
MEDICAL RESEARCH
MODIFIED NO. 7
كتابة: لجين بلال و سارة عمر
تدقيق:
الدكتور: منير أبو هلالة

Week 3: Part 2
Ready to use questionnaires

Dr Munir Abu-Helalah
MD MPH PHD
Associate Professor of Epidemiology and Preventive Medicine

Color code

- Slides
- Doctor
- Additional info
- Important

Introduction
we have discussed already to use question and some items related to assessment of the validity and reliability of the question to ensure that we are using the right questionnaire,

key thing to focus on before we start this part is that the more you practice writing questionnaires or proposal the more, you'll have the skills and you'll be solid in scientific research either writing a questionnaire for a survey study or part of questionnaire addition to a chart review form for cross-sectional study or for clinical trial or for case control study you need to practice more and more and read the critical appraisal tips that we discussed in the previous part because they help you to do critical appraisal for your own work and to help you also to do that for others like for your colleagues or work or if you have a project you want to a paper or any proposal that you would like to review

Ready to use questionnaires

- In general, for a tool to be validated for use in assessment, it should be:
- Valid
  - ✓ Assess clinical important difference: smallest improvement considered worthwhile by a patient
  - ✓ Tool sensitive for changes

- Reliable
- Precise
- Easy to administer
- Acceptable by the study population.

for example, I want to assist the stress level among medical students, instead of having new questions, there are tools there validated and they'll provide good and acceptable results we can use for assessment of the stress level for patients or students for different settings the same for if we want to assess the psychology of people from general population psychology of patients ,there are tools that we use actually, and they are good in assessment of different levels of quality of life for the physical mental and social health and also, they are good screening tools to assess subjects with potential psychological being impairment

THEYRE not diagnostic tool

diagnostic tool: for example, you are looking for depression these tools if they have good validity or reliability will pick a good number of subjects they should be in the IDEAL situation for example if you have a group from 100 patients **if the tool is perfect it will identify 20 patients with depression and if you do it as clinical judgment by a psychologist if he detected 20 subjects the questionnaire should detect 20 subjects**

in practice we don't have something like 100% a tool that will detect these ones we always hope comparing our questionnaire with a gold standard technique that's clinical assessment

for example you don't want to do TSH for patients with hypothyroidism and you have a questionnaire and those with certain symptoms you'll screen them.

- **VALID**

these days many medications are in the pipeline ,previously they focus on mortality from illnesses, now we have many new treatments for example for cancer different immune therapies and we have introduced percision medicine sometimes the outcome is not comparing the current medication with the new medication ,term of survival now maybe you look at complications or the quality of life symptoms cale Etc

these tools are good actually in looking for small changes with treatment and these changes should be clinically important differences that we use

- **RELIABLE**

either if you repeating the assessment yourself tomorrow or your colleague repeating it we should get the **same results**. **The results** for different items in the questionnaire should **be consistent**.

For example, if someone reports not socializing much or feeling low, they should respond consistently by indicating they feel depressed or unhappy, they should not answer that they're happy and outgoing but on item number 30 they indicate feeling low.

- **PERCISE**

The outcomes you select, such as identifying potential depression using the Hospital Anxiety and Depression Scale

- **EASY TO ADMINISTER**

should be **brief, clear, and focused NOT LENGTHy**. They **can be completed** either face-to-face or through a questionnaire

- **ACCEPTABLE**

shouldn't be sensitive, but sometimes we have to include sensitive questions

For instance, in the quality-of-life tool for breast cancer, sensitive questions **aren't placed at the beginning.** Validated questionnaires consider this when they're designed.

Definitions of terminology used in ready to use questionnaires selection

Instrument	**A questionnaire or interview or simple test (or some combination of these), used to measure and quantify health or disease status**
Domain	**An area or realm, one particular aspect within a broad assessment**
Measure	**A score, generally from a series of items designed to quantify some particular domain**
Item or indicator	**A single item, eg one question in a questionnaire**
Scale	**A simple test to quantify broad or single aspect of health using a numerical estimate from visual or numerical range**

There are terms you should be familiar with..

A **domain** is a smaller part, like in a quality-of-life tool: physical activity domain, mental health domain, or social health domain.

An **item** is a single question, like "How would you rate your overall health?"

A **scale** like rating your health status from 0 to 100.

Questionnaire needs to be adapted to study population

- Know the respondents
  - Language
  - Education
  - occupation
  - ethnic group
  - sensitive issues

For a questionnaire to be effective, it must be adapted to the study population.

The main thing you should keep in mind is that if you're using a ready to use questionnaire on quality of life, pain scale, psychological well-being is the language and education.

For example, if you have a questionnaire in English to give for patients, translating it yourself isn't recommended **unless you perform a validation study**. Many validated translations exist, like the Quality of Life scale, which is available in multiple languages and you can use it. So you need to validate the translation
Validation is also essential when choosing a questionnaire based on disease status (who are the participants). Some are specific to conditions like type 2 diabetes or hypertension but not valid for arthritis.You should select a questionnaire that's validated for your **target population and ensure it's available in the required language**.

We always write **In the methodology of any proposal**, the tool we are using (e.g., the Hospital Anxiety and Depression Scale in Arabic) has been validated for the Arab population, providing references. So to sum up, **you'll write the name of the tool and that is validated for your patients and the disease and the language**

Classification of ready to use questionnaires/scales

There are two main types of questionnaires (we have three but the main are two):
generic and **specific**.

A. Generic questionnaires:

- Developed to be applied for a **large range of populations** and health care problems

- They permit **comparisons between populations or other groups** of people, and also in the same group before and after an intervention.

Generic questionnaires **apply across broad populations** (applied on a wide range of population), allowing comparison across groups, like patients with asthma, COPD, or heart disease.
They're useful for comparing populations, like assessing the quality of life between Jordanians and Egyptians or between the general population and those with conditions like type 2 diabetes on their quality of life (I need to make sure that the questions are validated for this population).
Nothing new

A. Generic questionnaires:

- In order to apply any instrument for generic use, it **should be validated across different groups and should be acceptable by these groups.**

Examples of generic ready to use questionnaires

1. Short Form Survey (SF-36)
2. General Health Questionnaire
3. Sickness Impact Profile

tools of quality of life

they can be applied for Different diagnostic criterias or scores for pain scales and you should also ensure that they are valid for that population

A. Generic questionnaires:

- It should be always considered that these measures are less responsive changes in health when compared with disease specific questionnaires.

they will be less responsive to changes in health I can use them in cross-section study usually but I might have difficulties in using them in clinical trials unless you have in mind that these patient for example have poor quality of life and then these questioners will affect these changes in the the quality of life

- Therefore, if these are not used along with a disease specific questionnaire, it is advisable to choose a clinical outcome of direct relevance to the disease/health care problem under investigations.

I **always prefer when I have clinical trial particularly or even cross sectional studies to have both a generic one and specific ones**
generic one will help us to **look and compare** other illnesses with other populations and
the **specifics** one we can **detect changes and we can get more details** about the impact of quality of life for these questioners.

A. Generic questionnaires:

Limitations:

- They may be insensitive to subtle but important changes in status with respect to a specific disease.

so they might be insensitive to subtle changes we have patient with subclinical hypothyroidism for example I did clinical trial using some questioners for the quality of life and we had crossover clinical trial we give thyroxin to two groups patients SF36 actually **was unable to detect changes the clinical in the quality of life while other items like the general symptoms or the patient self-reporting other clinical outcomes** they were statistically significant for the thyroxin group compared with placebo

- They should be validated across a spectrum of different groups of people.

B. Disease or population specific questionnaires:

- They are designed to target **particular** population or patients group.

If you have a disease specific questionnaire it will go **deep and look at different symptoms and impact of these symptoms ,impact of different intervention and for that particular disease**

for example, if you have a cancer patient you look at the impact of chemotherapy while you can't use the same questioner this is the cancer specific questioner for patient type two diabetes we have different interventions and management there actually, for cancer patients we have general cancer tools and we have Site specific tools ,you have something in common between different groups of cancer patient and also you'll find **tools directed** only for colorectal cancer patients, for thyroid cancer for breast cancer Etc

- Examples of disease specific questionnaires. Asthma quality of life questionnaire Arthritis Impact Measurement scales Rand Diabetes Mellitus Battery

B.Disease specific questionnaires:

- These **quantify the severity** of individual symptoms (such as angina pectoris) or the impact of a disease on a person's overall quality of life.

these questioners will help us to quantify the severity of symptoms related to angina , RA Etc and will **detect changes and we can use them to look at temporal changes and to look for the impact of different interventions and we use them actually commonly in clinical trials**

- They are well suited to detecting important changes (over time or following treatment) in an individual or within a population.

B. Disease or population specific questionnaires:

- More sensitive than the generic measures to minor impairments and changes over time.

we mentioned in the generic questionnaire that they are wide and this is the opposite here ,yes they are sensitive to big changes and in minor impairments and they will achieve higher acceptability for the patients because they are directed to that particular disease or group

- They also are expected to achieve a **higher acceptability**.

B. Disease specific questionnaires:

- They are **not available for all diseases**
- They may be **lengthy and detailed.**
- They have **limited role in comparisons** with other conditions or with the general population.

not available for all diseases we need to ensure that if we have any tool that's **validated for that particular disease**

for example,when you give treatment for any illness our Target is to improve the quality of life and the patient will live as healthy as possible so to compare the result with the quality of life with general population we can't use that through the disease specific this is why we encourage having generic questioner and disease questionnaire but we need to ensure that generic questioner is valid for that particular disease

The European Organization for Research and Treatment of Cancer Quality of Life Questionnaires

- Cancer patients quality of life assessment through EORTC QLQ-C30
- Cancer site specific supplementary measures validated modules:

1. Lung (LC13)
2. Breast (BR23)
3. Bone Metastases (BM22)
4. Head & Neck (H&N35)
5. Oesophageal (OES18)
6. Ovarian (OV28)
7. Gastric (STO22)
8. Multiple Myeloma (MY20)
9. Cervical cancer module (CX24)
10. Oesophago-Gastric (OG25)
11. Prostate (PR25)
12. Colorectal Liver Metastases (LMC21)
13. Colorectal (CR29)
14. Brain (BN20)
15. Information module (INFO25)
16. Endometrial (EN24)

something very important this is a common problem that I've seen in the whole region ,if you have a tool validated, a ready to use questionnaire please look at the **copyright of that questioner** if I want to use the quality life is it available for **free for everyone** to use okay yes I will use it if I need to write an email to get permission **I need to write an email if I need to pay a fees for that I need to pay fees for that**

in the previous lecture we said that we need to write the proposal and the tools and everything before we write the questioner when you write the proposal you want to apply for a grant

for example you write in the grant I'm going to assist the quality of life using the GHQ generic questioner for patients from general population or the from the Primary Healthcare Services to look for depression in Jordan GHQ actually is paid questionnaire the Arabic and the English one so I should add in my budget fees for this questionnaire developers otherwise it be illegal for us to use it ,
very unfortunate that time there are **malpractices** that I'll have this tool I'll change some questions there and use it so I don't need to pay for that organization this i**s not acceptable BECAUSE you are using invalid questionnaire** the limitations for using that and you need to do a validity study for the modified questioner before you use it for your patients or subjects
European organization for research and treatment that're really excellent in preparing cancer questionnaires and we have also disease specific questioners I have used them across the region in the UK and they are really nice questions questioners and I encourage you guys to go through these questioners

C. Dimension specific scales/questionnaires:

- These focus on particular aspects of health
- The Beck depression inventory is an example
- Some questionnaires measure functional ability; the Barthel index and Townsend's disability scale are examples.

Dimension specific scales they are looking at one dimension a particular aspect of Health like depression big depression example. so we have disability scale for example like Barthal index, so we it will focus only at particular aspect of life not like the previous ones they are looking at different domains or aspects

How to select your questionnaire/tool?

- What is your research question?
- Who are the patients you are studying?
- What do you anticipate will happen?

how we select your question first question will be what is our my stud design then what is the research question who are the patients what do we participate to have that is it clinical trial or cross sectional study

How to select your questionnaire/tool?

- Appropriateness: of the measure to the question or issue of concern
- Correspondence between the content of the measure and goals of the study

Is it appropriate for that condition we should look at the components ,and we need to ensure that reliability and validity I will not accept any proposal approval for the use of any questionnaire without looking at the you have already written the reliability and validity and responsibility for that questionnaire suppose the questionnaire is reliable and valid and we have for example 200 questions there and the response rate is low I will not accept it

- Evidence in relevant populations of:
  Reliability, Validity, Responsiveness
  If no previous evidence, you need to assess them For translated questionnaires, the translated version should be validated.
- Practical considerations

as we said that we need to ensure that they are valid and if you make any change you need to get validation for that but please start your research life actually as a medical student and hopefully a future doctor and researcher with respecting the copyrights for those who have already developed some validated tools

Practical considerations:

- Mode of administration
- Time to administer
- Language
- Respondent burden
- Availability of supporting materials

sometimes we have tools in English and we want to talk to the patients in Arabic we need to ensure that we are having the right translation . this is why we need to look for the translated versions they have supporting materials you have in the budget plans to pay for the copyright of the questionnaire

Practical considerations:

- Patients controlled on medications are expected to have a quality of life close to normal population
- Exclusion criteria: exclusion of subject with psychological conditions. Sometimes it is secondary to the condition. Examples:
  - Depression secondary to hypothyroidism
  - Anxiety and depression secondary to erectile dysfunction

patient controlled on medications should **have quality life close to the normal population** and this can be evaluated through generic questionnaire I've seen some surveys actually we are looking at the psychological being of patients coming type two diabetes to the primary care unit and you can see some items

I exclude subjects with psychological conditions so

for example hypothyroidism there is depression secondary to hypothyroidism so if I exclude subjects with some impairment in the psychology I'm going to only to get those with good psychological Health so please **always consider these items**

we all ask also in the we have this ready to use questionere and we should have our interview questionere in addition to that Ready to use questioner what do we need to include in this R question that we are going to develop yourself we need to get the age gender medical history education level for example we need to get the drug history and you are asking about depression here or anxiety depression you ask about history of depression anxiety are they are taking medication for anxiety or depression have they been seen by psychologist recently do they have

for example recent traumatic life events, you are assessing quality of life for patients and some of them had very poor scores in depression scale simply because they have lost family or they have Financial concerns so we need to consider these factors in the analysis

extra slide

Some issues in use of questionnaires

- Newly diagnosed and sub-clinical disease
- Controlled patients, uncontrolled patients and newly diagnosed patients
- Healthy control versus general population sample
- General population might have highly prevalent chronic diseases such as ischaemic heart disease, AIDS, diabetes

we always look at the **duration of illness as predictor of quality of life so**
you might have diagnosed patient subclinical disease they don't have symptoms ,this is why we introduce actual **screen programs** because one of the screen criteria actually that the you are looking for the patients are screen subjects who have not sort medical advice so expect those with sub clinical or newly diagnosed some of them don't have symptoms and also we need to look at the control and uncontrolled disease I'm having a quality of life for Hypertension patients I need to look at the hypertension control or not
diabetes control or not so the **difference between healthy control and general populatio**n that these are subject without illnesses and you screen them for some illnesses and they don't have them
general population

for example you are doing a study comparing quality of life of diabetes with general population in Jordan you have high prevalence to other chronic illnesses especially for those above the age 50 and we'll have small differences there they are not simply because they are not due to that that for example diabetes or the hypoth they **don't have impact on the quality of life simply because you have confouding factor that general** population they have a high prevelance for other illnesses that will affect the quality of life so **difference in**
the score simply are due to the presence of other illnesses

extra slide

Validity and reliability

- What claims for validity have been made, and are they justified?

  In other words, what evidence is there that the instrument measures what it sets out to measure?

- What claims for reliability have been made, and are they justified?

  In other words, what evidence is there that the instrument provides stable responses over time and between researchers?

--
I want you guys to understand what are the **differences between validity and reliability** and why they are **important** to us
we ask a question what claims for **validity** have been made ,are they Justified is their any evidence that the tool we are using is **valid and is it reliable** (two things as reliablity: We can repeat the assessment under the same conditions getting the same results and also the responses in the questioner will be correlated to each other we have a direction toward depression we have a direction other questions toward depression)

validity as we discussed earlier that if I have **compared to the gold standard technique** for example you have generic questioner for the screen for depression in primary healthare if you have 100 subjects and they have been interviewed by psychiatrist and he picked 20 subjects with clinical depression your tool should detect 20 subjects or close to 20 subjects BUT if you pick only one or two you have issues with your questioner or tool
nothing added

اللهم إني ظلمت نفسي ظلما كثيرا و لا يغفر الذنوب إلا أنت
فاغفر لي مغفرة من عندك و ارحمني إنك أنت الغفور الرحيم

Validity and reliability

- Validity: question measure what you claim it measures
- Reliability: results are reproducible or consistent with similar

extra slide

Assessment of validity

- Validity is measured in four forms
  - Face validity
  - Content validity
  - Criterion validity
  - Construct validity

I would go deep with these ways of assessment of validity at least guys I need you to just to know how to differentiate them

extra slide

Face validity you just have quick look at the questionnaire
I see the questionnaire looking on the depression or anxiety
I'm okay with it

Face validity

The extent to which your questionnaire is measuring what it appears to be measuring

- Cursory review of survey items by untrained judges
  - Ex. Showing the survey to untrained individuals to see whether they think the items look okay
  - Very casual, soft
  - Many don't really consider this as a measure of validity at all

extra slide

Content validity

The extent to which items on the questionnaire are representative of the domain under study.

- Subjective measure of how appropriate the items seem to a set of reviewers who have some knowledge of the subject matter
  - Usually consists of an organized review of the survey's contents to ensure that it contains everything it should and doesn't include anything that it shouldn't
  - Still very qualitative

Content validity are **other items included in that question**
representative of depression
anxiety quality of life
we review the survey by expert or by someone to check that it's still very qualitative that just make our opinion we can't quantify that

Criterion validity

extra slide

- The extent to which the questionnaire is measuring similar to a 'gold' standard, another measure that has been used and accepted in the field. There are two types concurrent and
- Measure of how well one instrument stacks up against another instrument or predictor
  - Concurrent: assess your instrument against a "gold standard"
  - Predictive: assess the ability of your instrument to forecast future events, behavior, attitudes, or outcomes

criterion validity it's how we **compare our tool with the gold standard Technique we have a ready questionnaire** consists for example of 100 questions and gold standard technique for screening, for quality of life for psychology for pain score Etc, and you want to ensure questionnaire that will be shorter for 30 questions I want to compare our results with a newly developed questionnaire with the gold standard technique tool

I want to have a tool to screen for hypothyrodism in general practice we have for example High prevalence in Jordan or for depression we just have the questionnaire there and compare to the gold standard technique for example for hypothyroidism symptoms if you have it compared with the result of the TSH and ft4 and you'll see that your questionnaire is able to Detect those with abnormal thyroid function test or with results suggestive of hypothyroidism so we compare our result with the gold standard with investigations clinical judgment by expert or by previous tool that have been used widely for along time

Construct validity

extra slide

The extent to which an instrument measures the construct or trait under study.

It is present when there is a high correspondence between the scores obtained on a measure and definition of a construct it is designed to represent.

- Most valuable and most difficult measure of validity
- Basically, it is a measure of how meaningful the scale or instrument is when it is in practical use

if you high correspondent between the scores of different domains ,different items of the questionnaire this is the construct validity.

it's very difficult because we need to look at each item and colerate them with each other but it's very useful actually to assist the validity

Construct validity (2)

- Convergent: Implies that several different methods for obtaining the same information about a given trait or concept produce similar results

extra slide

we have something we call Convergent validity imply that several different methods obtaining same information what given TR or concept procedure they have different results face to face or if **you have different School different ways of having the data collection you will lead to the Same result**

extra slide

Construct validity (3)

- Divergent: The ability of a measure to estimate the underlying truth in a given area– must be shown not to correlate too closely with similar but distinct concepts or traits

Divergent **validity is that you compare your questionnaire with other questioners and it should be corelated to each other**

Reliability

extra slide

- ***Internal Reliability:*** Questions measuring the same construct are correlated to each other and not to other constructs.
- ***Test–retest Reliability:*** If you were to do the survey exactly the same way, under the same conditions you would get the same results.
- ***Responsiveness:*** The questions can detect change.

internal reliability that we have already discussed that questionnaire are measuring the same construct correlate with each other and not other constructs for example you have one direction towards good quality of life the item should be toward the same direction this is internal reliability .

test-retest Reliability test you compare it with yourself and the same conditions you compare it with colleague under the same conditions

extra slide

Assessment of reliability

- Reliability is assessed in 3 forms
  - Test-retest reliability
  - Alternate-form reliability
  - Internal consistency reliability

Test-retest reliability

extra slide

- Most common form in surveys
- Measured by having the same respondents complete a survey at two different points in time to see how stable the responses are
- Usually quantified with a correlation coefficient (*r* value)
- In general, *r* values are considered good if $r \geq 0.70$

extra slide

Test-retest reliability (2)

- If data are recorded by an observer, you can have the same observer make two separate measurements
- The comparison between the two measurements is intra observer (intra-rater) reliability

same Observer you have
intra-rater (compare with yourself)
inter-rater compare with other collegue

Test-retest reliability (3)

extra slide

- You can test-retest specific questions or the entire survey instrument
- Be careful about test-retest with items or scales that measure variables likely to change over a short period of time, such as energy, happiness, anxiety
- If you do it, make sure that you test-retest over very short periods of time

you should have under the same conditions happiness anxiety energy Etc you should consider that the subject should be under the same conditions if you have an assessment for example for depression you want to do it today and you change the medication for the patient and they will start tomorrow new medication you can't repeat the Judgment tomorrow because you'll have different conditions there

Test-retest reliability (4)

extra slide

- Potential problem with test-retest is the practice effect
  - Individuals become familiar with the items and simply answer based on their memory of the last answer
- What effect does this have on your reliability estimates?
- It inflates the reliability estimate

Example: Assessment of depression

extra slide

Circle one item

Version A:

During the past 4 weeks, I have felt downhearted:

Every day	1
Some days	2
Never	3

Version B:

During the past 4 weeks, I have felt downhearted:

Never	1
Some days	2
Every day	3

you'll put **them at different places actually and to see whether the responses were the same**

Example: Assessment of urinary function

extra slide

Version A:

During the past week, how often did you usually empty your bladder?

- 1 to 2 times per day
- 3 to 4 times per day
- 5 to 8 times per day
- 12 times per day
- More than 12 times per day

We used before number of times per day and here we used number of times per hour and in both we should have the same results by evaluating the same subject under the same conditions

also we change the wording guys I not want to go to deep in the relability and validity assessment at least to know how we assess the reliability what's the purpose of reliability what's the purpose of validity and what are the common types of validity assessment

Example of nonequivalent item rewording

extra slide

Version A:

When your boss blames you for something you did not do, how often do you stick up for yourself?

All the time
Some of the time
None of the time

we might change the question wording to see the responses we should have consistent results there

Version B:

When presented with difficult professional situations where a superior censures you for an act for which you are not responsible, how frequently do you respond in an assertive way?

All of the time
Some of the time
None of the time

Internal consistency reliability

extra slide

- Applied not to one item, but to groups of items that are thought to measure different aspects of the same concept
- Cronbach's coefficient alpha
  - Measures internal consistency reliability among a group of items combined to form a single scale
  - It is a reflection of how well the different items complement each other in their measurement of different aspects of the same variable or quality
  - Interpret like a correlation coefficient (≥0.70 is good)

we apply it to group of items you have domain hospital anxiety and depression scale you'll have quality of life domain social functioning you'll have group of items that you'll have alpha coefficient alpha to assist for reliability you should have more than7

you have a group combine **these questions together on a single scale we need to see how these different items coplement each other** in the measurement of these aspects you have a total score of 10 how **these items contributed to the total score** some items where one should have the direction should be the same

extra slide

Example: Assessment of physical function

	Limited a lot	Limited a little	Not limited
Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports	1	2	3
Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	1	2	3
Lifting or carrying groceries	1	2	3
Climbing several flights of stairs	1	2	3
Bending, kneeling, or stooping	1	2	3
Walking more than a mile	1	2	3
Walking several blocks	1	2	3
Walking one block	1	2	3
Bathing or dressing yourself	1	2	3

inspect there that if you have different items they should have the same direction if you have impairment there severe impairment in physical activity should be reflected in all the items together

extra slide

Internal consistency reliability (2)

- If internal consistency is low you can add more items or re-examine existing items for clarity

extra slide

Ready to use questionnaires

- Although there have been no agreed standards for these criteria, there have been various attempts to provide guidelines or cut off points for what is generally accepted targets

in summary read to use questionnaire are really excellent in term of assessment they have been validated in different illnesses that are ready to use for you and especially if they are valid for that population they have good reliability you can use them one of the difficulties actually that to compare the our tool with the gold standard technique not always we have a gold standard technique for assessment but please ensure that the questionnaire is valid for that particular disease or population and they are valid in that language as well

extra slide

Interobserver reliability

- How well two evaluators agree in their assessment of a variable
- Use correlation coefficient to compare data between observers
- May be used as property of the test or as an outcome variable

Concept	Comment
1. Validity	Ability to measure what it supposed to measure.
a.Face validity	Refers to the investigators' subjective assessment of the questionnaire: a reasonable measure and items appears to be measuring what they intend to measure
b. Cotent validity	More systematic and comprehensive assessment than the face validity . It examines that extent to which items on a questionnaire covers all aspects that they intend to measure.
C.Construct validity	Construct: hypotheses are generated, then the questionnaire is tested to determine if it reflect these hypothesis. There two types of construct validity: 1. Criterion validity: the extent that the results match with the pre-existing tools..[3] 2. Concurrent: when the new measure is administered at the same time with the pre-existing one
D. Convergent validity	The measure is correlated positively with other methods that measure the same concept.
E. Sensitivity (detection rate)	Proportion of actual cases. For example patients with clinical depression who score positive on measurement tool for depression
F. Specificity	It is the discriminative ability of a measure. Ie the proportion of people who are not cases and test negative on the measure

a summary for the things that we need to assess in ready to use **specificity and sensitivity** we'll discuss them in screen lecture but sensitivity detction rate

for example you are having a questionnaire screen for hypothyroidrism and you are looking for different symptoms detection rate is ability of your tool actually to pick patient with hyperism with the disease is present
for example you have 100 patients you should detect 100 patient in your questioner if you detect for example 98 your **sensitivity** will be 98% if you take 70 subjects your **sensitivity** will be 70%

if you have low **specificity** you should not use it this questionnaire as a screening tool 70% so if you have low specificity you should not use it this posted as screening tool
for example you are using JHQ questionnaire to look for depression or anxiety for patient for GENERAL population in primary healthcare clinics in some countries are using it while patients are using the waiting area to look at these questions to see whether these subjects they will be reffered for psychiatrists or psychological support we can give these questionnaires because they have **good sensitivity** detection rate

specificity is that we're able to label those who are healthy as healthy
for example you have 100 healthy subjects you labeled them as healthy although they are 10 of them are not this means that you manage to identify only 90 of the 100 as healthy these ten subjects are called the **false positive rate and the specificity is 1-false positive rate**

Concept	Comment
2.Responsiveness	Ability of an instrument to be responsive to actual changes that occurs over period of time.
3. Administration	Easy
4. Length	Not too long or too short.
5. Cost	Not expensive to obtain or to administer
6. Precision:	Ability to detect small changes
7. Reliabiliy:	The extent to which a measure yields the same number or score each time it is administered.
a.Internal consistency	A test for the homogeneity and extent to which items are correlated within the same scale or domains in the scale. Cronbach's alpha gives an estimate of reliability based on all possible correlations between all items in the scale. Researchers have regarded that 0.7 is the minimum acceptable level for internal consistency. [1,2]
Test-retest reliability	Relationship between scores obtained by the same person on two or more separate occasions. Kappa coefficient is used to test nominal data (ranging from -1 to 1,(0) if the agreement is not better than chance, negative if worse than chance and (1)if there is perfect agreement.

Cost we mentioned that we need to protect the copyright of the questionnaires or you need to make sure that enough resource is there precision especially clinical trials we need to look at questionnaires that detect small changes and this is the advantage of disease specific questionnaires than the generic ones

Measures of *validity* of a new instrument

Measure	Concept measured	How measured
Face validity	The investigators' subjective assessment of the instrument; whether it appears to be measuring what it is intended to measure and whether each indicator is a reasonable one	Judgement (superficial)
Content validity	The extent to which the items in an instrument covers all aspects of the attribute to be measured. More systematic and comprehensive assessment than face validity	Judgement
Criterion validity Concurrent validity	Validating an instrument by comparing it with a currently accepted reference measure[6] Term for criterion validity when the two scales are administered at the same time; used when attempting to replace an existing scale with a new one that has some advantage (eg simplicity)	Correlation coefficient, correlating the measure with some other accepted "criterion", ideally a gold standard[6]

Measures of *reliability* of a new instrument		
Measure	**Concept measured**	**How measured**
Internal consistency	**A test for the homogeneity, the extent to which the items within a domain (which broadly should measure the same thing) are correlated.**	**<u>Cronbach's alpha</u>, an average of the correlation coefficients between all items. Takes values between 0 and 1. A low value (<0.50) indicates that an item does not come from the same conceptual domain[5], a value of 0.7 has been judged the minimum acceptable level for internal consistency[6].** **<u>Split half reliability:</u> correlation of two summary scores (for example from odd- and even-numbered questions in a questionnaire)**
Test-retest reliability	**Relationship between scores obtained by the same person on two or more separate occasions.**	**<u>Kappa correlation coefficient:</u> Takes values between -1 and 1. A score of 1 indicates perfect agreement, 0 is the extent of agreement expected from chance, a negative score indicates worse agreement than would occur by chance**

Measures of *validity* of a new instrument

Construct validity	**Validating a new instrument by developing a hypothetical prediction of its performance, relevant where the variable of interest is abstract and cannot be directly observed[1]**	**For example a questionnaire for use in jaundice, measuring the extent of itching and excoriation, should show improvement when serum bilirubin decreases[1]**
Two subtypes:		
Convergent validity	**The measure is correlated positively with other methods accepted as measuring the same concept**	**Correlation coefficient**
Divergent or discriminant validity	**Lack of correlation with variables that measure a different unrelated topic**	**Correlation coefficient**

VERSIONS	SLIDE #	BEFORE CORRECTION	AFTER CORRECTION
V1→ V2			
V2→V3			

امسح الرمز و شاركنا بأفكارك لتحسين أدائنا !!