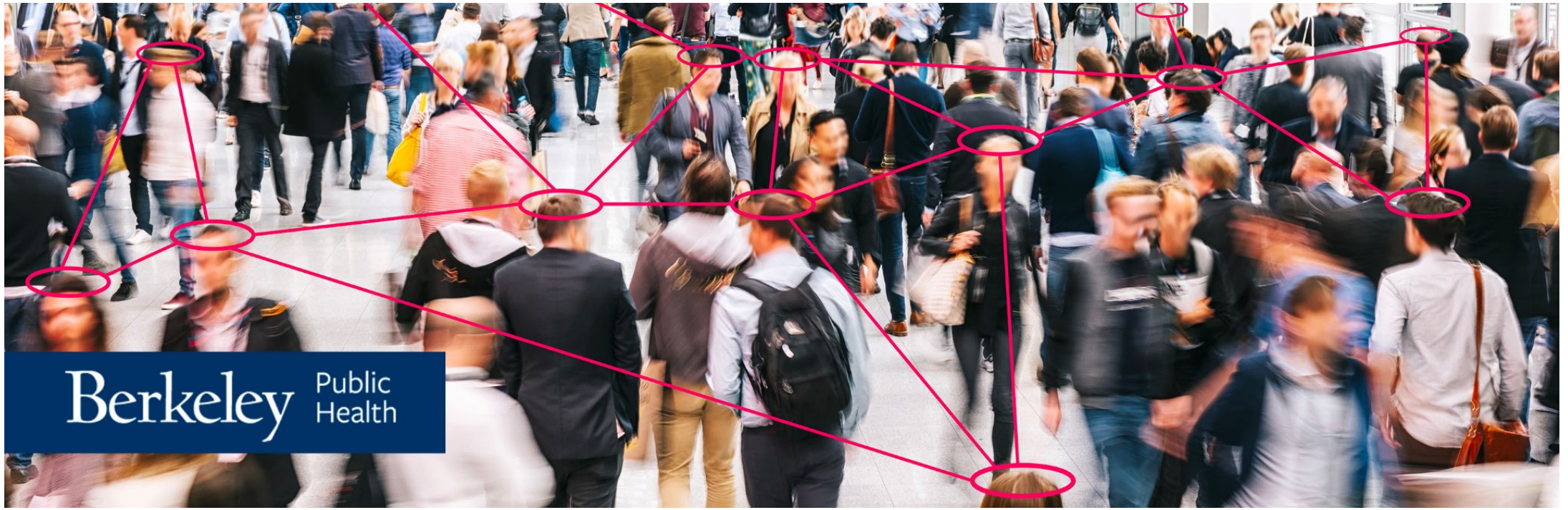


Epidemiologic Methods II

PHW250B

Week 5: Cohort Studies

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.

Agenda

- **Announcements**
  - Exam 1 open Wednesday - Sunday (9/24-9/28)
- **Review**
  - Open and closed populations
  - Types of cohort studies
- **Practice problems**
  - Refresher material tab (which focus on types of cohorts, misclassification)
  - 3 from tab 1 (focused on measures of association from various studies)
- **Q & A**

Exam 1…

YOU CAN DO IT!

Agenda

- **Announcements**
  - Exam 1 open Wednesday - Sunday (9/24-9/28)
- **Review**
  - Open and closed populations
  - Types of cohort studies
- **Practice problems**
  - Refresher material tab (which focus on types of cohorts, misclassification)
  - 3 from tab 1 (focused on measures of association from various studies)
- **Q & A**

Review: Open and Closed Populations

<u>Open population</u>
Characteristics: Migration in and out of populations may occur; loss to follow-up; competing risks; exposure status may change over time.

- Most likely scenario

- Measures of disease: Cumulative incidence via Kaplan Meier, actuarial, density method; directly estimate incidence density (events / person-time). *Cannot directly estimate prevalence.

- Examples: California Cancer Registry; residents of Berkeley, CA; pregnancies in a healthcare system during a defined study period

Review: Open and Closed Populations

Closed population

Characteristics: No migration in and out of populations may occur; no loss to follow-up; no competing risks; exposure status does not change over time.

- **Less likely scenario**

- Measures of disease: Directly estimate cumulative incidence (simple cumulative incidence method), incidence density, occurrence times, prevalence

- **Examples:** RCTs; Framingham Heart Study (common starting point for everyone in the study)

Review: Types of Cohort Studies

- You will see many different uses for “retrospective”, “prospective”, “concurrent”, and “non-concurrent” in regards to cohort studies.

- In this course, we will provide you with guidelines for how we think about these terms, but given the ambiguous terminology, we <u>will not test</u> you on identifying concurrent vs. non concurrent cohort design

    - You will see this question asked in practice problems. These are opportunities to think through these definitions but are in no way definitive answers.

When the investigator gets involved

This is an example of seeing different uses of concurrent/non-concurrent and retrospective/prospective.

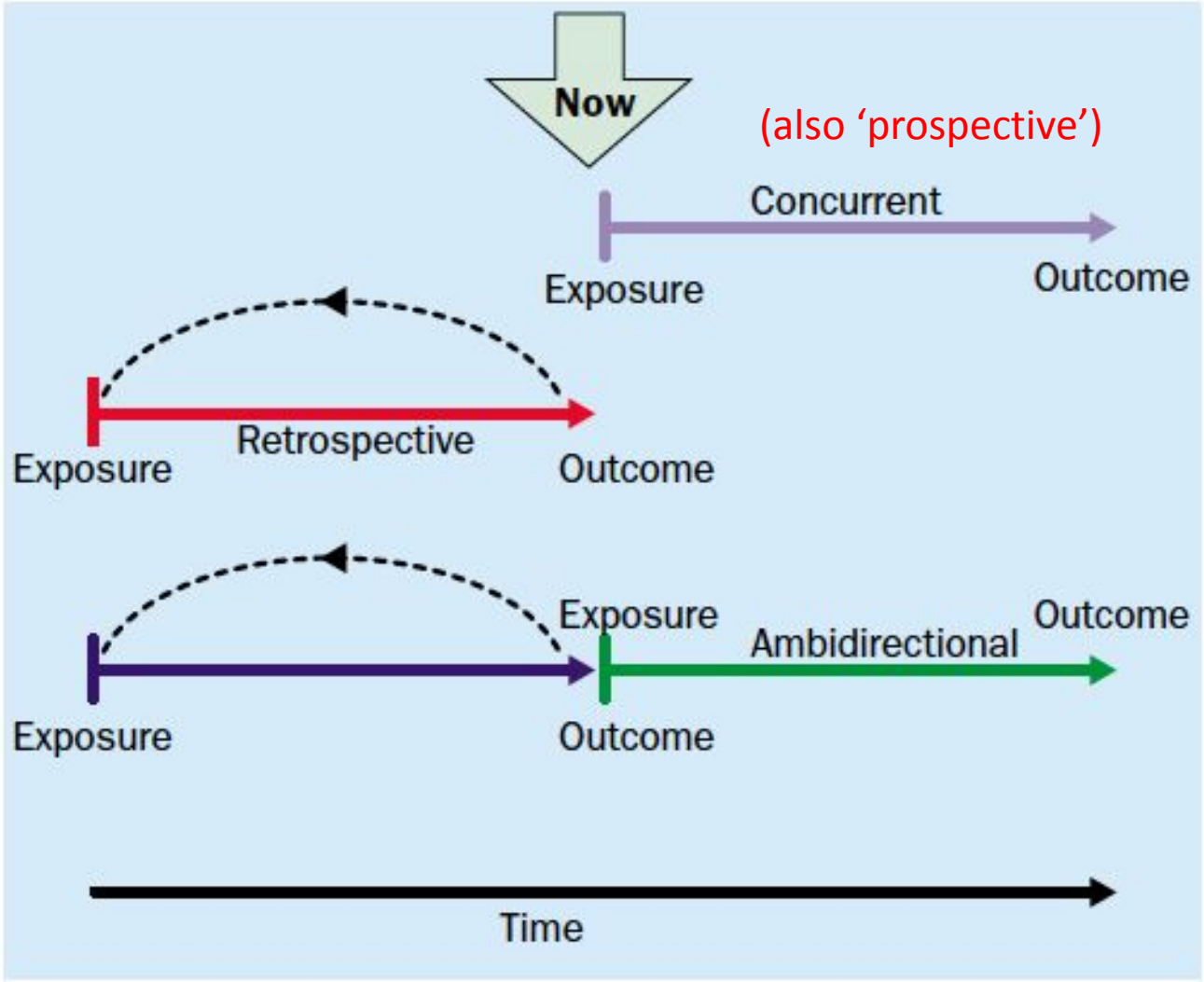


Figure 2: **Schematic diagram of concurrent, retrospective, and ambidirectional cohort studies**

Concurrent cohort study

Retrospective cohort study

Ambidirectional cohort study

Figure from Grimes & Schulz (2002)

Agenda

- **Announcements**
  - Exam 1 open Wednesday - Sunday (9/24-9/28)
- **Review**
  - Open and closed populations
  - Types of cohort studies
- **Practice problems**
  - Refresher material tab (which focus on types of cohorts, misclassification)
  - 3 from tab 1 (focused on measures of association from various studies)
- **Q & A**

Week 5, Refresher Tab Problems. #1

Due to concerns over polluted waterways in Rio, investigators conducted a study in 2017 to assess the incidence of antibiotic-resistant gastrointestinal disease among athletes who competed in water-related sports during the 2016 Rio Olympic Games. Is this an open or closed population? Explain your answer in no more than two sentences.

- *Is membership fixed (defined by an event or specific point in time)?*
- *Can anyone be added to the cohort?*

Week 5, Refresher Tab Problems. #1

Due to concerns over polluted waterways in Rio, investigators conducted a study in 2017 to assess the incidence of antibiotic-resistant gastrointestinal disease among athletes who competed in water-related sports during the 2016 Rio Olympic Games. Is this an open or closed population? Explain your answer in no more than two sentences.

This is a closed population because it is defined by a specific life event, and thus membership is permanent; once the water-related sports competitions in the 2016 Rio Olympic Games ended, no new members could be added to the population, and membership could only be lost by someone dying.

Week 5, Refresher Tab Problems. #3

OBJECTIVE: To investigate whether diabetes is associated with an increased risk of subsequent depression.

PARTICIPANTS: Claims data were randomly selected from 23 million people covered by the Taiwan National Health Insurance program. Using these data, we identified patients aged ≥20 years who were newly diagnosed with diabetes between 2000-2002. We then randomly selected non-diabetic subjects from 2000-2002 for comparison. Both groups were followed up through the end of 2007 and incident cases of diagnosed depression were identified in each group. Risks were compared between the two groups.

Full results and conclusions can be found in: Hsu YM, Su LT, Chang HM, Sung FC, Lyu SY, Chen PC. Diabetes mellitus and risk of depression. *Int J Nurs Stud*. 2011 Oct 7.

A. Identify (1) the exposure and outcome of interest, and (2) the study design. Please be as specific about the design as possible.

Week 5, Refresher Tab Problems. #3

A. Identify (1) the exposure and outcome of interest, and (2) the study design. Please be as specific about the design as possible.

Exposure: Diabetes
Outcome: Depression
Study design: Double concurrent cohort study

- **Concurrent:** Investigators start the study, person-time accumulates.
- **Prospective:** Exposure was measured before the outcome.
- **Double:** Two cohorts (exposed/unexposed) followed over time.

Week 5, Refresher Tab Problems. #4

OBJECTIVE: To investigate whether diabetes is associated with an increased risk of subsequent depression.

PARTICIPANTS: Claims data were randomly selected from 23 million people covered by the Taiwan National Health Insurance program. Using these data, we identified patients aged ≥20 years who were newly diagnosed with diabetes between 2000-2002 (the cases). We then randomly selected non-diabetic subjects from 2000-2002 for comparison. Both groups were followed up through the end of 2007 and incident cases of diagnosed depression were identified in each group. Risks were compared between the two groups.

Full results and conclusions can be found in: Hsu YM, Su LT, Chang HM, Sung FC, Lyu SY, Chen PC. Diabetes mellitus and risk of depression. *Int J Nurs Stud*. 2011 Oct 7**.**

A. Name 2 advantages of this type of study design.

Week 5, Refresher Tab Problems. #4

OBJECTIVE: To investigate whether diabetes is associated with an increased risk of subsequent depression.

PARTICIPANTS: Claims data were randomly selected from 23 million people covered by the Taiwan National Health Insurance program. Using these data, we identified patients aged ≥20 years who were newly diagnosed with diabetes between 2000-2002 (the cases). We then randomly selected non-diabetic subjects from 2000-2002 for comparison. Both groups were followed up through the end of 2007 and incident cases of diagnosed depression were identified in each group. Risks were compared between the two groups.

Full results and conclusions can be found in: Hsu YM, Su LT, Chang HM, Sung FC, Lyu SY, Chen PC. Diabetes mellitus and risk of depression. *Int J Nurs Stud*. 2011 Oct 7**.**

A. Name 2 advantages of this type of study design.

- Can capture dynamic nature of exposures and their relations in time to disease occurrence
- Can assess multiple health outcomes
- Participants can move between exposure groups over time
- Exposures are typically assessed before outcomes (ensuring causal "criterion" of temporality)
- You have direct measurement of disease rates (incidence density)

*Note that with double cohorts of this type, you may not be able to examine multiple exposures, since you have defined the exposure groups at the outset, so this answer was given partial credit.

Week 5, Refresher Tab Problems. #5

To clarify the possible role of postmenopausal estrogen use in coronary heart disease, we surveyed 121,964 female nurses, aged 30 to 55 years, with mailed questionnaires, beginning in 1976. Information on hormone use and other potential risk factors was updated and the incidence of coronary heart disease was ascertained through additional questionnaires in 1978 and 1980, with a 92.7 per cent follow-up. Cases of CHD were documented by medical records. (See results and interpretation in N Engl J Med 1985; 313:1044–9.)

a. Identify the study design. Be as specific as possible.

b. What were the exposures of interest in this study? How was exposure assessed?

c. List a potential concern with the method of exposure assessment.

Week 5, Refresher Tab Problems. #5

To clarify the possible role of postmenopausal estrogen use in coronary heart disease, we surveyed 121,964 female nurses, aged 30 to 55 years, with mailed questionnaires, beginning in 1976. Information on hormone use and other potential risk factors was updated and the incidence of coronary heart disease was ascertained through additional questionnaires in 1978 and 1980, with a 92.7 per cent follow-up. Cases of CHD were documented by medical records. (See results and interpretation in N Engl J Med 1985; 313:1044–9.)

a. Identify the study design. Be as specific as possible.

Concurrent, prospective single cohort study

b. What were the exposures of interest in this study? How was exposure assessed?

c. List a potential concern with the method of exposure assessment.

Week 5, Refresher Tab Problems. #5

To clarify the possible role of postmenopausal estrogen use in coronary heart disease, we surveyed 121,964 female nurses, aged 30 to 55 years, with mailed questionnaires, beginning in 1976. Information on hormone use and other potential risk factors was updated and the incidence of coronary heart disease was ascertained through additional questionnaires in 1978 and 1980, with a 92.7 per cent follow-up. Cases of CHD were documented by medical records. (See results and interpretation in N Engl J Med 1985; 313:1044–9.)

a. Identify the study design. Be as specific as possible.

Concurrent, prospective single cohort study

b. What were the exposures of interest in this study? How was exposure assessed?

Postmenopausal estrogen use via mailed questionnaires

c. List a potential concern with the method of exposure assessment.

Week 5, Refresher Tab Problems. #5

To clarify the possible role of postmenopausal estrogen use in coronary heart disease, we surveyed 121,964 female nurses, aged 30 to 55 years, with mailed questionnaires, beginning in 1976. Information on hormone use and other potential risk factors was updated and the incidence of coronary heart disease was ascertained through additional questionnaires in 1978 and 1980, with a 92.7 per cent follow-up. Cases of CHD were documented by medical records. (See results and interpretation in N Engl J Med 1985; 313:1044–9.)

a. Identify the study design. Be as specific as possible.
Concurrent, prospective single cohort study

b. What were the exposures of interest in this study? How was exposure assessed?
Postmenopausal estrogen use via mailed questionnaires

c. List a potential concern with the method of exposure assessment.
Exposure misclassification, recall bias due to self-report
E.g. People with different more risk factors for CHD may report higher exposure to postmenopausal estrogen.

Week 5, Refresher Tab Problems. #5

As compared with the risk in women who had never used postmenopausal hormones, the age-adjusted relative risk of coronary disease in those who had ever used them was 0.5 (95% CI 0.3,0.8), and the relative risk in current users was 0.3 (95 % CI 0.2, 0.6). (N Engl J Med 1985; 313:1044–9.)

d. Under what conditions would the 92.7% follow-up result in bias away from the null?

Week 5, Refresher Tab Problems. #5

As compared with the risk in women who had never used postmenopausal hormones, the age-adjusted relative risk of coronary disease in those who had ever used them was 0.5 (95% CI 0.3,0.8), and the relative risk in current users was 0.3 (95 % CI 0.2, 0.6). (N Engl J Med 1985; 313:1044–9.)

d. Under what conditions would the 92.7% follow-up result in bias away from the null?

- **Our study results show a protective effect (RR = 0.5 (95% CI 0.03, 0.08))**
- **Protective effect means that the risk of disease is lower in the exposed and higher in the unexposed.**
  - Risk in exposed = 0.4
  - Risk in unexposed = 0.8
  - Relative risk = 0.4 / 0.8 = 0.5
- **If our study results are biased away from the null, we can assume that the truth is closer to the null (null for RR = 1)**
  - Study results = 0.5
  - Truth > 0.5
- **Relative to our study results, the truth will have higher risk of disease in the exposed and lower risk of disease in the unexposed. In other words the study results will have a lower risk of disease in the exposed and higher risk in the unexposed.**
  - Truth:
    - Risk in exposed > 0.5
    - Risk in unexposed < 0.8

Week 5, Refresher Tab Problems. #5

As compared with the risk in women who had never used postmenopausal hormones, the age-adjusted relative risk of coronary disease in those who had ever used them was 0.5 (95% CI 0.3,0.8), and the relative risk in current users was 0.3 (95 % CI 0.2, 0.6). (N Engl J Med 1985; 313:1044–9.)

d. Under what conditions would the 92.7% follow-up result in bias away from the null?

Study results will have a lower risk of disease in the exposed and higher risk in the unexposed.

What decreases the risk of disease in the exposed?

People who are exposed and have the disease drop out.

What increases the risk of disease in the unexposed?

People who are unexposed without the disease drop out.

differential selection bias of exposure by disease status

	Disease	Not Disease
Exposed	Decrease in sample due to loss to follow-up	Increase in sample due to loss to follow-up
Not exposed	Increase in sample due to loss to follow-up	Decrease in sample due to loss to follow-up

Week 5, Refresher Tab Problems. #8

What is a primary issue that can lead to **selection bias** in cohort studies?

A. Differential recollection of exposure by diseased and non-disease individuals

B. Poor retention of unexposed individuals without disease to the end of the study

C. Higher likelihood of high-risk (exposed) individuals to report the outcome

D. Lower participation rates by those unexposed at baseline

Week 5, Refresher Tab Problems. #8

What is a primary issue that can lead to selection bias in cohort studies?

A. Differential recollection of exposure by diseased and non-disease individuals (Recall bias)

B. Poor retention of unexposed individuals without disease to the end of the study

C. Higher likelihood of high-risk (exposed) individuals to report the outcome (Nothing to do with selection)

D. Lower participation rates by those unexposed at baseline (not a primary issue)

Week 5, Refresher Tab Problems. #9

Why is it important to minimize loss to follow up?

Week 5, Refresher Tab Problems. #9

Why is it important to minimize loss to follow up?

Loss to follow-up decreases the number of individuals who can be included in the analysis and so reduces the statistical power of the study. Also, if those who are lost have different rates of disease than those who remain, the study results may be biased.

Agenda

- **Announcements**
  - Exam 1 open Wednesday - Sunday (9/24-9/28)
- **Review**
  - Open and closed populations
  - Types of cohort studies
- **Practice problems**
  - Refresher material tab (which focus on types of cohorts, misclassification)
  - 3 from tab 1 (focused on measures of association from various studies)
- **Q & A**

What can we estimate in a cohort study?

	Disease	No Disease	Total
Exposed	a	b	a+b
Unexposed	c	d	c+d
Total	?	?	?

Measures of disease/exposure

$R_e = P(D|E) = a/(a+b)$

$R_u = P(D|U) = c/(c+d)$

~~$R_t = P(D) = (a+c) / (a+b+c+d)$~~

~~$P_e = P(E) = (a+b) / (a+b+c+d)$~~

Measures of association

✓ $RD = R_e - R_u$

✓ $RR = R_e / R_u$

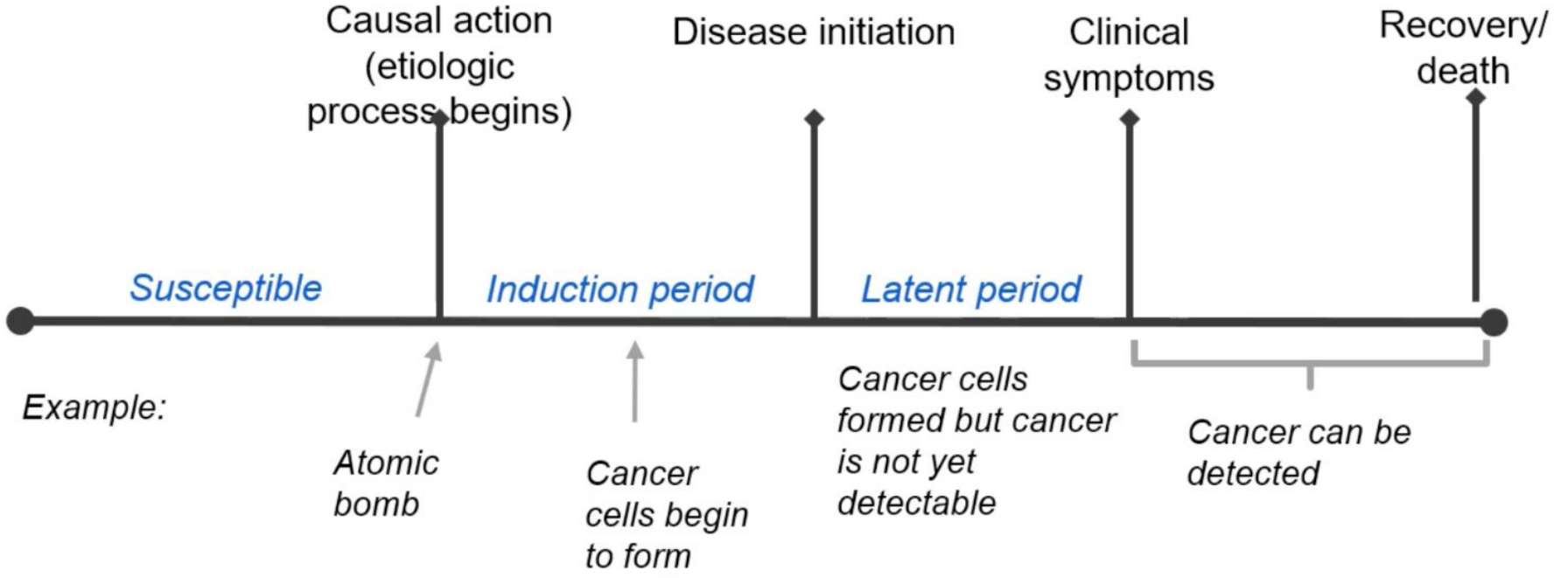
~~$PRD = (R_e - R_u) \times P_e$~~

✓ $AP_e = [(R_e - R_u)/ R_e] \times 100$

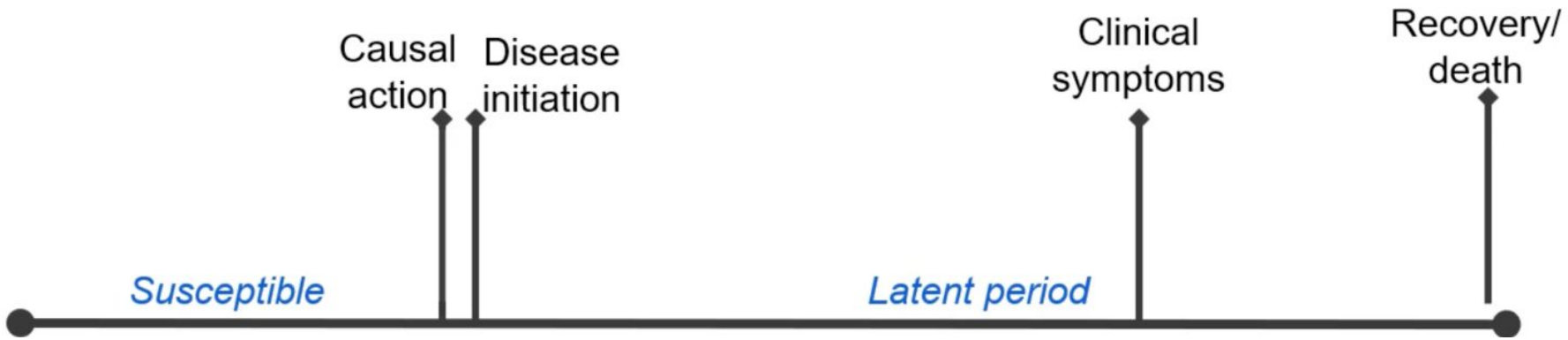
~~$AP_t = [(R_t - R_u)/ R_u] \times 100$~~

In a cohort study that samples on exposure status, we can only estimate measures of association that condition on exposure status.

Terminology: disease induction



In this example, a certain induction period is assumed, and follow-up only begins at the point when disease initiation is hypothesized to have occurred.



When follow-up starts immediately after the causal action, a short induction period and a long latent period is implied. This increased the amount of time a person is considered exposed.

Week 5, Tab 1 Problems. #1

You want to study the relationship between cigarette smoking and hypertension in the U.S. You follow a cohort of 4,916 adults for 5 years, and collect data on each individual at 6-month intervals.

A. During the study period, individuals who smoked contributed a total of 22,500 person-months, and non-smokers contributed a total of 176,805 person-months. Among smokers, 314 developed hypertension; among non-smokers, 786 developed hypertension. Set up a 2x2 table to present these findings.

Week 5, Tab 1 Problems. #1

You want to study the relationship between cigarette smoking and hypertension in the U.S. You follow a cohort of 4,916 adults for 5 years, and collect data on each individual at 6-month intervals.

A. During the study period, individuals who smoked contributed a total of 22,500 person-months, and non-smokers contributed a total of 176,805 person-months. Among smokers, 314 developed hypertension; among non-smokers, 786 developed hypertension. Set up a 2x2 table to present these findings.

	Cases	Person-months
Smokers	314	22,500
Non-smokers	786	176,805

Week 5, Tab 1 Problems. #1

You want to study the relationship between cigarette smoking and hypertension in the U.S. You follow a cohort of 4,916 adults for 5 years, and collect data on each individual at 6-month intervals.

	Cases	Person-months
Smokers	314	22,500
Non-smokers	786	176,805

B. What must be true in order for a given individual in the study population to contribute to person-time of observation?

Week 5, Tab 1 Problems. #1

You want to study the relationship between cigarette smoking and hypertension in the U.S. You follow a cohort of 4,916 adults for 5 years, and collect data on each individual at 6-month intervals.

	Cases	Person-months
Smokers	314	22,500
Non-smokers	786	176,805

B. What must be true in order for a given individual in the study population to contribute to person-time of observation?

Answer: Individuals must be at risk of developing hypertension in order to contribute to person-time.

Week 5, Tab 1 Problems. #1

You want to study the relationship between cigarette smoking and hypertension in the U.S. You follow a cohort of 4,916 adults for 5 years, and collect data on each individual at 6-month intervals.

	Cases	Person-months
Smokers	314	22,500
Non-smokers	786	176,805

C. Calculate the appropriate relative measure of association for these data.

Answer:

IDR = IDe / IDu = (314/22,500) / (786 / 176,805) = 3.14

Week 5, Tab 1 Problems. #1

You want to study the relationship between cigarette smoking and hypertension in the U.S. You follow a cohort of 4,916 adults for 5 years, and collect data on each individual at 6-month intervals.

	Cases	Person-months
Smokers	314	22,500
Non-smokers	786	176,805

IDR = IDe / IDu = (314/22,500) / (786 / 176,805) = 3.14

D. Interpret the measure calculated in Part C.

Week 4, Tab 1 Problems. #1

You want to study the relationship between cigarette smoking and hypertension in the U.S. You follow a cohort of 4,916 adults for 5 years, and collect data on each individual at 6-month intervals.

	Cases	Person-months
Smokers	314	22,500
Non-smokers	786	176,805

IDR = IDe / IDu = (314/22,500) / (786 / 176,805) = 3.14

D. Interpret the measure calculated in Part C.

Answer: During the study period, the rate of hypertension among smokers was 3.14 times the rate among non-smokers.

Week 4, Tab 1 Problems. #2

You conduct a study on the association between adverse childhood experiences and schizophrenia in a nationally-representative closed cohort of 1,000 adolescents, who were followed for 10 years. Assume no completing risks. You find that the rate of schizophrenia was 1.27 per 10,000 person-years. The American Psychiatric Association hears about your study and requests that you report your results in terms of a risk rather than a rate so that they can be more applicable to their patients.

A. Calculate the 15-year risk of schizophrenia in this population.

B. What <u>one</u> additional assumption needs to be met in order to extrapolate to 15 years? (1 point)

Week 4, Tab 1 Problems. #2

You conduct a study on the association between adverse childhood experiences and schizophrenia in a nationally-representative closed cohort of 1,000 adolescents, who were followed for 10 years. Assume no completing risks. You find that the rate of schizophrenia was 1.27 per 10,000 person-years. The American Psychiatric Association hears about your study and requests that you report your results in terms of a risk rather than a rate so that they can be more applicable to their patients.

A. Calculate the 15-year risk of schizophrenia in this population.

CI = 1 - e(-ID * Δt)

ID = 0.000127

CI = 1 - e(-0.000127 * 15) = 0.00190

Also accepted:

Because ID * Δt from the study is very small (.000127 * 10 = 0.00127): CI ≈ ID * Δt ID = 0.000127

CI = 0.000127 * 15 = 0.00191

B. What one additional assumption needs to be met in order to extrapolate to 15 years? (1 point)

Answer: The incidence density needs to be constant in this population over the time period (i.e., 15 years)

Agenda

- **Announcements**
  - Exam 1 open Wednesday - Sunday (9/24-9/28)
- **Review**
  - Open and closed populations
  - Types of cohort studies
- **Practice problems**
  - Refresher material tab (which focus on types of cohorts, misclassification)
  - 3 from tab 1 (focused on measures of association from various studies)
- **Q & A**

EXTRAS

(not covered in live session)

Week 4, Tab 1 Problems. #3

You are interested in studying whether exposure to air pollution, specifically particulate matter (PM-10), increases the risk of asthma among children. You enroll a cohort of children 5-15 years old in California and follow them from 1995-2000. Each month you contact the parents of each child to determine whether the child developed asthma in the last month. During these surveys, loss to follow-up was noted.

a. Why is incidence a more appropriate measure of disease for this study than prevalence? (multiple responses possible) (4 points)

i. The study has information about new cases of disease.

ii. Prevalence can only be calculated in steady state populations

iii. The study is interested in whether air pollution increases risk of asthma.

iv. Calculating prevalence requires knowledge on the duration of disease

b. Based on the information provided above, identify one method of estimating incidence that would be appropriate for this study (there may be more than one correct answer). Please provide both the name and the formula used for this method. (2 points)

Week 4, Tab 1 Problems. #3

You are interested in studying whether exposure to air pollution, specifically particulate matter (PM-10), increases the risk of asthma among children. You enroll a cohort of children 5-15 years old in California and follow them from 1995-2000. Each month you contact the parents of each child to determine whether the child developed asthma in the last month. During these surveys, loss to follow-up was noted.

a. Why is incidence a more appropriate measure of disease for this study than prevalence? (multiple responses possible) (4 points)

i. The study has information about new cases of disease.

ii. Prevalence can only be calculated in steady state populations

iii. The study is interested in whether air pollution increases risk of asthma.

iv. Calculating prevalence requires knowledge on the duration of disease

b. Based on the information provided above, identify one method of estimating incidence that would be appropriate for this study (there may be more than one correct answer). Please provide both the name and the formula used for this method. (2 points)

Incidence measured using the actuarial method (CI=I/(N-W/2) or the incidence density methods (I/PT) – any of the PT formulas were accepted. Kaplan Meier (CI=Ij / Nj) also accepted.

Week 4, Tab 1 Problems. #3

You are interested in studying whether exposure to air pollution, specifically particulate matter (PM-10), increases the risk of asthma among children. You enroll a cohort of children 5-15 years old in California and follow them from 1995-2000. Each month you contact the parents of each child to determine whether the child developed asthma in the last month. During these surveys, loss to follow-up was noted.

c. Given the data below, what **relative** measure of association/effect would you calculate to compare the low to high levels of exposure? Name this measure and calculate it. Be sure to include the appropriate units.

	Asthma	No asthma	Total person-months
Low exposure to PM-10	140	175	1,400
Moderate exposure to PM-10	175	80	1,670
High exposure to PM-10	230	80	1,550
Total	545	335	4,620

Week 4, Tab 1 Problems. #3

You are interested in studying whether exposure to air pollution, specifically particulate matter (PM-10), increases the risk of asthma among children. You enroll a cohort of children 5-15 years old in California and follow them from 1995-2000. Each month you contact the parents of each child to determine whether the child developed asthma in the last month. During these surveys, loss to follow-up was noted.

c. Given the data below, what **relative** measure of association/effect would you calculate to compare the low to high levels of exposure? Name this measure and calculate it. Be sure to include the appropriate units.

	Asthma	No asthma	Total person-months
Low exposure to PM-10	140	175	1,400
Moderate exposure to PM-10	175	80	1,670
High exposure to PM-10	230	80	1,550
Total	545	335	4,620

Incidence density ratio/ relative rate IDR=(230/1550)/(140/1400)=1.48

Week 4, Tab 1 Problems. #3

You are interested in studying whether exposure to air pollution, specifically particulate matter (PM-10), increases the risk of asthma among children. You enroll a cohort of children 5-15 years old in California and follow them from 1995-2000. Each month you contact the parents of each child to determine whether the child developed asthma in the last month. During these surveys, loss to follow-up was noted.

	Asthma	No asthma	Total person-months
Low exposure to PM-10	140	175	1,400
Moderate exposure to PM-10	175	80	1,670
High exposure to PM-10	230	80	1,550
Total	545	335	4,620

Incidence density ratio/ relative rate IDR=(230/1550)/(140/1400)=1.48

d. Interpret the measure you estimated in part c. (1 point)

Week 4, Tab 1 Problems. #3

You are interested in studying whether exposure to air pollution, specifically particulate matter (PM-10), increases the risk of asthma among children. You enroll a cohort of children 5-15 years old in California and follow them from 1995-2000. Each month you contact the parents of each child to determine whether the child developed asthma in the last month. During these surveys, loss to follow-up was noted.

	Asthma	No asthma	Total person-months
Low exposure to PM-10	140	175	1,400
Moderate exposure to PM-10	175	80	1,670
High exposure to PM-10	230	80	1,550
Total	545	335	4,620

Incidence density ratio/ relative rate IDR=(230/1550)/(140/1400)=1.48

d. Interpret the measure you estimated in part c. (1 point)

The rate of asthma per person-month among children 5-15 years old was 1.48 times higher in those who were exposed to a high level of PM-10 than those exposed to a low level of PM-10.

Week 4, Tab 1 Problems. #3

You are interested in studying whether exposure to air pollution, specifically particulate matter (PM-10), increases the risk of asthma among children. You enroll a cohort of children 5-15 years old in California and follow them from 1995-2000. Each month you contact the parents of each child to determine whether the child developed asthma in the last month. During these surveys, loss to follow-up was noted.

	Asthma	No asthma	Total person-months
Low exposure to PM-10	140	175	1,400
Moderate exposure to PM-10	175	80	1,670
High exposure to PM-10	230	80	1,550
Total	545	335	4,620

e. State the ideal counterfactual experiment for this study.

Week 4, Tab 1 Problems. #3

You are interested in studying whether exposure to air pollution, specifically particulate matter (PM-10), increases the risk of asthma among children. You enroll a cohort of children 5-15 years old in California and follow them from 1995-2000. Each month you contact the parents of each child to determine whether the child developed asthma in the last month. During these surveys, loss to follow-up was noted.

	Asthma	No asthma	Total person-months
Low exposure to PM-10	140	175	1,400
Moderate exposure to PM-10	175	80	1,670
High exposure to PM-10	230	80	1,550
Total	545	335	4,620

e. State the ideal counterfactual experiment for this study.

Ideally we would observe the risk or rate of asthma among the same children aged 5-15 in California from 1995-2000 if they had no exposure to PM-10, holding everything else constant.

Week 5, Refresher Tab Problems. #2

Background. …We investigated whether low birth weight or pre-term pregnancy outcomes were related to risk of subsequent IHD in the mother.

Methods. Routine maternity hospital discharge data were used to identify all singleton first births that occurred in Scotland between 1981 and 1985. These birth data were subsequently linked to hospital and mortality records for later admissions and deaths (1981-1999), providing 15–19 years of follow-up on all mothers. Risk ratios were adjusted for socioeconomic deprivation, maternal height and age at the time of the birth, and hypertension.

Findings and interpretation available in full article:
Smith G, Pell J, Walsh D. *The Lancet* 2001;357(9273):2002-2006.

a) Identify the study design – and be as specific as possible. *(2 points)*

When did the investigator begin the study (~2001) relative to the study period?
When was the exposure measured?

Week 5, Refresher Tab Problems. #2

a) Identify the study design – and be as specific as possible. *(2 points)*

Non-concurrent prospective single* cohort
**"Single" indicates there was only one cohort examined.*

Non-concurrent: Investigator started the analysis after the study period.
Prospective: Exposure was measured before the outcome.

Week 5, Refresher Tab Problems. #7

Indicate whether the following statements are true or false?

a) A retrospective cohort study is more efficient than a prospective cohort study for studying disease with a long latent and induction period

TRUE: A prospective cohort study would be inefficient because you would have to follow the cohort for a long time into the future.

b) Cohort studies are the most sensible design for examining many exposures in relation to a single disease.

FALSE: You define cohorts by the exposure. Cohort studies typically look at one exposure and multiple outcomes

c) The ideal comparison group for a cohort study would consist of exactly the same individuals in the exposed group had they not been exposed

TRUE: You want your unexposed (comparison) group to be as close to the counterfactural

d) Loss to follow-up can be a problem in a cohort study but not an experimental study.

FALSE: Experimental studies can also experience loss to follow-up

Week 5, Refresher Tab Problems. #6

A. An investigator in this study later tells you that 5% of those classified as not depressed in the unexposed group were actually truly depressed, while every case in the diabetic group was correctly diagnosed. (This happened because the diabetic group's physicians were monitoring that group's health more carefully.) The study data are below.

	Depression	No depression
Diabetic	240	1350
Not diabetic	150	1460

	D+	D-
Diabetic		
Not diabetic		

Draw a new 2x2 table and calculate (1) the cumulative incidence ratio from the misclassified data above, and (2) the CIR that *would have been obtained* if depression had been correctly diagnosed in both exposure groups.

Week 5, Refresher Tab Problems. #6

A. An investigator in this study later tells you that 5% of those classified as not depressed in the unexposed group were actually truly depressed, while every case in the diabetic group was correctly diagnosed. (This happened because the diabetic group's physicians were monitoring that group's health more carefully.) The study data are below.

	Depression	No depression
Diabetic	240	1350
Not diabetic	150	1460

	D+	D-
Diabetic	240	1350
Not diabetic	150+(1460*.05)= **223**	1460-(1460*.05= **1387**

Draw a new 2x2 table and calculate (1) the cumulative incidence ratio from the misclassified data above, and (2) the CIR that *would have been obtained* if depression had been correctly diagnosed in both exposure groups.

CIR_{biased} = (240/(240+1350)) / (150/(150+1460)) = 0.151/0.093 = 1.62
$CIR_{unbiased}$ = (240/(240+1350)) / (223/(223+1387) = 0.151/0.1385 = 1.09