted in Sweden and Denmark, countries that have population censuses as well as cancer registries that are thought to capture all cancer cases in the population. In addition, some studies have received contact information of cell phone users from cell phone companies, which allowed them to follow cell phone subscribers over time. Both case control studies and cohort studies have been used to explore this research question.

c. If the cell phone records do not directly allow you to assess exposure status, what is another way of gathering exposure information for cases and control? Discuss any potential disadvantages of the exposure assessment method you suggested.

These studies typically use self-administered questionnaires for exposure assessment of both cases and controls. Recall bias is a major issue. People have trouble remembering how much they used their phone on average in the past, and this may differ by case status if exposure assessment is retrospective.

Extras

Week 7, Additional Problems. #3

Read the abstract for the paper by Cohn et al. entitled “DDT and Breast Cancer in Young Women: New Data on the Significance of Age at Exposure” and answer the following questions.

Background. Previous studies of DDT and breast cancer assessed exposure later in life when the breast may not have been vulnerable, after most DDT had been eliminated, and after DDT had been banned.
Objectives. We investigated whether DDT exposure in young women during the period of peak DDT use predicts breast cancer.
Methods. We conducted a prospective, nested case–control study with a median time to diagnosis of 17 years using blood samples obtained from young women during 1959–1967. Subjects were members of the Child Health and Development Studies, Oakland, California, who provided blood samples 1–3 days after giving birth (mean age, 26 years). Cases (*n* = 129) developed breast cancer before the age of 50 years. Controls (*n* = 129) were matched to cases on birth year. Serum was assayed for *p,p′*-DDT, the active ingredient of DDT; *o,p′*-DDT, a low concentration contaminant; and *p,p′*-DDE, the most abundant *p,p′*-DDT metabolite.
Results. High levels of serum *p,p′*-DDT predicted a statistically significant 5-fold increased risk of breast cancer among women who were born after 1931. These women were under 14 years of age in 1945, when DDT came into widespread use, and mostly under 20 years as DDT use peaked. Women who were not exposed to *p,p′*-DDT before 14 years of age showed no association between *p,p′*-DDT and breast cancer (*p*= 0.02 for difference by age).
Conclusions. Exposure to *p,p′*-DDT early in life may increase breast cancer risk. Many U.S. women heavily exposed to DDT in childhood have not yet reached 50 years of age. The public health significance of DDT exposure in early life may be large.

Week 7, Additional Problems. #3

Read the abstract for the paper by Cohn et al. entitled “DDT and Breast Cancer in Young Women: New Data on the Significance of Age at Exposure” and answer the following questions.

A. What type of study design did the authors use?

B. Is this a concurrent or non-concurrent study? Was exposure information collected prospectively or retrospectively? (refer to the Methods Section: *Serum assays* if necessary)

Week 7, Additional Problems. #3

Read the abstract for the paper by Cohn et al. entitled “DDT and Breast Cancer in Young Women: New Data on the Significance of Age at Exposure” and answer the following questions.

A. What type of study design did the authors use?

Nested case-control using density sampling

B. Is this a concurrent or non-concurrent study? Was exposure information collected prospectively or retrospectively? (refer to the Methods Section: *Serum assays* if necessary)

Week 7, Additional Problems. #3

Read the abstract for the paper by Cohn et al. entitled "DDT and Breast Cancer in Young Women: New Data on the Significance of Age at Exposure" and answer the following questions.

A. What type of study design did the authors use?

Nested case-control using density sampling

B. Is this a concurrent or non-concurrent study? Was exposure information collected prospectively or retrospectively? (refer to the Methods Section: *Serum assays* if necessary)

This is a non-concurrent study. The investigators retrieved the serum samples to assess past exposure after the outcome was diagnosed, that is, after the exposure and outcome had happened (the methods section makes it clear that they collected cases of breast cancer at one time – 1998 – instead of collecting cases as they occurred).

Exposure information was collected prospectively via the collection of blood samples from the cohort from 1959-1967, before outcomes were recorded. The assays for DDT were run in 2000- 2001, but because the serum had been collected prior to the outcomes, we would still consider this prospective exposure assessment.

Week 7, Additional Problems. #3

Read the abstract for the paper by Cohn et al. entitled “DDT and Breast Cancer in Young Women: New Data on the Significance of Age at Exposure” and answer the following questions.

C. How did the authors choose the controls?

D. For what reason might the authors have chosen to use this method of control selection?

Week 7, Additional Problems. #3

Read the abstract for the paper by Cohn et al. entitled “DDT and Breast Cancer in Young Women: New Data on the Significance of Age at Exposure” and answer the following questions.

C. How did the authors choose the controls?

According to the abstract, controls were selected from the underlying cohort (members of the Child Health & Development Studies) and matched to cases on birth year. We can assume from the study design that they used risk-set sampling. If you ventured into the methods section you also learned controls were selected at random and were required to be disease-free at time of the case’s diagnosis.

D. For what reason might the authors have chosen to use this method of control selection?

Week 7, Additional Problems. #3

Read the abstract for the paper by Cohn et al. entitled “DDT and Breast Cancer in Young Women: New Data on the Significance of Age at Exposure” and answer the following questions.

C. How did the authors choose the controls?

According to the abstract, controls were selected from the underlying cohort (members of the Child Health & Development Studies) and matched to cases on birth year. We can assume from the study design that they used risk-set sampling. If you ventured into the methods section you also learned controls were selected at random and were required to be disease-free at time of the case’s diagnosis.

D. For what reason might the authors have chosen to use this method of control selection?

Risk-set sampling helps ensure that the exposure distribution among the controls is the same as it is in the person-time in the study base for the cases, and thus allows the authors to calculate a (reasonably) unbiased estimate of the IDR.

Week 7, Additional Problems. #1

Read the abstract for the paper by Cohn et al. entitled “DDT and Breast Cancer in Young Women: New Data on the Significance of Age at Exposure” and answer the following questions.

E. What does the calculated OR estimate?

F. What must be true for the OR to estimate this measure?

Week 7, Additional Problems. #1

Read the abstract for the paper by Cohn et al. entitled “DDT and Breast Cancer in Young Women: New Data on the Significance of Age at Exposure” and answer the following questions.

E. What does the calculated OR estimate?

The incidence density ratio (IDR).

F. What must be true for the OR to estimate this measure?

Week 7, Additional Problems. #1

Read the abstract for the paper by Cohn et al. entitled “DDT and Breast Cancer in Young Women: New Data on the Significance of Age at Exposure” and answer the following questions.

E. What does the calculated OR estimate?

The incidence density ratio (IDR).

F. What must be true for the OR to estimate this measure?

The control sampling fractions for exposed and unexposed controls must be equivalent to the corresponding exposure distributions within the PT of the study base for the cases. In other words, the ratio of unexposed controls to unexposed PT equals the ratio of exposed controls to exposed PT. See lecture videos or required reading (especially ME3 Ch. 8) for a more comprehensive explanation and equations if this is confusing.

Week 7, Additional Problems. #2

It is 1995, and you are working for the CDC. It is known that health-care workers (HCWs) are potentially at risk for human immunodeficiency virus (HIV) infection through occupational exposures to blood. Although prospective studies indicate that the estimated risk for HIV infection/seroconversion after percutaneous exposure (i.e., a puncture wound from needles or other sharps) to HIV-infected blood is approximately 0.3%, factors that influence this risk have not been determined. All occupationally acquired HIV infections are reported to a national surveillance system operated by the CDC. Assume that many health care institutions maintain worker surveillance systems with additional information about all health care workers that they employ.

Week 7, Additional Problems. #2

A. Briefly, design a case-control study to assess factors related to HIV seroconversion among HCWs after occupational percutaneous exposure to HIV-infected blood. Make sure to state the specific type of design, define exposure/non-exposure, define case/control status, and explain any sampling techniques that you would use and why.

Week 7, Additional Problems. #2

A. Briefly, design a case-control study to assess factors related to HIV seroconversion among HCWs after occupational percutaneous exposure to HIV-infected blood. Make sure to state the specific type of design, define exposure/non-exposure, define case/control status, and explain any sampling techniques that you would use and why.

This problem is based on an actual study that was conducted, but there are probably many different case-control studies one could design for this question. In the actual study, Case-HCWs had a documented occupational percutaneous exposure to HIV-infected blood, HIV seroconversion temporally associated with the exposure, and no other concurrent exposure to HIV. Control-HCWs had a documented occupational percutaneous exposure to HIV-infected blood and were HIV seronegative at the time of exposure and at least 6 months later. Cases were selected from the national occupational surveillance system, while controls were selected from a “passive surveillance project maintained by the CDC” that included data from ~300 healthcare institutions.

Week 7, Additional Problems. #2

Sampling techniques: For cases, you might decide to include all available cases, or a random sample. For controls, you could randomly select up to 4 controls from the healthcare institution surveillance system for each case. It is unclear from the actual study whether the investigators matched cases and controls on time of event; however, this would be an essential strategy if you were doing a density sampled case-control study.

To determine if this study might be a primary or secondary study base, the key question to consider is whether you can define the population of interest without reference to the cases. In this case, a plausible population of interest is all health-care workers in the United States who were exposed to percutaneous HIV-infected blood on the job and active in 1995 (or some period such as 1990 – 1995, 1980 – 1995). Since we can define this population in a way that does not mention cases, this is a primary study base. The questions are then how to sample from this base and how to identify cases. In the actual study, the investigators would have to assume that the controls/data used from the healthcare surveillance system are representative of all HCWs (i.e., they represent the source population for the cases). It might be possible to compare their demographics to data collected from HCWs (i.e., through professional associations or ongoing surveys of HCWs like the Nurses' Health Study) to test this assumption. In terms of cases, we are fortunate to have a complete registry of cases in this instance through the national surveillance system.

Sampling from this group could be challenging, but luckily CDC's passive surveillance system includes information on occupational incidents such as needle pricks.

Week 7, Additional Problems. #2

B. See the assigned Wacholder readings for the case-control study module. Which of the selection factors identified by Wacholder (1992 #1) as potentially affecting the study base principle do you think might pose a problem in your study?

Week 7, Additional Problems. #2

B. See the assigned Wacholder readings for the case-control study module. Which of the selection factors identified by Wacholder (1992 #1) as potentially affecting the study base principle do you think might pose a problem in your study?

It's plausible that under-ascertainment of cases would be a problem in this study, as not all HIV infections would be ascertained or reported to the occupational surveillance system. Wacholder doesn't discuss under-ascertainment of controls, but that is also possible since it is possible that not all needlesticks are reported.

Week 7, Additional Problems. #3

Describe how controls are sampled in a case control study that uses risk set or density sampling. What measure of association does the OR approximate?

Week 7, Additional Problems. #3

Describe how controls are sampled in a case control study that uses risk set or density sampling. What measure of association does the OR approximate?

Controls are sampled from those who are at-risk of disease in the cohort at the time each case is identified. Controls can themselves become cases at a later time point. The OR approximates the IDR if the controls are sampled at the same time as the cases.

Week 7, Additional Problems. #5

(Adapted from: Auvinen A, et al. Uranium and other natural radionuclides in drinking water and risk of leukemia: a xxxxx study in Finland. *Cancer Causes Control.* 2002;13(9):825-9.)
Read the abstract below and respond to the following questions about it.

Objective: We assessed the effect of uranium in drinking water on risk of leukemia.
Methods: The subjects (n=144,627) in the base cohort had lived outside the municipal tapwater system and were using drinking water from drilled wells during 1967–1980. A subcohort (n=274) was formed as a random sample of the base cohort. The study compared exposure among all cases in the base cohort with leukemia (n=35) with the subcohort. Concentration of uranium in the drinking water was analyzed.
Results: The odds ratio of leukemia for uranium was 0.91 (95% confidence interval 0.73–1.13).
Conclusions: Our results do not indicate an increased risk of leukemia from ingestion of natural uranium through drinking water at these exposure levels.

Week 7, Additional Problems. #5

(Adapted from: Auvinen A, et al. Uranium and other natural radionuclides in drinking water and risk of leukemia: a xxxxx study in Finland. *Cancer Causes Control.* 2002;13(9):825-9.)
Read the abstract below and respond to the following questions about it.

a. What type of study design is this? Be as specific as possible.

b. What type of study base is this?

Week 7, Additional Problems. #5

(Adapted from: Auvinen A, et al. Uranium and other natural radionuclides in drinking water and risk of leukemia: a xxxxx study in Finland. *Cancer Causes Control.* 2002;13(9):825-9.)
Read the abstract below and respond to the following questions about it.

a. What type of study design is this? Be as specific as possible.

Case-cohort.

b. What type of study base is this?

Week 7, Additional Problems. #5

(Adapted from: Auvinen A, et al. Uranium and other natural radionuclides in drinking water and risk of leukemia: a xxxxx study in Finland. *Cancer Causes Control.* 2002;13(9):825-9.)
Read the abstract below and respond to the following questions about it.

a. What type of study design is this? Be as specific as possible.

Case-cohort.

b. What type of study base is this?

Primary. They identified the entire base cohort – people who were not using the municipal tapwater system and then found all of the cases in this study base. Note that they excluded those using the tapwater system from both cases and controls because they are interested in people using well water – i.e., well water users are their target population. This was not an exclusion to create one cohort of tapwater vs. well water users.

Week 7, Additional Problems. #5

(Adapted from: Auvinen A, et al. Uranium and other natural radionuclides in drinking water and risk of leukemia: a xxxxx study in Finland. *Cancer Causes Control.* 2002;13(9):825-9.)
Read the abstract below and respond to the following questions about it.

c. Name the primary disadvantage/concern you might have with the type of study base you identified in part B.

d. What does the OR from this type of study estimate?

Week 7, Additional Problems. #5

(Adapted from: Auvinen A, et al. Uranium and other natural radionuclides in drinking water and risk of leukemia: a xxxxx study in Finland. *Cancer Causes Control.* 2002;13(9):825-9.)
Read the abstract below and respond to the following questions about it.

c. Name the primary disadvantage/concern you might have with the type of study base you identified in part B.

Correctly finding all of the cases that arose in this study base is the biggest challenge. If you answered secondary study base in question B, you could get credit for discussing the comparability of the controls to the study base.

d. What does the OR from this type of study estimate?

Week 7, Additional Problems. #5

(Adapted from: Auvinen A, et al. Uranium and other natural radionuclides in drinking water and risk of leukemia: a xxxxx study in Finland. *Cancer Causes Control.* 2002;13(9):825-9.)
Read the abstract below and respond to the following questions about it.

c. Name the primary disadvantage/concern you might have with the type of study base you identified in part B.

Correctly finding all of the cases that arose in this study base is the biggest challenge. If you answered secondary study base in question B, you could get credit for discussing the comparability of the controls to the study base.

d. What does the OR from this type of study estimate?

CIR. If you gave a different type of study design in question A but the measure you stated here was consistent (i.e., risk set sampled case-control and IDR), you could still get credit.

Week 7, Additional Problems. #5

(Adapted from: Auvinen A, et al. Uranium and other natural radionuclides in drinking water and risk of leukemia: a xxxxx study in Finland. *Cancer Causes Control.* 2002;13(9):825-9.)
Read the abstract below and respond to the following questions about it.

e. Some of the leukemia patients identified in the base cohort were also randomly selected into the subcohort. Would you analyze these people as cases or controls? Why?

Week 7, Additional Problems. #5

(Adapted from: Auvinen A, et al. Uranium and other natural radionuclides in drinking water and risk of leukemia: a xxxxx study in Finland. *Cancer Causes Control.* 2002;13(9):825-9.)
Read the abstract below and respond to the following questions about it.

e. Some of the leukemia patients identified in the base cohort were also randomly selected into the subcohort. Would you analyze these people as cases or controls? Why?

These people are retained as controls but are also cases; they are included as both controls and as cases in the analyses. In a case cohort study, every person in the cohort has an equal chance of being included in the study as a control, regardless of whether they later develop the study disease. Case-cohort studies are nested within a larger cohort and the control subjects are meant to representative of the baseline exposure distribution of the population which gave rise to the cases. This is what allows for the estimation of the CIR.

Week 7, Additional Problems. #5

(Adapted from: Auvinen A, et al. Uranium and other natural radionuclides in drinking water and risk of leukemia: a xxxxx study in Finland. *Cancer Causes Control.* 2002;13(9):825-9.)
Read the abstract below and respond to the following questions about it.

f. Suppose you are concerned that the exposure (levels of uranium in drilled wells) changes significantly over time, and that a person's current exposure status is more important than whether or not they were exposed at baseline. In 2-3 sentences, describe how you could sample controls to capture this.

Week 7, Additional Problems. #5

(Adapted from: Auvinen A, et al. Uranium and other natural radionuclides in drinking water and risk of leukemia: a xxxxx study in Finland. *Cancer Causes Control.* 2002;13(9):825-9.)
Read the abstract below and respond to the following questions about it.

f. Suppose you are concerned that the exposure (levels of uranium in drilled wells) changes significantly over time, and that a person's current exposure status is more important than whether or not they were exposed at baseline. In 2-3 sentences, describe how you could sample controls to capture this.

Density/risk-set sampling: exposure distribution in controls will resemble exposure distribution in total person-time in the base cohort. This is a better method to use if you want to account for changing exposure distribution over time.

Note that we are matching on time at risk, not exposure time.

Week 7, Additional Problems. #7

Read the following abstract and answer the related questions below:

Purpose: To test the hypothesis that exposure to poultry oncogenic viruses may be associated with increased risk of liver cancer for individuals working on poultry farms.
Methods: A case-cohort study of liver cancer within a cohort of 46,819 members of United Food and Commercial Workers (UFCW) unions from 8 cities across the United States.
Results: Farm workers who performed the following occupational tasks had elevated odds of liver cancer compared to farm workers who did not perform these tasks: slaughtering live poultry (OR = 8.9, 95% confidence interval [CI]: 2.7–29.3); catching live chickens (OR = 3.6, 95% CI: 1.2–10.9); and killing other types of animals for food (OR = 4.8, 95% CI: 1.5–16.6).
Conclusion: This study provides preliminary evidence that exposure to poultry oncogenic viruses may be associated with the occurrence of liver cancer.
(**Adapted from:** Felini M, Johnson E, Preacely N, Sarda V, Ndetan H, Bangara S. A pilot case-cohort study of liver and pancreatic cancers in poultry workers. *Ann Epidemiol.* 2011; 21(10):755-66.)

Week 7, Additional Problems. #7

Read the following abstract and answer the related questions below:

A. What kind of study base is this? What is the primary challenge in a study with this type of study base?

B. What measure of association is estimated by the odds ratio in this study design.

C. When designing the study, the researchers considered sampling controls randomly from UFCW union members in only 4 of the studied cities for logistical reasons.
(1) Explain why this may not be a good way to select controls, and (2) describe how controls should be selected in this case-cohort study.

Week 7, Additional Problems. #7

Read the following abstract and answer the related questions below:

A. What kind of study base is this? What is the primary challenge in a study with this type of study base?

This is a primary study base. (Because the population of interest can be defined as a distinct subset of the population – members of UFCW unions in 8 cities – without reference to case status.
The biggest challenge in a study with this type of study base is that it can be difficult to identify all cases within the study base.

B. What measure of association is estimated by the odds ratio in this study design.

C. When designing the study, the researchers considered sampling controls randomly from UFCW union members in only 4 of the studied cities for logistical reasons.
(1) Explain why this may not be a good way to select controls, and (2) describe how controls should be selected in this case-cohort study.

Week 7, Additional Problems. #7

Read the following abstract and answer the related questions below:

A. What kind of study base is this? What is the primary challenge in a study with this type of study base?

This is a primary study base. (Because the population of interest can be defined as a distinct subset of the population – members of UFCW unions in 8 cities – without reference to case status.
The biggest challenge in a study with this type of study base is that it can be difficult to identify all cases within the study base.

B. What measure of association is estimated by the odds ratio in this study design.

CIR

C. When designing the study, the researchers considered sampling controls randomly from UFCW union members in only 4 of the studied cities for logistical reasons.
(1) Explain why this may not be a good way to select controls, and (2) describe how controls should be selected in this case-cohort study.

Week 7, Additional Problems. #7

Read the following abstract and answer the related questions below:

A. What kind of study base is this? What is the primary challenge in a study with this type of study base?

This is a primary study base. (Because the population of interest can be defined as a distinct subset of the population – members of UFCW unions in 8 cities – without reference to case status.
The biggest challenge in a study with this type of study base is that it can be difficult to identify all cases within the study base.

B. What measure of association is estimated by the odds ratio in this study design.

CIR

C. When designing the study, the researchers considered sampling controls randomly from UFCW union members in only 4 of the studied cities for logistical reasons.
(1) Explain why this may not be a good way to select controls, and (2) describe how controls should be selected in this case-cohort study.

In case-cohort studies, it is important to select controls so that their exposure distribution accurately reflects the exposure distribution in the larger cohort at baseline. By selecting UFCW members from only 4/8 cities, the researchers may have controls whose exposure distribution is different from the exposure distribution in all 8 cities. They should instead select controls as a random sample from all 8 cities at baseline.

Week 7, Additional Problems. #8

Read the following abstract and answer the related questions below:

Background: High birthweight in female infants affects breast cancer risk as an adult. We explored whether mothers' weight gain during pregnancy, which influences infants' birthweight, is associated with risk of later breast cancer in their daughters.

Methods: The Nurses' Mothers case-control study of breast cancer was a density case-control study (nested in Nurses Health Study). The mothers of 814 nurses with and 1,809 nurses without breast cancer completed questionnaires with information on pre-pregnancy height and weight, pregnancy weight gain, and other aspects of their pregnancies with the nurse daughters.

Results: Mothers' weight gain during pregnancy was not associated with the daughters' odds of breast cancer. Compared to women whose mothers gained 20-29 pounds, women whose mothers gained 40 or more pounds had an OR of 0.82 for breast cancer (95% CI: 0.55-1.23).

Conclusions: There does not appear to be an association between mothers' weight gain during pregnancy and daughters' subsequent breast cancer development.

Adapted from: Wilson KM, Willett WC, Michels KB. Mothers' pre-pregnancy BMI and weight gain during pregnancy and risk of breast cancer in daughters. *Breast Cancer Res Treat.* 2011; 130(1):273-9.

Week 7, Additional Problems. #8

Read the following abstract and answer the related questions below:

A. What measure of association is estimated by the odds ratio (OR) in this study?

B. Which of the following conditions must be present for the authors to interpret the OR as the measure of association you identified in part A? Circle all of the correct answers.

a. The disease must be rare.
b. The exposure distribution in the controls must reflect the exposure distribution in the person-time in the cohort.
c. The exposure distribution in the controls must reflect the exposure distribution in the cohort at baseline.
d. Controls must be matched to cases on the time of case occurrence.

C. Give one reason why the researchers would have nested a case-control study in the ongoing cohort study to examine this question instead of making use of the full cohort.

Week 7, Additional Problems. #8

Read the following abstract and answer the related questions below:

A. What measure of association is estimated by the odds ratio (OR) in this study?

The authors state that this is a density case-control study, so the reported odds ratios should estimate IDRs. Since no explicit sampling information was given in the abstract to demonstrate that density or risk-set sampling was employed, you were given credit if you provided a different answer and justified it by noting the lack of sampling / person-time information.

B. Which of the following conditions must be present for the authors to interpret the OR as the measure of association you identified in part A? Circle all of the correct answers.

a. The disease must be rare.
b. The exposure distribution in the controls must reflect the exposure distribution in the person-time in the cohort.
c. The exposure distribution in the controls must reflect the exposure distribution in the cohort at baseline.
d. Controls must be matched to cases on the time of case occurrence.

C. Give one reason why the researchers would have nested a case-control study in the ongoing cohort study to examine this question instead of making use of the full cohort.

Week 7, Additional Problems. #8

Read the following abstract and answer the related questions below:

A. What measure of association is estimated by the odds ratio (OR) in this study?

The authors state that this is a density case-control study, so the reported odds ratios should estimate IDRs. Since no explicit sampling information was given in the abstract to demonstrate that density or risk-set sampling was employed, you were given credit if you provided a different answer and justified it by noting the lack of sampling / person-time information.

B. Which of the following conditions must be present for the authors to interpret the OR as the measure of association you identified in part A? Circle all of the correct answers.

a. The disease must be rare.
b. The exposure distribution in the controls must reflect the exposure distribution in the person-time in the cohort.
c. The exposure distribution in the controls must reflect the exposure distribution in the cohort at baseline.
d. Controls must be matched to cases on the time of case occurrence.

C. Give one reason why the researchers would have nested a case-control study in the ongoing cohort study to examine this question instead of making use of the full cohort.

Week 7, Additional Problems. #8

Read the following abstract and answer the related questions below:

A. What measure of association is estimated by the odds ratio (OR) in this study?

The authors state that this is a density case-control study, so the reported odds ratios should estimate IDRs. Since no explicit sampling information was given in the abstract to demonstrate that density or risk-set sampling was employed, you were given credit if you provided a different answer and justified it by noting the lack of sampling / person-time information.

B. Which of the following conditions must be present for the authors to interpret the OR as the measure of association you identified in part A? Circle all of the correct answers.

a. The disease must be rare.
b. The exposure distribution in the controls must reflect the exposure distribution in the person-time in the cohort.
c. The exposure distribution in the controls must reflect the exposure distribution in the cohort at baseline.
d. Controls must be matched to cases on the time of case occurrence.

C. Give one reason why the researchers would have nested a case-control study in the ongoing cohort study to examine this question instead of making use of the full cohort.

Collection of exposure data was lengthy and expensive. Assessing exposure on the whole cohort would thus have been costly and inefficient.