Issues in Epidemiologic Design and Analysis

Dan Wartenberg, PhD
David Ozonoff, MD, MPH

Problems in Conducting Studies

- Precision (lack of random error)
  - Depends on study size and efficiency
    - like coin flips; false positives; false negatives
- Validity (lack of systematic error)
  - Internal
    - Bias--playing favorites
      - selection, information, recall
    - Confounding--finding for the wrong reason
      - factor associated with disease and exposure
    - Misclassification
      - mislabeling, bad model, missing information
  - External
    - Generalizability

Internal Validity

- Bias is a distortion of study results
  - Can occur in any study design
  - Is an inherent flaw—cannot be fixed
- Some types of bias:
  - Selection, Information, Participation, Recall, Confounding

Selection Bias

- Selection—enrolling subjects
  - Differences between groups of subjects relative to that in the larger population—picking favorites
    - Does exposure status affect enrollment in a case control study?
    - Does disease status affect enrollment in a cohort study
      - Example: Hawthorne Effect
      - Example: Caffeine and Pancreatic Cancer
      - Example: Healthy Worker (Survival) Effect

Information Bias

- Information—measurement errors
  - Differential or non-differential
  - Disease and risk factors

Other Types of Bias

- Participation, Response, Loss to follow up
- Recall—memory trigger
- Berkson's Bias—differences of a hospital population
- CONFOUNDING

Confounding—1

- Mixing of effects (Rothman)
  - estimate of effect of exposure is distorted
  - mixed with the effect of an extraneous factor
  - Example
    - Add fluoride to drinking water
    - Implement dental hygiene education program
    - Dental caries decline
    - Which caused the observed effect?
  - Example: cigar smoking and baldness

Confounding—2

- Requirements for Confounding
  - Extraneous factor
    - must be predictive of disease (e.g., age, SES)
    - must be associated with exposure among cases
    - must NOT be intermediate step in causal path between exposure and disease

Confounding—3

- Confounding can occur even if there is no effect of exposure
  - Example: alcohol consumption and oral cancer
    - Association observed
    - Smoking is extraneous factor
      - Affects oral cancer cancer rate in non-drinkers
      - Is associated with alcohol drinking
    - More smokers among alcohol drinkers
    - The effect of alcohol is distorted by smoking
    - Size of effect depends on
      - Size of smoking effect
      - Strength of association between smoking and drinking

Confounding—4

- Confounding is a bias
  - prevent by design
  - remove (control) through analysis

Prevention of Confounding

- Randomization--experimental studies
- Restriction--limit subject population to those with in specified category(s) of extraneous factor(s)
- Matching--each pair has same value for extraneous factor
  - Expensive
  - Requires specialized analytic methods
  - Concern: overmatching

Matching: What is it?

- Selection of comparison or reference series that is identical (or nearly so) to the index series with respect to the distribution of one or more potentially confounding variables.
- Matching improves efficiency, not validity

Types of Matching

- Frequency Matching
  - define strata
  - estimate number of cases in each
  - select appropriate number of controls for each
- Individual Matching
  - match each case with one or more controls
  - Used historically
  - Has some methodologic problems

Matching: When to Use

- exposure disease association weak
- exposure rare
- only a few variables to match on
  - Otherwise cumbersome and expensive

Matching: Summary

- Select specified comparison subjects
- Advantages
  - gain in precision (more balanced design)
- Disadvantages
  - introduce confounding
  - limit analytic options
  - cannot assess effect of "matched" variable
  - more difficult and costly to implement

Control of Confounding

- Stratification
  - group data into (homogeneous) categories of extraneous factor
  - analyze for each category
  - combine for summary estimate
- Multivariate analysis
  - adjust through statistical modeling

An Example of Confounding

	Age <40		**Age 40-44**		**Totals**	
	User	**Non User**	**User**	**Non User**	**User**	**Non User**
Cases	**21**	**26**	**18**	**88**	**39**	**114**
Controls	**17**	**59**	**7**	**95**	**24**	**154**
OR	**2.8**		**2.8**		**2.2**	

Limitations of Stratification

- As number of confounders increase, the size of each stratum gets very small
- Difficult to discern confounding from effect modification with large number of variables

Effect Modification (Interaction)

- Change in the magnitude of an effect measure according to the value of an extraneous factor (i.e., heterogeneity)
- EM is a characteristic to be reported rather than a bias to be avoided
- Includes both synergy and antagonism Example: smoking (5x), asbestos (10x) and lung cancer

Data with Effect Modification

		Cases	Controls
Males	>1/2 pack	415	342
	<1/2 pack	232	280
Females	>1/2 pack	20	7
	<1/2 pack	21	21

From Doll and Hill 1950 BMJ

$\chi^2_{Wald} = 1.5$

$OR_{males} = 1.5\ (1.2,1.8)$

$OR_{females} = 2.9\ (1.0,8.2)$

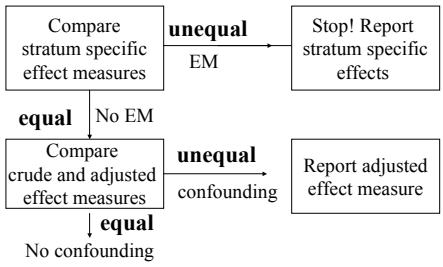
$OR_{MH} = 1.5\ (1.2,1.9)$

$OR_{C} = 1.5\ (1.2,1.9)$

Confounding vs. Effect Modification

- Either, both or none may be present
- Confounding
  - depends on distribution of factor among strata
  - a nuisance effect to be adjusted for
- Effect Modification
  - effect differs in size/direction among strata
  - an inherent feature of the strata to be described
- Effect Modification *supercedes* Confounding

EM/Confounding Flow Chart



Misclassification

- Differential vs. Non-differential
- Examples
  - Mislabeling
  - Bad model
  - Missing information

External Validity

- Generalizability
  - Issues
    - Representativeness of subjects
    - Subjects with appropriate characteristics
    - Extrapolation
    - Development of "universal" hypothesis

When is screening appropriate?

- If condition found, is treatment effective?
  - Efficacy, patient compliance, early treatment
- How great is the burden of suffering?
  - Death, disease, disability, discomfort, dissatisfaction, destitution
- How accurate is the screening test?
  - Sensitivity, specificity, simplicity, cost, safety (risks), acceptability, labeling effects
- How common is the disease?