CLINICAL RESEARCH METHODS AND STUDY DESIGN

Gregory W. Heath, D.H.Sc., M.P.H.
Department of Medicine
University of Tennessee College of Medicine, Chattanooga

Adapted from Jack M. Guralnik, M.D., PhD – NIA/NIH, 2004

Epidemiology

The study of the distribution and determinants of disease and injury in human populations

Fundamental Assumptions of Epidemiology

1. Human disease does not occur at random

2. Human disease has causal and preventive factors that can be identified through systematic investigation of different populations or subgroups of individuals within a population

Hennekens and Buring, 1987
Study Designs - Overview

Descriptive studies
- Populations: correlation (ecological) studies
- Individuals
  - Case reports
  - Case series
  - Cross-sectional surveys

Analytic studies
- Observational studies
  - Case-control studies
  - Prospective cohort studies – historical and longitudinal
- Intervention studies (clinical trials)

Case Reports and Case Series

Objective: To make observations about patients with defined clinical characteristics (e.g., patients with a certain disease or cluster of symptoms)

Design:
- simple description of clinical data
- data derived from a well-defined group of individuals
Observations should be comprehensive and adequately detailed

1. Clear definitions of phenomenon being studied

2. Definitions should be applied equally to all individuals in the series

3. Observations should be reliable and reproducible

Interpretation/Conclusion:
1. What observations have been made prior to this report?

2. What new phenomenon is illustrated?

3. What further studies should be done?

4. Is the study group representative of all patients with this disorder-- can conclusions be generalized?
Advantages

1. Useful in hypothesis formation, natural history studies, describing "clinical experience"

2. Easy and inexpensive to do in hospital settings

Disadvantages: Biased selection of study patients may lead to inability to generalize study results

1. Were only sickest or most typical patients included in the study?

2. Were the findings a chance happening or a characteristic of the disease?

Case report/Case series - Example


Initial report of five cases of pneumocystis pneumonia in previously healthy, homosexual men.
Cross-Sectional Studies (Prevalence Surveys)

A. Object: To make observations concerning the prevalence and characteristics of a disease and other participant characteristics in a well-defined population

Prevalence vs. Incidence

Prevalence – number with disease at one point in time
number at risk at that point

Incidence – number of new cases of disease over a period of time
number of people at risk during that period
Design

1. Define the population under study
2. Derive a sample of the population
3. Define the characteristics being studied

Observations

1. Should be standardized and clearly defined
2. Methods of data collection should be consistently applied to all study participants

Presentation of findings

1. Prevalence (%, cases per $10^5$, etc.) for the observation in the population
2. Mean or median levels of relevant factors in the population
3. Important subgroups may need separate data presentation (e.g., age, race, sex)
Conclusions

1. Descriptive:
   a. How common is the factor in the study population?
   
   b. What are the characteristics of the group of interest (those with disease, of given age, etc.) in the population?
   
   c. What are the distributions of factors of interest (age, blood pressure, vital capacity, etc.) in the study population?

Conclusions (con’t)

2. Associative:
   
   a. What are the relationships of the factors of interest to other factors in the study population?
   
   b. How do persons with the factor of interest differ from those without it?
Advantages

1. Inexpensive for common diseases
2. Provide more representative cases than do case series
3. Relatively short duration of the study
4. Can be addressed to specific populations of interest
5. Can examine wide variety of factors simultaneously

Disadvantages

1. Unsuitable for rare diseases
2. Bias may be operative
3. High refusal rate / low response rate
4. Generally more expensive and time-consuming than case-control studies
5. The disease process may alter measurements
6. No data are collected regarding temporal relationship
Cross Sectional Study - Example


Prevalence data on overweight and obesity using measured height and weight in National Health and Nutrition Examination Survey (NHANES)

Case-Control Study

Object: To make observations regarding possible associations between a disease and one or more hypothesized risk factors
Design

1. General strategy: To compare the prevalence or level of the possible risk factor between a representative group of disease subjects (cases) and a representative group of disease-free subjects (controls) derived from the same population.

Basic assumptions

a. Cases are representative of all patients who develop the disease.

b. Controls are representative of the general "healthy" population who do not develop the disease.

c. Information is collected from cases and controls in the same way.
Selection of cases

a. Should have standardized selection criteria from a well-defined population

b. Sources: case registries, admission records, pathology logs

c. Aim for as high a participation rate as possible

Selection of controls-- the most difficult issue

a. The perfect control group probably doesn't exist

b. Must have standard selection criteria from a well-defined population

c. Sources: sample of general population, neighborhood, families
d. Cost and accessibility should be considered in the selection of controls

e. Multiple control groups are considered to be methodologically superior

Observations: Data are collected "looking back" for possible exposures

1. All observations should be made using the same methods in cases and controls

2. Validity of measurement techniques should be established
Potential sources of bias: Selection and observation

1. Types of selection bias
   a. Prevalence-incidence bias
   b. Non-respondent bias:

Types of observational or interviewer bias

a. Diagnostic suspicion bias

b. Exposure suspicion bias
Types of observational or interviewer bias (continued)

c. Recall bias

d. Family information bias

Presentation of findings: The 2 X 2 table

<table>
<thead>
<tr>
<th>Characteristic/Exposure</th>
<th>Presence of Disease</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number with Disease</td>
<td>Number without Disease</td>
</tr>
<tr>
<td>Present</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Absent</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>Total</td>
<td>a + c</td>
<td>b + d</td>
</tr>
</tbody>
</table>
Advantages of a case-control study

1. May be the only way to study the etiology of rare diseases
2. Can study multiple etiologic factors simultaneously
3. Less time-consuming and expensive
4. If assumptions are met, associations and risk estimates are consistent with other types of studies

Disadvantages of a case-control study

1. Does not estimate incidence or prevalence
2. Relative risk is only indirectly measured
3. Both selection and information biases may give potentially spurious evidence of association between a factor and a disease
4. Usually cannot study rare exposures
5. Temporal relationship between exposure and disease can be difficult to document
Case-Control Study - Example


Case-control design was able to identify relationship of exposure to stilbestrol during mother’s pregnancy with occurrence of rare tumor in female offspring many years later.

Prospective or Longitudinal Cohort Studies

Object: To make observations concerning the association between a given exposure (risk factor) and subsequent development of a disease/outcome.

Study Design: To identify a group of persons exposed to a purported risk factor and a second similar group non-exposed to the risk factor and follow forward to compare incidence rates between groups.
Types of Prospective Studies

Concurrent prospective study (longitudinal study): a defined population at present time is surveyed to identify exposed and non-exposed individuals who are followed forward in time (e.g., several years) to define incidence rates

Non-concurrent prospective study (retrospective or historical cohort study): a defined population has had presence or absence of exposure ascertained in an accurate, objective fashion in the past and is surveyed at present for occurrence of disease to allow definition of incidence rates in exposed and non-exposed

TIMEFRAMES FOR HYPOTHETICAL CONCURRENT AND NON-CONCURRENT PROSPECTIVE STUDIES CONDUCTED IN 2003

Concurrent: Longitudinal

- Defined Population
  - Exposed
    - Disease 2023
    - No Disease 2023
  - Non-Exposed
    - Disease 2023
    - No Disease 2003

Non-Concurrent: Historical

- 1983
- 1993
3. Assumptions

a. Exposed and non-exposed groups are representative samples of a well-defined general population

b. Absence of "exposure" also well defined and assumed to be maintained in the non-exposed group during the course of the study

Observations

1. Definitions of disease outcome should be well determined prior to the study's inception and should not be changed during the course of the study

   a. Endpoints may vary in "hardness", e.g., from death to subjective symptoms

   b. Standard criteria should be applied to both exposed and non-exposed groups, i.e., there should be no bias in determining outcomes in exposed vs. non-exposed

2. Definitions of disease should be reliable and reproducible

3. Every effort should be made to minimize the "lost to follow-up" rate since large non-response rates (> 20%) raise questions as to the accuracy of incidence rates in exposed and non-exposed groups
Presentation of findings: The 2 X 2 table

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Presence Of Disease</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number With Disease</td>
<td>Number Without Disease</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>a</td>
<td>b</td>
<td>a + b</td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>c</td>
<td>d</td>
<td>c + d</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>a + c</td>
<td>b + d</td>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>

Conclusions

1. To what larger groups can the results be generalized?

2. Is the association significant? Is the association strong?
Advantages of the prospective study

1. Cases are incident cases and may be more representative of cases than in case-control studies.
2. Design provides more information about the natural history of the disease.
3. Incidence rates are available.
4. Relative risk is directly estimated.
5. Fewer sources of bias than retrospective studies.
6. Many diseases can be studied with regard to their relationship to the exposure.
7. Temporal relationships between exposure and disease firmly established.
8. Best to study effects of rare exposure with frequent cases among the exposed.

Disadvantages

1. Duration of the study may be exceedingly long, making difficult the maintenance of consistent study methods and enthusiasm of the staff.
Disadvantages (continued)

2. Follow-up of free-living populations may be very expensive

3. Large populations often required

4. Exposures can be studied only if baseline data are available

5. Rare diseases cannot be studied

6. Several types of bias may produce spurious association (bias of assessment, loss to follow-up)
Prospective Longitudinal Cohort Study - Example


Prospective cohort study that showed early increase in risk of lung cancer and heart disease mortality and confirmed this over 50 years of follow-up.

---

Prospective Historical Cohort Study - Example


Military medical records used to identify WW II head trauma exposure group and non-trauma comparison group who were traced and evaluated for dementia 50 years later.
Causal Inference in Observational Studies: Epidemiologic Criteria

A. Statistical significance
B. Strength of association (odds ratio, relative risk)
C. Dose-response relationships
D. Temporal sequence
E. Consistency of the association (internal "validity")
F. Replication of results (external validity)
G. Biological plausibility
H. Experimental evidence

Hierarchy of Study Design

Case reports
Case series
Cross-sectional surveys
Case-control studies
Prospective cohort studies
Clinical trials
In God We Trust.

All Others Must Bring Data.