Epidemiology: Overview of Key Concepts and Study Design

Polly Marchbanks
Lecture Outline (1)

• Key epidemiologic concepts
  - Definition
  - What epi is not…
  - What epi is…
  - Process of epi research
Lecture Outline (2)

- Study Design - general
- Descriptive studies
  - Case reports and case series
  - Descriptive incidence
  - Descriptive prevalence
  - Ecologic (correlational)
Lecture Outline (3)

• Analytic studies
  - Experimental
  - Observational
    - Cohort
    - Case-Control
    - Cross-sectional
Lecture Outline (4)

• Additional epi concepts
  - Bias or systematic error
    - Selection bias
    - Information bias
    - Confounding

• Summary, concluding remarks
EPI DEMIOLOGY is not...

The study of skin diseases

Local Atlanta Resident
EPI DEMIOLOGY is (hopefully) not...

The worst-taught course in school

Anonymous Student
EPI DEMIOLOGY is more than...

The science of long division

Statistician

\[ I_0 = \frac{(480)(2) \times 10^6}{(9.17)(0.953) + 5} \]
EPI DEMIOLOGY is more than...

The science of epidemics and epidemic diseases

Medical Dictionary
The Derivation of Epidemiology

epi = upon

demos = people

logy = study of
Fundamental Assumptions

• Non-randomness

• Preventability
Definition of Epidemiology

The study of the distribution and determinants of disease and health in human populations.
Unique Contribution of Epidemiology

- Focus is on populations
Epidemiology (Study)

The study of the distribution and determinants of disease and health in human populations.
Epidemiology (Distribution)

The study of the distribution and determinants of disease and health in human populations.
Epidemiology (Determinants)

The study of the distribution and determinants of disease and health in human populations.
Epidemiology (Disease And Health)

The study of the distribution and determinants of **disease and health** in human populations.
Applied Epidemiology

• Practical, action-oriented, relevant
• Provides data for decision-making
• Focused on prevention/intervention
Aims of Epidemiologic Research

- Describe
- Explain
- Predict
- Control
Levels of Epidemiologic Research

• Theoretical understanding
• Prevention/ intervention
Cholera in London, 1854
How to Find John Snow Pub

- London
- #39 Broadwick Street
- Corner of Broadwick & Lexington
- Can walk from theater district
- Connect these “dots” on a map: Oxford Circus, Soho Square, Leicester Square, Golden Square—Broadwick is near center of the box you have created
Exposure Variable - "E"

- Characteristic of interest
- Risk factor variable
- Predictor variable
- Independent variable
- Possible causal factor
Outcome or Disease Variable - "D"

- Health event of interest
- Illness, injury
- Response variable
- Dependent variable
- Effect variable
E-D Relationships: Examples

- Smoking - lung cancer
- Obesity - heart disease
- Dietary fat - stroke
- Income - malnutrition
- Prenatal care - birth defects
- Family history - diabetes
- Alcohol - motor vehicle injury
- Access to care - maternal mortality
Process of Epidemiologic Research

- Question
- Design
- Conduct
- Analysis
- Interpretation
- Recommendations
Process of Problem-solving

1. What is the problem?
2. How should the problem be approached?
3. Collect information
4. Organize and analyze the information
5. What does the information mean?
6. What are the next steps?

Process of Epidemiologic Research

1. Question
2. Design
3. Conduct
4. Analysis
5. Interpretation
6. Recommendations
The process of epidemiologic research is also a method of causal reasoning.
Epidemiology

- Theoretical
- Applied
- Consequential
Epidemiology: A Public Health Practice Summary Definition

- Quantitative basic science
- Method of causal reasoning
- Vehicle for clinical and public health action
- Can and should make a difference in the lives of people!
“Epidemiology is FUN...and if it’s not FUN, it’s not epidemiology!”

R. Stallones
Epidemiologic Study Design
Study Design: What Is It?

• Process of planning an investigation
• Link between concept and operation
• Bridge from ideas to action
Process of Epidemiologic Research

- Question
- Design
- Conduct
- Analysis
- Interpretation
- Recommendations
"...all study designs are potentially flawed...there is no such thing as a perfect study design; therefore, it becomes most important to understand the specific limitations of each design...no type of sophisticated statistical analysis will salvage a poorly designed study."

(KKM - pg. 36)
Categories of Study Design

- Descriptive
- Analytic
Descriptive Studies: Overview

• Patterns of occurrence
• Person, place, time
• Program planning
• Generate hypotheses
Descriptive Studies: Examples

- Trends in serum cholesterol
- Incidence of cervical cancer
- Secular trends in heart disease mortality
- Prevalence of domestic violence
- Incidence of rape among street youth
- Epidemiology of low birth weight
- Prevalence of smoking by age and gender
- Incidence and prevalence of dementia
Analytic Studies: Overview

- Determinants of disease
- Etiologic research
- Joint distribution of exposure (E) and outcome/disease (D)
- Test hypotheses
- Quantify association
Analytic Studies: Examples

- Diet - hypertension
- Exercise - depression
- Alcohol - liver cirrhosis
- Sun exposure - melanoma
- Smoking - heart disease
- Use of long acting reversible contraceptives (LARCs) - unintended pregnancy
Knowledge Continuum

Less                        More

Descriptive               Analytic

• Search for clues      • Clues available
Types of Descriptive Studies

- Case reports and case series
- Descriptive incidence studies
- Descriptive prevalence studies
- Ecologic (correlational) studies
Case Reports and Case Series

• Profile of a case or case series
• Generate new hypotheses
• Interface: medicine and epidemiology
• Numerator data only
• No measure of disease occurrence
Descriptive Incidence Studies (1)

• Patterns in occurrence of incident cases
• Defined population
• Specified period of time
• Distribution of cases by factors of interest
Descriptive Incidence Studies (2)

• Case ascertainment:
  – Ongoing reporting systems
  – Medical record review in highly select populations

• Denominators usually from census

• Numerator and denominator difficult to obtain for well-defined population
Descriptive Prevalence Studies (1)

- "Snapshot" of well-defined population
- Classify D and other variables at same time
- Captures all existing disease
- Also known as cross-sectional surveys
Descriptive Prevalence Studies (2)

- **Advantages:**
  - Quick, inexpensive, useful

- **Disadvantages:**
  - Survivor effect
  - Uncertain temporal relationship
Ecologic (Correlational) Studies (1)

- D in relation to E at aggregate level
- Data from groups not individuals
- Unit of observation is a population
Ecologic (Correlational) Studies (2)

• Ex: correlation between average cigarette sales and heart disease death rates in two counties
• Limitation: no individual link of E-D
Ecologic (Correlational) Studies (3)

• Advantages:
  – Quick, inexpensive, data available

• Disadvantages:
  – Aggregate may not = individual
  – Inadequate data on co-variables
  – Averages may mask complex relationships
Descriptive Studies - Summary

• Describe patterns of occurrence
• Four main types:
  - Case reports and case series
  - Descriptive incidence studies
  - Descriptive prevalence studies
  - Ecologic (correlational) studies
• Generate hypotheses for analytic study
Analytic Studies

- Joint distribution of E-D for individual
- Test hypotheses about E-D relationships
- Categories: experimental, observational
Experimental Studies (1)

- Assign E randomly, follow for D or other health outcome

- Types:
  - Clinical trial
  - Field trial
  - Community trial
Experimental Studies (2)

- Rolls Royce!!
- Control of extraneous variables, both known and unknown
- Limitations: ethical concerns, cost, length, not feasible for rare outcomes, volunteer effect
- Stopping rules
Observational Studies

• Nature prevails
• Three main types:
  – Cohort
  – Case-control
  – Cross-sectional
• Dimensions:
  – Direction, timing
Observational Studies - Direction

- Temporal relationship between our observations of E and D
- Forward: start with E
- Backward: start with D
- Nondirectional: E-D simultaneously
- Essence: start with E or D?
Observational Studies - Timing

• Chronological relationship between onset of study and occurrence of D

• Prospective: study onset --> D
  – D occurs after study begins

• Retrospective: D --> study onset
  – D occurs before study begins

• Essence: has the D occurred when the study begins?
Cohort Studies - Direction

• Subjects free of D selected on basis of E, followed forward in time for D
• Start with persons exposed (YES, NO)
• Follow forward for disease (YES, NO)
• Direction always forward
  − E ----> D
• Timing = prospective or retrospective
Cohort Studies - Prospective

- D has not yet occurred at study onset
  - E ---> Study Onset ---> D
  - Move from E to D through "real time"
  - D happens in the present, concurrently with calendar time
Cohort Studies - Historical (Retrospective)

- D has already occurred at study onset
  - E ---> D ---> Study Onset
- Direction still forward because moving from E to D
- Move from E to D through "historical time"
- D happened in the past, nonconcurrently with calendar time
Cohort Studies - Timeline

Prospective Cohort

<table>
<thead>
<tr>
<th>E</th>
<th>Study Begins</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>--------------</td>
<td>---</td>
</tr>
</tbody>
</table>

Historical Cohort

<table>
<thead>
<tr>
<th>E</th>
<th>D</th>
<th>Study Begins</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>--------------</td>
<td>---</td>
</tr>
</tbody>
</table>

CDC
Cohort Timeline - Example
Maternal Thalassemia (Anemia) And Fetal Death

**Anemia Study Death**

**Prospective Cohort**

**Anemia Death Study**

**Historical Cohort**
Cohort Studies - Review

- Subjects selected on basis of E
- Direction always forward
  - E --> D
- Timing
  - Prospective: "real time"
  - Retrospective: "historical time"
Cohort Studies - Flow Chart

Source Population

Study Group E

Study Group E+

D+

D
Cohort - Measures of Occurrence

- **THINK INCIDENCE!!**
- **Cumulative incidence - risk**
  - Number of new cases at end of follow-up divided by number of disease-free persons at start of follow-up
- **Incidence density - rate**
  - Number of new cases at end of follow-up divided by person-time at risk (e.g., person years of disease-free follow-up)
Cohort - Measures of Association (1)

- Think Relative Risk (RR)!!
- Risk ratio if cumulative incidence study
- Rate ratio if incidence density study
- Incidence of D in exposed divided by incidence of D in unexposed
Cohort - Measures of Association (2)

• RR = 1
  - No difference in incidence of D in exposed and unexposed
  - 1.0 = no effect
  - EX: 10/100 // 10/100 = 1.0
  • No association between E and D was observed - exposed and unexposed persons are equally likely to develop D
Cohort - Measures of Association (3)

• **RR > 1**
  - Suggests that E is a risk factor for D
  - **EX**: 50/100 // 10/100 = 5.0
  - Exposed persons are 5 times as likely to develop the D, compared with persons not exposed
Cohort - Measures of Association (4)

- **RR < 1**
  - Suggests that E protects against D
  - EX: 10/100 // 50/100 = 0.2
  - Exposed persons are 0.2 times as likely to develop the D, compared with persons not exposed
Cohort Studies - Major Advantages

• Logical temporal sequence
• Can measure incidence of D
• Well-suited for rare E
• Can study many effects of one E
Cohort Studies - Major Disadvantages

- Many subjects needed for rare D
- Follow-up: logistics, losses
- E status can change over time
- Prospective: time-consuming, costly, observation can influence behaviors
- Historical: requires suitable records
Case-control Studies - Direction

• Review direction: start with E or D?
• Subjects selected on basis of D
• Start with persons having D (YES, NO)
• Compare frequency of E in cases (D) with frequency of E in controls (non-D)
• Look backward for history of E (YES, NO)
• Direction is backward: D --- > E
Case-control Studies - Timing

• Review timing: has D happened when study begins?
• Basic Design: retrospective
  - D has occurred before onset of study
• Hybrid Design: prospective component
  - Study begins, enroll new cases of D
Control Selection– Guidelines

• Critical design issue
• No optimal group for all situations
• Controls should...
  – Represent source population from which the cases were derived
  – Represent persons who, if a case, would have been in the study
  – Be selected independently of E
Possible Sources of Controls

- Random sample of source population often best way to insure controls selected independent of E
- Can use other sources, but be cautious
  - Hospitals
  - Friends, relatives, spouses, neighbors
Review: Case-Control Studies

- Subjects selected on basis of D
  - Start with D
- Look backward for E
- Direction always backward
  - Disease ---> Exposure
- Timing in basic design is retrospective
  - D already occurred when study begins
Case-Control Studies - Flow Chart

Source Population

Cases D+

Controls \( \bar{D} \)

E+

E

E+

E
Case-control - Measures of Occurrence

- **Basic design:**
  - Usually not possible

- **Hybrid design:**
  - Incidence density rates possible if study is population-based
Case-control - Measures of Association

- THINK ODDS RATIO (OR)!!
- Odds ratio
  - Estimator of relative risk
  - Odds of E among cases compared with odds of E among controls
  - Also known as exposure odds ratio
  - OR = 1 - no association
  - OR > 1 - suggests E is a risk factor
  - OR < 1 - suggests E is protective
Case-control - Major Advantages

- Relatively quick and inexpensive
- Well-suited for rare D and D with long latency
- Requires fewer subjects at entry
- Can study multiple E
Case-control - Major Disadvantages

- Design "backward"
- Unsuitable for rare E
- Usually cannot measure D incidence
- Temporal E-D uncertainty
- Prone to selection and recall bias
### Breast Cancer

<table>
<thead>
<tr>
<th>OC Use</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>a</td>
<td>b</td>
<td>H1</td>
</tr>
<tr>
<td>No</td>
<td>c</td>
<td>d</td>
<td>H2</td>
</tr>
</tbody>
</table>

**V1** V2 **Total**
Example of Cohort Study: OC Use and Breast Cancer

- Start with OC users and nonusers
- Follow forward in time for breast cancer
- Compare incidence of breast cancer in OC users with incidence of breast cancer in nonusers
Example of Case-control Study: OC Use and Breast Cancer

• Start with cases of breast cancer and controls who do not have breast cancer
• Look backward in time for a history of OC use
• Compare frequency of OC use in cases with frequency of OC use in controls
Most Fundamental Difference

- Cohort
  - Start with exposure
- Case-control
  - Start with disease
Analytic Cross-sectional Studies (1)

- "Snapshot" in well-defined population
  - Assess E-D (prevalent) at same time
- Nondirectional design
- Analytic potential
Analytic Cross-sectional Studies (2)

• Measures of association
  – Prevalence ratio
    • Prevalence of D among E compared with prevalence of D among non-E
  – Prevalence odds ratio
    • Odds of D among E compared with odds of D among non-E
Analytic Cross-sectional Studies (3)

• **Advantages:** Quick, inexpensive, useful, suitable when E cannot change

• **Disadvantages:**
  - No measures of incidence
  - Survivor effect
  - Unsuitable for rare E, rare D, and D with short duration
  - Temporal E-D uncertainty
Basic Concepts of Bias
Accuracy

Lack of Error

/                               \
Validity                        Precision

Lack of systematic error (bias)  Lack of random error (chance)

- Function of study methodology  - Function of sampling variation
Bias

- **Bias** = Systematic error that results in a distortion of the E-D association
- **Validity** = lack of bias
- **Validity** is a function of study methodology
Bias = Inaccurate Results Due to:

- Manner in which study subjects are selected
  - Selection bias

- Manner in which necessary information is collected or coded (classified)
  - Information (misclassification) bias

- Admixture of effect - effect of E on D is mixed up with effects of other variables
  - Confounding
Selection Bias

Problem with who gets into the study
Information Bias

Problem with how information is collected or coded (classified)
Confounding Problem with a Third Factor (F)

- F must be associated with E
- F must be associated with D even in unexposed
- F must not be intermediate between E and D
Selection Bias

Information Bias

Confounding
Ways to Minimize Selection Bias

• Cohort
  - Select persons on basis of E independent of D
  - Strive for complete follow-up

• Case-control
  - Select persons on basis of D independent of E
Ways to Minimize Information Bias

- Obtain and classify all information as correctly as possible
- Obtain and classify information on E and D independently
Ways to Minimize Confounding

• Randomization
• Restriction
• Matching (partial restriction)
• Stratification
• Statistical modeling

*Note: Be sure to collect needed information on covariables
Summary (1)

• Key concepts
  - Definition of epidemiology
  - Process of epidemiologic research
Summary (2)

• **Types of study designs**
  - **Descriptive**
    - Case reports and case series
    - Descriptive incidence studies
    - Descriptive prevalence studies
    - Ecologic (correlational) studies
  - **Analytic**
    - Experimental
    - Observational
      - Cohort
      - Case-control
      - Cross-sectional
Summary (3)

- **Additional epidemiologic concepts**
  - Bias or systematic error
    - Selection bias
    - Information bias
    - Confounding
Study Design - Concluding Remarks

- Different ways to organize designs
- What design should I select?
- It depends!
- Must consider:
  - Objectives of study
  - Current knowledge about E-D
  - Ethical issues
  - Time, money, human resources
- Flexibility and creativity are KEY!
The End