

Yarmouk University

# Community Medicine

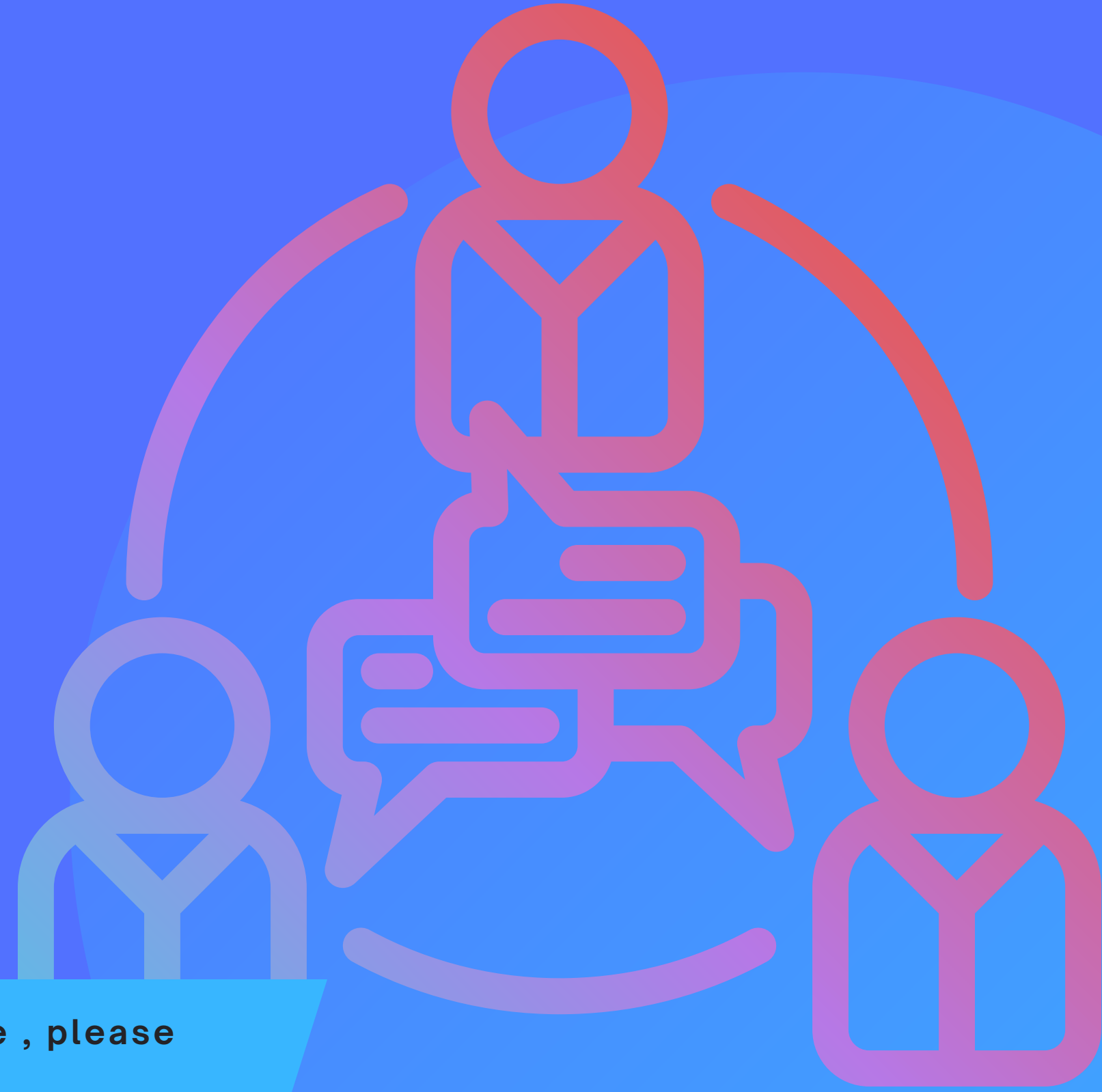
Lec. 4 - Data collection tool  
development, Sampling methods,  
quality assurance ( Part 1 )

Dr. Ola

Written By : Group C4



If you come by any mistake , please  
kindly report it to  
[shaghafbatch@gmail.com](mailto:shaghafbatch@gmail.com)





## Data collection tool development, Sampling methods, & Quality Assurance

MED 410 - Lec 4 (Part 1).

Dr. Ola Soudah

Scripted by: Motaz Hanandeh

كلام الدكتوراة مكتوب باللون الأحمر

# Data collection tool

- **Data:** means information collected as an image, text, vitals, numbers, or figures.
  - Any thing that can be collected and stored is named as data. For ex: Labs, radio images, examination, physician notes are data in hospitals and health care centers.
- **Data sources are classified into two types:**
  - **Primary data:** first-hand information collected by an investigator; It is **collected for the first time**.
    - **Pros:** more reliable, original, **more related to answer my research question** and **more comprehensive**.
    - **Cons:** **time consuming**.
  - **Secondary data:** refers to second-hand information; It is not originally collected and rather **obtained from already published or unpublished sources**.
    - **Cons:** 1. **the power (sample size) of the questionnaire isn't designed for your questionnaire → scientists take very large sample → the outcomes become more valid.**  
2. **may need time to restructure it to fit your research.**

# Cont...

- **Ex:**
  1. If you do your own questionnaire → **Primary data.**
  2. If a scientist made a questionnaire which is fit with your research, and you used it after taking permission → **Secondary data.**
- **Notes:**
  - **Validity:** it means accuracy and correctness of your data; you measure what you want to measure.
  - **Reliability:** it means consistency of your data; every time you ask a question, you will get the same answer.

## • The main sources of health data are:

- **Surveys** (the most commonly used).
- **Electronic Medical records (EMR):**
  - Everything happen to the patient in the hospital.
- **Claims data or administrative data.**
- **Vital records:**
  - a government document containing information about a person's important life events (**it is measured in real time (continuously)**). Ex: birth certificates, death certificates and daily coronavirus case.
- **Surveillance**, there are two types:
  - **Active:** ministry of health measures the numbers of certain diseases according to its interest. Ex: children obesity, purity of water, DM patient...
  - **Passive:** reporting to ministry of health about certain diseases. Ex: outbreaks, food poisoning...
  - The difference between surveillance and vital record: surveillance is done annually, every two years or according to agenda of ministry of health not in real time.
- **Disease registries:**
  - Every single thing about a specific disease will be recorded in a registry (diagnosis, treatment, labs...). Ex: In Jordan, Hussain cancer center records all cancer cases and their information in a registry.
- **Peer-reviewed literature:**
  - what are published like in **PubMed**.

• بمجرد ما عملت تحليل COVID رح يزيد الريكورد.

# Surveys

- Surveys are an important means of collecting health and social science information from a sample of people **in a standardized way** (tested, validated and accurate) to better understand a larger population.
- **Pros:**
  - Collect empirical data in a relatively **short period of time** (that's why it's the most used).
  - Surveys can collect data on a **representative sample of people**, particularly when samples are randomized, or purposive nonprobability sampling is used.
    - **If you take a larger sample size, it will be more representative for the population.**
- There are many **methods used to conduct surveys**, including **questionnaires** and **in-depth interviews** via phone, mail, email, and in-person.

# Electronic Medical records

- Electronic Medical records (EMR) or electron health records (EHR) track events and transactions between patients and health care providers, **real-time data**.
  - **EMR:** restricted for the hospital.
  - **EHR:** shareable with other health institutions.
- **EMR** Contain a patient's medical history, diagnoses, medications, treatment plans, immunization dates, allergies, radiology images, and laboratory and test results.
  - In order to use these data and to standardized it, we use coding systems. Medical coding is the transformation of healthcare diagnosis, procedures, medical services, and equipment into universal medical alphanumeric codes.
  - One of these coding systems is International Classification of Disease version 10 (**ICD10**).
  - **ICD10 is used only for diagnosis**, ex: Alzheimer disease is coded as AD; if you enter AD code to the system, all of Alzheimer disease patient's names will appear on the system, and again its only for diagnosis not treatment or investigation.
- Medical records help us measure and analyze trends in health care use, patient characteristics, and quality of care.
- **Pros:** The data are **automatically collected and usually accurate** and detailed because they come from health care providers.
- **Cons:** lack of documentations and miss of classification.

# Claims data or administrative data

- Claims databases collect information **on millions** of doctors' appointments, bills, insurance information, and other patient-provider communications.
  - **It is associated with insurance company.**
  - It is information found in medical billing **claims** forms filed on behalf of a group or population. This information is gathered from the medical bills or claims submitted by medical providers **to government and private health insurers.**
  - **Ex: Medicare in USA**
  - **It has bigger data than EMR.**
- **Pros:**
  - Like other medical records, they come directly from notes made by the health care provider, **real-time data.**
  - The large sample size of claims data, researchers can **analyze groups of patients with rare illnesses** and medical conditions.
- **Cons:** there may be **low validity** due to certain **illegal billing practices**, like **ordering unnecessary tests** or **billing** for services that were not provided.



# Vital Records

- **Vital records are collected** by the **National Vital Statistics System** and are **maintained by state and local governments**.
- **Vital records include** births, deaths, marriages, divorces, and fetal deaths. They also record information about the cause of death, or details of the birth.
  - **Daily number of COVID-19 cases are considered as vital record.**
- **Pros:** Vital records are useful because they **offer very detailed information (including social information)** and **include information about rare disorders that end in death.**
- **Cons:** Vital records **only provide information on diseases and illnesses that end in death.**

# Surveillance

- **Public health surveillance** is the ongoing systematic collection, analysis, and interpretation of data, closely integrated with the timely dissemination of these data to those responsible for preventing and controlling disease and injury.”
  - The surveillance is done according to **ministry of health’s interest**.
  - Active: ministry of health measures the numbers of certain diseases according to its interest. Ex: children obesity, purity of water...
  - Passive: reporting to ministry of health about certain diseases. Ex: outbreaks, food poisoning...
- These systems function through the efforts of local and state health departments, working in tandem with **a variety of health care providers** (laboratories, hospitals, private providers), **who are mandated by law to report cases of certain diseases (as a passive surveillance)**.
- **Surveillance data has a higher validity than surveys, because the data comes from lab tests, diagnoses, and other patient records.**

# Disease registries

- **Disease registries** are another type of public health surveillance.
- Registries are systems that allow people to collect, store, retrieve, analyze, and disseminate information about people with a **specific disease or condition (for only one disease)**.
  - Every single thing about a specific disease will be recorded in a registry (diagnosis, treatment, labs...). Ex: In Jordan, Hussain cancer center records all cancer cases and their information in a registry.
- **Pros:**
  - Disease registries let researchers estimate how large a health problem is, determine the incidence of the disease, study trends over time, and **evaluate the effects of certain environmental and occupational exposures (ex: asbestosis and lung cancer)**.
- We prefer to use surveillance rather than disease registries, because surveillance is directed to your interest (based on your interest not like disease registry; based on a specific disease).

# Peer-reviewed literature

- Peer-reviewed journals may include their articles data uploaded on some data repository, like Harvard COVID-19 dataverse.
  - The research of scholars who have collected their own data using an experimental study design, survey, or various other study methodologies.
  - We use it in systematic review and meta-analysis.
- They also present the work of researchers who have performed novel analyses of existing data sources.
- Survey and surveillance are primary data.

• الفكرة هون انه انا بوخذ ال primary data او ال analysis لباحث وبستخدمها ببيحثي ك secondary data.

# Selecting, designing, and developing your questionnaire

- Questionnaires offer an objective means of collecting information about people's knowledge, beliefs, attitudes, and behavior.
- **The questionnaire may be objective or subjective:**
  - **Objective:** yes or no questions. Ex: do you have DM? yes or no.
  - **Subjective:** vague; like in emotional or behavioral questions.



# Steps in questionnaire development

- **Before we develop a questionnaire:**

1. Research question should be clear.
2. The variables you need in your research that will answer your question should be clear (do variable list).

## Ask your self the following questions:

- **What information are you trying to collect?** Require a deep knowledge of the research topic (**do variable list to be easier to collect**).
  - How to do variable list? It's gathered from literature review.
  - Ex: does smoking increase prostate cancer? The cancer and smoking are the variables.
  - Ok, we know that age is associated with prostate cancer so, it's a variable and here its confounding variable (it is not the main variable of the research, we don't do research on it, but we take it into account).
- **Is a questionnaire appropriate?** Sometimes questionnaires were used inappropriately (Ex: Don't use questionnaire in rare diseases).
  - The questionnaire should be clear to avoid mis-perception.

## Rule of thumb (to know if the questionnaire is an appropriate tool or not):

don't use a questionnaire to assess **sensitive topics** (associated with stigma), can be mixed with people perceptions, a broad issue that can be assessed by a question.

### when its not proper to use a questionnaire?

1. sensitive topics (make it a self reporting and anonymous questionnaire).
2. Rare diseases (use EMR instead).
3. in things about emotions and behavior.

Broad area of research	Example of research questions	Why is a questionnaire NOT the most appropriate method?	What method(s) should be used instead?
Burden of disease	What is the prevalence of asthma in schoolchildren?	A child may have asthma but the parent does not know it; a parent may think incorrectly that their child has asthma; or they may withhold information that is perceived as stigmatizing.	Cross-sectional survey using standardised diagnostic criteria and/or systematic analysis of medical records.
Professional behaviour	How do general practitioners manage low back pain?	What doctors say they do is not the same as what they actually do, especially when they think their practice is being judged by others. <sup>w13</sup>	Direct observation or video recording of consultations; use of simulated patients; systematic analysis of medical records

- **Could you use an existing instrument?** Using a previously validated and published questionnaire will save you time and resources; you will be able to compare your own findings with those from other studies, you need only give outline details of the instrument when you write up your work, and you may find it easier to get published.
  - **If you take the questionnaire and the data of a research → secondary data.**
  - **If you take the questionnaire and recollect the data → primary data.**
- **Is the questionnaire valid and reliable?**
- A **valid questionnaire** measures what it claims to measure.
  - **The questions in the questionnaire really answer my research question.**
- **Reliable questionnaires** yield consistent results from repeated samples and different researchers over time.
  - **When I ask a question, it is understood in the same way in each time.**



# Questions

- **Types of questions:**
- Open ended: **I give you free space to answer.**

What habits do you believe increase a person's chance of having a stroke?

- Close ended: **options of answer is clear.**

Which of the following do you believe increases the chance of having a stroke?

	Yes	No	Don't know
Smoking	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Being overweight	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Stress	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Drinking alcohol	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>


# Open ended **VS** Close ended questions

	Pros	Cons
Open ended	Allow free expression Capture new response options, emotions, & thoughts	Take longer time Take more effort to finish (exhaustive) Need coding and analysis Rely on handwriting skills & clarity
Closed question	Easy and quick Clear and complete Suitable for self completion Easy to standardize	Guessed answers Ceiling – Floor effect: <b>واحد زهق من تعباية الاستبانة فراح عباهم كلهم. very bad.</b> Errors in filling it can't capture wider options

الجدول مهم

# How should you present your questions?

(Question formatting and response options)

Format	How it looks on a questionnaire	Uses and advantages
<b>Statements with tick box categories</b>	<p>Please tick the box that best matches your answer</p> <p>Yes <input type="checkbox"/> No <input type="checkbox"/> Don't know <input type="checkbox"/></p>	<p>General attitude measurement. Easily understood and quick to complete. Generates data suitable for non-parametric statistical analysis</p>
<b>Rating scales (see Sapsford<sup>3</sup> or Oppenheim<sup>2</sup> for details of different formats)</b>	<p>Please indicate how you feel about our new surgery opening hours by circling the number that best matches your opinion</p> <p>Find them convenient                      Find them inconvenient</p> <p>1      2      3      4      5</p>	<p>Quantifies attitudes on 5 or 7 point scale and differentiates between positive and negative. Good for participants who can conceptualise linear scales and numerical values. Generates data suitable for non-parametric statistical analysis</p>
<b>Visual analogue scales</b>	<p>On the line below please draw a cross to indicate how you've reacted to your new medication</p> <p>Reacted badly                      Reacted well</p> <p>_____X_____</p>	<p>Precise quantification of attitudes. Good for participants who can conceptualise linear scales and have good visual skills. Data must be transformed for statistical analysis</p>
<b>Symbols</b>	<p>The nurse has just given you a lesson in healthy eating. Look at the faces below and circle the one that best shows how you feel about the advice you have been given</p> <p></p>	<p>Similar to numerical rating scale and can be analysed using similar tests but easier to complete for children or those with visual or literacy problems</p>
<b>Open ended items</b>	<p>Do you think exercise and health are linked, and if so, how? Please write your response in the box below</p> <div style="border: 1px solid black; height: 80px; width: 100%;"></div>	<p>Allows creative expression but may not suit less forthcoming participants. Must be formally analysed with qualitative methods</p>

Satisfaction survey