Strengths and Weaknesses of Different Study Designs

Jo Freudenheim, PhD
Department of Social and Preventive Medicine
Disclosure Statement

- Jo Freudenheim, PhD
  - Honoraria: none
  - Consultancies: NCI
  - Common stock: none
  - Research: NIH, DOD, Avon
Goals and objectives

- Overall goal:
  Identification of strengths and weaknesses of study designs
  - External and internal validity
  - Confounding/Interaction
  - Bias
  - Incorporating different study designs
- External and internal validity
- Confounding/Interaction
- Bias
- Incorporating different study designs
Study Validity

- Our ability to make inferences based on the study findings
Study Validity: external validity

- The generalizability of the study findings
- To whom do the study findings apply
Study Validity: external validity

- Example: case control study of breast cancer among women aged 35-70 in Western New York
- How generalizable are those findings?
Study Validity: external validity

- Example: cohort study among postmenopausal women in the Women’s Health Initiative
- How generalizable are those findings?
Study Validity: internal validity

- Are the methods of the study appropriate, no major concerns with methodology
Study Validity: internal validity

- Are the methods of the study appropriate, no major concerns with methodology
  - Confounding/Interaction
  - Bias
- External and internal validity
- Confounding/Interaction
- Bias
- Incorporating different study designs
Causal Association: Direct

- The factor under study causes the disease
- The observed association is causal
- Ex: HPV causes cervical cancer
Causal Relations

- Indirect

Diagram:

- Factor A → Step 1 → Disease
Causal relations

- Factor is in the causal pathway
  - Not a confounder
  - Is another step on the path from exposures of interest to outcome
Types of Causal Relations

- Necessary and sufficient (Occurs rarely)
Types of Causal Relations

- Necessary but not sufficient

A + B + C + \ldots \rightarrow \text{Disease}

(Causal chain may also involve a specific temporal sequence)
(e.g. carcinogenesis)
Types of Causal Relations

- Sufficient but not necessary (also rare)

  - Factor A
    - or
    - Factor B
      - or
      - Factor C
        - or
        - Factor D

  - Disease
Types of Causal Relations

- Neither necessary nor sufficient (probably true for most of the diseases we study)

- Factor A + Factor B
- Factor C + Factor D
- Factor E + Factor F
- Factor G + Factor H

Disease
Non-Causal Association: Confounding

- The factor under study does not cause the disease even though it is observed to be associated with the disease.
- There is another factor that is correlated with both the factor under study and with the disease which is the true causal factor.
Non-Causal Association: Confounding

- People with yellow staining on their fingers are more likely to get lung cancer
- Does yellow stain cause lung cancer?
Non-Causal Association: Confounding

- It is reported that people who drink more coffee are more likely to get pancreatic cancer.
- Does coffee cause pancreatic cancer?
Non-Causal Association: Confounding

- It is reported that women with a history of herpes infection are more likely to get cervical cancer
- Does herpes cause cervical cancer?
Confounding

- Smokers have “yellow” finger
- Coffee drinkers are more likely to be smokers
- Those who are more sexually active, more likely to be exposed both to herpes and to HPV
- In both cases, causal agent is a different, associated factor
Non-causal association (confounding)

- Important to understanding public health significance of a finding
- If the prevalence of a risk factor in the population changes, will the incidence of disease change?
Confounding

- Smoking is a risk factor for Pancreatic cancer
- Smoking is associated with coffee drinking (but is not a result of coffee drinking)
Confounding

- Approaches to handling confounding
  - In designing and carrying out the study
    - Individual matching
    - Group matching
  - In the analysis of data
    - Stratification
    - Adjustment
### Example of Confounding

<table>
<thead>
<tr>
<th>Sucrose intake and risk of endometrial cancer: Swedish Mammography Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total sucrose g/d</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>≤ 15</td>
</tr>
<tr>
<td>16-25</td>
</tr>
<tr>
<td>26-35</td>
</tr>
<tr>
<td>≥ 36</td>
</tr>
</tbody>
</table>

*Adjusted for age, BMI, coffee, energy, diabetes

Friberg et al, CEBP 2011;20:1831-7
One More Concept . . .

- Interaction
  - When the incidence of disease in the presence of two or more risk factors differs from the incidence expected to result from their individual effects (cohort study)
  - 2 or more factors modify the effects of each other with respect to disease
  - Effect modification
Interaction

☐ Is there an association?

☐ Is the association due to confounding?

☐ Is the association equally strong in strata formed on the basis of the 3rd variable?
  ■ If no, then interaction is present
  ■ If yes, then no interaction is present
Interaction

- Is the association equally strong in strata formed on the basis of the 3rd variable?
  - If no, then interaction is present
  - If yes, then no interaction is present

- Interaction is a biological difference
  - The association is different in the different strata
Figure 6.1.10  BMI and breast cancer (age unspecified); cohort and case-control studies

Relative risk (95% CI)

Cohort
- Galanis 1998  1.11 (1.05–1.18)
- Mills 1989  1.11 (1.01–1.22)
- Key 1999  1.07 (1.00–1.08)
- Wolk 1998  1.04 (1.00–1.08)
- Silvera 2005  1.03 (0.99–1.07)
- Byrne 1996  1.02 (0.90–1.16)
- Overvad 1991  1.01 (0.85–1.21)
- Tornberg 1994  1.01 (0.97–1.05)
- Rapp 2005  1.00 (0.98–1.03)
- Zhang 2003  1.00 (0.96–1.05)
- Vatten 1992  0.96 (0.93–0.99)
- Wu 1999  0.96 (0.82–1.11)
- Wu 1999  0.94 (0.80–1.12)
- Kilkkinen 2004  0.92 (0.85–1.00)
- Rissanen 2003  0.90 (0.81–1.01)
- Knekt 1996  0.88 (0.79–0.99)
- Summary estimate  1.01 (1.00–1.02)
Figure 6.1.15
BMI and premenopausal breast cancer; cohort studies

Relative risk (95% CI)

Galanis 1998 1.09 (0.96–1.25)
Toniolo 1994 1.09 (0.96–1.23)
Manjer 2001 1.00 (0.89–1.13)
Tulinius 1997 1.00 (0.92–1.09)
Karaks 1998 0.99 (0.89–1.10)
Sonnenschein 1999 0.95 (0.84–1.07)
Huang 1997 0.95 (0.91–0.98)
Saadatien-Elahi 2002 0.94 (0.83–1.07)
Vatten 1992 0.94 (0.90–0.98)
Tehard 2004 0.93 (0.87–0.98)
Weiderpass 2004 0.92 (0.88–0.99)
Ahlgren 2004 0.92 (0.88–0.96)
Tornberg 1994 0.86 (0.80–0.92)
Rissanen 2003 0.75 (0.62–0.90)
Summary estimate 0.94 (0.92–0.95)
Confounding and study design

- Cross sectional studies
- Case control studies
- Cohort studies
- Randomized trials
Interaction and study design

- Cross sectional studies
- Case control studies
- Cohort studies
- Randomized trials
- External and internal validity
- Confounding/Interaction
- Bias
- Incorporating different study designs
Bias

- **Bias =**
  - Any systematic error in the *design, conduct or analysis of a study* that results in a mistaken estimate of an exposure’s effect on the risk of disease

  *(Schlesselman and Stolley, 1982)*
Sources and Types of Bias

- Selection bias
  Systematic differences in those included in a study and those who are not
Sources and Types of Bias

- **Selection bias:**
  - If selection of subjects does not reflect population:
    - Lack of generalizability or external validity
    - Not selection bias
  - If selection of subjects is different between groups within a study:
    - Lack of internal validity
    - Selection bias
Sources and Types of Bias

- **Information bias:**
  - Problems with the information obtained about subjects
    - Information about disease status
    - Information about exposure
Sources and Types of Bias

Information bias can lead to:

- **Misclassification bias:**
  - Non-differential
    - If the misclassification is not related to the exposure status or disease status
      - Similar misclassification of exposed and non-exposed or those with disease and without disease
    - RR or OR is biased towards 1.0
      - Less likely to detect a difference
Sources and Types of Bias

- Information bias can lead to:
  - **Misclassification bias:**
    - **Differential:**
      - If the misclassification is related to the exposure status or disease status
        - Differential misclassification in exposed and non-exposed, or in those with disease and without disease
      - RR or OR is biased away or towards 1.0
        - Less or more likely to detect a difference
The Born Loser

Listen to this, Brutus...

They now say a glass of wine a day is good for you!

Does that mean I can save up all week then drink them all on Saturday night?
Sources and Types of Bias

- Information bias:
  - Surveillance bias:
    - Diseases are more likely to be diagnosed in persons under medical surveillance
      - Example: Thrombophlebitis and oral contraceptive use. MDs may examine women using OCs more often and more thoroughly
    - Possible solutions:
      - Stratify cases and controls according to some index of medical care utilization
      - In a prospective study, systematically assess outcome in both exposed and non-exposed
Sources and Types of Bias

Information bias:

Observer bias:

- When exposure status is known to the observer, assignment to outcome (yes/no) may be biased
- Example: Knowledge of exposure to alcohol may increase likelihood of alcoholic cirrhosis dx

Possible solutions:

- Mask exposure status
- Standardize procedures
Sources and Types of Bias

- Information bias:
  - *Interviewer bias:*
    - Knowledge of case-control status may affect way in which questions on past exposure are asked
  - Possible solutions:
    - Mask disease status
    - Standardize procedures
    - Monitor quality during study (taping)
Sources and Types of Bias

- Information bias:
  - Recall bias:
    - Bias from a difference in the ability to recall past exposure between cases and controls
    - Example: OR of maternal rubella with congenital malformations in offspring were greater in the case-control studies than RR in prospective studies
  - Possible solutions:
    - Cohort studies
    - Validate responses
    - Nested case-control studies
Sources and Types of Bias

Information bias:
- In abstracting records
- Surrogate interviews
- Reporting bias
  - Perceptions, beliefs, attitudes
    - Over or underreporting
Bias and Confounding

- **Bias:**
  - Any systematic error in the design, conduct or analysis of a study

- **Confounding:**
  - Not an error in a study
- External and internal validity
- Confounding/Interaction
- Bias
- Incorporating different study designs
Studying Etiology of Disease

- Conclusions about what causes a disease is based on the totality of evidence
  - Animal models
  - In-vitro systems
  - Human population studies
- Our understanding of what causes a disease and also how best to prevent it continues to change and to be revised
Studying Etiology of Disease

- Animal models
  - Advantages
  - Disadvantages

- In-vitro systems:
  - Advantages
  - Disadvantages

- Human population studies
  - Advantages
  - Disadvantages
Studying Etiology of Disease

- **Animal models**
  - Controlled exposures
  - Detail regarding impact on tissues and specific organs
  - Between species differences

- **In-vitro systems:**
  - Frequently use human cells
  - Cellular level
  - Difficult to extrapolate to functioning body
  - Issues of dose

- **Human population studies**
Studying Etiology of Disease

- Animal models
- In-vitro systems
- Human population studies
  - Randomization not always ethical
  - Concerns of confounding and bias
  - Issues of measurement of exposure
    - Report of exposure
    - Changes in exposure
    - Other correlated exposures
1940’s, United Kingdom