Epidemiologic Study Designs

Dan Wartenberg PhD
Chief, Division of Environmental Epidemiology
Department of Environmental and Occupational Medicine
UMDNJ—Robert Wood Johnson Medical School
Piscataway, New Jersey 08854
USA

Thai Fogarty ITREOH Center
Chulalongkorn – Rutgers - UMDNJ

Overview

- Classic Study Designs
  - Cohort
  - Case Control
  - Cross-Sectional
  - Ecologic
  - Randomized Clinical Trials
- Newer Study Designs (later lecture)
  - Case-Crossover (self-matched case-control)
  - Time Series (ecologic)

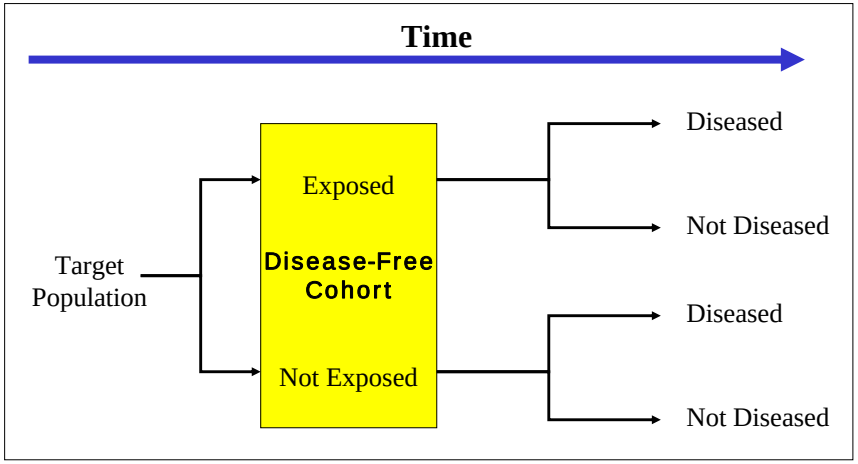
The Cohort Study

- The “What will happen to me?” study
  - Select exposed and unexposed healthy subjects
  - Follow and compare health status
    - example: Are dry cleaners more likely to get kidney cancer?
  - Good for rare exposures/risk factors (<5%)
  - Limitations
    - outcome misclassification
    - loss to follow up

Why Do A Cohort Study?

- Examine common diseases in the general population
  - Causes of death, cancer incidence, births, diabetes,…
- Study a rare exposure (risk) of interest, unusually high levels
  - workers, local pollution,
- Generalize association to the other populations
  - Flight attendants, professional sports, representative sample
- Study special resource groups
  - Doctors, nurses, alumni
- Compare geographical groups
  - Three mile island, specialty care hospitals, NJ

Single Sample Cohort Study Design

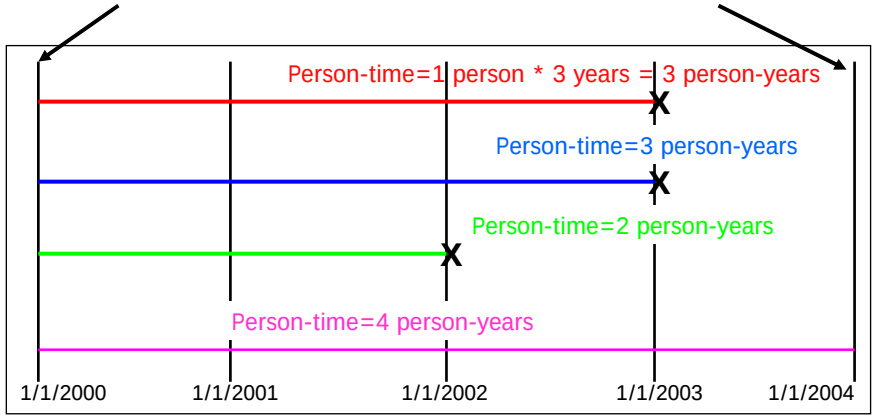


Incidence rate = Number of events that occur from t0 to t1 / Total person-time of observation

Person-time = Sum of total time spent in study for all study subjects

t0 = start of follow-up

t1 = end of follow-up



Determining Exposures/Risk Factors

- Questionnaires/Interviews
- Laboratory tests
- Physical measurements
- Special procedures
- Medical records
- Environmental measurements

Measuring Disease

- Acquiring the Data
  - Routinely collected records
    - Workplace
    - Registries/Surveillance
      - Hospital discharge data
      - National Death Index
      - Special collections: United States Renal Data System
  - Study specific examinations
  - Questionnaires
- Diagnosing Disease
  - Clear, consistent and workable definitions

Most Common Outcome Measures

- Rate Ratio (relative risk)
  - $IR_{Exposed}$ / $IR_{Unexposed}$
    - How many times worse is the rate among exposed than the background (unexposed) rate
- Rate Difference
  - $IR_{Exposed} - IR_{Unexposed}$
    - Size of the (exposed) rate above background (unexposed) rate

Interpreting Rate Ratios or Differences

- RR>1 or RD>0
  - Risk for disease is higher in exposed than in unexposed (or than expected)
  - Risk factor increases risk of disease
- RR=1 or RD=0
  - Risk for disease is equal in exposed and unexposed
  - Exposure is not a risk factor for this disease
- RR<1 or RD <0
  - Risk for disease is lower in exposed than unexposed
  - Risk factor reduces risk of disease (is protective)

The Framingham Study

- Important study of CVD begun in 1948
  - Suburb of Boston; population ~30,000
- Subjects: residents 30-62 years of age

	Men	Women	Total
– Random sample	3,074	3,433	6,507
– Respondents	2,024	2,445	4,469
– Volunteers	312	428	740
– Respondents free of CVD	1,975	2,418	4,393
– Volunteers free of CVD	307	427	734
– Total free of CVD	2,282	2,845	5,127

  - THE FRAMINGHAM STUDY GROUP
- Physical exam every 2 years, questionnaires, and hospital surveillance

Goals of Framingham Study

- Study designed to test:
  - Is an increase of CHD associated with age
    - Does CHD occur earlier and more frequently in males
  - Do persons with hypertension develop CHD at a greater rate than those who are normotensive
  - Is elevated blood cholesterol level is associated with an increased risk of CHD
  - Are tobacco smoking and habitual use of alcohol associated with an increased risk of CHD
  - Is increased physical activity associated with a decrease in development of CHD
  - Does an increase in body weight predispose to CHD
  - Does Diabetes mellitus increase rate of development of CHD

Framingham Exposure Examinations

- Smoking
- Alcohol Use
- Obesity
- Blood Pressure
- Blood Cholesterol
- Level of Physical Activity

- NOTE THAT EXPOSURES LIKELY VARIED CONSIDERABLY AMONG SUBJECTS

- If not, need to identify separate comparison group with lower exposures

Other Examples of Cohort Studies

- Administrative Data Collection
  - Vital records
  - Hospital records
  - Migrant Studies
- Active Data Collection
  - NHANES/HHANES
  - The British Doctors' Study
  - The Nurses Study
  - American Cancer Society CPS I, II
  - Worker cohorts
    - Steel workers
    - Nuclear workers
    - Atomic bomb survivors
    - College alumni

Cohort Study: Strengths and Weaknesses

- Weaknesses
  - Not good for rare disease
  - Misclassification of outcome
  - Information bias
  - Non-participation/non-response
  - Loss to follow up
  - Expensive, slow, logistically difficult
- Strengths
  - Exposure precedes disease
  - Can study rare exposures, multiple outcomes
  - Provide incidence rates as well as rate ratios

Types of Cohort Studies

- Prospective
  - Enroll/identify group of disease-free people
  - Collect exposure data prior to disease occurrence
  - Follow forward in time, noting both exposures and disease occurrence
- Historical (sometimes called retrospective)
  - Identify group of people who were disease-free at a prior, specified date
  - Determine who has disease and who does not
  - Then collect exposure/risk factor data

Example

- Mortality Data from British Physicians Study (Doll and Hill)

Cause of Death	Death Rate Among Non-smokers	Death Rate Among Smokers	Relative Risk	Attributable Risk	Attributable Risk Percent
Lung Cancer	0.07	2.27	32.4	2.20	96.9%
Other Cancers	1.91	2.59	1.4	0.68	26.6%
Chronic Bronchitis	0.05	1.06	21.2	1.01	95.3%
CVD	7.32	9.93	1.4	2.61	26.3%
All Causes	12.06	19.67	1.6	7.61	38.7%

Death rates per 1000 people; data from Doll and Hill

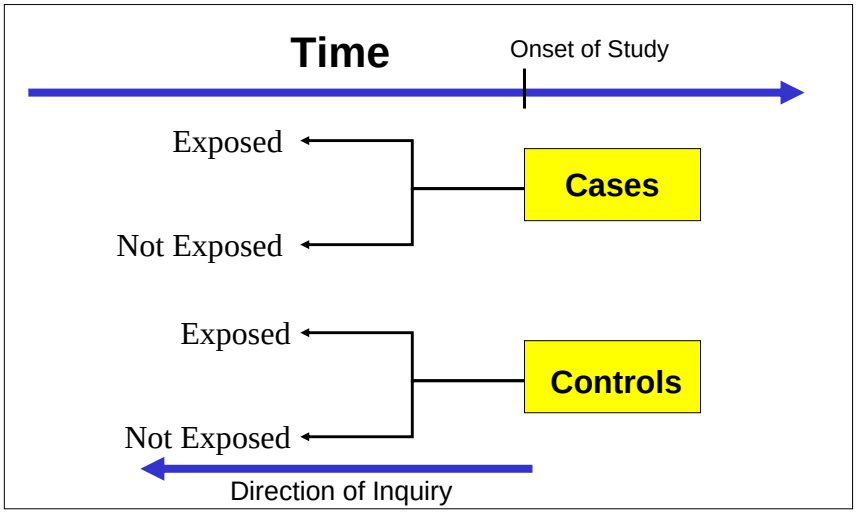
The Case Control Study

- The "Why me?" study
  - Select cases (ill) and controls (healthy) without knowledge of exposure/risks
  - Identify and compare various prior exposure/risks
    - Example: Are people who die from lung cancer more likely to have been smokers?
  - Good for rare diseases (<5%)
  - Limitations
    - exposure misclassification
    - possible control selection bias
    - no temporality

Why Do A Case Control Study?

- Examine a common exposures/risks of interest
  - Personal habits (e.g., smoking), occupation, environment
- Study a rare disease of interest
  - Childhood leukemia, ALS
- Generalize associations to the other populations
  - Power lines (emf), diet
- Why not study in a cohort design?
  - Can complete study more quickly
  - Is less expensive
  - Is easier to implement

Case Control Study Design



Selecting Cases

- Establish case definition
  - symptoms, disease code
- Selected cases should be representative of all cases
  - Typically use a sample
- Incidence cases are preferable to prevalent cases
  - Less likelihood of recall bias
  - Eliminates concern of duration of disease bias

Selecting Controls

- From similar/same population as cases but free of disease
  - (e.g., registry, hospital, lists)
- Sampled to represent exposure/risk distribution in population from which cases have been drawn
- Subjects chosen independently of exposure
- Can match on some variables to remove their influence
  - Cannot assess role of variables on which one matches

Determining Risk Factors/Exposures

- Questionnaires/Interviews
- Medical records
- Other historical records
  - environmental measurements
  - residential histories
  - occupational records

Measuring Exposures

- Acquiring the Data
  - Routinely collected records
    - Workplace monitoring
    - Environmental Surveillance
      - Ambient air quality
      - Drinking water contaminants
    - Study specific examinations
      - Power line location and load
  - Personal interviews/questionnaires

Most Common Outcome Measure

- Odds Ratio (estimate of relative risk)
  - Exposure $Odds_{CASES}$ / Exposure $Odds_{CONTROLS}$
  - How many times worse is exposure rate among cases than exposure rate among controls (background)

Interpreting Odds Ratios

- OR>1
  - Odds (risk) of exposure is higher in cases than in controls (or than expected)
  - Exposure increases risk of disease
- OR=1
  - Odds (risk) for exposure is equal in cases and controls
  - Exposure is not a risk factor for this disease
- OR<1
  - Odds (risk) for exposure is lower in cases than in controls
  - Exposure reduces risk of disease (is protective)

Case-Control Studies: Strengths and Weaknesses

- Weaknesses
  - exposure determination follows disease (temporality)
  - possible control selection, recall, and other biases
  - Inadequate information on confounding factors
  - Possible exposure misclassification
  - no rate information
  - study only one outcome (disease)
- Strengths
  - relatively cheap, easy, fast
  - good for rare disease (<5%)
    - fewer subjects needed
  - good for long latency diseases
  - can look at many exposures (causes)

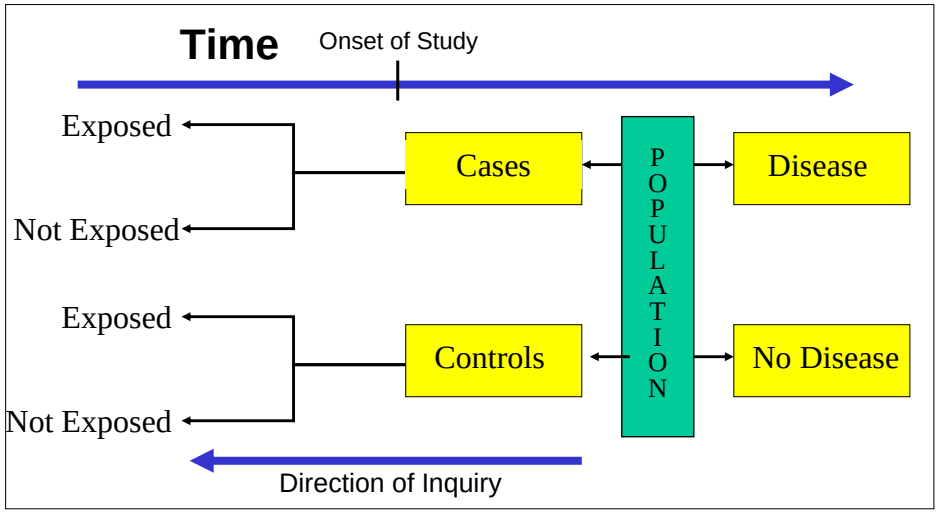
Cohort/Case-Control Comparison

	Cohort	Case Control
Event	Disease	Exposure
Population	Exposed	Diseased
Measure	Disease Rates	Exposure Rates
Advantages	temporality	cheap, quick
	multiple outcomes	multiple exposures
good for:	rare exposures	rare outcomes
	rate information	typically no rate info
Disadvantages	large, slow, costly	lack of temporality
	few exposures	few diseases

Nested Case Control Study

- Instead of Cohort, do Case Control Study within a Cohort
- Why
  - Too difficult/expensive to sample all (cases and) controls
- How
  - Identify cohort for study
  - Follow up outcome of interest
  - Select cases and controls from cohort
    - matched on time and other variables
    - i.e., risk of being a control is proportional to time in cohort—density sampling (via risk set sampling)
  - Assess exposure, analyze interpret
- More efficient than cohort in that only a sample of controls assessed for exposures and confounders

Case Control Study Design



Most Common Outcome Measure

- Odds Ratio (estimate of relative risk)
  - Exposure $Odds_{CASES}$ / Exposure $Odds_{CONTROLS}$
  - How many times worse is exposure rate among cases than exposure rate among controls (background)

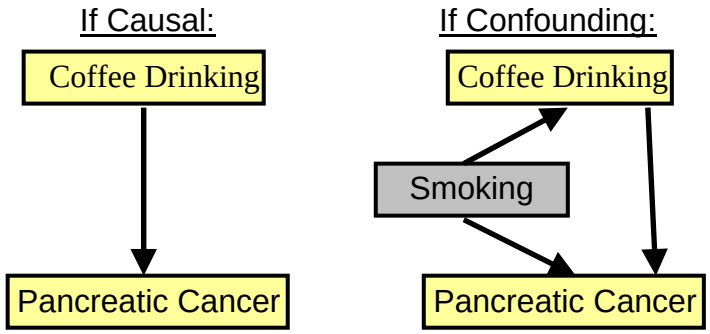
Basic Epidemiologic Questions

- What is the question under study?
- How were the study subjects selected?
- Did most subjects participate?
- How were risk factors assessed?
- Did exposure/risk precede disease?
- What was the size of the observed effect?
- What is the width of the 95% confidence interval?
- Was there adjustment for more than 1 risk factor?
- Are results similar to other studies?

Case-Control: Pancreatic Cancer

- Study by MacMahon et al. 1981
  - N Engl J Med 304: 630-633
- Chair of Epid at Harvard; Top Medical Journal
- Case control study of pancreatic cancer
  - Hypothesis: smoking and alcohol are risk factors
  - Cases diagnosed in any of 11 New England Hospitals
  - Controls other inpatients hospitalized by same doctors who hospitalized cases
  - Collected data on smoking, alcohol, coffee, tea
  - Found association with smoking and coffee
  - Possible explanations for coffee finding
    - Multiple comparison
    - Control selection bias
    - Confounding

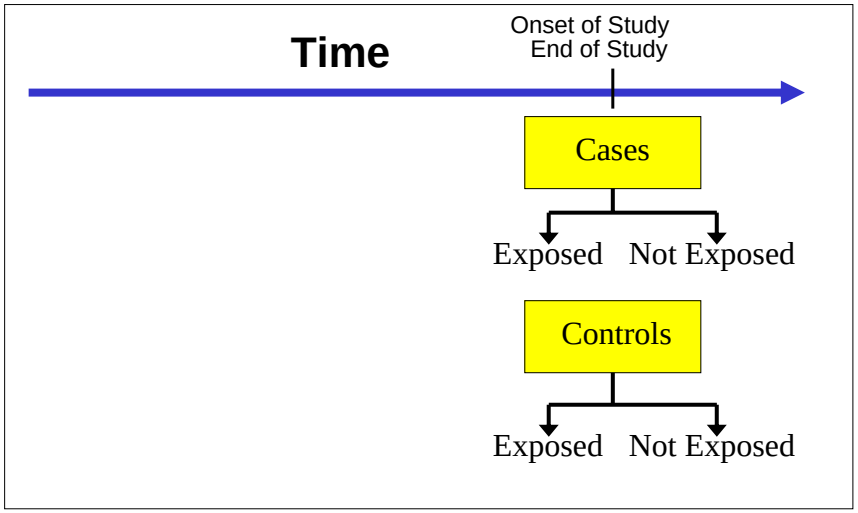
Coffee and Pancreatic Cancer
An Alternative Explanation



The Cross-Sectional Study

- The "Am I like my neighbors?" study
  - Compare health and risks simultaneously—"A Snapshot"
    - example: do vegetarians have lower blood pressure
    - Example: is high serum cholesterol associated with CHD
  - Quick, easy, inexpensive
  - Population-based
  - Often assess prevalence
  - Useful for health service planning
  - Sometimes is used in place of longitudinal data because recall bias is less likely (e.g., diet)
  - Is useful for things that do not change (e.g., blood type)

Cross-Sectional Study Design



Cross-Sectional Study: Strengths and Weaknesses

- Weaknesses
  - Does not address temporality
  - Does not allow for latency
  - Limited to prevalence evaluation—Cannot measure disease incidence
  - Selects for longer lasting diseases (i.e., Neyman bias)
- Strengths
  - quick, easy, inexpensive
  - population-based

The Ecologic Study

- The “Is my town like your town?” study
- Used Aggregate Data
- Group rather than individual is unit of analysis
- Typically cross-sectional (but can be other)
- Usually difficult to control for confounding
- Unable to control for within group variation
- Example: There are high levels of both toxic pollution and cancer in New Jersey, so the toxics must be causing the cancer.
- Example: International comparison of fat in diet and incidence rate of breast cancer
- **Heavily Criticized!!**

A Fully Ecologic Study Example

All data are aggregate (not individual)

Example: Data for Each Study Unit (e.g., town)

individual (*interior*)

	E	NoE	total
D	**4**	**4**	8
NoD	**6**	**16**	20
	10	20	30

OR=(4*16)/(4*6)=2.67

ecologic (*margins*)

	E	NoE	total
D	?	?	**8**
NoD	?	?	**20**
	10	**20**	**30**

D=8/30=27%; E=10/30=33%

Don’t know if those with disease were exposed

Partially Ecologic Study

- Similar to Fully Ecologic but:
  - Have individual outcome data
    - e.g., death certificates, cancer incidence
  - Have aggregate exposure data
    - e.g., regional air pollution monitors
- Most critics do not differentiate from fully ecologic

Strengths of Aggregate Analysis

- Enables analysis of large populations
  - Not easily collectable
  - Facilitates study of relatively small risks
  - Can assess public health impact of an intervention
  - Can be conducted easily and inexpensively with routinely collected databases (surveillance)
- *Statistical inference is valid in spite of biased estimates*
  - Useful for hypothesis generation and prioritization
  - Aggregate sampling variance biases results towards null

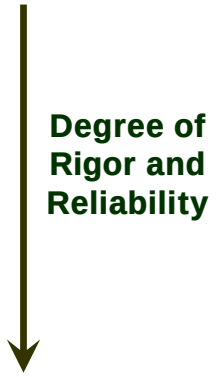
Thompson WD, Wartenberg D. 2007. Additive versus multiplicative models in ecologic regression. Stoch Environ Res Risk Assess (2007) 21:635–646.

The Ecologic Study: Strengths and Weaknesses

- Weaknesses
  - Uses aggregate data (don't know joint distributions)--*Ecologic Fallacy*
  - *Partially Ecologic Studies are better*
  - Data registries are not as reliable as individually collected data
  - Usually difficult to control confounding
  - Unable to control (or assess) within group variation
- Strengths
  - Allows studies of large populations
  - Utilize existing databases
  - Allows study of relatively small risks

Epidemiologic Study Designs

- Case Series (not discussed)
- Cross-Sectional
- Ecologic
- Case-Control
- Cohort
- Randomized Clinical Trial
  - Intervention (more general)



Randomized Clinical/Controlled Trials

- Often used to evaluate treatments
- Experimental study
  - Subjects divided into equivalent groups
    - If randomized, assume other factors equivalent across groups
  - Each group subjected to different treatment
  - Investigator controls intervention
  - Ideally, study is double blinded
  - Subjects followed through time, with outcomes monitored and counted
- Considered the "gold standard" in epidemiology

  (gold standard means best, that against which others are judged)

Example: The Physicians Health Study

- Randomized trial
  - 22,000 US male physicians ages 40-84
    - Aspirin—reduction of cardiovascular mortality
    - Beta-carotene—decrease cancer incidence
  - Large N; moderate risk; good responses
  - Knowledgeable enough to see side-effects
  - Concern about healthy volunteer effect
  - Early results
    - Jan. 1988—daily aspirin reduces risk of MI by 44% (RR=0.56, 95% CI 0.45-0.70)

Critical Issues in an RCT

- What is the hypothesis being studied?
- How was the study population selected?
- Is comparison group explicitly identified?
- Are patients allocated to treatment and control groups without bias?
- How is Intervention Administered?
- Is the Outcome Assessed without Regard to the Treatment Status?

Ethics of RCTs

- Must give people best known treatment
  - cannot withhold proven treatment
- Cannot test an adverse treatment
- Risks must be minimized
- Must be well designed
- Must obtain patients' written consent

Ethics of RCTs—2

- IRBs monitor patients' rights
  - is study scientifically sound
  - does patient understand risks
- Early stopping rules
  - is ethical to withhold beneficial treatment?
  - Is ethical to terminate a study prior to having conclusive scientific data?
    - e.g., AZT studies of AIDS

Elements of Informed Consent

- Research nature of project
- Explanation of procedures
- Explanation of risks and benefits
- Explanation of alternatives
- Explanation about confidentiality
- Name to contact with questions
- Participation voluntary
- Right to withdraw at any time
- Availability and cost of any care necessitated by any complications

Assessing RCTs

- Advantages
  - experimental
  - groups are treated equally
  - no selection bias
  - no confounding
- Disadvantages
  - randomization not perfect
  - blinding is difficult
  - need large sample
  - expensive
  - ethically difficult
  - logistically difficult