Epidemiologic Inference in Public Health I

340.721.60
Pablo Martinez Amezcua

Review for Final Examination 2019
Room Assignments for Final Exam  
Wednesday, October 23, 10:00am – 12:00pm

<table>
<thead>
<tr>
<th>If your last name begins with:</th>
<th>Go to:</th>
</tr>
</thead>
<tbody>
<tr>
<td>A – F</td>
<td>Becton-Dickinson Lecture Hall (room W1020)</td>
</tr>
<tr>
<td>G – L</td>
<td>Sheldon Hall</td>
</tr>
<tr>
<td>M – Z</td>
<td>Sommer Hall</td>
</tr>
</tbody>
</table>
Final Exam

• **Wednesday, October 23**
• 10 am-12 noon (CoursePlus enforces this!)
• Room assignments are posted
• You must complete the exam on your own. The exam is closed notes and books. You must adhere to the academic ethics code.
• **Bring a calculator.** You may not use the calculator on your cell phone or laptop.
• 20 multiple choice questions
Final Exam, Continued

• Exams are administered via Courseplus; you are required to bring a laptop to class in order to take the exam

• *Charge your laptop the night before the exam!*

• You will have a **paper copy** of the exam that you may use as scratch paper; this exam must be turned in at the end of the exam
Final Exam, Continued

• Bring to the exam:
  – Pen or pencil
  – Calculator
    • You are not allowed to use the calculator function on your cellphone or tablet.
    • You cannot use Excel or any other computer software.
  – Charged laptop computer
Final Exam

• During the exam:
  – The only application that should be open on your computer is an Internet browser, and the only tab that should be open on that browser is the examination in Courseplus

  – **Any failure to follow this rule will be considered an academic ethics violation**
Methods of Assessment

• 50% PRE-Activity Questions
• 25% Midterm Exam
• 25% Final Exam (cumulative)
Preparing for the Final Exam

• Consider
  – Reviewing your notes
  – Re-listening to online material
  – Listening to live recordings from lectures
  – Reviewing PRE-Activity Questions
  – Reviewing Activities
  – Reviewing Self-Assessments
  – Reviewing Practice Problems (old exam questions)
  – Re-Reading Textbook readings
  – Going to office hours to ask questions
What you learned in this course, in brief!

MODULE 1: Descriptive Epidemiology: Populations and Measurement

MODULE 2: Analytical Epidemiology: Study Designs

MODULE 3: Analytical Epidemiology: Comparisons and Inferences

MODULE 4: Synthesis and Applications
Module 1: Introduction to Epidemiology (Dr. Celentano)

Provided an overview of the role of epidemiology in disease control and the promotion of health.

1. Define epidemiology
2. Uses of epidemiology
3. Tools of epidemiology
Epidemiology

The study of how disease is distributed in populations and what factors influence or determine this distribution.

*Celentano & Szklo, 2019: 2*
Learning the epidemiologic tools of public health allows us to:

1. Measure disease burden
2. Assess risk factors for disease
3. Evaluate interventions
4. Provide the evidence-base for policy decisions
5. Communicate public health evidence
Module 1: Populations

- Defined target, source and study populations
- Described challenges in establishing a case definition and enumerating at-risk populations
- Defined key concepts for quantifying disease: at risk and person-time
Populations

- The group studied
- The group from whom the study population is drawn
- The group to whom inferences will be made
What Does it Mean to be At Risk?

Able to experience the outcome of interest during the follow-up period

At-risk Status Can Change Across Lifespan

Risk for pregnancy?

- Not at risk (not fertile)
- Menarche
- Not at risk (oral contraceptives: assuming proper use!)
- Menopause
- Not at risk (not fertile)
- Birth
- Death

Time
At-Risk Individuals Over Time

- What do the circles represent?
- What do the bars represent?
- What do the arrows represent?

EVENTS

CENSORING
The Concept of Person-Years

- Year 1: 5 py
- Year 2: 4 py
- Year 3: 3 py
- Year 4: 3 py
- Year 5: 2 py

Total = 17 py
Module 1: Epidemics (Dr. Celentano)

- Interpreted epidemic curves from an outbreak investigation
Endemic and Epidemic Disease

Number of Cases

Time

Endemic Epidemic
Three Important Variables in the Investigation of an Epidemic

1. Time of exposure
2. Time of disease onset
3. Incubation period
Module 1: Measures of Morbidity and Mortality (Dr. Deal)

- Defined and calculated indices of **morbidity**:
  - Prevalence
  - Incidence

- Indices of **mortality**:
  - Mortality rate (overall, age-specific mortality, causespecific mortality rate)
  - Case-fatality ”rate”

- **Adjusted rates**:
  - Direct
  - Indirect
Measures of Morbidity and Mortality

• Indices of morbidity:
  – Prevalence
  – Incidence
    • Cumulative incidence
      – Can be directly calculated in a closed cohort
      – In an open population, need other methods (e.g., Kaplan-Meier)
    • Incidence rate

  – Prevalence = Incidence x Duration of disease
Measures of Morbidity and Mortality continued

• Indices of mortality:
  – Mortality rate
    • Overall
    • Cause-specific

• To compare 2 populations that differ on one or more characteristics (e.g., age)
  • Direct adjustment
  • Indirect adjustment (SMR, SIR)
    – Adjusted rates are useful for comparing populations, but are NOT the true rates
Module 2: Validity & Reliability (Dr. Deal)

• Defined and calculated measures of validity, predictive value and reliability
• Described how the use of different cutpoints for a screening test affects the validity of the test
• Defined and calculated measures of reliability (overall percent agreement, positive percent agreement)
• Interpreted Kappa statistic
Validity: Sensitivity & Specificity

Validity = the ability of a test to distinguish between individuals who have a disease and individuals who do not have the disease

(1) **Sensitivity** = ability of a test to correctly identify individuals who have the disease

(2) **Specificity** = ability of a test to correctly identify individuals who do not have the disease
Comparing Cutpoints (Activity 3)

<table>
<thead>
<tr>
<th>Cutpoint of ≥80 mg/dL:</th>
<th>Cutpoint of ≥200 mg/dL:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Sensitivity = 100% (no FN)</td>
<td>• Sensitivity = low (lots of FN)</td>
</tr>
<tr>
<td>• Specificity = low (lots of FP)</td>
<td>• Specificity = 100% (no FP)</td>
</tr>
</tbody>
</table>

![Contingency Tables]

<table>
<thead>
<tr>
<th>Cutpoint</th>
<th>Test+</th>
<th>Test-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>80 mg%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes+</td>
<td>70</td>
<td>18</td>
<td>70</td>
</tr>
<tr>
<td>Diabetes-</td>
<td>504</td>
<td>465</td>
<td>510</td>
</tr>
<tr>
<td>Sensitivity = 100% (no FN)</td>
<td>Specificity = low (lots of FP)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cutpoint</th>
<th>Test+</th>
<th>Test-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>140 mg%</td>
<td>52</td>
<td>45</td>
<td>70</td>
</tr>
<tr>
<td>Diabetes+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes-</td>
<td>45</td>
<td>465</td>
<td>510</td>
</tr>
<tr>
<td>Sensitivity = low (lots of FN)</td>
<td>Specificity = 100% (no FP)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cutpoint</th>
<th>Test+</th>
<th>Test-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>200 mg%</td>
<td>26</td>
<td>44</td>
<td>70</td>
</tr>
<tr>
<td>Diabetes+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes-</td>
<td>0</td>
<td>510</td>
<td>510</td>
</tr>
<tr>
<td>Sensitivity = low (lots of FN)</td>
<td>Specificity = 100% (no FP)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Comparing Cutpoints (Activity 3)

### Cutpoint of $\geq 80$ mg/dL:
- Sensitivity = 100% (no FN)
- Specificity = low (lots of FP)

### Cutpoint of $\geq 200$ mg/dL:
- Sensitivity = low (lots of FN)
- Specificity = 100% (no FP)

### Which cutpoint is better?

**Depends on the importance of False Positives and False Negatives**

### Consequences of FP:
- Emotional cost
- Financial cost to re-test
- More invasive test

### Consequences of FN:
- Missed opportunity to treat
Predictive Value

• What proportion of individuals who test positive (or negative) actually have (or do not have) the disease?

• **Positive & Negative Predictive Value:**
  – Are important for the patient
  – Depend on disease prevalence & validity of the test
Validity vs. Reliability

• **Validity**
  – Are the test results correct?
  
    – *Analogous to accuracy*

• **Reliability**
  – Are the test results the same when the same test is repeated in the same individual under similar conditions?
  
    – *Analogous to precision*
Public Health Surveillance (Dr. Castillo)

The WHO Definition

- Surveillance is the “systematic collection and use of epidemiologic information for the planning, implementation, and assessment of disease control” (WHO, 1968)

- “Public health surveillance is the continuous, systematic collection, analysis and interpretation of health-related data needed for planning, implementation, and evaluation of public health practice”* (WHO, 2012)

Types of Surveillance

- **Passive surveillance**
  - Passive surveillance often gathers disease data from all potential reporting health care providers—most common type of surveillance

- **Active surveillance**
  - Active surveillance provides stimulus to health care providers in the form of individual feedback or other incentives

- **Sentinel surveillance**
  - Instead of attempting to gather surveillance data from all health care providers, a **sentinel surveillance system** selects, either randomly or intentionally, a small group of health providers from whom to gather data
Module 2: Epidemiologic Study Design I
(Dr. Celentano)

• Reviewed ecologic and cross-sectional study designs and appreciate how they inform other types of observational studies
• Described the structure and design of case-control studies, including selection of cases and controls and the process of matching
• Discussed potential sources of bias, and distinguish between nested and non-nested designs
Descriptive Epidemiology

• Distribution of disease and its determinants is the domain of descriptive epidemiology

  Analysis of disease patterns according to the characteristics of the person, place and time

  Who is getting the disease?
  Where is it occurring?
  How is it changing over time?

Descriptive study designs

- Case reports and case series
- Cross-sectional surveys
- Exploratory ecological designs
Analytical Epidemiology

- Determining risk factors and causes of disease
- Evaluating preventive and therapeutic interventions that alter the course of disease

Analytic Study Designs

- Case-control studies
- Cohort studies
- Randomized trials
Design of a Case-Control Study

Then Determine Exposure History:

Start With:
- People with the Disease: ‘Cases’
- People without the Disease: ‘Controls’

Determine Exposure History:
- Were Exposed
- Were Not Exposed

Advantages of a Case-Control Design
- When the disease is rare (low prevalence)
- Relatively short study time
- Relatively inexpensive
- Possible to study associations of a disease with several exposures
Matching In Case-Control Studies

Group matching
(frequency matching, stratification)

Individual matching
(matched pairs)
Design of a Nested Case-Control Study

TIME 1

YEARS

TIME 2

Study Population

obtain interviews, bloods, urines, etc.

Develop Disease

Do NOT Develop Disease

CASES

CONTROLS

NESTED CASE-CONTROL STUDY
Advantages of Nested Case-Control Studies

• Possibility of recall bias is eliminated, since data on exposures are obtained before disease develops.

• Exposure data are more likely to represent the pre-illness state since they are obtained years before clinical illness is diagnosed.

• Costs are reduced compared to those of a prospective study, since laboratory tests need to be done only on specimens from subjects who are later chosen as cases or as controls.
Module 2: Epidemiologic Study Design II (Dr. Celentano)

• Described the basic components and structure of cohort studies

• Described the key features of randomized trials
  - Distinguished experimental from observational designs
  - Described benefits of randomization
  - Described basic design elements of randomized trials
Design of a Cohort Study

Start with:

Exposed

Then, follow-up:

Disease

No Disease

Not Exposed

Disease

No Disease

For a defined standard period of time
Prospective / Retrospective

- **Prospective** (concurrent)
  - *Cohort is assembled NOW* and followed into
  - Obtain exposure information prior to determining endpoint status

- **Retrospective** (nonconcurrent/historical)
  - *Cohort was assembled in the PAST* and followed forward in time up to the present
  - Obtain endpoint status prior to or at the same time as exposure information
Advantages and Disadvantages of Cohort Studies

• **Advantages**
  - Can assess several outcomes simultaneously
  - Control of time and outcome measurements
  - Less potential for bias than case-control studies but same potential for confounding

• **Disadvantages**
  - Usually requires large samples
  - Often requires long follow-up
  - Not very efficient for rare outcomes if exposure is low
  - Financially costly
  - Cohort members “choose” exposures themselves
Design of a Randomized, Controlled Trial

**Why is randomization important?**

- **Primary**: to remove the potential for bias in the choice of treatment
- **Secondary**: to increase the likelihood of balance between groups for known and *unknown* risk factors

---

** Defined Population **

**NEW Treatment**

- Improved
- NOT Improved

**CURRENT/NO Treatment**

- Improved
- NOT Improved
# Types of Outcome Data from Different Study Designs

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Design</th>
<th>Outcome Data &amp; Measures of Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized trial</td>
<td>Randomize participants to treatment and then follow up to see who develops the outcome</td>
<td>Incident cases of disease; Incidence (risk); incidence rate</td>
</tr>
<tr>
<td>Cohort</td>
<td>Exposure measured and then participants followed up to see who develops the outcome</td>
<td>Incident cases of disease; Incidence (risk); incidence rate</td>
</tr>
<tr>
<td>Case-control</td>
<td>Participants selected into study based on outcome; then go back and assess previous exposure</td>
<td>Mix of prevalent and incident cases (can design to limit to incident cases); odds</td>
</tr>
<tr>
<td>Cross-sectional</td>
<td>Exposure and outcome measured at the same time</td>
<td>Prevalence</td>
</tr>
</tbody>
</table>
## Types of Outcome Data from Different Study Designs

<table>
<thead>
<tr>
<th>Measures of disease frequency</th>
<th>Type of study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incidence (risk)</td>
<td>Randomized trial; Cohort study</td>
</tr>
<tr>
<td>Incidence rate</td>
<td>Randomized trial; Cohort study</td>
</tr>
<tr>
<td>Prevalence</td>
<td>Cross-sectional</td>
</tr>
<tr>
<td>Odds</td>
<td>Case-control</td>
</tr>
</tbody>
</table>
Module 3: Measures of Association, Part 1 (Dr. Celentano)

- Described the importance of making comparisons to determine associations

- Learned how to calculate and interpret measures of association
  - Risk or rate difference
  - Relative risk
  - Odds ratio
Comparison Group

• The purpose of selecting a comparison group is to obtain an estimate of the disease risk experience that the exposed group would have had, had they not had been exposed.

  – The comparison group should be as similar as possible to the exposed group on all factors except the exposure.

Disease Risk Experience Comparisons

Cohort Studies

Start with

EXPOSED AND NOT EXPOSED

Follow

DISEASE NO DISEASE DISEASE NO DISEASE

Calculate risk of disease in those who are EXPOSED

Calculate risk of disease in those who are NOT EXPOSED
Comparing Risks between Groups

• Identify the exposed and comparison groups.
• Then, make comparisons in the risk experiences between the groups.
  – Calculate measures of association
    • Difference in risks
    • Ratio of risks

Calculating Measures of Association

1. Difference in Risks*
   (Risk in Exposed) - (Risk in Non-Exposed)

2. Ratio of Risks*
   \[
   \frac{\text{Risk in Exposed}}{\text{Risk in Non-Exposed}}
   \]

*Can also calculate for rates

Let’s Calculate and Interpret RR

Use the 2x2 table:

<table>
<thead>
<tr>
<th></th>
<th>D+</th>
<th>D-</th>
</tr>
</thead>
<tbody>
<tr>
<td>E+</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>E-</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>
Relative Measures of Association

- **Relative Risk**
  - Calculated in cohort studies

- **Odds Ratio**
  - Calculated in case-control studies

**Interpreting Risk or Rate Ratios**

- **RR = 1**
  - Risk in exposed = Risk in non-exposed
  - (no association)

- **RR > 1**
  - Risk in exposed > Risk in non-exposed
  - (positive association; ? causal)

- **RR < 1**
  - Risk in exposed < Risk in non-exposed
  - (inverse association; ? protective)

**Odds Ratio in a Case-Control Study**

<table>
<thead>
<tr>
<th>History of exposure</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>History of exposure</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>No history of exposure</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

\[
\text{Odds Ratio} = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a \times d}{b \times c}
\]

Turns out the same as in a cohort study.

**Interpreting Risk or Rate Ratios**

- **Risk or rate ratios:**
  - **RR = 1**
    - Risk in exposed = Risk in non-exposed
    - (no association)
  - **RR > 1**
    - Risk in exposed > Risk in non-exposed
    - (positive association; ? causal)
  - **RR < 1**
    - Risk in exposed < Risk in non-exposed
    - (inverse association; ? protective)
Module 3: Measures of Association, Part 2 (Dr. Deal)

- Identified when the OR is a reasonable estimate of the RR.
- Reviewed the calculation of the OR for unmatched studies and calculate the OR for matched case-control studies.
- Understood and applied the concepts of attributable risk percent and the population attributable risk percent.
When is the OR a Reasonable Estimate of the RR?

• Both the OR and RR are measures of association.

• What are there circumstances under which the OR approximates the RR?

  – **Cases are representative**, with regard to history of exposure, of all people with the disease in the source population
  
  – **Controls are representative**, with regard to history of exposure, of all people without disease in the source population
  
  – When the risk of disease is low – “rare disease assumption”
A study of 10 cases and 10 unmatched controls with the following exposure status (E=exposed, N=non-exposed)

6 of the 10 cases were exposed.

3 of the 10 controls were exposed.

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>N</td>
<td>4</td>
<td>7</td>
</tr>
</tbody>
</table>

ODDS RATIO (unmatched pairs) = \[
\frac{6 \times 7}{4 \times 3} = 3.5
\]
Calculate the OR for a Matched Case-Control Study

In a matched pairs study, we calculate the OR as the ratio of the discordant pairs.

\[ \text{ODDS RATIO (matched pairs)} = \frac{b}{c} \]

Real study data do not automatically appear in a 2x2 table format!
**OR in a Case-Control Study: Matched**

A study of 10 matched pairs
10 cases and 10 controls with the following exposure status
(E=exposed, N=non-exposed)

<table>
<thead>
<tr>
<th>CASES</th>
<th>CONTROLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>N</td>
</tr>
<tr>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>E</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>E</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>E</td>
<td>N</td>
</tr>
<tr>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>E</td>
<td>N</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cases</th>
<th>Exposed</th>
<th>Not exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Not exposed</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

ODDS RATIO (matched pairs) = \(\frac{4}{1} = 4\)
Motivating the concept of attributable risk: Risks in the Exposed and Non-Exposed

That is, the “Risk Difference”

Remember, right before this slide we stretched because we were shifting topics!
AR and AR% 

**Attributable Risk in the EXPOSED**

1. Risk Attributable to Exposure in the Exposed

\[
= \left( \text{Risk in the Exposed} \right) - \left( \text{Risk in the Non-Exposed} \right)
\]

Attributable Risk (AR)

**Attributable Risk Percent in the EXPOSED**

2. Proportion of Risk Attributable to Exposure in the Exposed

\[
= \left( \frac{\text{Risk in the Exposed}}{\text{Risk in the Exposed}} \right) - \left( \frac{\text{Risk in the Non-Exposed}}{\text{Risk in the Exposed}} \right)
\]

When multiplied by 100, Attributable Risk Percent (AR%)
### Population Attributable Risk

3. Risk Attributable to Exposure in the Total Group (PAR)

\[
PAR = \left( \frac{\text{Risk in Total Group}}{\text{Total Group}} \right) - \left( \frac{\text{Risk in the Non-Exposed}}{\text{Total Group}} \right)
\]

Total group = exposed plus the non-exposed

Still the same comparison group

Population Attributable Risk (PAR)

### Population Attributable Risk Percent

4. Proportion of Risk Attributable to Exposure in the Total Group (PAR%)

\[
\text{PAR\%} = \left( \frac{\text{Risk in Total Group}}{\text{Total Group}} \right) - \left( \frac{\text{Risk in the Non-Exposed}}{\text{Total Group}} \right)
\]

When multiplied by 100, Population Attributable Risk Percent (PAR%)

### Calculating Risk in the Total Group

Risk in the Total Group =

\[
\left( \frac{\text{Risk in the exposed}}{\text{Risk in the total group}} \right) \times \left( \frac{\% \text{ exposed in the total group}}{100} \right) + \left( \frac{\text{Risk in the non-exposed}}{\text{Risk in the total group}} \right) \times \left( \frac{\% \text{ non-exposed in the total group}}{100} \right)
\]

Simply, the weighted average risk of the disease in the exposed and non-exposed.

### Interpretation of the PAR%

Interpretation:

- Proportion of the disease risk in the population that is due to the exposure.
- Proportion of the disease risk in the population could avoided if no one in the population were exposed.
Module 3: Bias (Dr. Deal)

- Defined and distinguished between non-differential versus differential bias.

- Defined selection bias and information bias in association studies, and identified their sources.

- Described approaches to minimize bias in association studies.
Recall a Goal of Epidemiologic Research

• Goal: To identify exposure and disease associations.

• Observational epidemiologic methods cannot determine whether an association is causal, though.

  A \[\rightarrow\] B

  “A causes B”

• Several possible explanations exist for any epidemiologic association observed, only one of which is cause.
Possible Explanations for Associations

• Bias
  – Information bias
  – Selection bias

• Confounding

• Chance

• Cause
Effects of Bias in Epidemiology

Recall...Populations Lecture

Internal Validity

Bias in how study population was obtained
Measurement error in data collection

Internal validity = Degree to which a study is free from bias or systematic error

Role of bias in epidemiology

- In observational epidemiologic studies
  - Bias exists
- After the results of a study are obtained, identify type, likelihood, direction and magnitude of bias
  - Assess the influence of bias on the inferences, including causal inferences, that may be drawn from the study
- Best approach is to prevent or minimize bias in the first place
  - Design
  - Analysis
Non-Differential Information Bias

• Misclassification of or measurement error in the exposure and/or outcome, but the extent of the error does not differ between the groups being compared.
  
  – Cohort study – The extent of the misclassification or measurement error of the disease does not differ between the exposed and the unexposed.
  
  – Case-control study – The extent of the misclassification or measurement error of the exposure does not differ between cases and controls.
Differential Information Bias

- Differential bias results when the inaccuracy of the classification or measurement differs between:
  - Cases and controls in a case-control study
  - Exposed and nonexposed in a cohort study

**Consequence of differential bias**

- Can bias measures of association in any direction
  - Can lead to apparent association even if one does not really exist
  - Can lead to an apparent lack of association even if one does exist
- The magnitude can be substantial, even with small differences in sensitivity or specificity

**Example of Differential Misclassification in a Case-Control Study**

<table>
<thead>
<tr>
<th></th>
<th>TRUE OR:</th>
<th>BIASED OR:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases</td>
<td>Controls</td>
</tr>
<tr>
<td>Exposed</td>
<td>80</td>
<td>50</td>
</tr>
<tr>
<td>Not Exposed</td>
<td>20</td>
<td>50</td>
</tr>
</tbody>
</table>

\[
\text{OR} = \frac{ad}{bc} = \frac{80 \times 50}{50 \times 20} = 4.0 \quad \text{OR} = \frac{ad}{bc} = \frac{72 \times 50}{50 \times 28} = 2.6
\]
Summary of Information Bias

- Information bias is an over- or underestimation of an association caused by measurement error.
- A non-differential measurement error usually biases the association towards the null.
- A differential measurement error can lead to an over- or underestimation of an association.

Minimize information bias in a case-control study:
- Mask (blind) participants to study hypothesis.
- Mask (blind) interviewers to case or control status of the participants.
- Have trained interviewers interview same proportion of cases and controls.
- Use objective measures of exposure.
- Use of multiple control groups (Activity 7).

Minimize information bias in a cohort study:
- Mask (blind) participants to the study hypothesis when relying on self-report of disease.
- Mask (blind) medical record reviewers to exposure status of participants.
- Use objective measures of disease and with confirmation.
Selection Bias (always differential)

**Selection Bias**

Selection bias: distortion of a measure of association in study population from what would be observed in the source/target population due to the choice of participants

- Selection bias is an error occurring in the design phase of the study

**Classic example of selection bias**

Coffee and cancer of the pancreas

**Primary study results**

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥1 cup day</td>
<td>207</td>
<td>275</td>
</tr>
<tr>
<td>None</td>
<td>9</td>
<td>32</td>
</tr>
</tbody>
</table>

Drinking 1 or more cups of coffee per day is associated with 2.7 times the odds of pancreatic cancer compared to drinking 0 cups of coffee per day.

\[
OR = \frac{ad}{bc} = \frac{207 \times 32}{275 \times 9} = 2.7
\]

**Selection bias in a case-control study**

- Distortion of a measure of association from what would be observed in the source population due to the way the cases and controls were selected in a case-control study

- The purpose of the controls in a case-control study is to provide an estimate of the prevalence of the exposure in the source population for the cases

**Minimizing Selection bias in a case-control study**

- Do not put eligibility/eligibility criteria on the controls that are not also put on the cases

  - Can distort the exposure prevalence in the controls relative to the source population and thus, result in selection bias

  - e.g., Patients with diseases known to be associated with smoking or alcohol consumption were excluded from the control group in the MacMahon study, but they did not exclude smokers and alcohol drinkers from the cases
Module 3: Confounding & Effect Modification (Dr. Celentano)

- Identified and learned how to eliminate or reduce confounding in the design and analysis of observational epidemiologic studies.
  
  A type of bias

- Identified and learned how to highlight the presence of effect modification.
  
  Not bias, possibly biology
Confounder – Classic Definition

• A confounding variable or confounder is a 3rd factor that is
  (1) A risk factor for the disease,
  (2) Associated with the exposure, and
  (3) Not a factor in the causal pathway from exposure to disease (i.e., not a mediator).

• If any of these criteria is not satisfied, then the 3rd factor is not a confounder.

B is a confounder
Empirical Way to Identify a Confounder
Using the actual data from your study:

a) Calculate crude RR

\[
\begin{array}{c|c|c|c}
& D+ & D- \\
\hline
E+ & & \\
\hline
E- & & \\
\hline
\end{array}
\]

\[RR_{\text{crude}}\]

Compare crude and stratum-specific RRs

- If they differ by more than 10-20%, then confounding is likely present.

b) Stratify by potential confounder and calculate stratum-specific RRs

\[
\begin{array}{c|c|c|c}
& D+ & D- \\
\hline
C+ & & \\
\hline
C- & & \\
\hline
\end{array}
\]

\[
\begin{array}{c|c|c|c}
& D+ & D- \\
\hline
E+ & & \\
\hline
E- & & \\
\hline
\end{array}
\]

\[RR_1\]

\[
\begin{array}{c|c|c|c}
& D+ & D- \\
\hline
E+ & & \\
\hline
E- & & \\
\hline
\end{array}
\]

\[RR_2\]
Ex: Assess Presence of Confounding

- Is smoking a confounder of the alcohol-MI association?
  - Yes, the stratum-specific ORs differ from the crude OR

- Is alcohol drinking associated with MI?
  - No, the un-confounded, stratum-specific ORs=1.0

Adapted from Schlesselman 1982
Methods to Minimize Confounding

- In the study design
  - Restrict
  - Match

- In the analysis
  - Stratify (or restrict to one stratum)
  - Adjust

Residual Confounding

- Despite controlling for confounding, some confounding can remain because the confounder isn't measured perfectly or the method used to control for confounding is inadequate.
  - In other words, the adjusted RR still has some bias.

- No foolproof method to identify or quantify residual confounding.
  - Consider how well the confounder has been measured or modeled.
Effect Modification

• In contrast to confounding, effect modification is not an error in the design or analysis of the study.

• Effect modification is present when the association between an exposure and outcome differs between categories of a third factor.
  – It results due to the biological *interaction* between two factors

![Diagram of Effect Modification](image)

*What Does Effect Modification Look Like?*

The association between exposure E and disease D differs in the presence and absence of modifier M.
Your Perfect Example of Effect Modification!

- Inherit a faulty gene for metabolizing lactose (sugar in milk) and drinking milk in the production of GI discomfort (e.g., gas and bloating).

Lactose intolerance (-) Lactose intolerance (+)
Your Perfect Example of Effect Modification!

- Inherit a faulty gene for metabolizing lactose (sugar in milk) and drinking milk in the production of GI discomfort (e.g., gas and bloating).

The effect of milk on GI discomfort differs by the presence of a genetic mutation.

Lactose intolerance (-)  Lactose intolerance (+)
Other Examples of Effect Modification

• The effect of insulin (as a treatment) on hypoglycemia is higher for patients with chronic kidney disease

• The effect of falling on a bone fracture is stronger for people with osteoporosis

• The effect of the physical condition of neighborhoods on physical activity differs by age
Stratifying to Identify Effect Modification

1. Calculate the RR for the exposure-disease association within each stratum of the possible effect modifier

2. Compare stratum-specific RRs to each other
   • If the same $\rightarrow$ no effect modification
   • If different $\rightarrow$ effect modification
     • How different?
       • That’s the art of epidemiology.
Quantitative vs Qualitative Effect Modification

Quantitative

M+

RR=4.0

RRs in same direction.

M-

Qualitative

M+

RR=0.25

RRs in different direction.

M-
Interaction

• Effect modification occurs because of biological interaction between two factors in the production of disease.

• Interaction occurs when the risk of the outcome in people who have two factors is greater or lesser than would be expected based on the independent risks of the outcome in people with one, but not both factors.
  – Greater – positive interaction or synergy
  – Lesser – negative interaction or antagonism
Joint Production of Disease Can be Additive or Multiplicative

• *In the absence of interaction*, when two factors act together, based on biology they can follow

  – An *additive model* – the independent risks from the two factors *add* to each other

  – A *multiplicative model* – the independent risks from the two factors *multiply* each other
Interaction under Additive and Multiplicative Models

**Additive Model**

**Additive model:**
Incidences for participants exposed to neither, one, or both risk factors

<table>
<thead>
<tr>
<th>FACTOR A</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>+</td>
<td>15</td>
<td>50</td>
</tr>
</tbody>
</table>

Under the additive model, expected rate = (9-3)+(15-3)+3 = 21. Observed 50.
50 > 21, so interaction is present.

**Multiplicative Model**

**Multiplicative model:**
RRs for participants exposed to both risk factors is higher than expected based on the independent RRs

<table>
<thead>
<tr>
<th>FACTOR A</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>1.0</td>
<td>3.0</td>
</tr>
<tr>
<td>+</td>
<td>5.0</td>
<td>40.0</td>
</tr>
</tbody>
</table>

Under the multiplicative model, expected RR=15, observed RR=40.
40 > 3.0*5.0, so interaction is present.
Module 3: Causal Inference (Dr. Deal)

- Risk Factors
  - Direct vs. Indirect
Module 3: Causal Inference (Dr. Deal)

- Risk Factors
  - Necessary vs. Sufficient

### Examples – Necessary vs. Sufficient Cause

<table>
<thead>
<tr>
<th>Necessary</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>YES</td>
<td>RV virus for rabies</td>
<td>Extremely high doses of radiation for leukemia</td>
</tr>
<tr>
<td></td>
<td>Trisomy 21 for Down’s syndrome</td>
<td></td>
</tr>
<tr>
<td>NO</td>
<td>HPV for cervical cancer</td>
<td>Smoking for lung cancer</td>
</tr>
<tr>
<td></td>
<td>Phenylalanine in the diet for PKU</td>
<td>Most chronic diseases</td>
</tr>
</tbody>
</table>
Key concepts in addressing public health problems using epidemiology

• Recognizing and describing public health problems

• Generate hypotheses
  – Patterns of disease occurrence by person, place, and time

• Develop a research question to address public health problem based on a causal model
  – Define population-at-risk, exposure, outcome

• Test hypotheses based on the causal model using epidemiologic approaches
  – Make comparisons between groups

• Draw inferences and revise the causal model
  – Establish temporality of the exposure-disease relationship
  – Compare results among studies in epidemiology and other fields
  – Determine if risk factors are causal

• Intervene based on revised causal model to prevent the public health problem
Causal Models (aka Causal Pathways)

- Describing causal models or pathways based on associations observed can help us to identify where to intervene.
  - This is the simplest causal model or pathway:

  A ➔ B
  "A causes B"

- Recall: Causal Models (aka Causal Pathways)
  - What are some ways to characterize causal relationships in causal models or pathways?
    - Necessary and sufficient causes
    - Direct and indirect causes
What is Causal Inference?

“The thought process and methods that assess or test whether a relation of cause to effect does or does not exist.”

- Miguel Porta, 2008
Causal “Criteria”

Hill’s aspects of an association that should be considered when assessing causation

- **Strength**
  - How big is the measure of association?

- **Consistency**
  - Are findings from studies conducted using different designs in different populations similar?

- **Specificity**
  - Is the factor only associated with the endpoint of interest and is this the only factor that has been found to be associated with the endpoint of interest?

- **Temporality**
  - Was the factor experienced before the endpoint occurred?

- **Biological gradient**
  - Does the magnitude of the measure of association increase (or decrease) with increasing extent of the factor?

Hill’s aspects of an association that should be considered when assessing causation

- **Plausibility and Coherence**
  - Is the association supported or refuted by the contemporary biologic literature?

- **Experiment**
  - Does the magnitude of the measure of association decrease (or increase) after changing the factor status?

- **Analogy**
  - Have associations between similar factors and similar endpoints been observed?

Hill’s Causal Guidelines

1. Temporal relationship
   • The only guideline that must be met

2. Strength of the association
3. Consistency
4. Specificity
5. Biological gradient
6. Biological plausibility
7. Coherence
8. Experiment
9. Analogy

If met, it is less likely that an observed association is due to confounding or bias (or chance)

Module 4: Evaluating Screening Programs & Health Interventions (Dr. Celentano, online material)

• Described the natural history of disease and screening
• Reviewed criteria to evaluate screening effectiveness
• Reviewed designs to evaluate:
  – Screening programs and outcomes
  – Health services and clinical interventions
• Evaluated sources of bias in screening
• Described how measurement of a health intervention can affect how we determine effectiveness
Natural History of Disease

- Preclinical Phase
  - Biologic Onset
  - Detectable Preclinical Phase (DPCP)

- Clinical Phase
  - Symptoms
  - Diagnosis
  - Therapy

- Outcome

- Disease First Detectable by Screening
Natural History of Cervical Cancer

- **Primary Prevention**
  - Health education
  - HPV vaccine

- **Secondary Prevention**
  - At risk
    - Exposure to HPV serotypes 16,18
  - Pre-clinical phase
    - HPV infection, tumor growth and dissemination
  - Disease onset
  - Pap smear
  - HPV testing

- **Tertiary Prevention**
  - Surgical treatment (conization, among others), radiation/chemotherapy, palliative care

- **Clinical Phase**
  - Abnormal vaginal discharge and bleeding, symptoms of invasion to surrounding tissues (colon) and metastasis

- **Time**

- **Severity of Disease**

- **Diagnosis & Treatment**

- **Chronic/Sequela**

- **Recovery/Cure**

- **Clinical horizon**
Assessing the Effectiveness of a Screening Program Using Outcome Measures

1. Reduction of mortality in the population screened
2. Reduction of case-fatality in screened individuals
3. Increase in percent of cases detected at earlier stages
4. Reduction in complications
5. Prevention of or reduction in recurrences
6. Improvement of quality of life in screened individuals
Design of a Randomized Trial of the Benefits of Screening: HIP Study

HIP Enrollees
~ 62,000

Randomized

Screening Including Mammography
~31,000

Breast Cancer

No Breast Cancer

Regular Care
~31,000

Breast Cancer

No Breast Cancer

Compare Mortality
10/20/2015, the American Cancer Society released updated guidelines for breast cancer screening for women at average risk.

Experts, advocates criticize new mammography advice

Conflicting suggestions on when to have tests could confuse women

BY MEREDITH COHN
The Baltimore Sun

The American Cancer Society now recommends that women wait to begin annual mammograms until age 45 rather than 40, advice that conflicts with other recommendations, leading some to worry that it will confuse women who might need the test.

While several professional societies continue to recommend annual tests starting at age 40, the U.S. Preventive Task Force recommends biennial exams beginning at age 50.

The newest advice released Tuesday prompted a sharp reaction from some experts and advocates who fear it could prompt women to put off mammograms used to detect breast cancer, the most common cancer found in women, which kills about 40,000 a year in the United States.

“We’ve spent years telling women to come get mammograms on a regular basis, and this [new recommendation] makes me crazy,” said Dr. Michael Schultz, a breast cancer surgeon and medical director of the breast center at the University of Maryland St. Joseph Medical Center. “It’s confusing enough for women.

The cancer society’s recommendation prompted Sen. Barbara A. Mikulski, a Maryland Democrat, to renew her call Tuesday for a bill she introduced in August to protect women’s access to mammograms starting at age 40.

“Women’s access to free preventive mammograms must not be impeded, discouraged or eliminated,” Mikulski said.

The cancer society based its new See MAMMOGRAMS, page 17

Baltimore Sun Oct 21, 2015
Possible Biases in Evaluating Screening Programs

• **Volunteer bias** - People who choose screening may have health-seeking behaviors in general or may have poorer risk profiles (e.g., family history) when compared to those who do not choose screening. Thus, will bias toward screening appearing beneficial or not appearing beneficial, respectively. *Concern in observational studies of screening (not randomized trials). Solution: Randomize.*
Possible Biases in Evaluating Screening Programs

• **Lead time bias** - Survival appears longer for those screened compared with those not screened because their diagnosis was advanced in time by screening. *Concern if compare survival times between cases detected in a screening program and cases detected outside of a screening program. Solution: Compare cause-specific mortality rates.*
Possible Biases in Evaluating Screening Programs

• **Length bias** - More likely to screen/detect cases with a longer than shorter detectable preclinical phase (DPCP), and cases with a longer DPCP are likely those that are more slowly progressing and thus, likely have a better prognosis. *Concern if compare case-fatality rates or the 5-year survivals between cases that were detected by screening and those not detected by screening. Solution: Compare cause-specific mortality rates.*
# Health Intervention Evaluation

<table>
<thead>
<tr>
<th>Evaluation of Health Services</th>
<th>Evaluation of Health Services</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Efficacy</strong></td>
<td><strong>Effectiveness</strong></td>
</tr>
<tr>
<td>The extent to which a specific intervention or service produces a beneficial result under ideal conditions (based on results of an RCT)</td>
<td>The extent to which a specific intervention or service, as deployed in the field, does what it is intended to for a defined population</td>
</tr>
</tbody>
</table>
Health Outcomes Research

- Morbidity
- Mortality
- Quality of life (Euro-QAL)
- Functional status (ADL, IADL)
- Patients’ perceptions of health status
  - Symptom recognition
  - Patient satisfaction and pain levels (Smiley faces)
Health Outcomes Research – Non-randomized (In vogue right now!)

Use existing electronic medical records to address key clinical questions, including about interventions (e.g., treatment effectiveness in the setting of PRECISION MEDICINE).

**Advantages of Using Large Data Sets**

- Data refer to real-world populations, so the problem of external validity is minimized
- Analysis can usually be completed quickly because the data have already been collected
- Sample size is usually not a problem except in specific subgroups

**Disadvantages of Large Data Sets - I**

- Because data were gathered for fiscal or administrative purposes, they are often not well suited for research and may be incomplete
- Data on independent and dependent variables may be very limited

**Disadvantages of Large Data Sets - II**

- Data on disease severity, details of intervention and diagnostic coding may be inconsistent
- Data relating to possible confounders may be inadequate or missing
Final Exam Emphasis

• The emphasis of the final exam questions will be on concepts/methods covered in greater detail or covered multiple times in the course lectures, Activities, Assignments, textbook and other course materials.

• Cumulative
Genetic Epidemiology

”Most diseases have both genetic and environmental components. Unless we consider both, we will never get it right”!

Multifactorial Traits (Complex Traits)

- Two genes additive effects
- Gene-environment additive effects
- Gene-environment interaction
Genetic Epidemiology Questions

Is there familial clustering?

Where is the disease gene?

How does this gene contribute to disease in the general population?

<table>
<thead>
<tr>
<th></th>
<th>Disease</th>
<th>No disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family history, positive</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Family history, negative</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

SNP4 (t/g)

SNP3 (a/c)

SNP2 (a/c)

SNP1 (t/c)

Good Luck!

WHEN YOU STUDY FOR THE EXAM

BUT YOUR MIND GOES BLANK AS SOON AS THE PROF PUTS IT IN FRONT OF YOU.
Good Luck!

When you study for the exam,

But your prof does it as soon as the prof puts it in front of you.