Observational Study Designs

Denise Boudreau, PhD
Center for Health Studies
Group Health Cooperative

Today

- What and why on observational studies
- Measures of disease occurrence
  - Prevalence
  - Incidence
- Cohort studies
  - Design
  - Risk estimate
  - Prospective versus retrospective
  - Strengths and limitations
  - Examples

- Epidemiology: Study of how health related states or events are distributed in a population and what factors influence or determine the distribution.
  - Examples of questions it aims to answer:
    - Why does disease develop in some people but not others – or what are risk factors that increase a person's risk for a disease
    - What is the natural history and prognosis of disease
    - How does new modes of prevention, treatment, or health care delivery impact health outcomes

- Pharmacoepidemiology: Study of the use of and effects of drugs in the population
  - Effects may refer to a variety of outcomes such as disease, adverse events, or health care utilization and costs
  - borrows its focus of inquiry from clinical pharmacology (i.e., effects of drugs in humans) and;
  - borrows methods from epidemiology

Study Designs

Experimental Clinical Trials
- Study outcomes after randomize exposure

Observational studies
- Ecologic
  - Compare group characteristics
- Cross-sectional
  - Study individuals at one point in time
- **Case-control**
  - **Study exposure by outcome**
- **Cohort**
  - **Study outcomes by exposure**

Efficacy
- Shows how intervention works in ideal conditions
- Generally healthy people
- Prevent drop-outs and non-compliance
- Less generalizable to other individuals outside study population

Effectiveness
- Shows how intervention or treatment works in practice
- Generally less healthy people
- Takes into account dropping out
- Observational

Study Designs

Experimental Clinical Trials
- Study outcomes after randomize exposure

Observational or epidemiologic studies
- Ecologic
  - Compare group characteristics
- Cross-sectional
  - Study individuals at one point in time
- **Case-control**
  - **Study exposure by outcome**
- **Cohort**
  - **Study outcomes by exposure**

Measures of Disease Occurrence

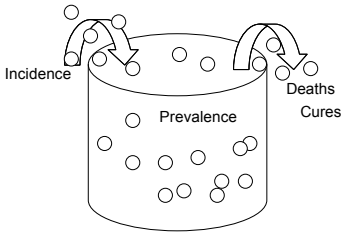
Measures of Disease Occurrence

- Prevalence = $\frac{\text{No. of cases of a disease in the population at a specified time}}{\text{Total population during same time}}$
- Proportion with no units
- Numerator includes new and ongoing cases
- Represents a cross-sectional "snapshot" of the population that estimates the burden of disease
- Does not estimate risk of developing disease

Examples of prevalence

- HT use before and after WHI results in 5 health plans*
  - 24,682 of 169,586 women were using HT in September, 1999
    - Prevalence = 14.6%
  - 11,825 of 149,607 women were using HT in December, 2002
    - Prevalence = 7.9%
- NSAID use is 10-15% in persons 65+ years
- 6-10% of primary care patients suffer from major depression**

*Obstet Gynecol 2004;104:1042-50.
**Psychiatry, 1992, 14(4): 237-47



Measures of Disease Occurrence

- Cumulative Incidence = $\frac{\text{No. of new cases of disease during a period of time}}{\text{No. persons at risk of developing the disease during same time period}}$
- Proportion with no units
- Probability of developing disease
- Measure of risk
- Can be measured only in closed population
- Assumes all subjects followed until develop disease or observation period ends

Examples of cumulative incidence

- Among 21,011 women continuously enrolled in GHC and undergoing at least two mammography screens during 1998 – 2002, 2,258 have positive 2nd screen
  - CI=2,258 / 21,011 or recall rate of 10.8%
- Among all LBW babies born in a Boston hospital during 2004, the proportion who develop pneumonia 6-weeks after birth

Who is "at risk"?

Persons are at risk if they do not have the disease of interest and are capable of developing the disease

Examples:

- Study of statin use and ovarian cancer risk
  - Exclude women with prior oophorectomy
- Study of prednisone use and flu risk
  - Exclude vaccinated subjects
- Study of SSRI use and breast cancer recurrence
  - Include women with prior diagnosis of breast cancer

Measures of disease occurrence

- Incidence = rate $\frac{\text{No. of new cases of disease during a period of time}}{\text{Person-time of observation among persons at risk during same time period}}$
- Average rate at which disease develops in a population
- Actual rate with units of time^{-1}
- Accounts for differing rates of follow-up so don't need closed population
- Also referred to as incidence density, hazard rate, and mortality rate

Person-time

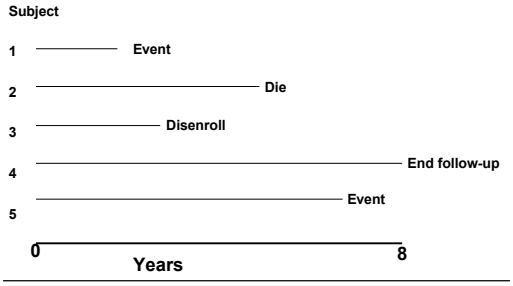
= sum, over all individuals, of time at risk until the date of the event of interest or date of censoring (i.e., death, end of follow-up, disenrollment from health plan, drop-out)

Example: 8 year follow-up study

Subject	Years Follow-up	Event	Died	Disenroll
1	2.0	1	0	0
2	5.2	0	1	0
3	3.5	0	0	1
4	8.0	0	0	0
5	5.9	1	0	0

Person time = 24.6 yrs

Person-time



Examples of incidence rate

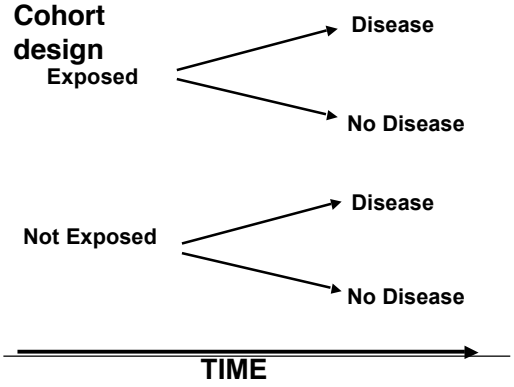
- Incidence rate = 2 events / 24.6 person-years
  - 0.08 per py
  - =80 per 1000 py
- Incidence rate of stroke is 6.4 per 1000 py among MI patients using statins & 11.1 per 1000 py among MI patients untreated

Ann Pharmacother 2002;36:751-7 .

Cohort study

- Exposed and non-exposed individuals are followed over time to determine whether they experience the outcome of interest
- Examples of exposure: medication use, environmental factor, condition, procedure
- Examples of outcome: disease, death, health care utilization, costs

Cohort design



Cohort

	Disease	No Disease	
Exposed	a	b	a + b
Not Exposed	c	d	c + d
	a + c	b + d	

Relative risk = incidence of disease in exposed Compared to the incidence of disease in unexposed
= (a/a+b) / (c/c+d)

Relative risk (Risk ratio)

- Ratio of disease incidence among exposed to disease incidence among non-exposed
- Quantifies magnitude of the association between exposure and disease
- Varies from 0 to infinity
- RR=1: no association
- RR>1: exposure is a risk factor for disease; increases risk for disease
- RR<1: exposure decreases the risk for disease
- Example:
  - RR=2.0 can be interpreted as two fold increase in risk
  - RR=0.7 can be interpreted as 30% decrease in risk

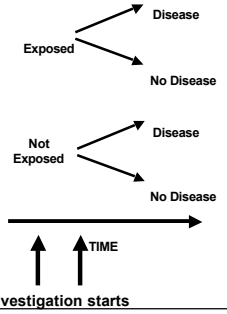
Cohort studies

- Aka: longitudinal study, follow-up study, observational study
- Disease free subjects chosen on exposure
  - Unexposed group should be comparable to exposed population except without exposure
  - Information obtained should be comparable for exposed and unexposed populations
- Types of cohort studies
  1. Prospective
  2. Retrospective (historical cohort study)

Prospective cohort study

Example:

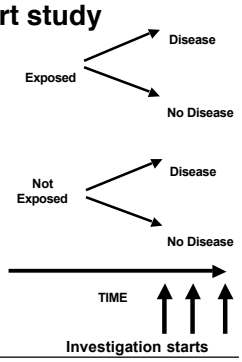
- Nurses Health Study
  - Assemble cohort in 1976 and follow through 1994
  - Prospectively collect data via mailed questionnaire during study period



Retrospective cohort study

Example:

- Klungel et al. article
  - Begin study in 2000 using data already collected via health plan
  - Identify cohort surviving MI b/n 1986-1996 and assess for lipid lowering therapy use and events during 6-months following MI



Ann Pharmacother 2002;36:751-7.

Cohort Studies

- Study outcomes by exposure
- Subjects are disease free at start
- Pros:
  - Can often show temporality of relationship
  - Less bias due to prospective evaluation of exposures
  - Can evaluate multiple diseases
- Cons:
  - Can often span many years
  - Need motivated cohort of people who will be repeatedly evaluated
  - Analysis can become complex
  - $$$$

Prospective Cohort

Limitations

- Loss to follow-up
- Misclassification of disease or exposure status
- If large number of subjects is required or long follow-up = $$$ or logistically challenging – especially for prospective design
- Hard to study rare diseases
- Changes over time in staff/methods
- Little control over nature and quality of data in retrospective designs

Strengths

- Can establish time order
- Can obtain incidence rates
- Can study more than one disease or outcome
- Minimizes bias in ascertainment of exposure status and covariates – especially if collecting data prospectively
- Efficient for rare exposures
- No controls, so no bias in control selection

Prospective vs. Retrospective

- Cost
- Latency of disease
- Loss to follow-up
- Availability and quality of data
- Importance of scientific question

Bias?

- Deviation of results or inferences from “truth”
- Systemic error in the design, conduct, or analysis of a study that results in a mistaken estimate of the association between an exposure and outcome
- Major issue in any epidemiologic study design

Types of bias

- Selection bias
  - Ex: Women with family history of breast cancer more likely to participate in study of breast cancer
- Surveillance bias – disease ascertainment differs in the monitored population
  - Ex: OC use and thrombophlebitis
- Information bias
  - Ex: Recall bias, interviewer bias, non-response bias
- Misclassification
  - Incomplete disease capture
- Confounding
  - Ex: Impact of HT on breast density differs in younger versus older women

Biases in Cohort Studies

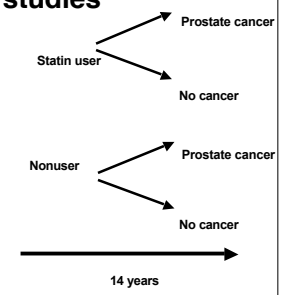
- Selection bias - less of a problem than case-control studies
- Information bias/misclassification
  - Degree of accuracy of classification of exposure, confounders and disease status
- Loss to follow-up (affects validity)
- Nonresponse (limits generalizability, not validity)
- Confounding

Analysis in cohort studies

- Cox Proportional Hazards (survival analysis)
- Logistic or linear regression
- Poisson regression
- Measure time from exposure to outcome
- Can look at time-dependent exposures and covariates

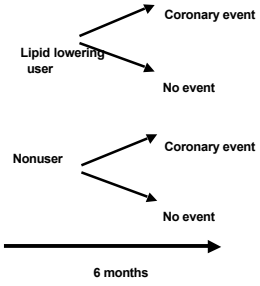
Example of cohort studies

- The association between statin use and prostate cancer risk
  - All men 45+ years enrolled in GHC for at least 2 years during 1990-2004
  - Exposure is statin use, which may change over 14 years
  - Follow 14 years until develop prostate cancer, die, or disenroll from GHC
  - Each subject will contribute person-time to follow-up
  - Survival analysis to account for time varying exposure, adjust for other risk factors, & account for censoring



Examples of cohort studies

- Lipid lowering use and cardiovascular events after MI
  - GHC enrollees surviving MI b/n 1986-1996
  - Exposure is lipid lowering use during 6-months after MI
  - Follow 6-months until event, die, or disenroll from GHC
  - Each subject will contribute person-time to follow-up
  - Survival analysis to account for time varying exposure, adjust for other risk factors, & account for censoring



Ann Pharmacother 2002;36:751-7.

Things to consider when reading cohort studies

- Appropriate population
- Are exposed and non-exposed similar with exception of exposure
  - Subject characteristics, data quality, follow-up
- Appropriate data collection
- Appropriate follow-up
- What are the potential biases
- Were potential confounders considered
- Generalizability of results
- Does study make sense and is enough information provided