Observational Study Designs

Denise Boudreau, PhD
Center for Health Studies
Group Health Cooperative

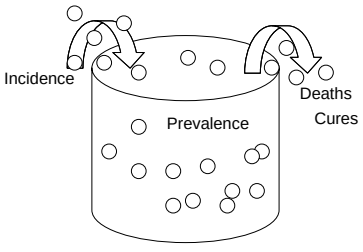
Today

- Review cohort studies
- Case-control studies
  - Design
  - Identifying cases and controls
  - Measuring exposure
  - Risk estimate
  - Strengths and limitations
  - Biases
  - Examples
- Take home points

Review

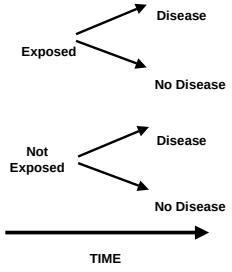
Measures of disease occurrence

- Prevalence = number of disease/population at risk
- Cumulative incidence = new cases/population at risk
  - Closed population
- Incidence rate = new cases/person time of observation among at risk



Cohort Studies

- Study outcomes by exposure
- Subjects are disease free at start
- Can measure incidence
- Can estimate risk
- Two types: prospective and retrospective cohorts
- Bias and confounding always possibility



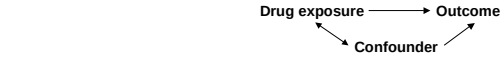
Cohort Studies

- Pros:
  - Incidence and risk estimates
  - Ideal for rare exposures
  - Can often show temporality of association
  - Less bias due to prospective evaluation of exposures
  - Can evaluate multiple diseases
- Cons:
  - Need motivated cohort of people who will be repeatedly evaluated if collecting prospective data
  - $$ if prospectively collecting primary data over long period of time
  - Not efficient design for rare outcomes
  - Feasibility for studies of long duration?
  - Analysis can become complex
  - Bias always possible
    - Confounding
    - Misclassification of disease or exposure
    - Loss to follow up
    - Non-response if collecting primary data
    - Selection bias

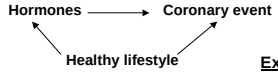
Bias

- Selection bias
- Surveillance bias
- Information bias and misclassification
- Confounding

Confounding



Example of confounding



Example of intermediate factor (not a confounder)

How can one deal with confounding?

- Randomization
  - Assign persons to drug or placebo
  - Not always feasible or ethical
- Matching
  - Equal no. exposed & unexposed in each stratum of confounder
  - Risk of overmatching
  - Can be time-consuming, expensive, and reduce power
- Restriction
  - Look only in specific subgroup(s)
  - Generalizable only to subgroups
- Adjustment
  - Standardization, stratification, multivariate analysis, propensity scores

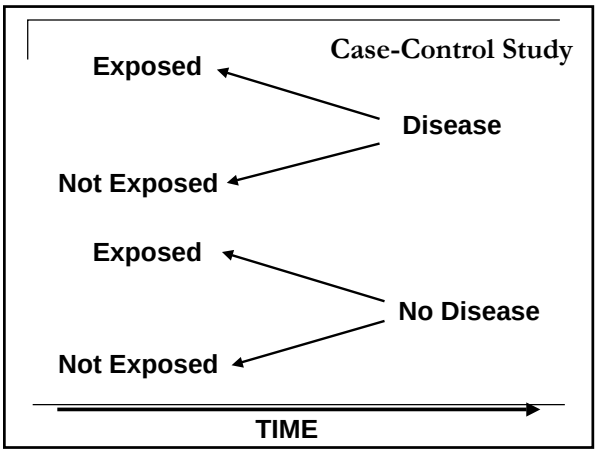
Example of confounding by indication

- Study of SSRI use and risk of suicide
  - Cases – teenagers who committed suicide in WA state during 2001-2002
  - Controls – random sample of teenagers from school rosters, frequency matched on county and age
  - Results
    - Unadjusted OR=2.0 or 2 times increase risk of suicide among SSRI users compared to non-users
    - Adjust results by depression and depression severity and OR=1.1

Case-control studies

Case-Control Studies

- Identify individuals with disease/outcome of interest and a comparable control group
- Look back in time to determine proportion of cases and controls that were exposed and non-exposed
- Examples of exposure: medication use, environmental factor, condition, and procedures
- Examples of cases: cancer cases, adverse event, diabetic patients, patients undergoing CABG



Selection of cases and controls

- Define source population
  - General population
    - Ex: social security recipients, 3 western WA counties
  - Hospital, clinic, business
- Eligibility criteria – often specific to study question and ensures "at risk"
  - Inclusion criteria
    - Ex: women 65-79 years residing in 1 of 3 WA counties
  - Exclusion criteria
    - Prior diagnosis of breast cancer

Identifying Cases

- Ideally, random sample of everyone who develops disease or all diseased in sample population
  - Example: all breast cancer cases diagnosed in 3 WA county during 1997-1999
- Type of cases
  - Incident (newly diagnosed) – generally preferred
  - Prevalent
- Sources
  - Registries
  - Hospital records
  - Clinics
  - Health care utilization data
  - Work place

Selection of Controls

- Should be selected from same population that gives rise to cases
- Similar to cases in all respects other than having disease
  - Patient characteristics, co-morbidities, etc.
  - Follow-up time
  - Ascertainment of exposure
- Selected independent of exposure
- Ex. sources
  - Population – preferred
    - School rosters, insurance lists, random digit dialing
  - Hospital or clinic based
  - Friend or sibling

Matching controls to cases

- To ensure that cases and controls are comparable with respect to major risk factors for disease that not interested in evaluating
  - May also need to match on follow-up time
- Individual matching
  - For each case, a control is selected who is similar in terms of matching characteristic (e.g., age)
- Frequency matching
  - Proportion of controls with matching characteristic is identical to proportion of cases with the same characteristic
- Match at ratio of 1:1 - 4:1
  - Increase statistical power
  - At times, may want to use more than 1 type of control group

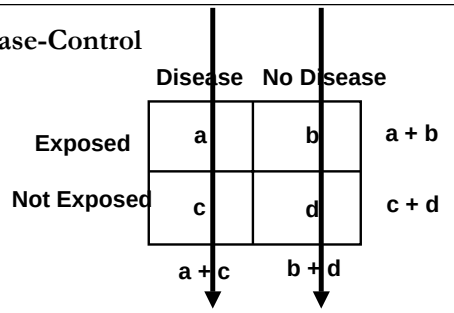
Measuring medication exposure in observational studies

- Self-report
  - Mailed questionnaire
  - Phone or in-person interview
- Medical records
- Automated health plan data
- Retail pharmacy records
- Medication inventory

Case-Control Studies

- Cannot provide direct estimate of incidence or prevalence of disease
- If incident cases are used, study provides information about relative incidence of disease in exposed vs. non-exposed
- If prevalence cases are used, study provides information about relative prevalence of disease in exposed vs. non-exposed

Case-Control



Odds Ratio = odds of exposure given you have the disease compared to the odds of exposure given you are disease-free

OR= a/c / b/d

Odds Ratio

	Cases	Controls
Exposed	75	100
Not exposed	125	500

OR = 75/125 / 100/500 = 3.0

Odds Ratio

- Ratio of exposure among cases to exposure among controls
- Answers: What are the odds that the case was exposed?
- Measure of the association between exposure and disease
- Varies from 0 to infinity
- OR=1: no association
- OR>1: exposure is a risk factor for disease; increases risk for disease
- OR<1: exposure decreases the risk for disease
- Example:
  - OR=3.0 can be interpreted as 3 fold increase in risk
  - OR=0.5 can be interpreted as 50% decrease in risk

When does odds ratio approximate relative risk?

- The disease is rare
- Exposure frequency among cases studied is representative of that among all cases in population
- Exposure frequency among controls studied is representative of that among all non-cases in population

	Disease	No Disease
Exposed	a	b
Not Exposed	c	d

Odds Ratio=a/c /b/d

Relative Risk =a/a+b /c/c+d

If disease is rare
-a will be small relative to b
-c will be small relative to d
= a/b / c/d = a/c / b/d

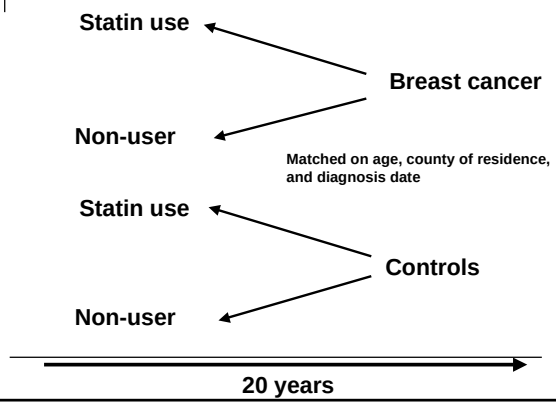
Case-Control Studies

- Pros:
  - Excellent for rare diseases
  - Look at the effect of multiple exposures in relation to disease
  - Usually inexpensive and fast
- Cons:
  - Can't directly estimate incidence or prevalence
  - Bias from inappropriate controls
  - Self-reported exposure data subject to recall bias
  - Only look at one disease

Example of case-control study

- Association between statin use and breast cancer risk
  - Cases – 975 women 65-79 years diagnosed with primary invasive breast cancer (SEER registry) in 1997- 1999 whose names appeared on list of social security recipients (CMS) & resided in 1 of 3 western WA counties
  - Controls – 1007 women whose names appeared on CMS list where matched to cases on age, diagnosis year, and county
  - Data collected by in-person interview
  - Logistic regression used to evaluate risk of breast cancer in statin users
  - Adjusted for matching variables
  - Adjusted for other confounders (ex: body mass index, HT use)
  - Long term statin users (>5 years) had reduced risk of breast cancer (OR=0.7) compared to non-users

*Cancer. 2004;100:2308-16.



Association between statin use and breast cancer (cont.)

- Limitations
  - Recall bias
    - Validated self-reported medication use with health plan data
    - Used pictures of medications & life calendar to improve recall
  - Missing data
  - Limited follow-up period
  - Residual confounding?
  - Generalizable to postmenopausal women in WA
- Strengths
  - Histologically confirmed breast cancer cases
  - High response rate
  - Data on many potential confounders

Example of case-control study

- Antibiotic use and breast cancer risk
  - Cases: 2266 primary invasive cases between 1993-2001 among GH enrollees
  - Controls: matched 3:1 on age and duration of enrollment for a total of 7,953
  - Data from automated health plan records
  - Cumulative days of antibiotic use associated with increased risk of breast cancer
    - Clear dose response

*JAMA 2004;291:827-35.

Antibiotic use and breast cancer risk (cont.)

- Limitations
  - Confounding by indication - no data on indication for use
  - Residual confounding and missing data
  - Conducted within one health plan in WA state
- Strengths
  - Stable population with complete data on all health care utilization (diagnoses, pharmacy, procedures, etc.)
  - Data on risk factors for breast cancer from breast cancer surveillance program
  - Complete capture of incident cases

Biases in case-control studies

- Selection bias – distort true association due to selection of subjects who either are not generalizable and/or who have unequal relationship between exposure and outcome
  - Ex: Cases that die soon after diagnosis are not represented and therefore study is only generalizable to survivors
  - Ex: Response rate varies by exposure
  - Ex. Hospital controls more likely to be sicker and more or less likely to be exposed to medications
  - Ex. Controls with disease rule out more or less likely to be exposed
    - OC use and thromboembolism

Biases in case-control studies (cont.)

- Information
  - Measurement error
    - Ex: Self reported weight underestimates true weight by average of 0.5kg
  - Differential misclassification
    - Ex: Cases recall exposure better than controls
    - Ex: Interviewers probe cases more than controls
  - Non-differential misclassification
- Confounding – distortion of association due to imbalance between exposed and non-exposed with regard to risk factor(s) for disease
  - Ex: Confounding by indication for a drug

Things to consider when reading case controls studies

- Are cases and controls similar with exception of disease?
  - Subject characteristics, data quality, follow-up
- Appropriate data collection or data source
- Appropriate time frame
- What are the potential biases?
- Potential confounders considered
- Generalizability of results
- Does study make sense and is enough information provided?

Inferring causal relations in observational studies

- Association is strong (risk estimate)
- Association makes biological sense
  - Plausible explanation
  - No plausible non-causal explanation
- Suspected cause precedes the presence of disease
- Magnitude of association is strongest when predicted to be so
  - Dose response

Measures of risk	Depends on:	Addresses:
Relative risk or risk ratio **Odds ratio** **Hazards ratio**	Influence of exposure (E) on relative incidence of disease (D) in E persons	Does E cause D?
Risk difference or attributable risk (AR) **= Incidence in E minus incidence in non-E**	Influence of exposure (E) on absolute disease (D) in exposed persons	How much of D is E responsible for? Should anything be done to modify E?
Population attributable risk (PAR) = Incidence in population minus incidence in non-E	Risk ratio Frequency of exposure Frequency of disease	What rate of D in population is caused by E? Should resources be allocated to control E relative to other E causing greater health problems?
Statistical significance or P value	Risk ratio Sample size Frequency of disease or exposure, depending on design	Could association between E & D have occurred by chance?

Attributable risk is also important

- RR is large but overall incidence of disease is low
  - Ex: RR=5.0; I_e=100/100,000 py; I_o= 20/100,000 py
    - AR = 80/100,000 py
- RR is modest but overall incidence of disease is high
  - Ex: RR=1.5; I_e=125/1000 py; I_o= 85/1000 py
    - AR=40/1000 py or 4000/100,000 py

Things to know..

- Different measures of disease occurrence
- Cohort and case-control studies
  - Design features
  - Selecting subjects
    - Exposed and non-exposed
    - Cases and controls
  - Strengths and limitations
  - Common biases
  - Be able to identify confounding and ways to deal with confounding
  - Interpretation of risk estimates (odds ratio and relative risk)