

# *Week 4: Sampling techniques*

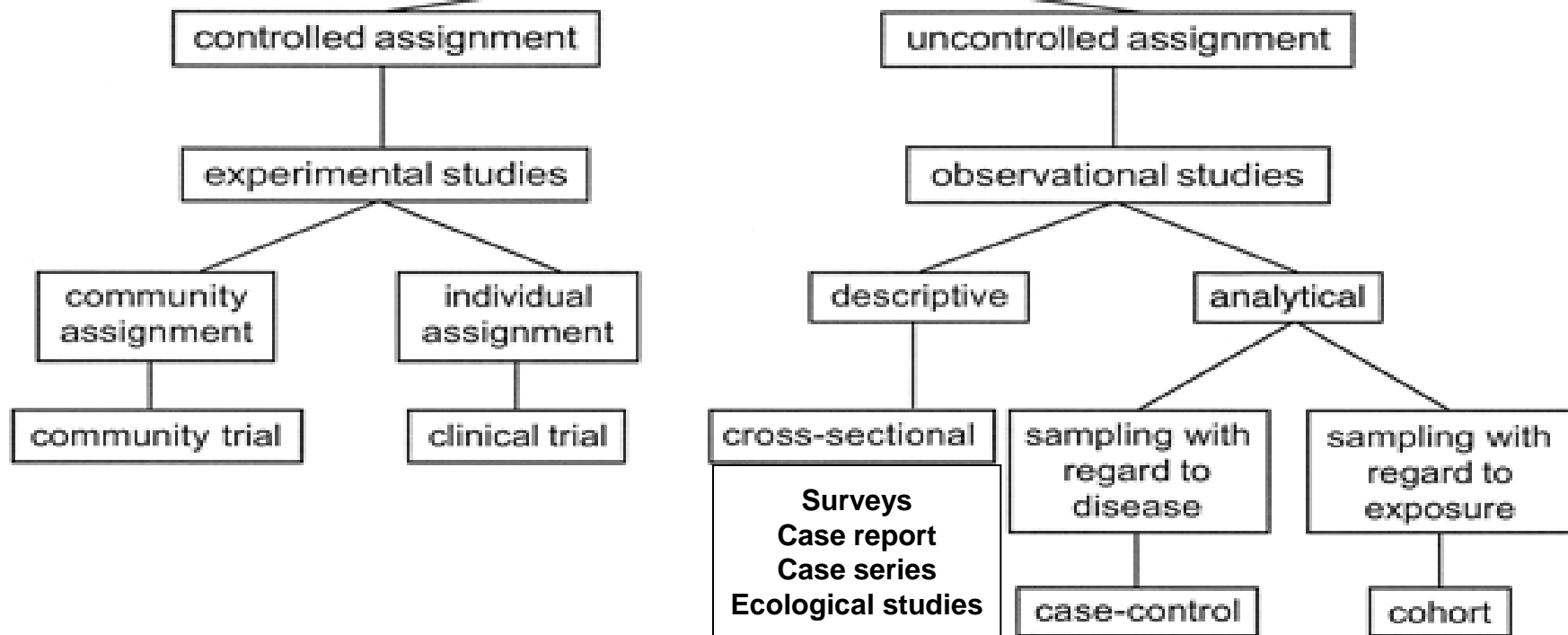
Dr Munir Abu-Helalah  
MD MPH PHD

Associate Professor of Epidemiology  
and Preventive Medicine

# Study Design: Definition

A study Design is a specific plan or protocol for conducting the study, which allows the investigator to translate the conceptual hypothesis into an **operational** one.

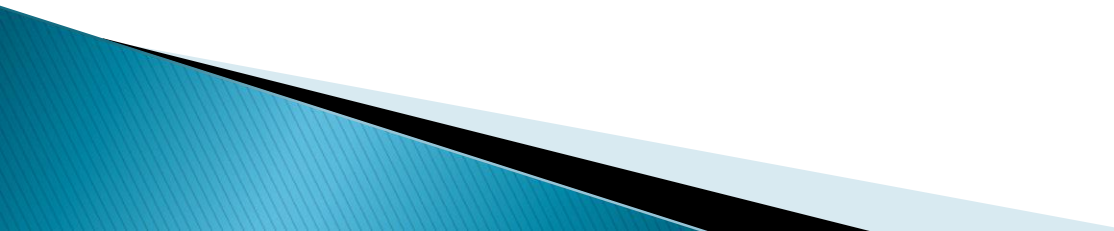
# Epidemiological Study Designs



Source: Waning B, Montagne M: *Pharmacoepidemiology: Principles and Practice*: <http://www.accesspharmacy.com>

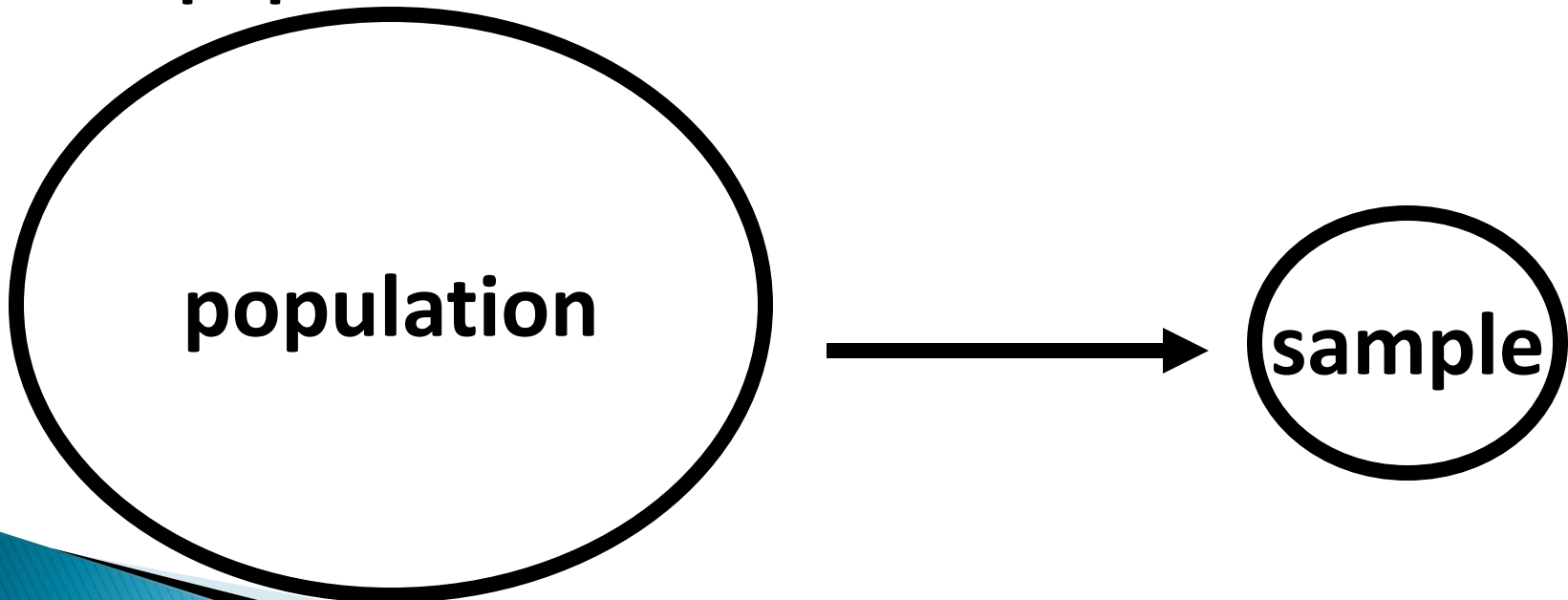
Copyright © The McGraw-Hill Companies, Inc. All rights reserved.

# Sampling techniques

- ▶ A Word About Sampling...
  - ▶ The **population** is all the members of the group you are researching (e.g., all youth in our city)
  - ▶ The **sample** is the selection of the population who will be asked questions
  - ▶ To **generalize** is to state that what you say about your sample can be applied to the rest of the population
- 

# Sampling

**Sampling is a process by which we study a small part of a population to make judgments about that population.**



# Selection of samples

Types of sampling most frequently used in health surveys

- Complete or comprehensive survey of each unit in the population (e.g. nurses in a single hospital)
  
- Probability sample survey
  - Systematic sampling
    - Record reviews
    - Studies of health care workers
  - Cluster sampling
    - Used in surveys of widely dispersed populations

# Definitions

**A study unit** may be a person, a health facility, a prescription, or other such unit.

**The study population**, sometimes called the reference population, is the collection of the entire population of all possible study units. Again, this population may be people, health facilities, prescriptions or other such units.

**A representative sample** has all the important characteristics of the population from which it is drawn.



# SAMPLING METHODS

**A sampling frame is** a list of all of the available units in the study population. If a complete listing is available, the sampling frame is identical to the study population. The method of sampling depends on whether there is a sampling frame available. If a sampling frame exists, or if it can be created, **probability sampling** is used. If there is none available, probability samplings cannot be used.



# non-probability sampling

using non-probability methods is likely to be less representative than a **probability sampling** and so study results are less valid.

# NON -PROBABILITY SAMPLING METHODS

## 1. Convenience Sampling

is a method by which, for convenience sake, the study units that happen to be available at the time of data collection are selected in the sample. This is the least representative sampling method.

# NON -PROBABILITY SAMPLING METHODS

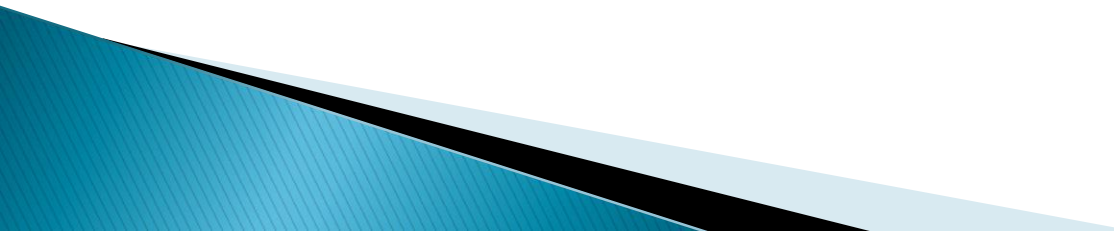
## Quota sampling

is a method by which different **categories** of sample units are included to ensure that the sample contains units from all these categories. For example, a quota sample of patients from a health center that might included 10 patients with diabetes, 10 with diarrhea, and 10 with malaria.

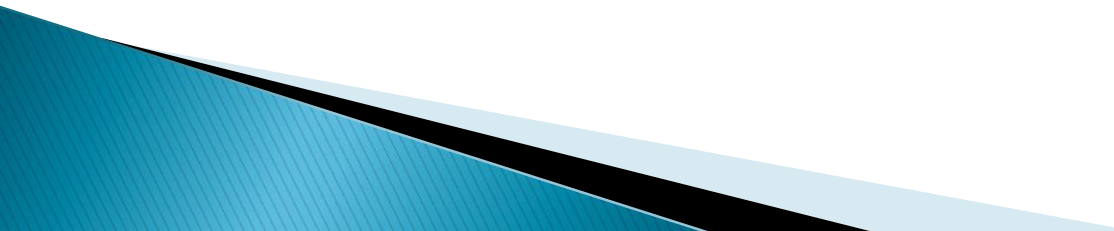
Quota sampling is a method of sampling widely. Interviewers are each given a quota of subjects of specified type to attempt to recruit for example, an interviewer might be told to go out and select 20 adult men and 20 adult women, 10 teenage girls and 10 teenage boys so that they could interview them about their television viewing.

It suffers from a number of methodological flaws, the most basic of which is that the sample is not a random sample and therefore the sampling distributions of any statistics are unknown.

# Types of Probability Samples

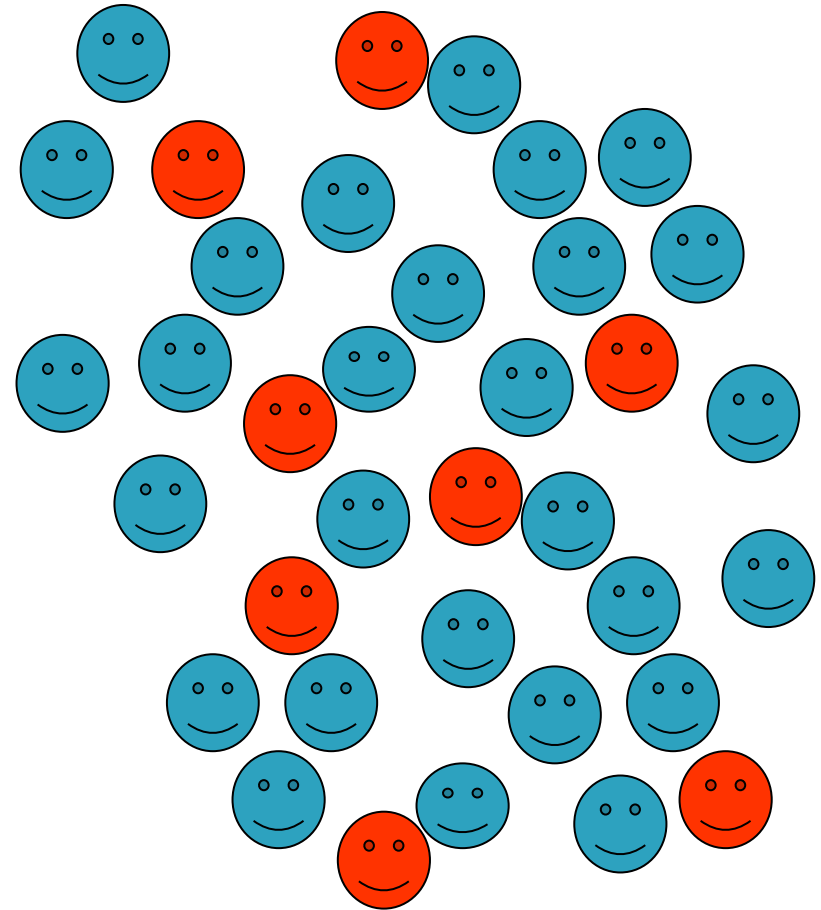
- **Simple Random**
  - **Systematic Random**
  - **Stratified Random**
  - **Random Cluster**
  - **Stratified Cluster**
  - **Complex Multi-stage Random**  
(various kinds)
- 

# 1. Simple Random Sampling-1

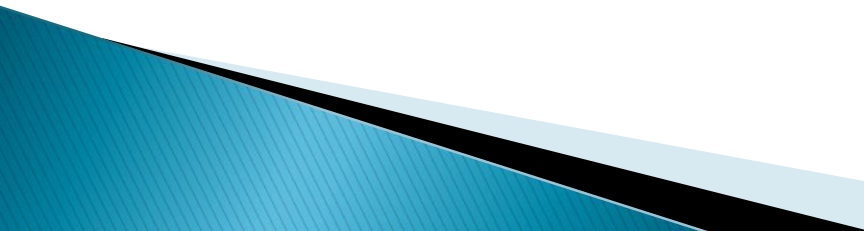
- a) Make a **numbered list** of all units in the reference population from which you will select the sample (for example, a list of all the health centers in the country).
  
  - b) Decide on the size of the sample (say 20 facilities).
  
  - c) Choose the facilities to include by a lottery method. (For example the numbers of all the facilities can be placed in a **box** and drawn, a **random number table** can be used, or random numbers can be generated using a **spreadsheet** or **calculator**.)
- 

# Simple Random Sampling-2

- Each element in the population has an equal probability of selection AND each combination of elements has an equal probability of selection
- Names drawn out of a hat
- Random numbers to select elements from an ordered list



# How to select a random number?

- Flip a coin
  - Choose a number from a 'hat'
  - Bank note
  - Calculator
  - Computer
  - Table of random number
- 

## 2. Systematic Sampling

In systematic sampling, sample units are selected from a numbered list of all units in the study population by using a regular interval, starting from a random sampling starting point.

To calculate the **sampling interval**,

- Determine the total number of units in the population

- Determine the sampling interval

$$= \frac{\text{number of units}}{\text{desired sample size}}$$

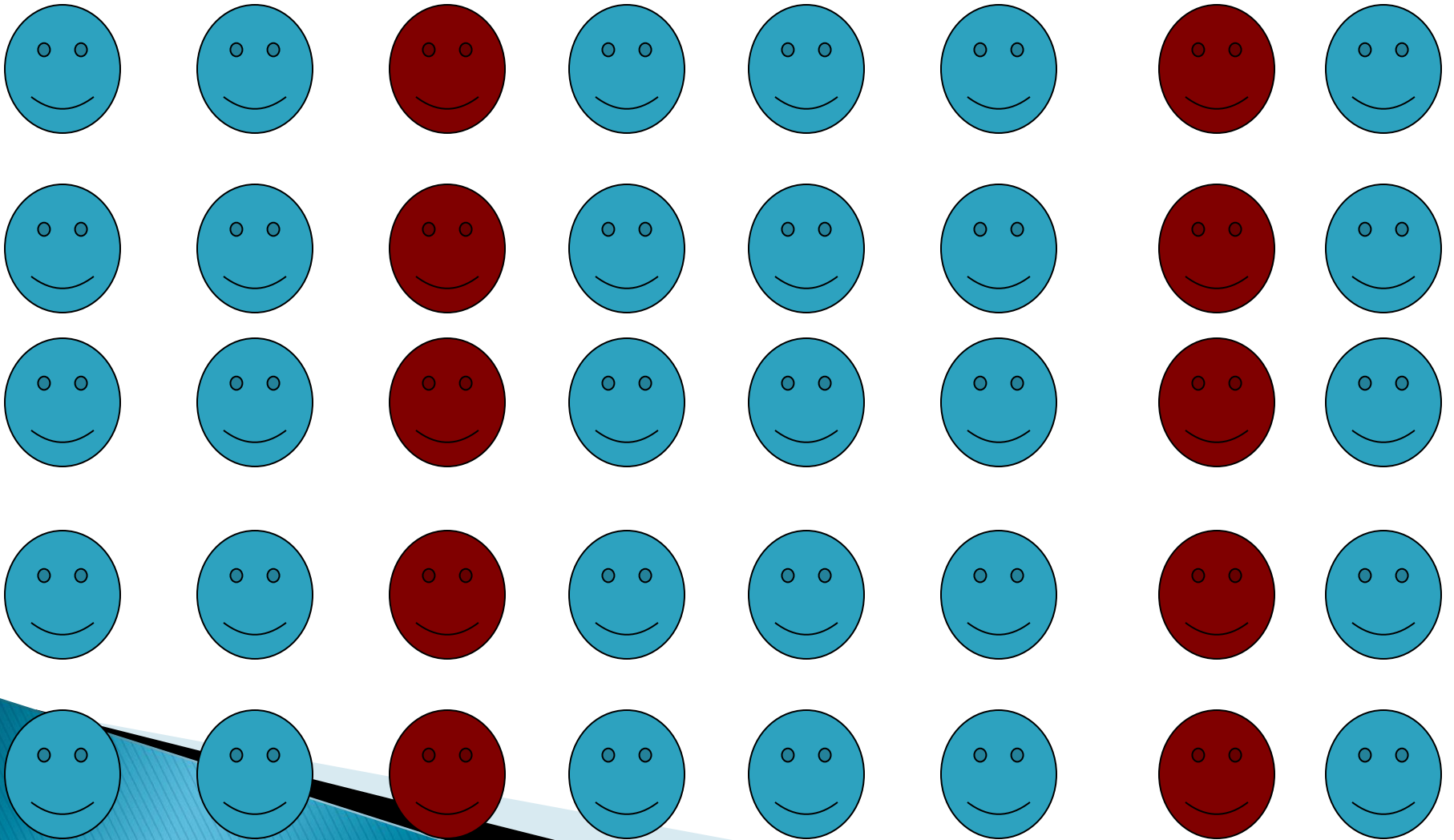
For example, if we want to select 20 health centers from a list of 46 in our sampling frame, our sampling interval would be  $46/20 = 3$ .



**The first facility chosen in this case can be 1, 2 or 3, which are all the possible sampling units within the first sampling interval. This is selected by choosing a random number with one digit less than or equal to the sampling interval.**

**Later facilities are selected by adding the sampling interval to the previous result. If the first result was 3. then the next facilities selected would be facility 6, 9 and so forth. The method just described gives every unit an equal chance of being selected.**

Sample 12, sampling interval=48/12=4



**Example:**

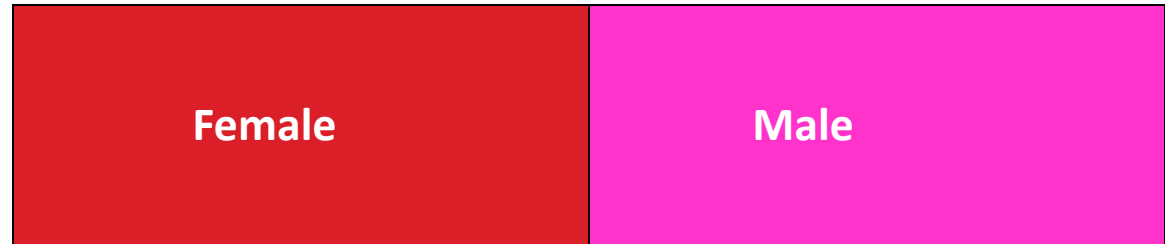
**Assume you are doing a study involving children under 5. There are a total of 1500 households, and you have a required sample size of 100 children. From a preliminary study you have done, there is one child every 2.5 households. Therefore you would need to visit  $100 \times 2.5$  or 250 households to find the required 100 children.**

***sampling interval =  $1500 / 250 = 6$  (Visit every 6<sup>th</sup> household)***

- **select a number between 1 and the sampling interval**
- **add the sampling interval to the chosen starting point to obtain the second sampling unit, add the interval to the second unit ....**

# 3. Stratified Sampling-1

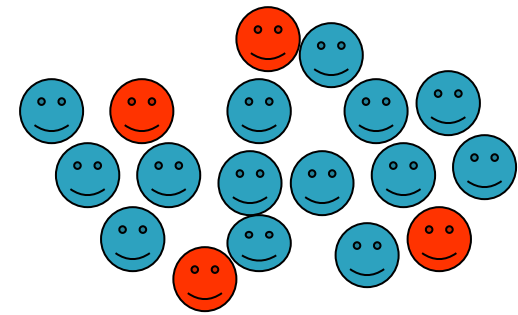
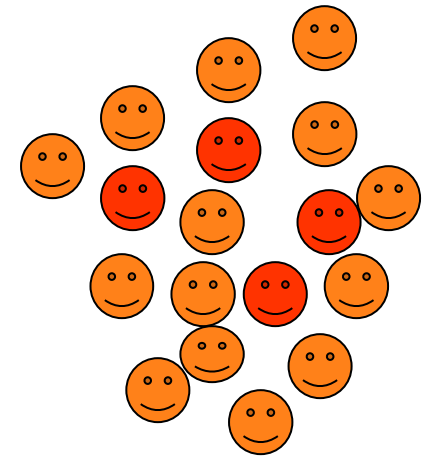
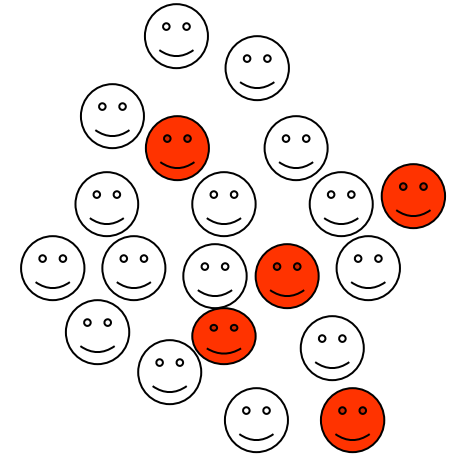
- Stratified sampling is used when the reference population contains clearly different sub-populations, which should be considered separately.



- When stratified sampling is used, the sample frame (the list of the overall population) is sorted into two or more groups. These different strata (groups) may then be sampled either randomly or systematically.
- Basis for grouping must be known before sampling
- Select random sample from within each group

# Stratified Sampling-2

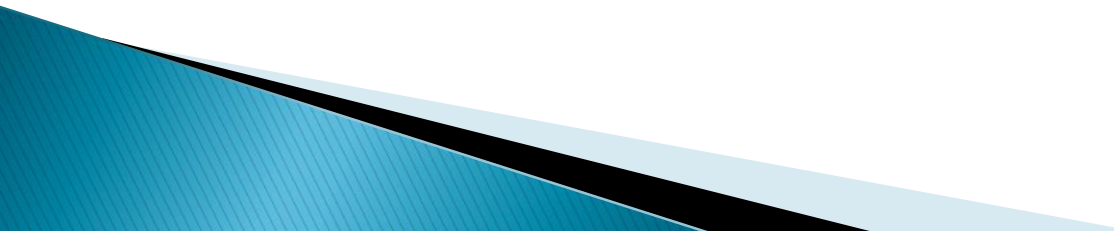
- For a given sample size, reduces error compared to simple random sampling IF the groups are different from each other
- Tradeoff between the cost of doing the stratification and smaller sample size needed for same error
- Probabilities of selection may be different for different groups, as long as they are known
- Over sampling small groups improves intergroup comparisons



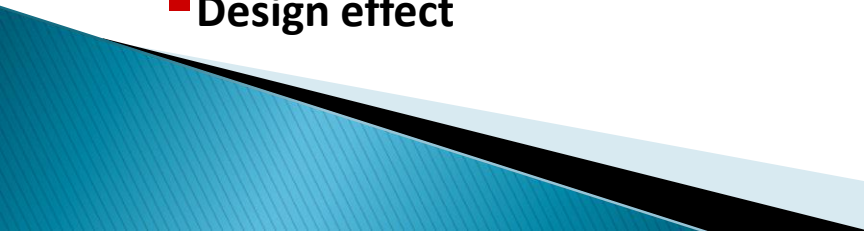
# 4. Cluster Sampling

- **Cluster sampling: Dividing the population into subgroups called clusters (not as homogeneous as strata), randomly sampling clusters, and then possibly selecting a random sample of people in each cluster.**
- **In a cluster sample, a group of sample units is selected together, rather than each unit being selected separately.**  
(Sampling unit is a group of individuals) e.g.
  - Households
  - Health centers
  - Schools
  - Village
- **Selection with probability proportional to size**

e.g. EPI WHO sampling procedure of selecting 30 groups of 7 children is a common cluster sampling method.

- The main **advantage** of cluster sampling is that the method is easy to use and often logistically simpler to organize.
  - The **disadvantage** is that the samples selected may be less representative especially when the number of clusters selected is small. As a rough guide, double the sample size if cluster sampling is used.
- 

# Steps in selecting a cluster sample

- Enumerate all population concentrations in the sampling universe
  - Draw up a cumulative population list
  - Determine the sampling interval
  - Pick a number between 1 and the sampling interval from a random number table
  - Add the sampling interval to the chosen starting point to obtain the second cluster, add the interval to the cluster
  - Select individuals =  $\text{sample size} / \text{number of clusters}$
  - Design effect
- 



# Stratification vs. Clustering

## Stratification

- Divide population into groups different from each other: sexes, races, ages
- Sample randomly from each group
- Less error compared to simple random
- More expensive to obtain stratification information before sampling

~~Less feasible than clustering~~

## Clustering

- Divide population into comparable groups: schools, cities
- Randomly sample some of the groups
- More error compared to simple random
- Reduces costs to sample only some areas or organizations

More feasible but less representative when compared with stratified sampling

## 5. Multistage Sampling

In multistage sampling, the methods described above can be combined. For example, we might wish to select 32 health facilities in a country containing 56 districts, each of which contains a number of health facilities. From the 56 districts, **16 districts** would first be selected. In each district **two health facilities** would then be randomly selected. This would be two-stage random sampling.

# Example 1

- ▶ Prevalence of stress among medical students

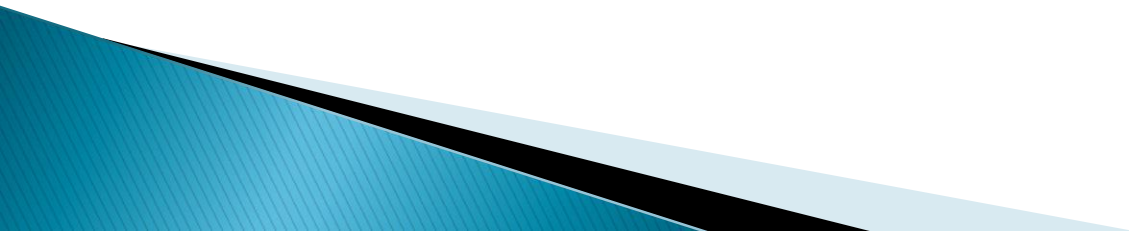
# Example 2

- ▶ Satisfaction of postgraduate students with studying at Jordan University

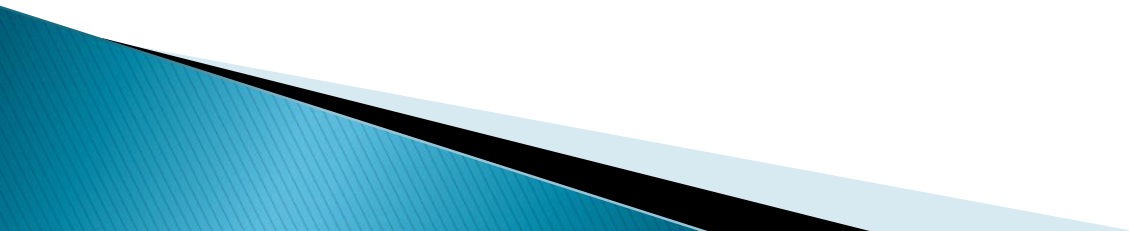
# Example 3

- ▶ Prevalence of adult hypothyroidism in Jordan

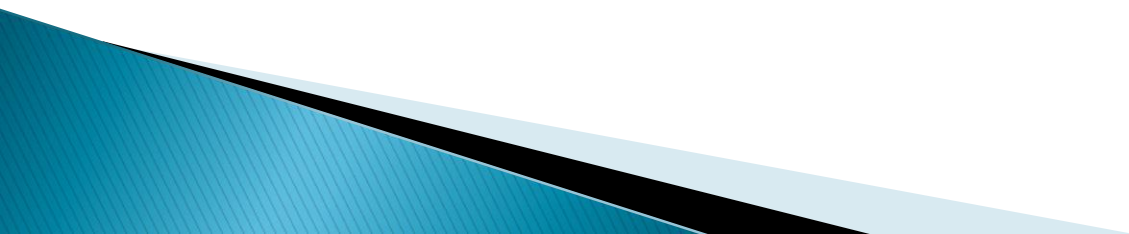
# Complications rate of type II diabetes mellitus in Jordan



# Stress level among medical residents at Jordan University Hospital



# Satisfaction of healthcare professionals working at Jordan University hospital





**Thank you!**

