Yarmouk University

Community Medicine

Lec. 4 - Data collection tool development, Sampling methods, quality assurance (Part 2)

Dr. Ola

Written By : Group C4 - Mahmod Omari

If you come by any mistake , please kindly report it to shaghafbatch@gmail.com

Wording

Every word that you put in the question is important

It should be clear words (doesn't have more than 1 meaning)

- **Clarity.** Make questions as clear and specific as possible.
    - For example, asking, “How much exercise do you usually get?” is less clear than asking: “During a typical week, how many hours do you spend in vigorous walking?” -the answer could be non specific like : alot
    - here we defined the type of exercise → “vigorous walking”
    - And ”usually get” → by giving time frame (typical week)
    - And defined the frequency by how many hours

*Emotions and behaviors are subjective somehow which are much harder to ask for

Wording

- **Simplicity.** Use simple, common words (every one can understand and doesn't have ambiguity) and grammar that convey the idea, and avoid technical terms and jargon.
    - For example, it is clearer to ask about "drugs you can buy without a prescription from a doctor" than to ask about "over-the-counter medications." this in US

    ** jargon : special words or expressions used by a profession or group that are difficult for others to understand.

    ** in reality antibiotic is not an OTC

- **Neutrality.** Avoid "loaded" words and stereotypes that suggest a desirable answer. *don't lead him to the answer
    - Asking "During the last month, how often did you drink too much alcohol?" may discourage respondents from admitting that they drink a lot of alcohol.
- Leading here is " too much alcohol" ** I should ask → how many cups of alcohol do you drink during an average week

Setting the Time Frame

- To measure the frequency of the behavior(emotion or diet) it is essential to have the respondent describe it in terms of some **unit of time.**
- Why the time is important→ because the emotions/ behaviors are changeable from time to time so by giving a time framing the answer could be giving by averaging the behavior or emotions during this period of time

During the last 7 days, how many cigarettes did you smoke (one pack is equal to 20 cigarettes)?

[ ] cigarettes in the last 7 days

*this is an open question and the data type is continuous (continuous data give you more info)

* Try to shorting the time frame to reduce recall bias (not remembering)

- The investigator must first decide what aspect of the behavior is most important to the study: **the average** or **the extremes**.

Avoid Pitfalls

- **Double-barreled questions.**(asking about 2 things in the same question) Each question should contain only one concept.
  - "How many cups of coffee or tea do you drink during a day?" ask separately
- **Hidden assumptions.** Sometimes questions make assumptions that may not apply to all people who participate in the study.
  - For example, a standard depression item asks how often, in the past week: "I felt that I could not shake off the blues even with help from my family."
  - Shake the blue : get rid of depression
  - Here the assumption is that his family helped them , maybe not his family his friend , or he doesn't even have a family.

What should the questionnaire look like?

- In general, questions should be short

and to the point (around 12 words or less).

- Sometimes its ok to have long explanation for interview

- Physical layout of their questionnaire

as the font size, color, or question order. → بساعد البارتيسيانت
انه يعبيه بسرعه وسهوله وما يزهق

Summary points

Questionnaire studies often fail to produce high quality generalisable data

When possible, use previously validated questionnaires

Questions must be phrased appropriately for the target audience and information required

Good explanations and design will improve response rates

For new questions and scales

- To make sure your Questionnaire is good
- **Pretest :** by small sample group(pilot studying) and take there feedback for any problems or any unclear questions
- Pretest the instrument clarity and timing.

- **Validate**
- Questionnaires and interviews can be assessed for validity (an aspect of accuracy) and for reproducibility (precision)
- Can be assessed by interviews ,expert views , focus group , or pilot study ,etc.
- I can take more than one method to assess it

Questionnaire administration methods

- Self reported by mail or online.
- Interview by face to face, telephone based , or virtual.
- Group or focus group interview.

- Self → I give you a paper or online link and you fill it by your self
- interview→ I read the question for you and I report the answer (similar for taking history)

Population Sampling

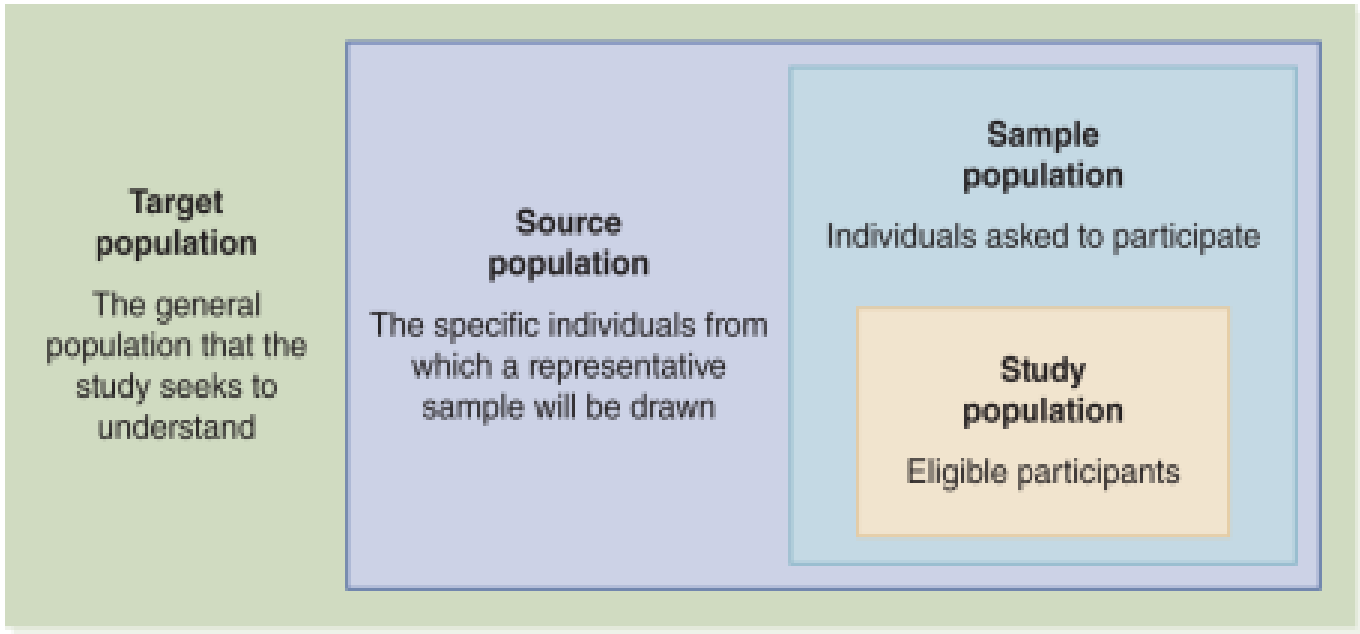
الصراحه الدكتوره شرحتها بطريقه أخرى بس بتوقع فيها أخطاء والكتاب طريقه ثالثه
وشككككرننن ☺

- We cant sample the entire population …. For a study.

- The source population,

sometimes called a **sampling frame**, is a well-defined subset of individuals from the target population from which potential study participants will be sampled.

FIGURE 19-1 Types of Populations



From the target population we take the source sample according(to the research question exclusion and inclusion criteria **not sure**) and for whom I have access

Then Sampling methods are applied for the source sample to extract the sample population (Probability or non probability based)

Who accept to be part of the study are the study population

For any participation I should get a consent from the participant

most important characteristic of the sample is being representative of the target population

I cant compare all the variables between the sample and the target population, so I compare the basic characteristics like avg age, gender, income, etc.

- **Sampling methods classification:**
- Non-Probability Sampling Methods
- Probability Sampling Methods

Probability : the chance of getting or selection of one individual

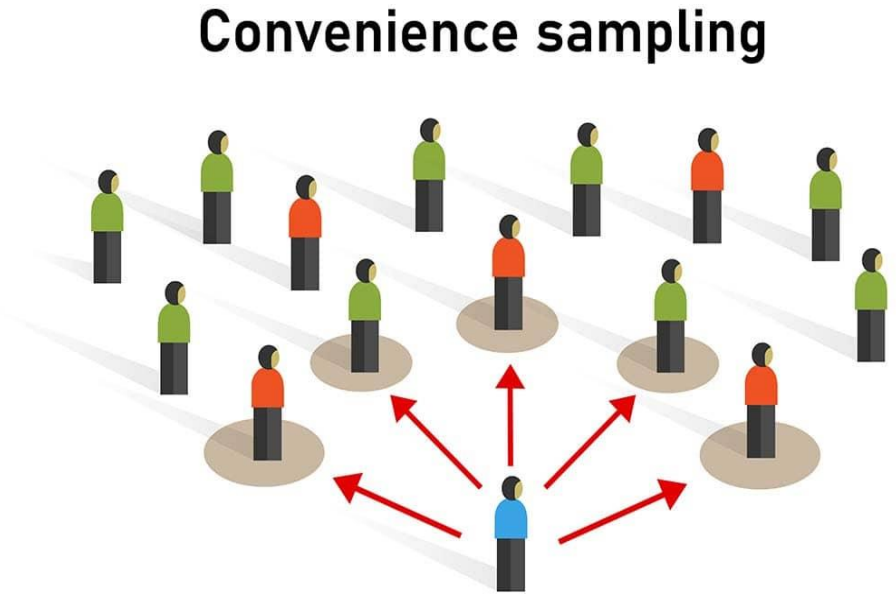
Non-Probability Sampling Methods

- Convenience sampling
- Quota sampling
- Purposive Sampling
- Snowball sampling

Convenience sampling

- You chose based on convenient (you chose the sample that is easy for access)
- Select any one available.

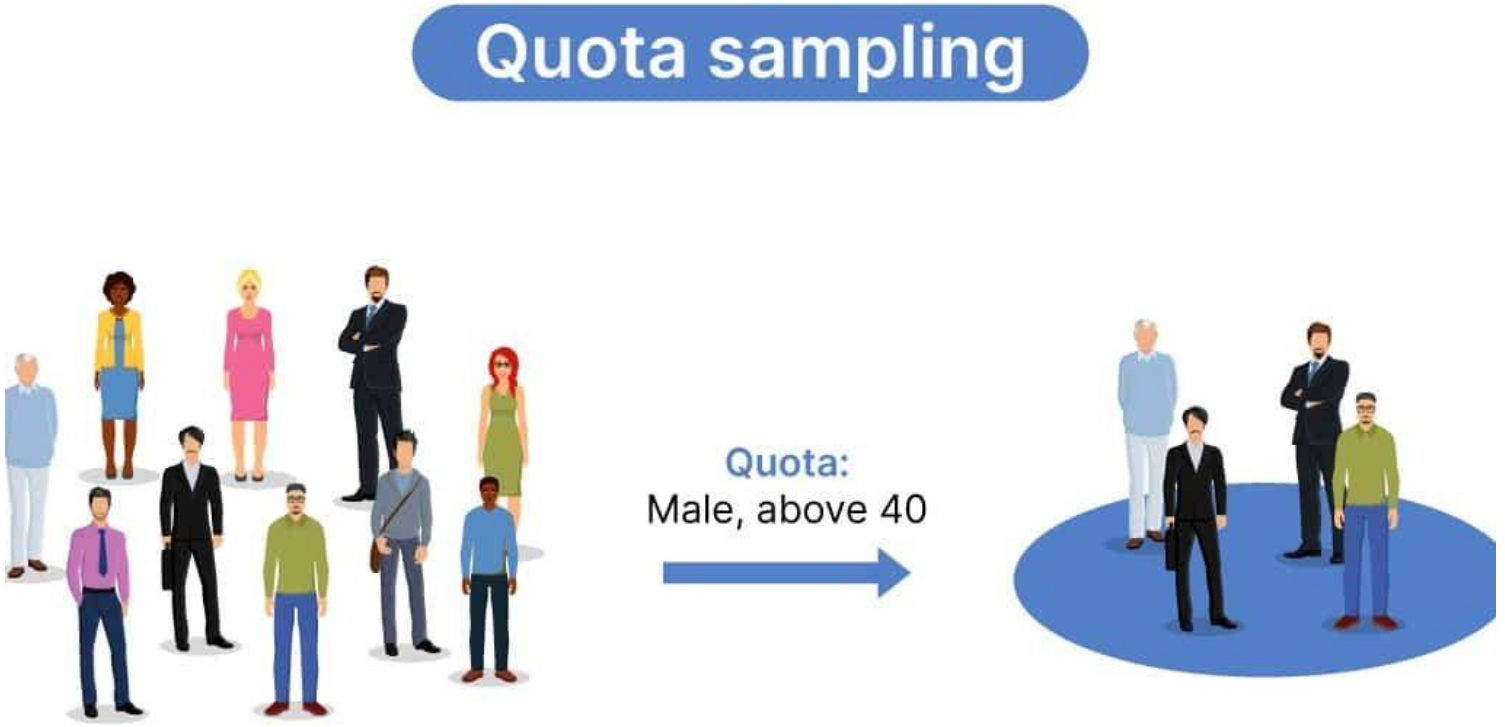
على سبيل المثال لو انا بقاعه بختار الي قدام عشانهم اقرب الي



Quota sampling

I have some general restrictions of my sampling criteria like age gender ethnicity

على سبيل المثال بختار جنيرال كركترسيتك (انثى) وبختار منهم عشوائي



Purposive Sampling

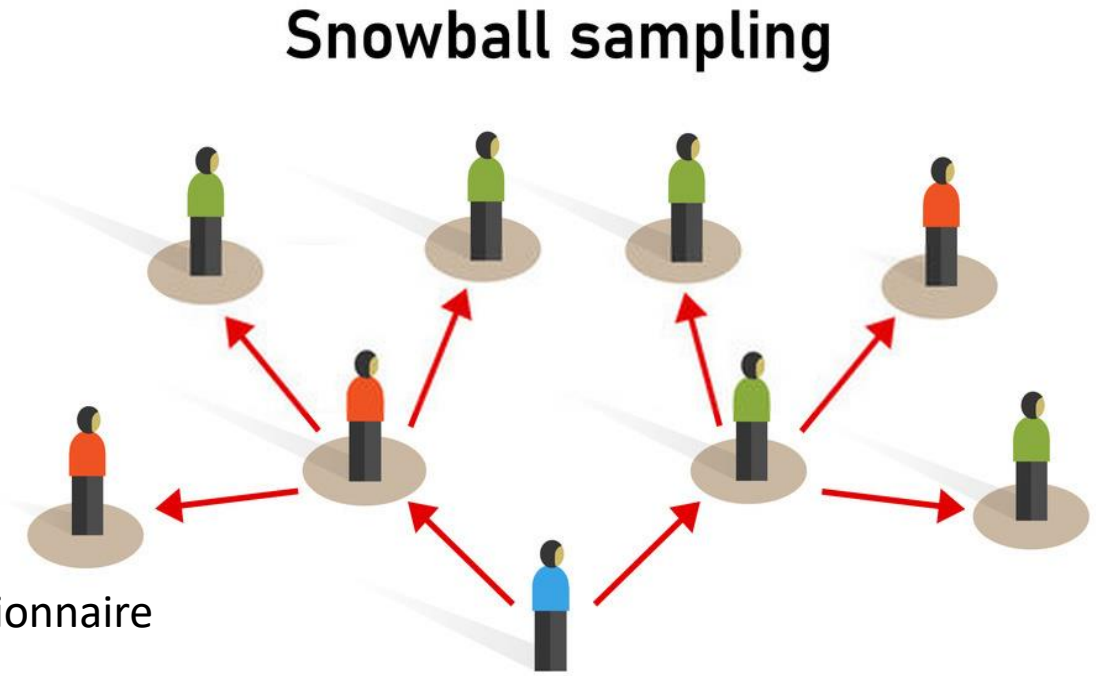
Selective sample based on a purpose

على سبيل المثال بختار حدا لانه خبير بالموضوع تبع الريسيرش
او بختار الناس الي بالمناطق الفقيره لانه دراستي عن النظافه

Snowball sampling

Like social network

Existing subjects are asked to nominate further subjects known to them, so the sample increases in size like a rolling snowball.



Questionnaire

It start by one we call him seeder
And the act of distribution
called seeding

مثلا بحكي لصاحبي عبيلي الكوستشنير
وخلي اصحابك يعبوه واحكي لاصحابك همه
كمان همه يوزعوه وهكذا

Snowball sampling

- Most online questionnaire especially during COVID was snowball sampling

- Pros: you can reach a large group of people
- Cons: it depends on the seeder (if he is socially active he will get more participant)

- All Non-Probability Sampling technique are biased without any exceptions
- We try to minimize bias by having a large sample size

Probability Sampling Methods

- Simple random sampling
- Systematic sampling
- Stratified sampling
- Clustered sampling
- Multi-stage sampling

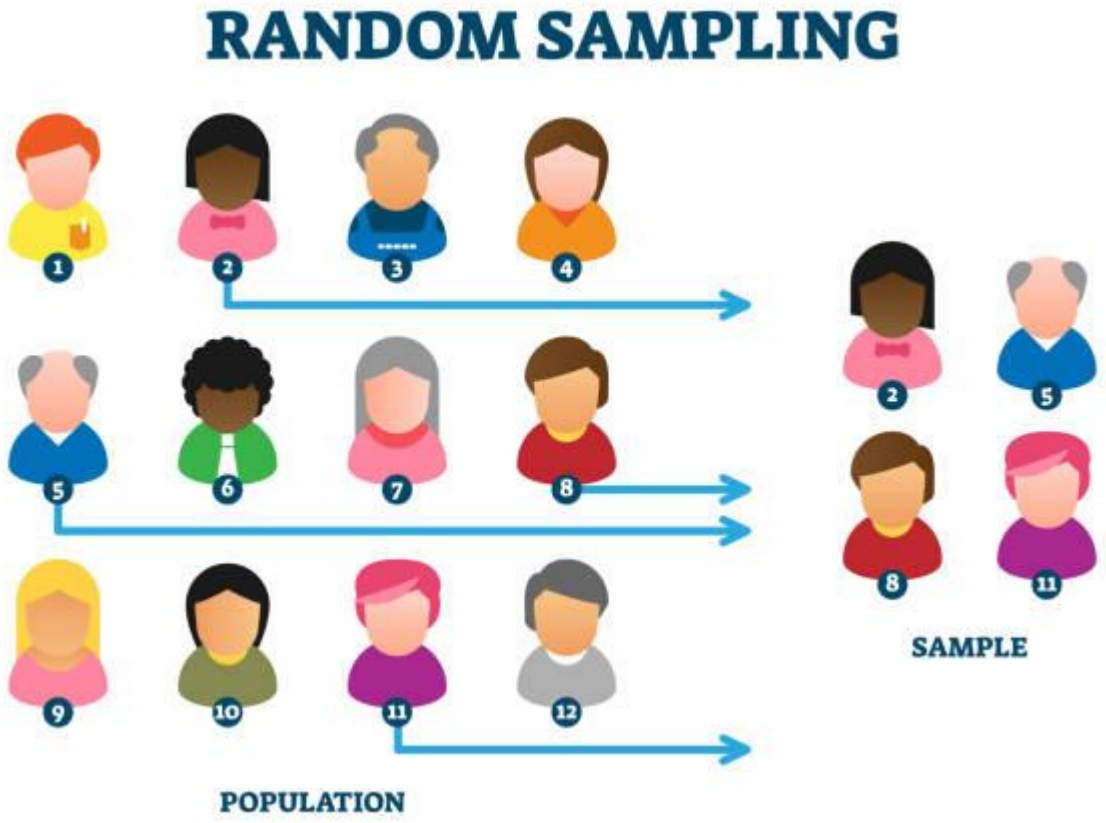
- Should consider the selection probability (your chance of selection a certain individual)

- survey weight = mean * probability of selection

Simple random sampling

The easiest one

should have list of the whole population and chose from them by random way



In this case each individual is chosen entirely by chance and each member of the population has an equal chance, or probability, of being selected.

Method to choose by chance:

- Toss a coin.
- Random Number generator software.

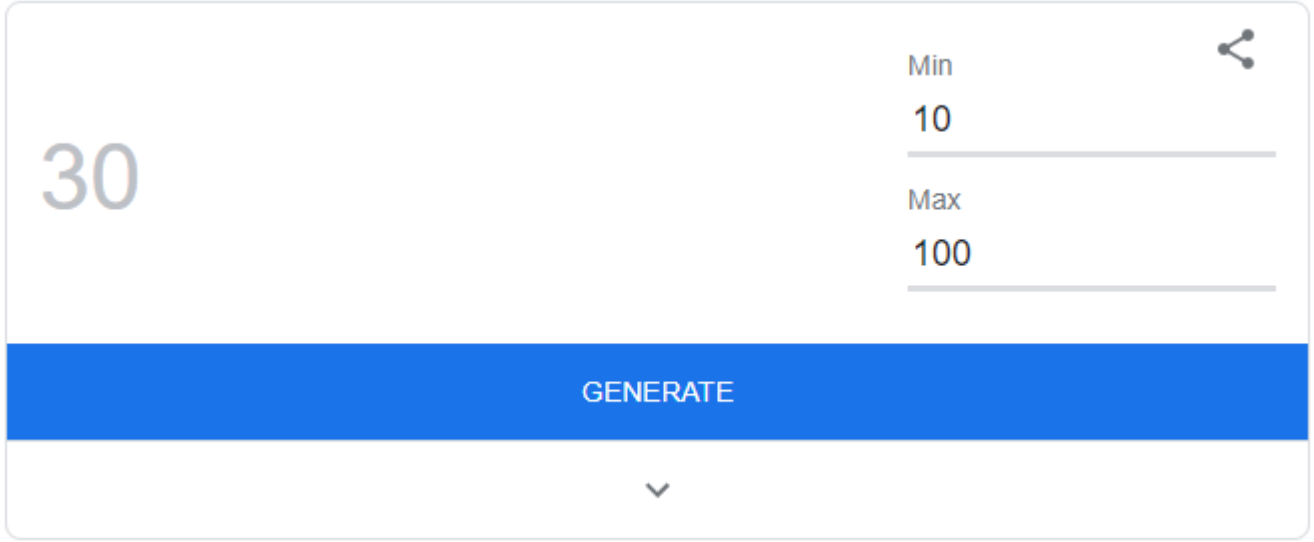
Random number generator

Have a list of individuals name (total).

Determine your sample size (n).

Press generate, pick the person with the number shown.

بدخل البيوليشن وبختار عدد السامبل الي
بدي إياها وهو بختارلي عشوائي

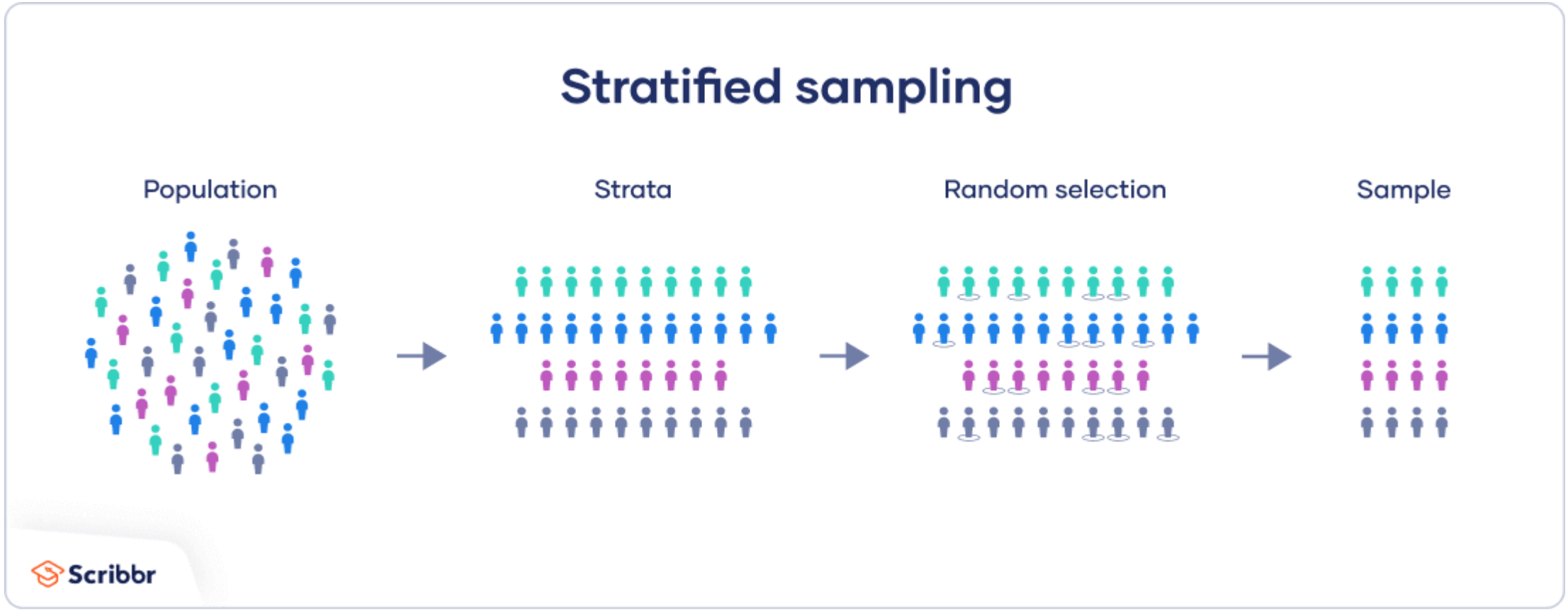


Systematic sampling

for example I take the list then I take 1 and skip 2
mostly you end with representative sample

Stratified sampling

this method is useful when there is underrepresented group (minority)



we divide the population into homogenous subgroup (strata) then we chose randomly within each stratum

مثلا بقسم الناس ل ناس اردنين، ناس مش اردنين، ناس اهاليهم مسافرين وبختارمن كل مجموعه عشوائي

Clustered sampling

- In a clustered sample, subgroups of the population are used as the sampling unit, rather than individuals.

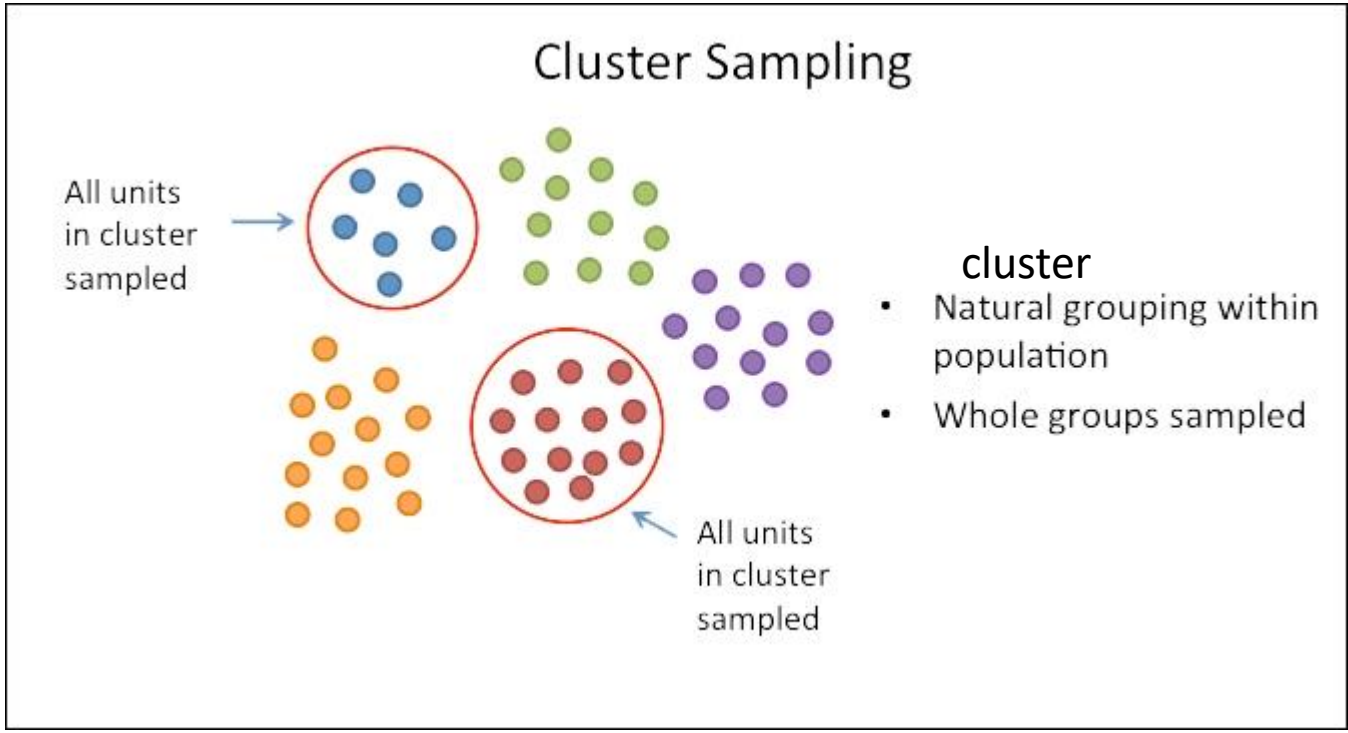
useful in income study because there is natural cluster

في مناطق غنية ومناطق متوسطه ومناطق فقيره فمثلا بوخذ جزء من المناطق الغنية وجزء من الفقيره وجزء من الوسط

so it give you more representative sample for clusters

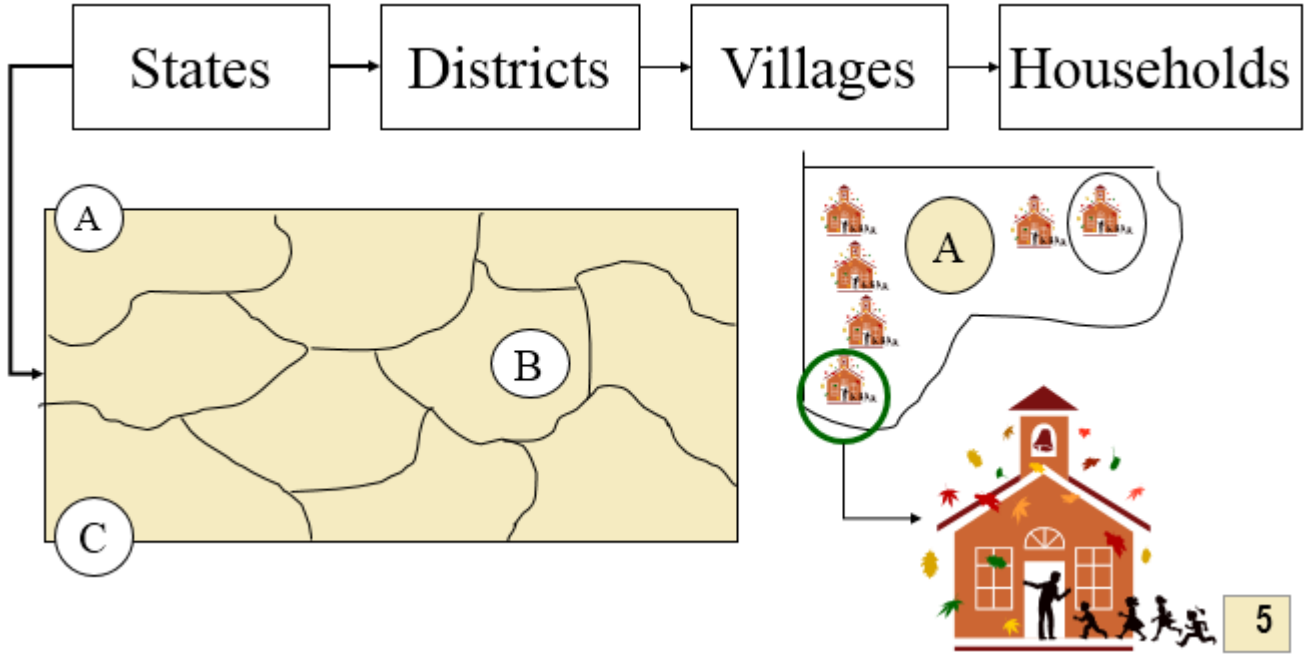
**hospital, school, university based sampling have cluster effect

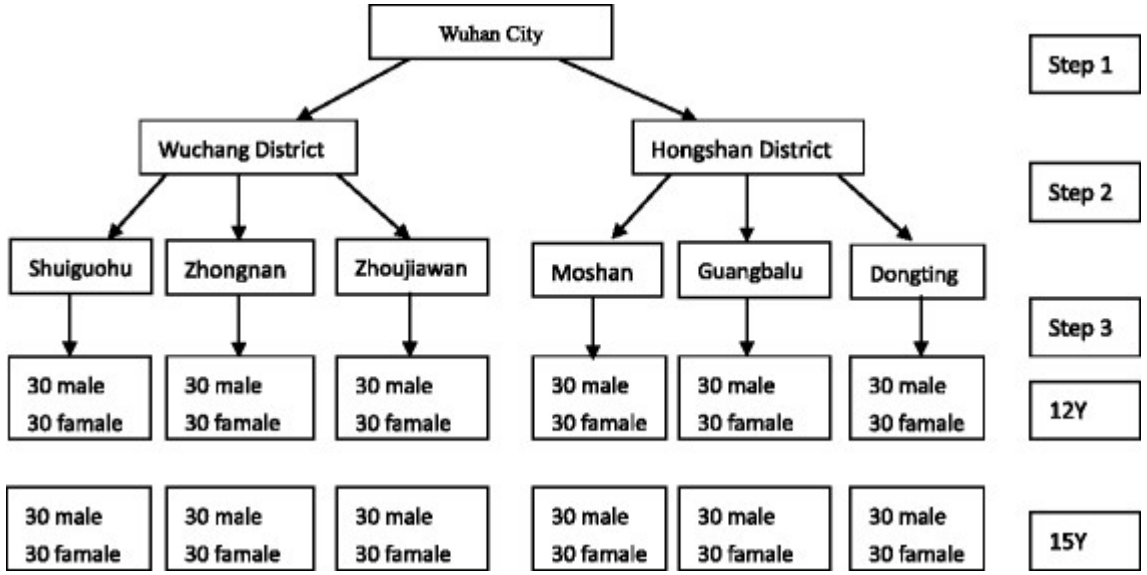
**I can take the whole group or just a part



Multi-stage sampling

in survey mostly we don't have one sampling stage
the total probability multiplied by every step probability to calculate the sampling weight





لو بدي اخذ سامبل ببلش بال (state) الأردن بعدين ببلش بال district المحافظات بختار منهم عشوائي مثلا بدي 100 من اربد و 100 من الزرقا و100 من عمان , بعديها بنتقل للvillage مثلا الويه وبرضه بختار عشوائي ومن الالويه بختار عشوائي households من كل بيت بوخذ شخص او 2

every step in random way

Sample Quality

- **Sample bias:** the sample is not representative for the population. Sample is different in its' characteristics compared to the population.

Like non-probability sampling

- **Sampling bias may be introduced when:** there is an error in the sampling technique
- Poor recruitment methodology
- Poor sampling methodology

- **How to avoid:** is to stick to probability-based sampling methods.
- Or take very large sample

Types of sampling bias

لسا رح نوخذهم بعد بس احفظ type

Type	Explanation	Example
Self-selection	People with specific characteristics are more likely to agree to take part in a study than others.	People who are more thrill-seeking are likely to take part in pain research studies. This may skew the data.
Non-response اخترته ولكن رفض يقابلني او انسحب من الدراسه	People who refuse to participate or drop out from a study systematically differ from those who take part.	In a study on stress and workload, employees with high workloads are less likely to participate. The resulting sample may not vary greatly in terms of workload. →
Undercoverage	Some members of a population are inadequately represented in the sample.	Administering general national surveys online may miss groups with limited internet access, such as the elderly and lower-income households.

We should account for there
I should study the age and gender characteristic and compare with my sample
If there is a significant difference
It can affect my result

Survivorship	Successful observations, people and objects are more likely to be represented in the sample than unsuccessful ones.	In scientific journals, there is strong publication bias towards positive results. Successful research outcomes are published far more often than null findings.
Pre-screening or advertising	The way participants are pre-screened or where a study is advertised may bias a sample.	When seeking volunteers to test a novel sleep intervention, you may end up with a sample that is more motivated to improve their sleep habits than the rest of the population. As a result, they may have been likely to improve their sleep habits regardless of the effects of your intervention.
Healthy user	Volunteers for preventative interventions are more likely to pursue health-boosting behaviors and activities than other members of the population.	A sample in a preventative intervention has a better diet, higher physical activity levels, abstains from alcohol, and avoids smoking more than most of the population. The experimental findings may be a result of the treatment interacting with these characteristics of the sample, rather than just the treatment itself.

Questions

Note in EMR we use random sample technique or clustering by clinic or Stratified sampling