Study Design II

Alison Merikangas
23 September 2009
Outline

• Types of Studies
  – Randomized Trial
  – Case-Control Studies
  – Cohort Studies
• Relative Risk & Odds Ratios
• Bias & Confounding
• Validity & Precision
• Data Management & Storage
• Literature Example
• Further Reading
Types of Studies
Randomized Trial

Study Population

Randomly Assigned

Current Treatment
  - Improve
  - Do Not Improve

New Treatment
  - Improve
  - Do Not Improve
Randomization

• Randomization is the process by which allocation of subjects to treatment groups is done by chance, without the ability to predict who is in what group
• Prevent bias in allocating subjects to treatment groups (avoid predictability)
• Achieve comparability between the groups (there is no guarantee)
• How to assign random numbers:
  – Table of random numbers
  – An allocation scheme
  – Computers or calculators
  – Random number websites (random.org)
Case-Control Studies

Have the Disease

Were Exposed

Were Not Exposed

Do Not Have the Disease

Were Exposed

Were Not Exposed
Case & Control Selection

“Total” Population

Reference Population

Cases

Controls
Matching

• Intended to circumvent selection bias – match those with similar characteristics that might affect the study outcome (e.g. age, sex, and race)
  – Individual matching (matched pairs)
    • Must use matched tests
  – Group matching (frequency matching)
• Cannot study the effect that the “matching characteristic” has on the outcome
• Matching on many variables may make it difficult to find an appropriate control
Blinding

- Blinding is used to increase the objectivity of the persons dealing with the randomized study
  - Mask subjects (& caregivers)
    - Placebo or sham treatment
  - Mask observers
  - Mask data collectors
  - Mask data analysts
### Odds Ratio in a Case-Control Study

The odds ratio (OR) is calculated as:

\[
OR = \frac{ad}{bc}
\]

#### First Select

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exposed</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exposed</td>
<td>(a)</td>
<td>(b)</td>
</tr>
<tr>
<td>Not Exposed</td>
<td>(c)</td>
<td>(d)</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>(a + c)</td>
<td>(b + d)</td>
</tr>
<tr>
<td><strong>Proportions exposed</strong></td>
<td>(\frac{a}{a + c})</td>
<td>(\frac{b}{b + d})</td>
</tr>
</tbody>
</table>
Prospective & Retrospective Cohort Studies

Defined Population

Prospective

2009

Exposed

Disease

No Disease

Disease

No Disease

Retrospective

2029

Not exposed

2029

2019

2019

2009

NOT Randomly Assigned
Relative Risk in a Cohort Study

\[
RR = \frac{\frac{a}{a + b}}{\frac{c}{c + d}}
\]

<table>
<thead>
<tr>
<th>Disease</th>
<th>No Disease</th>
<th>Totals</th>
<th>Incidence Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>a</td>
<td>b</td>
<td>(a + b)</td>
</tr>
<tr>
<td>Not exposed</td>
<td>c</td>
<td>d</td>
<td>(c + d)</td>
</tr>
</tbody>
</table>

First select

Then follow

RR and Association with Disease

- Relative risk expresses the magnitude of the association between the exposure and the disease, a key consideration for inference about etiology. It does not depend on exposure prevalence.
When RR and OR are used

• Relative risk
  – Randomized controlled trials
  – Cohort studies

• Odds ratio
  – Case-control studies
  – Retrospective studies (Cohort studies)
Meaning of RR/OR

• An RR/OR of 1
  – no difference in risk between the two groups

• An RR/OR of < 1
  – the event is less likely to occur in the experimental group than in the control group

• An RR/OR of > 1
  – the event is more likely to occur in the experimental group than in the control group
If a rare disease...

Both \( \frac{a}{a + b} \) and \( \frac{c}{c + d} \) will be very small, so

\[
\frac{R}{O} = \left( \frac{\frac{a}{a+b}}{\frac{ad}{bc}} \right) \cdot \left( \frac{\frac{c}{c+d}}{\frac{bc}{ad}} \right) = \left( \frac{a}{a+b} \right) \times \frac{bc}{ad} = \left( \frac{abc}{a+b} \right) = \left( \frac{b}{a+b} \right) \approx 1
\]

But what if the disease is common?
The relationship between risk ratio (RR) and odds ratio by incidence of the outcome.

Other Study Designs

• Nested Case-Control Studies
• Case-Cohort Studies
• Case-Crossover Studies
• Cross-Sectional Studies
  – AKA Prevalence Study
Other measures of association

- Relative rates
- Relative odds
- Relative hazards
- Attributable risk
  - In the exposed
  - Population
Sample Size

To estimate sample size:
1. The difference in response rates to be detected
2. An estimate of the response rate in one of the groups
3. Level of significance ($\alpha$)
4. The value of the power desired ($1-\beta$)
5. Whether the test should be one-sided or two-sided
# Statistics - Hypothesis Test

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>Null Hypothesis True</th>
<th>Null Hypothesis False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject Null Hypothesis</td>
<td>Type I Error $(\alpha)$ Significance Level</td>
<td>Correct $(1-\alpha)$ Confidence Level</td>
</tr>
<tr>
<td>Fail to Reject Null Hypothesis</td>
<td>Correct $(1-\beta)$ Power</td>
<td>Type II Error $(\beta)$</td>
</tr>
</tbody>
</table>
Bias & Confounding
Information Bias

• Distortion of the estimate of effect due to measurement error or misclassification of subjects on one or more variables
  – Recall bias
  – Reporting bias
  – Instrument bias

• When information is incorrect, there is misclassification
  – Differential misclassification occurs when the level of misclassification differs between the two groups
  – Non-differential misclassification occurs when the level of misclassification does not differ between the two groups
Selection Bias

• Fundamental bias of assignment
• Distortion in the estimate of effect resulting from the manner in which subjects are selected for the study population
Confounding

Due to Confounding

Factor X

Disease

Causal

Observed Association
Validity & Precision
Accuracy vs. Precision

High accuracy, but low precision

High precision, but low accuracy
Random Error

• A problem of precision
• Essentially attributable to sampling variation
  – Depends on study design
    • E.g. sample size considerations
  – Statistical characteristics of the estimator
    • E.g. variance
Systemic Error

• A difference between what the estimator is actually estimating and the true effect measure of interest

• Attributable to:
  – Methodological aspects of study design or analysis
Internal vs. External Validity

Population
- Study
- Actual
- Target
- External

Validity
- Sample
- Inference
- Internal
- External
Extrapolation

• Conclusions for a study are drawn for the target population
• The target population may be similar to those included in the investigation, or may include a range of data not represented in the study sample
• Often we want to map results onto the population at large
## At-risk Groups

$$AR\% = \frac{RR - 1}{RR} \times 100$$

<table>
<thead>
<tr>
<th>Relative Risk</th>
<th>Attributable Risk %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>75</td>
</tr>
<tr>
<td>10</td>
<td>90</td>
</tr>
<tr>
<td>20</td>
<td>95</td>
</tr>
</tbody>
</table>
## Populations or Communities

\[
PAR\% = \frac{b(RR - 1)}{b(RR - 1) + 1} \times 100
\]

- PAR = population attributable risk
- RR = relative risk
- b = proportion of population with the risk factor

<table>
<thead>
<tr>
<th>RR</th>
<th>b</th>
<th>PAR%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.01</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>0.01</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>0.10</td>
<td>9</td>
</tr>
<tr>
<td>20</td>
<td>0.10</td>
<td>65</td>
</tr>
<tr>
<td>2</td>
<td>0.50</td>
<td>33</td>
</tr>
<tr>
<td>20</td>
<td>0.50</td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
<td>50</td>
</tr>
<tr>
<td>20</td>
<td>1.00</td>
<td>95</td>
</tr>
</tbody>
</table>
Data Management & Storage

• It is ideal to plan for data management before you begin a study
  – How data is stored
  – Type of variables
  – How data will be analyzed
Population-based linkage analysis of schizophrenia and bipolar case–control cohorts identifies a potential susceptibility locus on 19q13

C Francks¹, F Tozzi¹, A Farmer², JB Vincent³, D Rujescu⁴, D St Clair⁵ and P Muglia¹

¹Division of Medical Genetics, GlaxoSmithKline, Verona, Italy; ²MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, London, UK; ³Centre for Addiction and Mental Health, University of Toronto, Toronto, ON, Canada; ⁴Division of Molecular and Clinical Neurobiology, Department of Psychiatry, Ludwig-Maximilians University, Munich, Germany and ⁵Department of Mental Health, University of Aberdeen, Aberdeen, UK
Further Reading

• Textbooks

• JHU Opencourseware
  – Sukon Kanchanaraksa, PhD, Johns Hopkins University
  – Marie Diener-West, PhD, Johns Hopkins University

• Articles