Confounding

A variable that (a) is causally related to the disease under study (or is a proxy for an unknown or unmeasured cause) and (b) is associated with the exposure under study (Kesley)

- Any risk factor for a disease is a potential confounder
- Wholly or partially accounts for apparent effect of exposure on disease (either direction)
- Occurs in nature, not due to study design or execution
Confounding
Examples of Confounding

- Lighters and Lung Cancer
- Breast Cancer Prevention
  - Breast Feeding
  - ? Parity
  - Age at first pregnancy
- Coffee Drinking and Myocardial Infarction

\[
\text{OR} = \frac{ad}{bc} = \frac{(90)(60)}{(60)} \div (60) = 2.25
\]
Controlling confounding through stratified analysis

<table>
<thead>
<tr>
<th></th>
<th>Smokers</th>
<th></th>
<th>Non-Smokers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MI</td>
<td>No MI</td>
<td>MI</td>
<td>No MI</td>
</tr>
<tr>
<td>Coffee</td>
<td>80</td>
<td>40</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>No Coffee</td>
<td>20</td>
<td>10</td>
<td>40</td>
<td>80</td>
</tr>
<tr>
<td>Totals</td>
<td>100</td>
<td>50</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

OR = 1.0

OR = 1.0
Controlling Confounding

A. Controlling by Design
- a) randomization – assures same # with and without any potential confounder in both groups
- b) restriction – only allow into study if fall into specific groups
- c) matching – for every person with a factor in case have person without in controls

B. Controlling by Analysis
- a) *stratified analysis* – make groups homogenous
- b) multivariate analysis – most popular
1. Definitions
   - a) traditional (statistical)
     - Risk of Lung Cancer
   - b) biological
   - c) public health

2. Additive vs. Multiplicative (lack of one implies the other)
New thinking about interaction

- synergy – parallelism = positive additive interaction = \( R(AB) - R(B) - R(A) + R(ab) \)
Bias

- A systematic error in the *collection or interpretation of data* in an epidemiologic study. (Henneken) Any systematic error in the *design, conduct or analysis* of a study resulting in a mistaken estimate of an exposure effect. (Schlesselman)
- Found in the design or conduct of study, as opposed to confounding which is found in nature
Types of Bias

1. Recall Bias
   - Particular problem in case-control studies

2. Diagnosis Bias
   - Knowledge of E may influence Dx (e.g. BCPs and PE)

3. Hawthorne Effect
   - General Electric plant in Hawthorne, NY
   - Productivity tied to ↑ (and ↓) in lighting
   - Called ‘placebo’ effect in medicine; participants and researchers ‘blinded’ to actual treatment status
Selection Biases

“a distortion in the estimate of effect resulting from the manner in which subjects are selected for the study” (KKM)

- Detection Bias – differential surveillance based on exposure status
  - Surveillance Bias (Schlesselman) – BCPs and endometrial CA (Feinstien)
  - Greater in ‘milder’ diseases picked up on routine visits
Selection Biases

- Loss to Follow up (Non-response Bias)
  - Cohort Studies
  - Compliant participants tend to be healthier
- Healthy Worker Bias
  - Even 23 years after d/c soldiers healthier
  - Caution comparing work cohorts to general population
- Volunteer bias
  - ↓ smokers, ↑ exercise,
HRT ➔ CAD Controversy

- Observational Studies
  - HRT Protective for CAD
    - Tended to be studies of volunteer worker cohorts

- Randomized Trials
  - Slight increase in risk
Incidence-Prevalence Bias

- Incidence – all *new* cases of disease in a time period
  - Tend to be acute
- Prevalence – *existing* cases of disease at one point in time
  - Tend to be chronic
- Cross-sectional studies tend to pick up chronic cases
Direction of Incidence-Prevalence Bias Depends on Population

- Hospital-based study of depression
  - Systematically miss patients who improved (or committed suicide)

- In-patient study of MI patients
  - Systematically miss sudden deaths and those successfully thrombolysed and released

- Studies of schizophrenia
  - Bias can be in either direction. Prognosis fairly bright (60-80% go on to productive lives) if based on outpatient population; fairly grim if based on in-patient population (DSM)
Does public assistance breed dependency?

<table>
<thead>
<tr>
<th></th>
<th>1-2 yrs</th>
<th>3-7 yrs</th>
<th>&gt;7 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>% who have ever received AFDC</td>
<td>30%</td>
<td>40%</td>
<td>30%</td>
</tr>
<tr>
<td>% receiving AFDC at particular time</td>
<td>7%</td>
<td>28%</td>
<td>65%</td>
</tr>
</tbody>
</table>

Long-term recipients more likely to be picked up in a cross-sectional survey
Berkson’s Bias: A selection bias due to differing rates of hospitalization

<table>
<thead>
<tr>
<th>VAGINAL BLEEDING</th>
<th>TYPE OF CANCER</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Endometrial</td>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>100</td>
<td>100</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>900</td>
<td>900</td>
<td>1800</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1000</td>
<td>2000</td>
<td></td>
</tr>
</tbody>
</table>

\[ \text{OR} = \frac{(100)(900)}{(100)(900)} = 1.0 \]

In the general population, there is no association between vaginal bleeding and endometrial cancer.
Probability of admission varies: vag bleed = 70%, endometrial CA = 10%, other Cancer = 50%

Now, OR = (73)(9450)/(85)(90) = 4.3

Spurious association

<table>
<thead>
<tr>
<th>VAGINAL BLEEDING</th>
<th>Endometrial</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>73</td>
<td>85</td>
</tr>
<tr>
<td>No</td>
<td>90</td>
<td>450</td>
</tr>
<tr>
<td></td>
<td>163</td>
<td>535</td>
</tr>
</tbody>
</table>
How to address selection biases?

- if $a$, $b$, $c$, $d$ represent selection probabilities for the cells in a 2x2 table, ensure $ad/bc = 1$

- Overestimate:
  $ad/bc > 1$

<table>
<thead>
<tr>
<th></th>
<th>D</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>e</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>