CHAPTER 14

MEASUREMENT TOOLS INSTRUMENTS AND QUESTIONNAIRES

Anne-Marie Schott, Francois Chapuis, Gilles Landrivon, Eric-Nicolas Bory, Francois Delahaye

The formulation of the research question requires a good definition of the factors studied and the judgment criteria. The latter will define, within the framework of the protocol, the data that will be collected and their method of collection. Data collection involves measurements and questionnaires.

The evaluation of measuring instruments is essential at an early stage of research. Indeed, the flaws of an instrument not only have consequences on the power of the study and the size of the sample required, but they can also compromise the very validity of the conclusions of the study by introducing irreversible errors into the estimation of variables (factors studied, judgment criteria, confounding factors).

The questionnaire is a special mode of data collection. Initially used in sociology, psychology and psychiatry, questionnaire technologies are now of interest to all medical specialties.

The quality of a questionnaire depends on the quality of the transmission, storage and interpretation of the information collected. Through the questionnaire, the investigator makes a measurement of the phenomenon studied. A bad questionnaire, or badly asked questions, can induce measurement bias: there can then be an over- or under-estimation of the importance of the phenomenon observed, totally invalidating the conclusions of the study.

All measurements are subject to some degree of error. It is fundamental to know the different types of errors, their origins, how to quantify them and reduce them.

There is no universally good measuring instrument. The choice of an instrument, among those that exist and have already been used, depends on several elements, including the following three that must be taken into account: the objective of the research, the disease studied and the target population.

I - CHOICE OF A MEASURING INSTRUMENT

A - Definition of the objective

1 - Cross-sectional use for diagnostic or prognostic purposes

We may want to measure a symptom or a characteristic allowing us to classify a population into several groups of subjects (screening). This involves determining the difference between the subjects, for a given variable, in a cross-sectional manner.

For instance:

- to know the prevalence of a condition in a specific population or to detect a condition in a specific population, in epidemiology;
- to decide at the individual level of a therapeutic intervention or to determine if an individual is eligible in a clinical trial.

Thus, the cervical smear is used to screen for cervical cancer and to assess its prevalence in a given population.

Some examinations are intended to assess the risk of the development of a disease in healthy subjects or the worsening of a disease in patients. In these cases, it is a question of measuring risk and prognosis factors. For example, several types of antibodies are sought in patients with hepatitis B. The presence and level of these antibodies make it possible to assess the evolutionary potential of the disease and thus to establish its prognosis.

2 - Longitudinal use to assess the effectiveness of an intervention

It is a question of determining the evolution of a subject in a longitudinal way by comparing two states of the same subject. Most often, the evaluation is done before and after a therapeutic intervention whose effectiveness is to be evaluated.

B - Search for existing instruments

This step is obligatory for all types of measurement.

Paradoxically, when one is interested in a specific objective for a specific population, it becomes very difficult to find the appropriate instrument. One is then tempted to be very critical of existing instruments and to underestimate the difficulty of developing a new instrument. A common mistake clinicians make is to too easily discard existing scales and embark on the development of a new instrument with the optimistic idea that they can do better. In fact, the development of a measuring instrument requires a considerable investment of time and money. The stage of exhaustive review of all the instruments developed in the field is therefore mandatory.

Conversely, the other frequent error is to consider that a scale is good because it has been "validated", without taking into account the objectives and the conditions under which the validation was carried out. A validated scale for determining the level of cognitive functions in elderly people is certainly not valid for measuring cognitive status in 20-year-old students.

Once this stage of research in the literature and with experts has been completed, it is then necessary to choose among the existing instruments the one that seems most suited to the objective that has been set. This requires determining their relevance to what you want to measure. It should also be investigated whether the accuracy and reproducibility of these instruments have been assessed correctly, in the same context and for similar purposes.

C – Define the phenomena to be measured

The phenomenon that we want to measure can be expressed according to various types of scale:

- Qualitative data: they are expressed in categories that cannot be ordered in relation to each other (eye color, professional categories, etc.). When there are only two possible categories, the data is said to be dichotomous or binary (vital status, sex).
- Ordinal data: unlike the previous ones, these data can be sorted in ascending or descending order and combined like numbers (for example, multiplied or added to form indexes).

• They can be discontinuous (or discrete), if they take only certain integer values, which are either categories (different stages of dyspnea, successive grades of evolution of a cancerous tumor), or values arranged regularly along a scale with a constant interval between each value (number of epileptic seizures per month, number of inflammatory joints).

• They can be continuous, if they can take virtually all possible values between the two extreme values of the response scale (blood pressure, weight). The intervals are constant and known.

The choice of the type of measurement scale depends on the variables concerned and the objectives. In a questionnaire designed to collect information on calcium consumption, we can quantify the milk consumed and construct a calcium consumption index in mg per day. If the objective is simply to divide individuals into groups, the results can be expressed in ordered categories (less than 500 mg/d, from 500 to 1000 mg/d, more than 1000 mg/d). The categories have the advantage of having practical meaning in the clinic. However, they are more arbitrary in nature and can end up ignoring significant differences between the groups if they have not been chosen judiciously.

You have to make sure that the variables you are measuring are appropriate for the purpose of the research. For this, we consult a group of experts, then a sample of the population on which we wish to apply the instrument.

Once the instrument has been chosen, it is necessary to evaluate its qualities within the framework of the particular objectives of the proposed research.

II - EVALUATION OF THE REPRODUCIBILITY OF AN INSTRUMENT

Two major types of characteristics associated with a measuring instrument are traditionally distinguished: reproducibility and accuracy *(table 1)*.

The reproducibility of an instrument is its capacity to provide an identical measurement repeatedly (capacity of the thermometer to indicate the same temperature repeatedly, etc.).

Reproducibility is essentially related to random error. The smaller the random error, the better the reproducibility.

The accuracy of an instrument is its ability to provide an exact measurement of the phenomenon to be measured (ability of the thermometer to indicate the exact temperature, etc.). Accuracy is related to both random error and bias. Accuracy (also called "validity") is defined by sensitivity and specificity and is developed in chapter IX.

Table 1 - The two main types of characteristics associated with a measuring instrument

Bias	systematic error	accuracy (or validity)
Chance	chance error	reproducibility (or reliability)

A - Assessment of reproducibility

It includes the following steps:

- List the sources of potential errors (inclination of the radiation with respect to the patient's body for x-rays, time of day for weight or height, season for the evaluation of physical activity, conditions under which the blood pressure measurement takes place, etc.).
- Classify these potential sources of error in decreasing order of importance, based on expert opinion, data from the literature, and the conditions in which you want to use the instrument. In a multicentre trial, it is essential to study the variation between centers during a pilot study. On the other hand, if the planned study takes place in a single centre, this source of error does not exist.
- Determine whether these variations actually exist in practice, and if necessary measure their importance and try to reduce them as much as possible.

It is impossible to establish here an exhaustive list of all the possible sources of variation. Some are specific to the type of measurement, others exist, to varying degrees, with almost all types of measurement:

- inter-investigator variation (variation between measurements of the same phenomenon by different investigators);
- intra-investigator variation (variation between measurements of the same phenomenon by the same investigator at different times);
- intra-patient variation (variation of the measurement or of the phenomenon itself in the same patient at two different times).

To assess these potential sources of error, repeated measurements are performed by varying one source of error at a time. It must be ensured that between these repeated measurements, the actual value has not changed. This has an implication in the choice of the time elapsed between two repeated measurements:

- if labile variables are measured, a very short time interval must be chosen between successive measurements in order to avoid a real change in value;
- on the other hand, for very stable variables such as height in adulthood or bone density, the time elapsed between two measurements may be longer, while maintaining the assurance that there has been no change of real value.

Take the example of the radiological diagnosis of lung cancer. To assess the inter-investigator variability, each radiograph is read by each of the blinded radiologists and the agreement of their conclusions is measured. Intra-investigator variability is estimated by showing the same X-ray several times to the same investigator. Intra-patient variability is estimated by taking several X-rays of the same patient and having them interpreted by a single radiologist.

B - Measurement of reproducibility

In the field of biology, laboratories continuously perform reproducibility measurements by dividing the sera in two and measuring the variability of the measurement for the same serum. This variability is expressed in standard deviation around the mean. For example, the reproducibility for the sodium assay is ± 2.3 mmol/l. Since normal values are between 130 and 150 mmol/l, it is easy to judge the acceptability of this error.

In clinical medicine, we often have to work with discontinuous data. Suppose chest x-rays are taken from 110 exposed subjects to determine the presence or absence of signs of pneumoconiosis. The radiographs are shown successively to two radiologists and the extent to which their opinions agree *(table 2)*.

Table 2 - Concordance of the opinions of two radiologists for 110 subjects suspected of pneumoconiosis

		Radiologist #2		Total
		Present	Absent	
Radiologist #1	Present	14	7	21
	Absent	8	81	89
Total		22	88	110

In this table 2x2, the percentage of concordance between the radiologists is represented by the box at the top left (present-present) and that at the bottom right (absent-absent).

That is: $\frac{14 + 81}{110}$, or 86%

There can be more than two categories. Take the example of 100 women with breast cancer examined independently by two clinicians who must assign a degree of severity on a scale from I to IV according to predefined criteria *(Table 3)*.

Table 3 - Concordance of the opinions of two clinicians on the severity of the disease in 100 women with breast cancer

		Clinician #B				
		I	II	III	IV	Total
Clinician # A	I	**25**	7	2	1	35
	II	4	**14**	5	0	23
	III	3	6	**17**	3	29
	IV	0	1	2	**10**	13
Total		32	28	26	14	**100**

The perfect agreement between clinicians is:

$$\frac{25 + 14 + 17 + 10}{100}, \text{ or } 66\%,$$

which does not look very good. However, when looking at the 34 cases in which clinicians disagreed, the difference was most often (27 times) only one stage.

For this type of data with more than two response categories, perfect agreement is not a very good reflection of the situation, because it gives no indication of the importance of the discrepancy. It is also too dependent on the number of categories: the greater the number, the lower the chances of perfect matching.

Finally, the simple calculation of the percentage of concordance does not take into account the cases where the concordance is due to chance. To overcome these drawbacks, several statistical concordance indexes have been proposed (*see below*).

These concordance indices can only be applied in simple studies, and when using discontinuous variables. When the study design is more complex, when there are more than two observers involved or when the sources of variation are more numerous, it is preferable to use appropriate mathematical methods, making it possible to study several sources of variation at that time.

These methods, based on the analysis of variance, evaluate the part of the variation linked to the real differences - either between the subjects, or in the same subject after a treatment or a sufficiently long time interval - and the part of the variation due measurement errors.

The principles of the analysis of variance are also used when the variables studied are continuous. From these methods are derived statistics used specifically for the evaluation of measuring instruments.

Once the sources of errors have been identified and their importance assessed, it is mandatory to reduce them. In the previous example, the error due to inter-investigator variations can be reduced by using only one radiologist to read all the radiographs, or by performing identical training for all radiologists. The error due to intra-patient variation can be reduced by applying strict rules and standards for the realization of X-rays (positioning of patients, radiological constants, etc.).

C - Importance of reproducibility

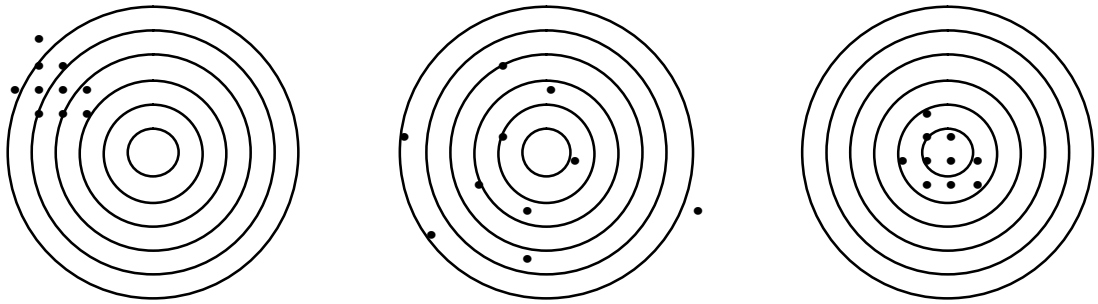
The accuracy of an instrument is limited if its reproducibility is not sufficient *(fig. 1)*. One must therefore first assess the reproducibility before embarking on a full validation study because a non-reproducible measurement is useless.

The lack of reproducibility can have serious consequences on all types of studies:

- in therapeutic trials, it decreases the power of the study by increasing the variance of the variable of interest;
- in epidemiological studies of etiology or causality, the role of a factor or exposure in the occurrence of a disease is sought. If the measurement of the risk factor or that of the disease studied is not reproducible, their association is underestimated. This can mask a

real and significant relationship between risk factor and disease (thereby increasing the risk of not detecting a real relationship). When one wants to measure confounding factors, this reduces the ability to control for these factors, thus biasing conclusions in unpredictable ways.

Fig. 1 *- Accuracy and reproducibility: the classic illustration of the target*
On the center target, the shot is subject to random error and all the impacts are randomly distributed around the center: the shot is not very reproducible, and therefore not very exact. On the left target, the impacts are grouped together but systematically next to the desired goal: the shot is very reproducible but very inaccurate. On the right target, the impacts are grouped around the center: the shot is accurate and reproducible.



D - Importance of sampling

Certain rules must be observed concerning the choice of subjects and the sampling technique.

The reproducibility and accuracy of an instrument can vary depending on the type of subjects on which it is determined. They must therefore be evaluated on subjects similar to those to whom the instrument is to be applied. Bone density measurements are less reproducible in osteoporotic subjects than in healthy subjects, because the vertebrae are less clearly visible, and the measurement of their contour is marred by a greater error. If we want to use bone densitometry to distinguish healthy subjects from osteoporotic subjects, we must assess the reproducibility of this examination in the general population. On the other hand, if we want to distinguish, among osteoporotic subjects, different stages of severity, we must evaluate the reproducibility and accuracy of the device in a sample of osteoporotic subjects and not in the general population.

E - Statistical methods for evaluating the reproducibility of measuring instruments

To study the reproducibility of an instrument, we use the concordance between values obtained by the same instrument under different conditions. To study the variation between different radiologists studying the same radiographs, the results obtained by each radiologist are compared, and the extent to which these results agree using statistics of concordance.

Concordance between two variables must be distinguished from other more vague and less strict relationships such as a simple association or correlation. This important distinction between concordance and association is the basis of the general principles used for the statistical analysis of measurements.

1 - General principles

To describe the relationship that exists between two variables in a given population, the indexes usually used represent the extent to which the variations of one of the variables are similar to those of the other variable. These two variables may or may not be expressed in similar measurement scales.

Examples:

- If we are interested in the relationship between cholesterol levels and age (cholesterol being a continuous variable and age being expressed in 10-year categories), an index of association describes the extent to which cholesterol levels rise (or goes down) from decade to decade. These association indices are the linear correlation coefficient (r), the regression coefficient, Spearman's rho coefficient (ρ), Kendall's tau coefficient (τ).

- We want to study the relationship between the hemoglobin level and the creatinine level in patients with renal failure. The regression coefficient indicates how strongly the serum creatinine value influences that of the hemoglobin level (the regression coefficient corresponding to the slope of the regression line), and the correlation coefficient r determines the linear relationship that exists between these two variables, that is, how much these two variables vary together and in the same way.

Regarding measuring instruments, when two instruments are supposed to measure the same phenomenon, it is not enough to provide proof that their measurements are simply correlated beyond mere chance. These measurements should in theory be identical if there were no measurement error. Thus the relationship that we want to highlight must be very strong, it is not a question of a correlation but of a concordance. Conventional association indexes are unsuitable for describing concordance because two variables can have a very close relationship without ever being in perfect agreement. Two instruments A and B measuring bone density can vary in the same direction and in the same proportion even if A always gives higher measurements than B. On the other hand, two variables can have a very strong negative correlation, which corresponds to a very poor agreement despite the very strong correlation.

Several concordance indexes have been proposed to assess the reproducibility and accuracy of instruments. Only the kappa coefficient is studied here.

In order to use concordance indices, the two variables must be expressed in the same unit. If we have to compare two variables that are expressed on different scales, we cannot use concordance indices, but only measures of association.

2 - Type of data

The concordance indexes used depend on the type of data.

a - Qualitative data

We can simply calculate the percentage of concordance. However, this measure encompasses the percentage of agreement that would exist by chance, even if the two variables were totally different and independent. The kappa coefficient is more appropriate because it measures the concordance between the variables taking into account the effect of chance.

Failure to take into account the effect of chance can lead to false conclusions by overestimating the true concordance. For a binary variable, the concordance predictable by pure chance is given by the formula:

$$[p1x\ p2] + [(1-p1)\ (1-p2)]$$

where p1 is the proportion of people with a characteristic according to a measure X, and p2 the proportion of people having the same characteristic according to a measure Y.

The kappa coefficient is defined as:

$$\text{Kappa} = \frac{\text{observed agreement - predicted agreement}}{\text{1 - predicted agreement}}$$

When two measurements agree to a degree no greater than pure chance, the value of kappa is zero.

When the two measurements agree perfectly, kappa is equal to 1.

Example: a study compares information on compliance with reserpine treatment obtained by interviewing patients on the one hand and by medical records on the other *(Table 4)*.

Table 4 - Concordance of information on compliance with reserpine treatment obtained by interviewing patients on the one hand and by reading medical records on the other, in 217 patients

		Medical file		Total
		Yes	No	
History taking	Yes	**14**	7	21
	No	25	**171**	196
Total		39	178	217

$$\text{Percentage agreement due to chance} = \frac{(21x39) + (196x178) + 81}{217^2} = 0.7583$$

$$\text{Observed percent agreement} = \frac{14 + 171}{217} = 0.8525$$

$$\text{Kappa} = \frac{0.825 - 0.7583}{1 - 0.7583} = 0.39$$

In this example, not taking chance into account would have resulted in significantly overestimating the concordance (85% instead of 39%).

b - Ordinal data

Also in this case, kappa is the statistic of choice but it can be changed to weighted kappa. The principle is to take into account the importance of the discrepancy by considering all the different degrees of partial concordance.

For example, if there are four response categories *(Table 3)*: a discrepancy between a stage I according to observer A and a stage II according to observer B is less serious than if observer A estimates the stage at I and observer B the stage to IV.

c - Continuous data

The indices to be used are the intra-class correlation coefficients which associate a measure of correlation with a statistic testing the difference between the means. In terms of linear regression, this means that these coefficients not only test the similarity of the slopes of the regression lines, but also that the line passes through 0.

The use of an analysis of variance with more than two entries makes it possible to study the variations coming from several sources simultaneously and corresponds to methods which are beyond the objectives of this chapter.

III – QUESTIONNAIRE: KNOWING HOW TO EXPRESS WHAT YOU WANT

Questionnaire: what is it? The printed document? All the questions put to an individual? Or the overall concept of "interview-questions-answers-recording of answers" which ultimately leads to the development of the printed document?

In a clinical trial, the paper or electronic form of the questionnaire, is intended to store the information collected during the study. This version is all the more easily elaborated since it follows the chronological sequence of the trial (for example, one questionnaire per medical visit, known as the visit 'dispatch slip') and since it is intended to record data that is not subject to interpretation (for example, diastolic and systolic blood pressure values).

It is quite different from a questionnaire intended for the evaluation of subjective data concerning the experience of patients, their pain, their quality of life. How to objectively measure the subjective? Are we measuring what we want to measure? Is the vocabulary used understandable by all and in the same way (in Central Africa, fever means both malaria and fever)?

The questionnaire, which is a communication exercise for both parties, can only be developed when one of the parties knows what it wants to communicate and obtain from the other. It is not possible to build it before the research question to be answered, or before the plan of the study is known precisely.

A - Definition and validation of the exploration field

Well before writing the questions, it is essential:

- to precisely define what will be measured. A questionnaire is first intended to measure the element of interest (for example: pain, satisfaction, etc.);
- and to ensure that a term used covers the same concept and the same definition for everyone.

The correct definition of terms represents such an obstacle that a multidisciplinary group of experts is often called upon when constructing a new questionnaire.

The study of the literature is of invaluable help because many questionnaires have already been validated. They can, under certain conditions, be re-used.

B - The various types of questions

There are three broad, non-exclusive categories of questions. The choice depends on the nature of the information sought and how the subsequent analysis is envisaged.

1 - The open question

Example: "How would you describe your pain?"

It is the least restrictive for the questioned subject. He is able to use his own words to express the full panoply of his feelings. This is particularly important when the experimenter seeks above all to define the extent of what is felt by the patient, the variety of opinions of the healthcare team…

Synthesizing such a diverse set of responses to a single question for coding and statistical analysis purposes is complicated. It can lead to misinterpretations. The great latitude left to the subject in his answer exposes the investigator to his goodwill and/or his ability to express and transmit what he feels. Coding, which involves a reduction (through a process of standardization) of information, leads to an impoverishment of the information initially collected.

2 - The semi-open question

Example: "During your last medical consultation, did your doctor also examine one (or more) member(s) of your family: father, mother, brother(s), sister(s), child(ren) , others) ?"

With this type of question, the memory effort is minimal. The answer is suggested and directed.

3 - The closed question

Example: "Were you born on a Friday the 13th?"

It imposes an unequivocal answer among those proposed by the team that drafted the questionnaire (yes, no, don't know). No effort (or almost) of memorization is required. Coding the response is easy and (within the risk of error) reliable. On the other hand, no choice is possible apart from the proposed answers. This is not harmful for the example given above. It certainly is for the evaluation of emotions, behavior, caloric intake, pain…

C - Appropriate vocabulary

Are there words to avoid or others to use more specifically? Every single, familiar word referring to a clear concept can be used. Otherwise, it should be avoided.

How to know? An experienced team chooses the vocabulary according to the target audience and avoids, during the construction of a questionnaire intended for the "average patient", the use of culturally marked, new terms, precise but too specialized scientific or medical terms, terms insufficiently widespread in the population, or conversely too vague or covering a broad and imprecise meaning, or variously interpretable, ... An individual may prove unable to describe his "marital status" when he is certainly able to say if he is single, cohabiting, married, divorced, or widowed.

A simple, clear and precise vocabulary whose meaning is identical for any potential subject regardless of their socio-economic status, socio-professional category or level of education is a guarantee of the quality of the questionnaire.

D - Wording of the question

We must strive for clarity and the absence of ambiguity. Above all, it is a matter of avoiding formulations that are too general, complex, ambiguous and contentious. Some examples are given in *Table 5*.

Table 5 *- Wording of questions to avoid*

- the double negative: do you think that people who don't have a job aren't happy?
- the double questioning (two questions at the same time): are you in favor of the systematic addition, for 2 euros per month, of fluoride in drinking water to prevent cavities in your children?
- the strong suggestion despite the interrogative form: don't you find that the systematic fluoridation of drinking water is an attack on individual freedom?
- the combination of the two previous ones (supported suggestion and two items explored by the same question): do you prefer a clean source of energy such as nuclear power to a polluting coal-fired power station?
- the word(s) with an imprecise meaning: are you often constipated?
- the imprecise question: would you be in favor of a moderate reduction in caloric intake in inactive people?

E - Organization of questions

The purpose of a questionnaire is to collect quality data, freely communicated by the subject who participates in the study. The concern is to prevent the questionnaire from being incorrectly or incompletely completed.

You have to interest the patient, make him want to answer, and therefore focus on designing the content as much as the form of the questionnaire:

- at the beginning are the questions to which one absolutely wants to have an answer, and to which the patient wants to answer. It is the initial hook, attractive, with guaranteed response;
- then come the questions by theme, grouped together, as this greatly facilitates the conduct of the interview (have you ever had prostate surgery? If so, answer the 6 questions that follow, otherwise go directly to question 22);
- the most personal questions generally take place at the end, when the subject is confident and seeks less than at the beginning to organize or control the information he delivers.

F - Coding

Converting the answers into figures, which are easier to use for analysis purposes, involves the use of a conversion system.

Two possibilities are available to the researcher:

- the development of a personal coding system, essential if the field of study is new or very specialized,
- or the application of pre-established and already tested systems: knowing the prevalence of diseases and their complications in a hospital can be done by applying a national or international classification, for example the International Classification of Diseases of the World Health Organization, whose codes are accessible to all.

The coding depends on the type of question.

A closed question, imposing an answer among those proposed, allows easy coding from the outset.

On the other hand, an open question does not make it possible to predict all the answers a priori. The answers are then grouped by logical categories, in a subjective way, at the end of the study, by the investigator.

Where to put the coding instructions? Two options are available to the investigator:

- on the questionnaire itself

At each question level, the coding values for each variable are shown. It is useful for the epidemiologist, in the management of the variables, as well as later for the statistician during the analysis, to locate the relative place occupied by the coded value of the variable in the whole of the questionnaire *(Table 6)*. The coding is permanently visible, the risk of transcription error is reduced. But the coding space requires an increase in the length of the questionnaire, particularly when the coding instructions are long and complex. The volume, the weight as well as the generated cost are generally higher.

- in an appended document (table 7)

The questionnaire is simplified in its presentation: only the question appears. But the coding information is less easily and less quickly available, an annex document gets lost, …

Coding also depends on the research question and the planned statistical analysis. It is traditional to reserve certain codes for particular values: "no" often takes the value 0, "yes" the value 1, the missing values the values 9, 99,...

The research question may, however, condition the values of the variables in a particular way. The statistician can, during the analysis, recode variables (by transformation) if the initial coding turns out to be inadequate.

Consulting an epidemiologist or biostatistician at this stage is certainly wise.

***Table 6** - Example of coding included in the questionnaire*

N° Question	Wording of the Question	Coding	Identification
06	What year were you born?	|__|__|__|__|	(58-61)
	(indicate full year)		
07	Were you born on a Friday the 13th?	|__|	(62-62)
	(no = 0, yes = 1, don't know = 9)		
08	Were you born in metropolitan France?	|__|	(63-63)
	(no = 0, yes = 1, don't know = 9)		
09	If yes, in which department?	|__|__|	(64-65)
	(Indicate the department number)		

***Table 7** - Minimum information when the coding rules appear in a document annexed to the questionnaire*

- variable number
- his full name (for example, day of birth)
- its abbreviation: most often a maximum of eight symbols, for reasons of compatibility between computer files (for example, DAYBIRTH)
- a brief description (for example, day of the week during the birth)
- its type: character, numeric, logical, date, … (for example, numeric)
- its width once coded, possibly including the comma and the decimals (for example, 4 characters - 2 digits, comma, 1 decimal - for the "body temperature" variable - 37.5°C)
- the page (and possibly the reference) of the questionnaire to which it refers
- the database (computer file) to which it belongs, in case there are several files
- last but not least, the coding methods (for example, no = 0, yes = 1, don't know = 9)

G - Use of a foreign questionnaire

The researcher may be tempted to use a questionnaire developed abroad for a similar problem. In addition to the difficulty of a perfect translation, there is the question of the transfer and applicability of the cultural concepts which governed the development of the questionnaire.

The demand expressed by a patient with low back pain following an accident at work is not the same depending on whether he is happy or not in the performance of his daily work. The researcher must take this into account when analyzing the data he has collected on this topic. But did he initially take care to verify that the same request was expressed in the same way by the subjects of the study population? Patients from different cultures do not express similar difficulties through the same symptoms. Conducting research in this area requires taking these cultural particularities into account right from the questionnaire development phase.

Similarly, a questionnaire designed to assess various aspects of the state of health of British patients, although it has been validated in other northern European countries and has led to similar results, would probably not reflect not the same concept if it were applied as it is in the Latin countries of Southern Europe: we do not culturally have the same idea of health in each of our European countries. It is the same between Americans and French.

Why then are foreign questionnaires sometimes used?

Because this is unavoidable in certain circumstances, for example when a multicentre study is taking place simultaneously in several countries with different languages, or when it is necessary to base a new study on the methodology and the questionnaire of a previously published study, considered as the reference on the subject, in order to replicate it and compare the results. A few steps are necessary for the proper transposition of a questionnaire from one country (one culture) to another (Table 8).

Table 8 *- Stages of "cross-cultural" validation*

- selection of the foreign questionnaire most likely to meet the need
- parallel and independent double translation (from the original language to the current language), avoiding the trap of literal translation, by two people familiar with the idea underlying each question (in general, by at least one bilingual researchers on the team)
- double reverse translation (from the current language to the original language), by other translators than the first, in order to verify that the concepts conveyed by each question are still present
- arbitration of differences during a consensus meeting
- pre-test of the questionnaire with potential future respondents
- final changes

IV - ACCURACY AND REPRODUCIBILITY

Two qualities are essential to a questionnaire: its accuracy and its reproducibility.

Accuracy is the ability of the questionnaire to provide an exact measure of what is to be measured, such as the ability of the archery champion to hit the center of the target, or that of a thermometer to indicate the exact temperature.

By analogy with a diagnostic test, the perfectly accurate questionnaire would be one with perfect sensitivity and specificity (100%), and for which the positive and negative predictive values would also be 100%. Has such a test ever existed? The questionnaire administered to the candidates of the traffic code examination is not only supposed to screen the candidates

who know the traffic code well (sensitivity of the test) and eliminate those who know it insufficiently (specificity of the test), but also to predict that a candidate who passed the test knows his rules of the road well (positive predictive value) or that another candidate failed because he knew them badly (negative predictive value).

If you do not know which measure to choose, because several can apply to this concept, you must use several. If we do not know which dimension of the concept to measure, we must measure them all.

Reproducibility is the ability of the questionnaire to provide an identical measurement repeatedly, as for the archer to put all his arrows in the same place, or for the thermometer to indicate the same temperature repeatedly. The condition of reproducibility is verified as soon as the same test, applied several times in succession under the same conditions, leads each time to the same result. It is therefore a very different quality from the previous one but just as important.

Figure 1 illustrates these two concepts.

How to ensure the quality of a questionnaire before its use in the study?

- By testing it during a break-in period (pre-test) and applying it some time away (re-test). Agreement between responses should be high, if the information sought had no reason to change. This tests reproducibility.
- Then by comparing the results obtained by its own questionnaire with those of a reference questionnaire previously published by another team in similar circumstances on a similar population. This tests for accuracy.
- Or by using a measurement method already validated by another team: there are sometimes dozens of them for the same situation to be assessed. It is nevertheless necessary to re-test the measurement method in this new context since it has been validated in different circumstances.

V - PRESENTATION OF THE QUESTIONNAIRE

All questionnaires must be very clearly identified.

The name of the study (e.g. CST2I - Comparison STudy of 2 Interventions) and the title of the questionnaire (e.g. "Assessment of the quality of life at 6 months") must appear in large print on the cover as well as center code number (in a multicenter study), the date at which it was completed, and the patient identification.

If several different questionnaires are used in the same patient, each questionnaire must be distinguishable, for example thanks to a specific color, an easily recognizable specific identification (for example, | Q | U | A | V | 0 | 6 | and | Q | U | A | V | 2 | 4 | represent respectively the quality of life assessment questionnaires at 6 and 24 months).

Questionnaires of the same type, used in different patients, are identified by a unique code which respects anonymity while avoiding confusion between questionnaires. We often uses the following format: | K | E | R | C | E |, where the first three letters represent the first three of the surname and the following two the first two of the first name.

Questions should be numbered in sequential order, with no omissions. The same is true for pages.

Any symbol that can facilitate the rapid understanding of the questionnaire, without altering its meaning, is welcome. This is particularly true of an arrow which avoids the use of a sequence of words, of a point (•) or an asterisk (*) placed in front of each idea in order to distinguish it from the preceding one. It is the same, to avoid getting the wrong line, for the dotted lines which connect the end of the question to the box in which the answer will appear.

Born on a Friday the 13th? (no = 0; yes = 1; don't know = 9) …… |__|

Readability and understanding are always facilitated by sufficient spacing between questions.

VI - THE PRE-TEST

The pre-test represents the final phase of the preparation of the questionnaire. It can be difficult to live because the questioning is sometimes painful! However, it is an essential step.

It is conducted under conditions identical to those of the real study and must relate to at least 20 to 30 individuals. How many pre-tests? It is rare for the questionnaire to be perfectly rectified from the outset: a second pre-test is probably to be planned as soon as the study is planned.

On the occasion of this pre-test are obtained for the first time:

- quantitative data, such as the time needed to complete the questionnaire (average duration, minimum and maximum duration), the number of questions that do not get an answer, the number of inconsistent answers (disagreeing with an answer to a previous item), the existence of any uncompleted questionnaires, etc.
- qualitative data, perhaps the most important, such as the general impression of ease, or not, in completing the questionnaire, or the difficulties linked to the cultural level of the respondents (problem more often linked to the vocabulary used than to the concept addressed by the question), or relating to the logical sequence of the questions, …

The comments of the questioned are very rich in information. Never neglecting them and trying to obtain as many as possible and of the best quality are two essential rules. These comments can lead to the resurgence of concepts that were initially abandoned or ignored, or to the emergence of new concepts. They may also concern the readability of the questionnaire, the repetition of questions, the insufficient space left for the answer to open questions...

CONCLUSION

Measurement instruments play an important, sometimes underestimated, role in the quality of a study. To accurately determine the relationships between studied factors and judgment criteria, these factors must be measured in a reproducible and exact way. The validation of measuring instruments is therefore essential and corresponds to a delicate and often very long process.

The clinician or research team using a questionnaire pursues two goals:

- obtain the information necessary for their study or investigation,
- measure this information with the maximum quality (in terms of accuracy and reproducibility),

while respecting the dignity of patients. The golden rule is certainly to establish a relationship of trust from the first questions.