STUDY DESIGN
CASE SERIES AND CROSS-SECTIONAL

Daniel E. Ford, MD, MPH
Vice Dean for Clinical Investigation
Johns Hopkins School of Medicine
Introduction to Clinical Research

STUDY DESIGN

- Provides "differential diagnosis" of a study's strengths and weakness
- Determines confidence in results of study
- Facilitates critical appraisal of the medical literature
- Linked to research question

STUDY DESIGNS AND CORRESPONDING QUESTIONS

- Ecologic
- What explains differences between groups?
- Case Series
- How common is this finding in a disease?
- Cross-sectional
- How common is this disease or condition?
- Case-control
- What factors are associated with having a disease?
- Prospective
- How many people will get the disease?
- Randomized trial
  If we change something does the outcome change
  What factors predict development?
### 2 x 2 TABLE

<table>
<thead>
<tr>
<th>Risk Factor (Exposure)</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>A</td>
<td>B</td>
<td>A + B</td>
</tr>
<tr>
<td>No</td>
<td>C</td>
<td>D</td>
<td>C + D</td>
</tr>
<tr>
<td>Total</td>
<td>A + C</td>
<td>B + D</td>
<td></td>
</tr>
</tbody>
</table>

Relative Risk = \( \frac{A(A + B)}{C(C + D)} \)

Odds Ratio = \( \frac{AD}{BC} \)

### STUDY DESIGN DEFINITIONS

- Based on sampling strategy, i.e., how we choose who gets into the study
- Sampling with regard to disease: cross-sectional and case-control studies
- Sampling with regard to exposure or treatment: prospective studies

### CRITERIA FOR CAUSAL INFERENCE

- Experimental evidence
- Temporality
- Strength of the association
- Dose-response relationship
- Consistency in different populations
- Specificity: exposure leads to only 1 disease
- Biologic plausibility
- Coherence
- Analogy
Not all study designs are created equal!

HIERARCHY OF STUDY DESIGNS

- RCTs
- Prospective Studies
- Case-controlled Studies
- Cross-sectional Studies
- Ecologic Studies

STUDY DESIGN EXAMPLE

- Does higher dose of dialysis (Kt/v) result in lower mortality in hemodialysis patients?
ECOLOGIC STUDIES

- Sometimes called correlational studies
- Compares outcomes between groups, **not** individuals
- Useful to examine trends over time or to explain differences between groups

2 x 2 Table

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>a</td>
<td>b</td>
<td>a + b</td>
</tr>
<tr>
<td>No</td>
<td>c</td>
<td>d</td>
<td>c + d</td>
</tr>
<tr>
<td>Total</td>
<td>a + c</td>
<td>b + d</td>
<td>N</td>
</tr>
</tbody>
</table>

* Letters represent group rates or means, not individuals
* Exposure can be assessed before or after outcome

Kt/V AND MORTALITY IN 100 DIALYSIS UNITS

<table>
<thead>
<tr>
<th>Mean Clinic Kt/V</th>
<th>Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Acceptable</td>
<td>20</td>
</tr>
<tr>
<td>Low</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>40</td>
</tr>
</tbody>
</table>
VITAL STATISTICS

- Common data source for ecologic studies
- Describes disease patterns in entire geographic or political populations
- Routinely collected information from birth and death certificates; allow comparisons between countries over time
- Comparison by age, race, sex, geographic areas and time period

VITAL STATISTICS

ADVANTAGES

- Inexpensive
- Representative of large groups and large geographic areas
- Available over long periods of time
- Uniform coding rules

VITAL STATISTICS

DISADVANTAGES

- Group "ecologic" data -- not individual
- Uncertain accuracy of diagnoses
- Changes in ICD codes
- Variability in coding practices
- Limited to available data
- Mortality may not reflect incidence
ECOLOGIC STUDIES
DISADVANTAGES

- Subject to ecologic fallacy
- Lead to unusual conclusions if not testing biologically plausible hypotheses
- Usually done early in the investigation of a research question when cohort studies or clinical trials not available

STUDY DESIGNS AND CORRESPONDING QUESTIONS

- Ecologic
  - What explains differences between groups?
- Case Series
  - How common is this finding in a disease?
- Cross-sectional
  - How common is this disease or condition?
- Case-control
  - What factors are associated with having a disease?
- Prospective
  - How many people will get the disease? What factors predict development?
CASE REPORTS

- Make observations about medical phenomena in an individual patient
- Simple description of clinical data without comparison group
- Observations should be comprehensive and adequately detailed

Kt/V AND MORTALITY CASE REPORT

- 55 year old man has been on dialysis for 35 years
- On home dialysis daily during that time
- No evidence of hypertension, cardiovascular disease, LVH

CASE REPORTS ADVANTAGES

- Easy and inexpensive to do in hospital
- Provides information on new disease or new therapy
- Useful in conveying "clinical experience"
- Helpful in hypothesis formation
CASE REPORTS
DISADVANTAGES
• Biased selection of subjects so that conclusions are difficult to generalize

• Were the findings a chance happening or characteristic of the disease?

• Is the "exposure" really higher than a comparison group?

CASE REPORTS EXAMPLES
• Asbestos and mesothelioma

• Pneumocystis pneumonia

• Legionnaire's Disease

DECIDING TO PUBLISH
• What observations have been made prior to this report?

• What new phenomenon is illustrated?

• What further studies should be done?
CASE SERIES

- Group of patients with a disease or outcome
- Usually consecutive series
- Detailed observations
- No comparison group – difficult to address etiologic questions

2 x 2 TABLE

Case Series

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>a</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>No</td>
<td>c</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>Total</td>
<td>a+c</td>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>

DISTRIBUTION OF Kt/V IN 100 PATIENTS WHO DIED DURING THE FIRST YEAR OF DIALYSIS

N = 100

Low Kt/V 70%
Acceptable Kt/V 30%

- review records on 100 patients who died
CASE SERIES
OBSERVATIONS
• Should have clear definitions of the phenomena being studied
• These same definitions should be applied equally to all individuals in the series
• All observations should be reliable and reproducible (consider blinding)

CASE SERIES
PRESENTATION OF FINDINGS
• Proportions (% per 10^5, etc.) of the study populations with the outcome, confidence intervals
• Means, standard errors for continuous variables
• Are there important subgroups that need data presented separately?

CASE SERIES
ADVANTAGES
• Informs patients and physicians about natural history and prognostic factors
• Easy and inexpensive to do in hospital settings
• Helpful in hypothesis formation
• Can help answer the question of why this outcome occurred
CASE SERIES
LIMITATIONS

- Cases may not be representative
- Outcome may be a chance finding, not characteristic of disease
- Begs the question “Compared to what?”

STUDY DESIGNS AND CORRESPONDING QUESTIONS

- Ecologic  • What explains differences between groups?
- Case Series  • How common is this finding in a disease?
- Cross-sectional  • How common is this disease or condition?
- Case-control  • What factors are associated with having a disease?
- Prospective  • How many people will get the disease? What factors predict development?

CROSS-SECTIONAL STUDIES

- Make observations concerning the prevalence and characteristics of a disease in a well-defined population during a defined period of time (period prevalence)
- Estimate prevalence
- Examine characteristics associated with condition or disease by comparing cases to noncases
2 x 2 TABLE
Cross-sectional Study

<table>
<thead>
<tr>
<th>Disease</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure Yes</td>
<td>a</td>
<td>b</td>
<td>a + b</td>
</tr>
<tr>
<td>No</td>
<td>c</td>
<td>d</td>
<td>c + d</td>
</tr>
<tr>
<td>Total</td>
<td>a + c</td>
<td>b + d</td>
<td>N</td>
</tr>
</tbody>
</table>

* Draw a 1% random sample of all hemodialysis patients dialyzed in 1996
* Assess Kt/V in all patients and their vital status at the end of 1996

PREVALENCE OF LOW Kt/V AND MORTALITY

<table>
<thead>
<tr>
<th></th>
<th>Dead</th>
<th>Alive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Kt/V</td>
<td>400</td>
<td>1,000</td>
<td>1,400</td>
</tr>
<tr>
<td>High Kt/V</td>
<td>350</td>
<td>1,250</td>
<td>1,600</td>
</tr>
<tr>
<td></td>
<td>750</td>
<td>2,250</td>
<td>3,000</td>
</tr>
</tbody>
</table>

Odds Ratio = \( \frac{(400)(1,250)}{(350)(1,000)} = 1.4 \)


Burt, Hypertension, 1995
Study Design

Number of Medicare ESRD Patients on Dialysis in the United States

Physician Visits Per Person in the Last Six Months of Life: 2002-2003

Table 2. Odds Ratios of Elevated C-Reactive Protein Level by Recent History of Major Depression in Men and Women*
Table 2: Odds Ratios for Elevated C-Reactive Protein Level by Sex: Trend in Men and Women

<table>
<thead>
<tr>
<th>Sex</th>
<th>Outcome</th>
<th>Odds Ratio (95% Confidence Interval)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>Abnormal</td>
<td>1.5 (1.0-2.4)</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>Single status</td>
<td>1.10 (0.72-1.69)</td>
<td>.66</td>
</tr>
<tr>
<td></td>
<td>Divorced</td>
<td>2.00 (1.00-4.00)</td>
<td>.01</td>
</tr>
<tr>
<td>Women</td>
<td>Abnormal</td>
<td>1.00 (0.59-1.75)</td>
<td>.57</td>
</tr>
<tr>
<td></td>
<td>Single status</td>
<td>0.93 (0.59-1.47)</td>
<td>.73</td>
</tr>
<tr>
<td></td>
<td>Divorced</td>
<td>0.74 (0.40-1.41)</td>
<td>.89</td>
</tr>
</tbody>
</table>

*Note: odds ratio estimates 95% confidence interval for having 1 or more episodes of low-grade depression.

SAMPLING

- Process of obtaining a sample of a population for study
- In clinical research, goal should always be a representative sample
- Variety of methods available

CROSS-SECTIONAL STUDIES

SAMPLING THE POPULATION

- Derive a sampling “frame”
- Choose a sampling strategy
- Maximize response rate
CROSS-SECTIONAL STUDIES
TYPES OF SAMPLING

- Simple random—each individual has the same probability of being chosen
- Stratified random—first create strata and then select randomly within strata. If most variance is between strata, gives lower sampling variance
- Systematic—ex., select every 4th person, used commonly in clinical research, akin to stratified random sample if list is ordered
- Cluster

RESPONSE RATES AND SAMPLING

- Sample size of 500
  - 5% of 10,000=500
  - 75% of 666=500
- Which study provides the most valid causal inference?
- Are persons who do not respond (can’t be found or say no) likely to be different than those who do?

NONRESPONSE IN SAMPLING CROSS SECTIONAL STUDIES

- Minimize non-response
  - smaller sample size allows more intensive recruitment
  - collect data on non-responders, if possible
  - intensively recruit a sub-sample of non-responders
CROSS-SECTIONAL STUDIES
ADVANTAGES
• Inexpensive for common diseases
• Should be able to get a better response rate than other study designs
• Relatively short study duration
• Can be addressed to specific populations of interest

CROSS-SECTIONAL STUDIES
DISADVANTAGES
• Unsuitable for rare or short duration diseases (prevalence = incidence x duration)
• High refusal rate may make accurate prevalence estimates impossible
• More expensive and time consuming than case-control studies
• Disease process may alter exposure – reverse causality
• No data on temporal relationship between risk factors and disease development

Defining Cross-Sectional Studies
• How short is the assessment period?
  – Symptom questionnaire and then physical exam
• Cases accumulated over long period of time
• Time trends of multiple cross-sectional studies (smoking rates in population over time)
STUDY DESIGNS AND CORRESPONDING QUESTIONS

- Ecologic  • What explains differences between groups?
- Case Series  • How common is this finding in a disease?
- Cross-sectional  • How common is this disease or condition?
- Case-control  • What factors are associated with having a disease?
- Prospective  • How many people will get the disease?
  What factors predict development?