Choosing Study Subject
Specification and Sampling

Prof. Saleh Bawazir
Dept. of Clinical Pharmacy
College of Pharmacy
King Saud Univ. Riyadh
Saudi Arabia
• A good choice of study subject assuring that the findings accurately represent what is going in the population.

• **Conflicting goals challenging the judgment of investigator:**
  
  • The protocol must **specify a sample** of subject that can studied at an **acceptable cost in time and money**.
  
  • Then to **control random error** in generalizing the study finding.

• This presentation focus on the process of **specifying** and **sampling** the kinds of subjects who will be representative and feasible.
- **Population**: a complete set of people with a specified set of characteristics.

- **Sample**: a subset of the population.

- **Geographic**: a characteristics that define a population in the commoner usage (e.g. the population of Saudi Arabia).

- **Geographic, clinical, demographic, and temporal**: characteristics used in research.
The research characteristics are used to define two kinds of population:

1) Clinical and demographic define the **Target population**: the large set of all patients throughout the world to which the results will be generalized. (all children with asthma).

2) Geographic and temporal define the **Accessible population**: the subset of the target population that is available for study. (children with asthma living in Riyadh in 1989).
Choosing study subject that represent the population

- **TRUTH IN THE UNIVERSITY**
  - Research question
  - Phenomena of interest

- **ERROR**
  - Design
  - Intended sample

- **TRUTH IN THE STUDY**
  - Intended variables

- **ERROR**
  - Implement
  - Actual subject

- **FINDINGS IN THE STUDY**
  - Actual measurements

- **EXTERNAL VALIDITY**

- **INTERNAL VALIDITY**
Generalizing the study findings

**TRUTH IN THE UNIVERSE**

**Target population**
- (GENERALIZATION FAIRLY SECURE)
  - Same association exists in all suburban U.S. adults

- (GENERALIZATION LESS SECURE)
  - Same association exists in:
    a) Other U.S. adults (inner city blacks)
    b) People living in other countries
    c) People living in the 21st century
    d) etc

**Accessible population**
- Same association in the all Framingham adults

**Intended sample**
- Same association exists in the designed sample of Framingham adults

**INTERNAL VALIDITY**

**INFERENCE #1**

**EXTERNAL VALIDITY**

**INFERENCE #1**

**Actual Subjects**
- Association between hypertension and CHD observed in the actual sample of Framingham adults
Steps in designing the protocol for acquiring study subjects

- Investigator who is planning a study reverses the previous sequence:

  1. Specifying the clinical and demographic characteristics of target populations that will serve the research question well.

  2. Uses temporal and geographic criteria to specify the choice of an accessible population that is representative and practical.

  3. Scientific and efficient approach to sampling from the accessible population.
Three steps in designing the protocol for choosing the study subject

**RESEARCH QUESTION**
(Truth in the Universe)

**STEP # 1:**
Target population
Specify clinical And demographic Characteristics
CRITERIA
Well suited to the Research question

**STEP # 2:**
Accessible population
Specify temporal And geographic Characteristics
CRITERIA
Representative of Target population And easy to study

**STEP # 3:**
Intended sample
Design an approached to selecting the sample
CRITERIA
Representative of accessible population and easy to do

---

**STUDY PLAN**
(Truth in the study)

---

**Specification**

**Sampling**
• When investigator wants to do a study he will creating two sets of selection criteria inclusion and exclusion that define the populations to be studied.
Establishing inclusion criteria:

define the main characteristics of the target and accessible populations.

- The task of specifying the clinical characteristics involves difficult judgments about what factors are important to the research question and how to define them.

- The task of specifying the demographic characteristics (age, sex, and race) is even tougher, often involving trade-offs that balance generalizability against efficiency.

- The task of specifying geographic and temporal characteristics of the accessible population also involve trade-offs between scientific and practical goals.
• Establishing exclusion criteria:
  Indicate subsets of individuals who meet the eligibility criteria, but are likely to interfere with the quality of the data or the interpretation of the findings.
• May improve the feasibility of a study at the cost of generalizability, so the investigator should use them sparingly.
• Some exclusions are mandated by ethical considerations, or by the patient's unwillingness to participate.
### Designing Inclusion and Exclusion Criteria

<table>
<thead>
<tr>
<th>Inclusion criteria (be specific)</th>
<th>Considerations</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Specifying the characteristics that define populations that are relevant to the research question and efficient for study:</td>
<td>A 5-year trial of calcium supplementation for preventing osteoporosis might specify that the subjects be:</td>
</tr>
</tbody>
</table>
# Designing Inclusion and Exclusion Criteria

<table>
<thead>
<tr>
<th>Considerations</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target population Demographic</td>
<td>white females age 45-50</td>
</tr>
<tr>
<td>characteristics</td>
<td></td>
</tr>
<tr>
<td>Clinical characteristics</td>
<td>In good general health: no known life-threatening disease.</td>
</tr>
<tr>
<td></td>
<td>not paraplegic or taking long-term corticosteroids</td>
</tr>
</tbody>
</table>
### Designing Inclusion and Exclusion Criteria

<table>
<thead>
<tr>
<th>Considerations</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessible population</td>
<td>Geographic (administrative) characteristics</td>
</tr>
<tr>
<td></td>
<td>Patients attending the medical clinic at the investigator’s hospital</td>
</tr>
<tr>
<td></td>
<td>Temporal characteristics</td>
</tr>
<tr>
<td></td>
<td>between January 1 and December 31, 1989</td>
</tr>
</tbody>
</table>
### Designing Inclusion and Exclusion Criteria

<table>
<thead>
<tr>
<th>Considerations</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exclusion criteria</strong>&lt;br&gt;(be parsimonious)</td>
<td>specifying subsets of the population that will not be studied because of:</td>
</tr>
<tr>
<td></td>
<td>the calcium supplementation trial might exclude subjects who are:</td>
</tr>
<tr>
<td></td>
<td>A high likelihood of being lost to follow-up</td>
</tr>
<tr>
<td></td>
<td>Alcoholic or plan to move out of state</td>
</tr>
<tr>
<td></td>
<td>An inability to provide good data</td>
</tr>
<tr>
<td></td>
<td>Disoriented or have a language barrier</td>
</tr>
</tbody>
</table>
Designing Inclusion and Exclusion Criteria

<table>
<thead>
<tr>
<th>Considerations</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exclusion criteria</strong></td>
<td></td>
</tr>
<tr>
<td>(be parsimonious)</td>
<td></td>
</tr>
<tr>
<td>Ethical barriers</td>
<td>Kidney stone formers (which would contraindicate oral calcium)</td>
</tr>
<tr>
<td>The subjects refusal to participate</td>
<td>unwilling to accept possibility of random allocation to placebo group</td>
</tr>
</tbody>
</table>
Choosing the accessible population

• Presented by two main options to the investigator:

1- Clinic-based samples:
   Select subjects who come to the hospital or clinic.
   Are inexpensive and easy to recruit, but selection factors that determine who comes to the hospital or clinic may have an important effect.

2- Population-based samples:
   Select subjects in their homes producing a sample that is representative of a specified region.
   Useful for guiding public health and clinical practice in the whole community.
• Usually the accessible population is too large or too spread out over time, and there is a need to select a smaller group of individuals for study

• There are two main classes of sampling designs, probability and nonprobability.
• **Probability sampling:**
  - Uses a random process to guarantee that each unit of the population has a specified chance of selection.
  - Statistical significance tests are based on the assumption.
  - There are several versions of the probability approach:
    1) **Simple random sampling** is the process of enumerating every unit of the accessible population, and then selecting the sample at random.
    2) **Systematic sampling** involves selecting by a periodic process, as in the Framingham approach of taking every second person from a list of town residents. It offers no logistic advantages over simple random sampling, and in clinical research it is rarely a better choice.
3) **Stratified random sampling** involves dividing the population into subgroups according to characteristics such as sex or race, and taking a random sample from each of these "strata".

4) **Cluster sampling** is the process of taking a random sample of natural groupings of individuals in the population. Very useful when the population is widely dispersed and it is impractical or costly to list and sample from all of its elements.
Nonprobability sampling:

- more practical than probability designs for many clinical research projects.
- objective is to produce a facsimile, for the search question at hand, of a probability sample.
- There are three main non-probability sampling designs:
  1) **Consecutive sampling** involves taking every patient who meets the selection criteria over a specified time interval or number of patients. The best of the nonprobability techniques, and is very often practical.
2) **Convenience sampling** process of taking those members of the accessible population who are easily available.
   - Widely used in clinical research because of its obvious advantages in cost and logistics.
   - An acceptable choice for some research questions

3) **Judgmental sampling** involves hand-picking from the accessible population these individuals judged most appropriate for the study
Choosing the sample design for selecting study subject from the accessible population

- **Consecutive sample**
  - Often the best option

- **Sample random sample**
  - (To reduce the number in the sample if a consecutive sample would be too large)

- **Stratified random sample**
  - (To increase the size of specified subgroup in the sample)

- **Convenience sample**
  - Judgmental sample
  - (To draw an easy, inexpensive sample when almost any sample will be representative)

- **Cluster random sample**
  - (To draw an inexpensive but representative sample from a large population that is widely spread or difficult to enumerate)
The goals of recruitment

(1) to recruit enough subjects to meet the sample size requirements of the study
(2) to recruit a sample that is unbiased

• The commonest problem of clinical research are:
  1- Falling short in the number of subjects.

Solution:
  a) **Assume** the number of subjects who meet the entry criteria and **agree** to enter the study will be **Fewer**, often by several fold, than the number projected at the outset
• **Solution cont:**

  **b)** Estimate the magnitude of the recruitment problem empirically with a pre-test
  **c)** Tabulate the achievements in meeting the recruitment goals.
  **d)** Examining the proportions of potential subjects lost to the study at various stages can lead to strategies for reducing some of these losses
2- Biased Sample.

Solution:

a) Choose populations and sampling methods wisely.
b) Minimize bias due to technical error or non-response

The response rate:

– The proportion of subjects who agree to enter the study among those who are selected.
– Influences the validity of inferring that the sample represents the population.
Enhancing the response rate:

1. People who do respond
2. People refuse to participate

The level of it that will seriously compromise the generalizability of the study depends on:

– The research question.
– Reasons for not responding.

*The best way to deal with nonresponse bias, is to minimize it at the outset:*

- Improving the efficiency and attractiveness of the initial encounter (*using brochures and individual discussion to allay anxiety and discomfort*).
Providing incentives (reimbursing the costs of transportation and providing the results of tests).

If barriers is the language – by using bilingual staff and translated questionnaires.

3. People difficult to reach
   - Reduced by designing a systematic series of repeated contact attempts, and by using alternative methods (mail, telephone, home visit).
• General recruitment approaches

1- Selecting patients who are already known to the members of the research team, *(e.g. study of a new treatment in patients who are attending the investigators clinic).*

  - The chief concern is to present the opportunity for participation in the study fairly, making clear the advantages and disadvantages.
2- Involves contacting populations that are not known to the members of the research team.

- Many approaches for contacting the prospective subjects:
  - Screening in work settings or public places (shopping malls; using mailings and telephone calls).
3- It is necessary to prepare for recruitment in advance by getting the support of important organizations; for example:

- Meeting with hospital administrators to discuss a clinic-based sample.
- Meeting with the leadership of the medical society and county health department to plan a community screening operation.

4- Endorsements should be obtained in writing and included as an appendix in any application for funding.
5- Endorsements should be obtained in writing and included as an appendix in any application for funding.

6- For large studies it may be useful to create a favorable climate in the community by giving:
   - Public lectures.
   - Advertising through radio, TV, newspaper, or mass mailings.
Errors can occur in either the design or implementation stages.

**Design Errors:**

1. The target population is not well suited to the research question.

**preventive Strategies:**

- Design inclusion criteria that specify the age, sex, and clinical characteristics of appropriate subjects.
- Design a parsimonious set of exclusion criteria that eliminate unwanted individuals.
- Discuss alternatives with colleagues.
2-The accessible population does not sufficiently represent the target population.

preventive Strategies:
• Design inclusion criteria that specify the time frame and the geographic characteristics of a suitable accessible population.
• Discuss alternatives with colleagues.
Design Errors cont...

3- The intended sample does not sufficiently represent the accessible population because of biased sampling design.

**preventive Strategies:**
- Use a consecutive or probability sample whenever practical.
- Otherwise, use good judgment in designing a convenience sample.
- Draw a pretest sample to confirm the characteristics and availability of the subjects.
Implementation Errors:

The actual sample does not sufficiently represent the intended sample:

1- Due to random sampling error (chance).

preventive Strategies:
• Increase the number of subjects
• Consider stratified sampling to enlarge specified subgroups.
Implementation Errors- cont…

1- *Due* to systematic sampling error (bias) Nonresponse bias.

   a) Failure to make Contact

*preventive Strategies:*
- Use repeated attempts and/or alternative approaches.
b) subject refuses to participate

preventive Strategies:

• Provide 'written orientation in advance, and discuss potential source of anxiety.
• Consider incentives such as transportation, reimbursement, and test results.
• Consider special steps to inquire about the nonrespondents.
Implementation Errors- cont…

Mistakes

preventive Strategies:
• Before the study, pretest the sampling methods.
• During the study, control the quality of the recruitment process.
THANK U.