**Instructions**

Please complete the following lab assignment. You may work on the assignment in groups or on your own. However, to get credit, you must submit your own answers in Canvas. This lab is open note and open book. You may also ask the instructor and the TA questions. Please note that in most cases we will try to guide you towards answering your own question rather than directly providing you with an answer.

# Calculations

[R calculations are available here](https://www.dropbox.com/s/gn19tm3657bf9i5/lab_effect_modification.Rmd?dl=0).

[Google sheet calculations are available here](https://docs.google.com/spreadsheets/d/1QK2ZH9YAopJ-YBWL3E_MbjZnNFBP4Taks2WPbqHesvs/edit#gid=961627830).

# Q1. Multiple exposures and outcomes

[Select choice]

A hallmark characteristic of traditional case-control studies is the sampling method used to enroll participants. Specifically, participants are primarily sampled [strategies].

|  |  |
| --- | --- |
| ✅ | based on their "disease" status. |
|  | based on their "exposure" status. |
|  | at random. |

# Q1. Feedback

A hallmark characteristic of case-control studies is the sampling method used to enroll participants. Specifically, participants are primarily sampled based on their "disease" status.

"In contrast to a cohort study, in which exposed and unexposed individuals are compared with regard to the disease incidence (or some other mean value for the outcome), a case-control study compares cases (usually diseased individuals) and controls (e.g., nondiseased individuals) with respect to their level of exposure to a suspected risk factor."

- Szklo, Moyses, Nieto, F. Javier. Epidemiology (p. 24).

Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

Click the "Next" button below to move on to the next question.

# Q2. Strengths of case-control studies

[Check all that apply]

Which of the following are advantages of case-control studies relative to cohort studies? Please check all that apply.

|  |  |
| --- | --- |
|  | Case-control studies tend to be more efficient when the primary exposure of interest is rare. |
|  | When analyzing data from a case-control study, the disease (outcome) incidence rate can be estimated directly. |
| ✅ | Case-control studies require relatively less follow-up time, thus optimizing speed and efficiency. |
|  | Case-control studies tend to be much less prone to selection bias. |

# Q2. Feedback

# Case-control studies do require less follow-up time relative to cohort studies because the outcome is known at the beginning of the study by definition. Therefore, well-done case-control studies tend to yield results more quickly and more efficiently than cohort studies.

# Case-control studies DO NOT tend to be more efficient when the primary exposure of interest is rare. Cohort studies tend to be more efficient when the primary exposure of interest is rare because participants are often sampled (i.e., included) because of their exposure status.

# When analyzing data from a case-control study, the disease (outcome) incidence rate CANNOT be estimated directly because the risk of the outcome within each level of exposure is generally not proportionate to the risk in the population the participants were sampled from.

# Case-control studies DO NOT tend to be much less prone to selection bias. All studies are potentially susceptible to selection bias. In case-control studies, when cases and controls are not selected from the same (or similar) reference population(s), selection bias may ensue (Sklo et al., p. 26).

# Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

# Click the "Next" button below to move on to the next question.

# Q3. Sklo 7c. pg. 47

[Multiple Choice]

A population-based case-control analysis was conducted to evaluate whether dietary patterns influence the risk of a rare disorder, classic Hodgkin lymphoma (cHL) in younger or older adults.

Cases of incident cHL were recruited from the greater Boston metropolitan area of Massachusetts and the state of Connecticut from August 1, 1997, to December 31, 2000. Eligible patients were aged 15 to 79 years, living within the target geographic area and without human immunodeficiency virus (HIV) infection at diagnosis. Cases were identified by using the rapid case ascertainment systems of Harvard and Yale universities with additional support from the Massachusetts and Connecticut state tumor registries. Six hundred seventy-seven eligible cases were invited to participate in the study, and 84% ( n = 567) consented. Certain data used in this study were obtained from the Connecticut Tumor Registry in the Connecticut Department of Public Health. Population-based controls without a history of cHL were frequency matched to cases by age (within 5 years), sex, and state of residence (Massachusetts or Connecticut). In greater Boston, controls were identified through the “Town Books,” annual records documenting all citizens aged ≥ 17 years, which are 90% complete. Of 720 invited controls in Massachusetts, 51% ( n = 367) consented. In Connecticut, 450 eligible controls aged 18 to 65 years were identified by random-digit dialing, and 61% ( n = 276) consented. Of 69 eligible controls in Connecticut aged 66 to 79 years identified through the Health Care Financing Administration (Medicare), 52% ( n = 36) consented to participate.

What is the study base of this study?

|  |  |
| --- | --- |
|  | Individuals aged 15–79 years old living in the greater Boston metropolitan area and the state of Connecticut |
|  | All individuals living in the greater Boston metropolitan area and the state of Connecticut |
|  | All individuals aged 15 -79 |
| ✅ | Individuals aged 15–79 years old living in the greater Boston metropolitan area and the state of Connecticut living with classic Hodgkin lymphoma |

# Q3. Feedback

# What is the study base of this study?

# Individuals aged 15–79 years old living in the greater Boston metropolitan area and the state of Connecticut

# Szklo, Moyses, Nieto, F. Javier. Epidemiology (pg. 531).

# Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

# Click the "Next" button below to move on to the next question.

# Q4. Sklo 5a. pg. 46

[Multiple choice]

In planning an individually matched case-based case-control study to test the hypothesis that air pollution (measured by individually placed monitors) is related to a certain type of respiratory cancer, the investigators decide to match cases and controls on age, gender, ethnic background, and smoking (yes or no).

In addition to general logistical difficulties usually associated with matching, what is the main undesirable consequence that may result from matching cases and controls in this study?

|  |  |
| --- | --- |
|  | This scenario would likely increase statistical power. |
| ✅ | This scenario may lead to overmatching. |
|  | This scenario would likely result in confounding by age. |

# Q4. Feedback

# In addition to general logistical difficulties usually associated with matching, what is the main undesirable consequence that may result from matching cases and controls in this study?

# By matching on ethnic background, the investigators would also match on variables related to ethnic background, which may include residence area. As a result, cases and controls may be overmatched; that is, they may be matched on the exposure of interest.

# Szklo, Moyses, Nieto, F. Javier. Epidemiology (pg. 530).

# Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

# Click the "Next" button below to move on to the next question.

# Q5. Sklo 5b. pg. 46

[Multiple choice]

In planning an individually matched case-based case-control study to test the hypothesis that air pollution (measured by individually placed monitors) is related to a certain type of respiratory cancer, the investigators decide to match cases and controls on age, gender, ethnic background, and smoking (yes or no).

Because the disease of interest is rare, the investigators decide to individually match 10 controls for each case. Is this a reasonable strategy considering the additional costs involved and the tight budget to conduct this study?

|  |  |
| --- | --- |
|  | Probably, yes. |
| ✅ | No, probably not. |
|  | There isn't enough information given to make a determination |

# Q5. Feedback

# Because the disease of interest is rare, the investigators decide to individually match 10 controls for each case. Is this a reasonable strategy considering the additional costs involved and the tight budget to conduct this study?

# Probably not. Little additional efficiency (statistical power) is achieved when the ratio of controls to cases is greater than 4:1 or 5:1.

# Szklo, Moyses, Nieto, F. Javier. Epidemiology (pg. 531).

# Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

# Click the "Next" button below to move on to the next question.

# Q6. Matching on many variables

[True/False]

If an attempt is made to match on too many characteristics, it may prove difficult or impossible for investigators to adjust for all of these characteristics during data analysis.

|  |  |
| --- | --- |
| ✅ | True |
|  | False |

# Q6. Feedback

# This is true. If an attempt is made to match on too many characteristics, it may prove difficult or impossible for investigators to adjust for all of these characteristics during data analysis.

# Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

# Click the "Next" button below to move on to the next question.

# Q7. Frequency matching

[True/False]

Frequency matching is typically preferred over individual matching when there is a limited number of potentially matching control variables and/or when multiple variables are to be matched.

|  |  |
| --- | --- |
| ✅ | True |
|  | False |

# Q7. Feedback

This is true.

"Individual matching may be logistically difficult in certain situations, particularly when there is a limited number of potentially eligible controls and/or if matching is based on multiple variables. An alternative strategy is to carry out frequency matching , which consists of selecting a control group to balance the distributions of the matching variable (or variables) in cases and controls but without doing a case-by-case individual matching."

Szklo, Moyses,Nieto, F. Javier. Epidemiology (Kindle Locations 1057-1060). Jones & Bartlett Learning. Kindle Edition.

Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

Click the "Next" button below to move on to the next question.

# Q8. When to use frequency matching

[Multiple choice]

Under which of the following conditions would frequency matching likely be preferred over individual matching?

|  |  |
| --- | --- |
|  | When controls will be sampled from a geographically defined area. |
|  | When you are concerned about the number of eligible cases. |
| ✅ | When you are concerned about the number of eligible controls. |
|  | When matching is based on a single primary characteristic of interest. |

# Q8. Feedback

The correct answer is: When you are concerned about the number of eligible controls.

"Individual matching may be logistically difficult in certain situations, particularly when there is a limited number of potentially eligible controls and/or if matching is based on multiple variables. An alternative strategy is to carry out frequency matching, which consists of selecting a control group to balance the distributions of the matching variable (or variables) in cases and controls but without doing a case-by-case individual matching."

- Szklo, Moyses, Nieto, F. Javier. Epidemiology (pg. 36-37).

Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

Click the "Next" button below to move on to the next question.

**Q9. Purpose of matching in case-control**

[True/False]

The primary aim of matching in case-control studies is to make cases and controls as similar as possible with regard to potential confounding factors.

|  |  |
| --- | --- |
| ✅ | True |
|  | False |

# Q9. Feedback

The primary aim of matching in case-control studies is to make cases and controls as similar as possible with regard to potential confounding factors.

This is true. Matching of controls to cases is one strategy for the controlling of confounding as it makes both groups similar with regard to the potential confounder. The nested case-control is a special example of matching controls to cases by time.

Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

Click the "Next" button below to move on to the next question.

**Q10. Matching on exposure**

[True/False]

Matching on factors strongly related to the exposure may introduce selection bias in population-based case-control studies.

|  |  |
| --- | --- |
| ✅ | True |
|  | False |

**Q10. Feedback**

This is true. Matching on factors strongly related to the exposure may, indeed, introduce selection bias in population-based case-control studies.

Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

Click the "Next" button below to move on to the next question.

**Q11. Sklo 4a. pg. 46**

[Multiple choice]

A case-control study is conducted within a well-defined cohort. The reason for this is that expensive additional data collection is needed and the budget is not sufficient to obtain these data from all cohort participants.

What type of case-control study within this cohort would be ideal to study multiple outcomes?

|  |  |
| --- | --- |
|  | Case-based |
| ✅ | Case-crossover |
|  | Nest case-control |
|  | Case-cohort |

**Q11. Feedback**

A case-cohort design would be ideal, as the same control group, represented by a sample of the total cohort at baseline, could be used as a comparison group to multiple case groups. If a nested case-control study design were chosen, a different control group would have to be selected for each case group, which would not be cost-effective.

Szklo, Moyses, Nieto, F. Javier. Epidemiology (pg. 530).

Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

Click the "Next" button below to move on to the next question.

**Q12. Sklo 7a. pg. 47**

[Multiple choice]

A population-based case-control analysis was conducted to evaluate whether dietary patterns influence the risk of a rare disorder, classic Hodgkin lymphoma (cHL) in younger or older adults.

Cases of incident cHL were recruited from the greater Boston metropolitan area of Massachusetts and the state of Connecticut from August 1, 1997, to December 31, 2000. Eligible patients were aged 15 to 79 years, living within the target geographic area and without human immunodeficiency virus (HIV) infection at diagnosis. Cases were identified by using the rapid case ascertainment systems of Harvard and Yale universities with additional support from the Massachusetts and Connecticut state tumor registries. Six hundred seventy-seven eligible cases were invited to participate in the study, and 84% ( n = 567) consented. Certain data used in this study were obtained from the Connecticut Tumor Registry in the Connecticut Department of Public Health. Population-based controls without a history of cHL were frequency matched to cases by age (within 5 years), sex, and state of residence (Massachusetts or Connecticut). In greater Boston, controls were identified through the “Town Books,” annual records documenting all citizens aged ≥ 17 years, which are 90% complete. Of 720 invited controls in Massachusetts, 51% ( n = 367) consented. In Connecticut, 450 eligible controls aged 18 to 65 years were identified by random-digit dialing, and 61% ( n = 276) consented. Of 69 eligible controls in Connecticut aged 66 to 79 years identified through the Health Care Financing Administration (Medicare), 52% ( n = 36) consented to participate.

What type of study design is this?

|  |  |
| --- | --- |
| ✅ | Individually matched case-based case-control study |
|  | Unmatched case-based case-control study |
|  | Individually matched case-cohort study |
|  | Risk-set matched case-control study |

**Q12. Feedback**

What type of study design is this?

This is an individually matched case-based case-control study

Szklo, Moyses, Nieto, F. Javier. Epidemiology (pg. 531).

Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

Click the "Next" button below to move on to the next question.

**Q13. Sklo 7b. pg. 47**

[Multiple choice]

A population-based case-control analysis was conducted to evaluate whether dietary patterns influence the risk of a rare disorder, classic Hodgkin lymphoma (cHL) in younger or older adults.

Cases of incident cHL were recruited from the greater Boston metropolitan area of Massachusetts and the state of Connecticut from August 1, 1997, to December 31, 2000. Eligible patients were aged 15 to 79 years, living within the target geographic area and without human immunodeficiency virus (HIV) infection at diagnosis. Cases were identified by using the rapid case ascertainment systems of Harvard and Yale universities with additional support from the Massachusetts and Connecticut state tumor registries. Six hundred seventy-seven eligible cases were invited to participate in the study, and 84% ( n = 567) consented. Certain data used in this study were obtained from the Connecticut Tumor Registry in the Connecticut Department of Public Health. Population-based controls without a history of cHL were frequency matched to cases by age (within 5 years), sex, and state of residence (Massachusetts or Connecticut). In greater Boston, controls were identified through the “Town Books,” annual records documenting all citizens aged ≥ 17 years, which are 90% complete. Of 720 invited controls in Massachusetts, 51% ( n = 367) consented. In Connecticut, 450 eligible controls aged 18 to 65 years were identified by random-digit dialing, and 61% ( n = 276) consented. Of 69 eligible controls in Connecticut aged 66 to 79 years identified through the Health Care Financing Administration (Medicare), 52% ( n = 36) consented to participate.

What is a common reason for using an individually matched case-based case-control study like the one described above?

|  |  |
| --- | --- |
|  | When the disease of interest develops very quickly. |
|  | When researchers have access to a preexisting cohort with previously ascertained disease status. |
|  | When data need to be collected across a large geographic area. |
| ✅ | When the disease is rare. |

**Q13. Feedback**

What is a common reason for using an individually matched case-based case-control study like the one described above?

When the disease is rare.

Szklo, Moyses, Nieto, F. Javier. Epidemiology (pg. 531).

Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

Click the "Next" button below to move on to the next question.

**Q14. Residual confounding**

[Multiple choice]

When we categorize continuous variables (like age) and match controls to cases using those categories, differences within categories can sometimes still remain. What is this phenomenon called?

|  |  |
| --- | --- |
|  | Information bias |
| ✅ | Residual confounding |
|  | Heterogeneity of effects |
|  | Selection bias |

**Q14. Feedback**

The correct answer is: Residual confounding.

"When matching is conducted according to categorical definitions of continuous or ordinal variables, residual differences between cases and controls may remain (residual confounding)."

- Szklo, Moyses, Nieto, F. Javier. Epidemiology (pg. 40).

Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

Click the "Next" button below to move on to the next question.

**Q15. Case-cohort study**

[Check all that apply]

Which of the following statements are true in the case-cohort study? Check all that apply.

|  |  |
| --- | --- |
| ✅ | The control group is a sample of the baseline cohort. |
|  | Problems due to information bias occurs less than the nested case-control study design. |
| ✅ | The control group can be used to study other outcomes. |
|  | The study design allows for direct estimation of risk ratio. |

**Q15. Feedback**

"Controls are selected as a random sample of the total cohort at baseline. In this design, known as case-cohort, the control group may include individuals who become cases during the follow-up... An important advantage of the case-cohort design is that a sample of the baseline cohort can serve as a control group for different sets of cases occurring in the same cohort."

Szklo, Moyses, Nieto, F. Javier. Epidemiology (pg. 28).

Further, "If everyone in the study base has been followed for the same period of time, then the case-cohort study provides an unbiased estimate of the ratio of incidence proportions that would be measured if the entire cohort had been studied."

Lash TL, VanderWeel TJ, Haneuse S, Rothman KJ. Modern Epidemiology. fourth. Wolters Kluwer; 2021 (pg. 168).

Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

Click the "Next" button below to move on to the next question.

**Q16. Nested case-control study**

[Multiple choice]

From 1983 to 1988, blood samples were obtained from 3,450 HIV-negative men in the Multicenter AIDS Cohort Study (MACS) and stored in a national repository. In 2010, a researcher was interested in examining the association between levels of inflammation and HIV infection. Of the 3,450 men, 660 men were identified as HIV-infected cases. The researcher investigated the association between C-reactive protein (CRP) and HIV infection among these 660 cases and 660 controls, matched to the cases by age and ethnicity, who did not become infected with HIV. The researcher used the stored blood samples to measure the serum level of CRP, a marker of systemic inflammation.

This 2010 study is an example of a:

|  |  |
| --- | --- |
| ✅ | Traditional case-based case-control study |
|  | Case-cohort study |
|  | Nested case-control study |

**Q16. Feedback**

The correct answer is a traditional case-based case-control study. It isn't a case-cohort study because the entire baseline cohort (or a random sample of the entire baseline cohort) wasn't used as the control group. It isn't a nested case-control study because there is no mention of risk-set sampling (i.e. matching to participants who have not yet experienced the outcome at the time the case experienced the outcome).

Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

Click the "Next" button below to move on to the next question.

**Q17. Nest cc vs. case-cohort**

[True/False]

Case-cohort studies are very similar to nested case-control studies. The main difference between a nested case-control study and a case-cohort study is the way in which controls are chosen.

|  |  |
| --- | --- |
| ✅ | True |
|  | False |

**Q17. Feedback**

This is true. Nested case-control studies sample (choose) controls using incidence density or risk-set sampling. Case-cohort studies sample (choose) controls as a random sample of the total cohort at baseline.

Please make sure you understand why this is the correct answer. You may use the "Previous" button below to update your answer if your original answer was incorrect.

Click the "Next" button below to move on to the next question.

**Feedback**

**Optional**: Please feel free to leave any comments below about the usefulness of this lab. Which parts were helpful? What could I do to improve it? What is still unclear?