A QUALITY OF LIFE TOOL KIT INVENOUS DISORDERS

Robert Launois\(^{(1,2)}\)

\(^{(1)}\) REES France - 28, rue d’Assas - 75 006 Paris – Email : reesfrance@wanadoo.fr - Site Internet : http://www.rees-france.com
\(^{(2)}\) UNIVERSITE Paris XIII – Faculté de médecine Léonard de Vinci – 74, rue Marcel Cachin - 93017 Bobigny – France

Correspondence and reprint address :

Professeur Robert LAUNOIS
REES France
28, rue d’Assas
75006 PARIS, France
Tel. +33 1 44 39 16 90
Fax. +33 1 44 39 16 92
Email : launois.rees@wanadoo.fr
Acknowledgments and funding

This study was sponsored by Laboratoire Servier, France
Cover title: A QUALITY OF LIFE TOOL KIT

4418 words

Key words: Chronic veinous insufficiency questionnaire - Generic and specific Quality of life scales – Utility - Preference approach
ABSTRACT

Individual preferences are now at the center of the medical decision-making process. Different experimental methods are being used to reveal them. Psychometricians use direct observation of a subject’s reactions in a particular pathological situation: patients are asked to arrange the intensity of the impacts on numeric ordinal scales, but an actual metrical measure is not yet available. Traditionally, economists believe that in a market, only the consumer’s choices enable us to estimate his level of satisfaction. In the health care field, where market mechanisms are not fully operational, they tend to extract the patients preferences through forced choices between hypothetical health states. A real metrical measure is thus obtained. The objective of this article is to exhibit these two methods: psychometric and economic, and to show how they have been implemented in medical research.
CONCEPTS

The first stage of any study examining the quality of life is to define the universe of the area to be studied. Once defined, the universe must be categorized to define specifics domains to be quantified. In order to assess these domains, a number of criteria or indicators must be available in order to quantify them and appropriate scaling procedures must be selected. Finally, development of a definitive indicator system must consider objectives for which it has been designed, without which results are meaningless.

Quality of life is such a unifying concept that ultimately all facets of the being may be included: environmental factors, behavior, and lifestyle. This discussion shall be restricted to those factors which influence patients’ quality of life as a result of disease or its treatment.

Life may not, however, be assessed generally: at best, different aspects of life may be assessed. Categorization of health is a difficult stage. The specific nature of these dimensions is controversial. Factor analysis has supported the validity of four distinct areas: symptoms, functional well being, emotional stability, and appropriate social integration. Up to this point these four domains are only concepts, that is, abstract principles. Measurement of these domains must be performed using solid recordable parameters. For each domain, a number of items has to be defined which will act as intermediaries between the abstract characteristics to be inferred and either objective or subjective measures. Choices depend on the approach chosen to assess health problems. For some authors, the definition of health may be restricted to the absence of clinical symptoms or biological abnormalities. Other workers distinguish between those diseases which may be defined by the profession, and sickness expressed in terms of behavior. A number of definitions stress the patient's perception of illness, that is, they are based primarily on a patient's individual satisfaction or lack of satisfaction with his well-being. The intensity of
symptoms, degree of incapacity, or level of dissatisfaction depend both on the absolute severity of the phenomenon and on the degree to which it interferes with daily life.

The scoring procedures, that is, the conventions which govern allocation of values for different indicator positions, are a primary feature of standardization required for the measurement instrument. They package empirical interpretation into a unit which may be used and dictate the method of statistical assessment of findings. The best characteristic of a scale is its invariance, that is, the degree to which it can be manipulated without distorting its structure. In the ordinal scale all transformations which preserve the order in the scale do not change available information. Such a scale is said to be preserved following monotonic transformation. In an interval scale all numbers on the scale may be multiplied by a constant factor, or the origin shifted by a constant number without changing results. Such a scale is said to be preserved by affine transformation \((y = mx + c)\). Finally, in a ratio scale the relationship between values is not changed if they are multiplied by the same constant factor. It is preserved by linear transformation \((y = mx)\). The more precise the information contained in the scale the more restricted the ability to modify the scale without changing the information contained therein becomes.

If scales are to be used as measurement instruments, they must be applied appropriately if they are to produce reliable results; in other words, they must measure what they were designed to assess. A discriminative tool constructed to identify a problem is not necessarily that which allows changes to be followed, and the tool used to follow changes may not be useful to assess if the allocation of resources is right. The choice of a method requires initial definition of users' needs: identification of a problem, assessment of change in response to treatment, or greater and more coherent use of scarce resources. An economist needs overall results. He requires a common measurement to compare the effects of decisions high up the health care system: techniques and equipment available to a statistically average population. A clinician's aim is to
apply techniques and equipment available to him, to draw from them the maximum possible benefit for the patient. All features of the disease must be approached, explaining why assessment has remained a multidimensional problem. These two different attitudes have produced two different approaches in the development of questionnaires. The quality of life may be assessed in two ways globally, by examining the whole health states status combining but not categorizing partial observations; or by composition, by constructing the general from the specific.

The overall approach is based on the supposition that quality of life exists as a continuum from good health to death. Most often several dysfunctions appear in combination. The entire range of quality of life alterations must therefore be situated on an interval scale by reasoning in terms of stereotypes or on the basis of a health status classification system. The assessment involves determining values of coefficients between zero and one attributable to each of these typical states. These weighting factors constitute adjustment factors of the quantity of life according to its quality, hence their name of "quality of life (QOL) coefficients". The product of the number of years or fractions of a year spent in a particular health status and the corresponding quality of life coefficient transforms the time spent in poor health into equivalent fractions of years of good health. Repetition of this operation over time for various stages provide the number of years left to live adjusted as a function of the quality of life, called QALY (quality adjusted life years). The cost of the treatment can then be divided by the QALY result in order to compare the respective merits of the two treatments.

In the compositional approach a number of items are combined either simply or by ad hoc weighting of selected variables into subscales. These scores by dimensions may or may not aggregated to produce an overall score. When they remain as distinct entities in the final assessment mechanism they produce profiles. If they are combined into a single figure, the term...
index or combined indicator is used. In all cases, the method chosen will reflect the complexity of the situation. This is a method which has been used since the 19th century by psychometrists endeavouring to impose "the discipline of measurement and figures to aspects of the spirit." This concept was also put forward by Alvan Fenstein and the psychopathologists in recommending grading of clinical judgments.

**UTILITY PREFERENCE APPROACH**

The methods used to record individuals' preferences are highly varied: "standard gamble," "time trade-off," and "category rating." The first of these methods was traditionally used to assess key preferences in situations of uncertainty. Because of this it is considered to be particularly appropriate in medical fields.

The protocol on which it is based is simple: three states of health (S1, S2, and S3) are carefully detailed and shown to a subject who must choose between the following options: Either treatment A which guarantees situation S2, treatment B which may have two possible outcomes: state S1 of probability p or state S3 of probability 1 - p. States S1, S2, and S3 are arranged in a hierarchy with S2 occupying a position between S1 and S3. When the value of p is varied from zero to one this produces a threshold value where the patient is unable to decide between the two options. This value may be used to assess the utility of the first of these therapeutic possibilities.

The dilemma faced by patients suffering from varicous disease highlights the use of such a system. Mrs. X suffering from varices may be offered two possibilities: either long-term therapy or the risks of a stripping. The outcome of the first choice in the short term is without doubt: she will live. The second choice is more risky as the chances of surgical success without anesthesia trouble have been estimated by his general practitioner to be 99.09% in this case. The patient is caught between two possible courses of action. She may either choose the risky situation which
includes an unavoidable risk of death or adopt the secure option but, by definition give up any possibility of improving her functional and esthetic situation. A problem then arises in that if the patient opts for the secure course, he will be better off than if the worst outcome of the risky approach were to occur, that is, death, but worse off than if the operation succeeds. In order to decide she must assess the relative desirability of remaining in her present state with angina compared to the best and worst possible outcomes following the higher risk option.

The dilemma may be solved using a standard gamble based on population statistics. The structure of the gamble is identical to that of the initial problem. Choice is limited to a certain outcome and a risk outcome; survival without sequelae or death. Two differences exist, however, by comparison with the initial dilemma:

1. The decision rests on a hypothetical situation removing emotional overlay which played a part in the initial problem, and
2. Risk calculation is not based on personal assessment but on objective measurement.

By varying probabilities attributable to the higher risk situation it is possible to assess the psychological value which the subject attributes to the certainty situation. Where the chances of success of the risky approach is equal to 99.09% the patient must choose between the certainty of living with varicosis or the risk of undergoing an operation which may not to succeed due to the anesthesia risk. The risk may not worth the gamble and the patient chooses the safe option.

If, in contrast, however, the anesthesia risk is very low (<0.01%) the probability of surviving the operation is raised and the patient in this situation will opt for the gamble.

Where the chances of success are low, the patient will favor the status quo. In the contrary situation, he will tend to lean toward the higher risk approach. The only difference between
these two situations is in the value $p$, the probability of success. As this increases the subject is less likely to choose the safe option and more likely to take the higher risk option. Finally, there is a threshold coefficient value where the patient is unable to choose between the two options. This value may be used to assess the current quality of life of the patient. If pain is severe or frequent, the value of the threshold coefficient is low.

If the patient will undergo anything to escape his current condition, the operation proposal is accepted even where the chances of success are limited, confirming the patient's poor state of health. If the pain is mild, the critical value for the coefficient is higher, the patient's present condition approaches that of good health; the patient does not accept the operation proposal unless it is almost certain to succeed.

The utility/preference approach has a number of advantages. First, this method produces a detailed measurement which combines mortality, morbidity, resultant physical sensory, and socio-emotional and cognitive effects, symptoms of the disease, and secondary effects of treatment into one single score. It allows calculation of a weighted life expectancy as a function of quality of life, which may not be done with specific profiles used to study the multiple effects of disease over time. Results and costs may be brought together when they may be related to a fundamental domain. Secondly, the score directly reflects patient preference and is not influenced by weighting factors defined by the healthy population or by the practitioners caring for the patient. The instrument may be specific for the disease if appropriate parameters are chosen to define the areas to be addressed. The method has an undisputed scientific basis: decision in the face of uncertainty, described by Von Neumann and Morgenstern. Despite the indisputable applications of this mechanism, it cannot be denied that there are restrictions to its use. First, replies vary as a function of the context in which questions are set and second, it is not
always possible to identify clinical variables which form the basis of the overall score. Finally, the sensitivity of a given indicator must be demonstrated in different disease states.

**GENERIC QUALITY OF LIFE SCALES**

Generic quality of scales use a single self-completed questionnaire said to apply to all diseases. The best known are the sickness impact profile (SIP) the Nottingham Health Profile (NHP) and the SF 36.

The SIP\(^{32}\) consists of 136 questions grouped into two domains: physical and psychological state, and five specific independent categories. The ensemble may be used to provide a global score. Each question assesses change in behavior and measures intensity of the upset. An interval scale using apparently equal gradations is used to assess the relative severity of each functional problem. This system was presented in 1975 to 108 Seattle HMO members and 25 health professionals. Each point was scored between zero and 15. Subdomain and overall global scores were calculated by dividing the sum of individual scores into the maximum possible score.

The NHP\(^{33}\) uses a two-part questionnaire. The first part consists of 38 questions with "Yes" or "No" responses, covering six domains: sleep, physical mobility, pain, effective reactions, social isolation, and emotional reaction. The second part assesses seven independent variables: work, salary, domestic work, interpersonal relationships, social life, family life, and sexual life, holidays and pastimes. Results are scored zero or one. Domains are not grouped together but points assessing each domain are weighted as a function of their relative severity. The reference technique used is pair comparison: each item in a domain is compared successfully to all other points within that domain. The subjectively more severe point is noted in each case. This system was used on a pilot group of 1,200 laymen without medical training to assess the frequency of
points deemed more severe than others. Symptoms and problems were graded in a hierarchy, comparing mean standard deviation to frequency.

The SF-36 questionnaire or “SF-36”, is a generic quality-of-life scale widely used throughout the world (11). It consists of 36 items divided among eight dimensions: “Physical Functioning” (10 items), “Physical Role” (4 items), “Bodily Pain” (2 items), “General Health” (5 items), “Vitality” (4 items), “Social Functioning” (2 items), “Emotional Role” (3 items), “Mental Health” (5 items). The higher the score, the better the patient’s health status.

Profiles are not without merit: their reproducibility and validity have been well established. They also allow assessment of different domains of quality of life in one combined scale without using multiple measurement scales. This is easier both for investigators and patients. They do have problems, however, notably, they do not allow the specific consequences of a given disease on quality of life to be assessed. In venous diseases, walking, far from being a handicap is considered to be therapy, whereas standing upright and immobile, which is not listed in the NHP or the SIP, is a real problem for patients suffering from venous disease. The failure of the parameters used to relate to the specific problem leads inevitably to sensitivity failings or even validity problems as the functional defects explored may not be relevant. The main drawback of the generic scales is their failure to identify small but significant clinical changes over time. Disease-specific scales seem to be better suited to discriminating between the benefits of a particular treatment in venous insufficiency. It therefore appears important to develop a specific quality of life indicator for venous insufficiency which takes into account the patients’ point of view and provides the attending physician with fine measurement of the functional and psychosocial consequences of the disorder.
SPECIFIC QUALITY OF LIFE SCALE

Proponents of specific quality of life scale\textsuperscript{35,36,37} advocate initial identification of relevant domains based on information reported in the literature and interviews with experts and patients. Signs and symptoms gathered may be combined to assess the impact of disease on the daily life of the patient. An initial questionnaire is designed to scale the indicators and to select the most relevant items. Given that the items pool is designed to provide the basis for construction of the final questionnaire it is important to list many more parameters than will be used in the definitive version. For each item two types of questions are formulated. The first concerns the presence and intensity of the impairment experienced and the second the importance attributed to it by the patient. Authors diverges at this stage of the analysis some only include in the final questionnaire those items with the highest product between frequency and importance\textsuperscript{38}. The instrument implicitly integrates patient preferences as these constitute the foundations of its structure and are at the basis of the choice of the items. However, once the questionnaire has been constructed, the items are equally weighted. The other school eliminates redundant parameters by principal component analysis; and regroups items according to their contribution (loading) to different factors.

The \textbf{Chronic Venous Insufficiency Quality of Life Questionnaire (CIVIQ)}\textsuperscript{39,40,41} is the first quality of life (QoL) questionnaire specific to chronic venous insufficiency (CVI) based on those rules. This 20-item questionnaire providing a Global Index and a profile on four QoL dimensions: Psychological (9 items), Pain (4 items), Physical (4 items) and Social (3 items) was initially developed in French (CIVIQ-2) Items on the CIVIQ scale were scored from 1 to 5. A low score for items represented greater patient comfort The recall period was the previous 4 weeks. In order to facilitate the interpretation of the results were reversed The highest figure was allocated to the lowest response option and vice-versa such as to obtain a score directly
proportional to the quality of life The score for each dimension was obtained by adding scores for each constituent item and the Global Index was obtained by summing the 20 items. Items were weighted equally The extreme possible minimum and maximum values which the scales have is dependent on the number of articles used in each of the dimensions and on the number of levels or categories for each item. For example, the score for a scale constructed from 2 items, each of 5 levels, can range from 2 points (2 * 1) to 10 points (2 * 5). The score of a scale which includes 5 items of 5 points each ranges from 5 points (5 * 1) to 25 points (5 * 5). It is, therefore, not straightforward to identify different orders of magnitude when calculating mean scores for different dimensions affected. In order to compare mean scores between dimensions or scales absolute scores were then converted into an index The method chosen was the one described by John E. Ware for the SF-36\textsuperscript{10}. For each dimension, we calculated $S$, the sum of scores for the patients' answers to the questions, $m$, the minimum theoretical value if all of the answers were on the first level of the scale for all of the items belonging to the dimension, and $M$, the maximum theoretical value if all of the items were scored at the maximum level on the scale for all items belonging to the dimension. The standardised score for each dimension was obtained by applying the equation: 

$$ \frac{S - m}{M - m} \times 100. $$

For each dimension, we therefore obtained a result from 0 to 100. According to this scoring method, improvement in quality of life between D0 and D28 is represented by an increase in the score. The difference is therefore positive in this situation.

CIVIQ2 was translated in English before subsequent use in a prospective, multicenter, international study, the RELIEF* study, assessing QoL in CVI patients (with and without venous reflux). 4,048 adult patients (46.6% with venous reflux) from 18 countries This process requires the identification of cross-cultural equivalence of the concepts or constructs measured so that comparisons may be made between populations of different cultures. Thus the content of the
questionnaire was first submitted to the opinion of the different countries involved. The linguistic validation was therefore performed in two steps: first, confirmation of the cultural relevance of the questionnaire’s content; second, the translation process. Cultural adaptations into English, Italian, Polish, Portuguese and Spanish were achieved. Additional versions were then made available in Arabic, Czech, Hungarian, Russian and Slovakian. During this process some conceptual changes were made to refine the content validity of the Social dimension of the CIVIQ. The item #11 “to practice a sport”, was translated in order to reflect a physical concept more than a social concept. This explains why it tends to group to the Physical dimension in the international sample. The same choice was made for the item #10 “to go out in the evening”, with similar results. The items #8 “to travel (car, bus, plane)”, #9 “to do the housework” and #15 “must take precautions” faced specific translation problems which result here in a questionable cross-cultural conceptual equivalence. In all, 9 versions were used in the RELIEF study.

METRIC PROPERTIES TO BE VALIDATED

Quality of life scales must be validated before being used in clinical trials. The scales must have specific metrological properties\(^{42-43-44-45-46-47}\) which have to be confirmed in a validation study; these properties are face validity, content validity precision, accuracy and sensitivity.

✓ Face Validity

The face validity of a questionnaire depends on the quality of its preparation: are the questions precise enough to specify the domains explored. Do they relate to a well defined period of time? Is the aggregation procedure adequate? Sentences are better to be phrased directly on patients’ descriptions. When patients’ own vocabulary is used to compile the questionnaire, a simple construction is obtained, without which it would be impossible for the patients to participate actively in the survey.
✓ **Content validity**

Content validity requires two conditions to be fulfilled; the entire range of patients’ complaints must be included and the items chosen must be shown to be representative. The content validity of the CIVIQ 2 is supported by the fact that the process used to select items was based on input from direct definitions of complaints obtained from patients suffering from veinous insufficiency. The use of factorial analysis enabled redundant items to be identified and excluded.

✓ **Reliability**

A scale is reliable if in measuring the same phenomenon on a number of occasions it produces similar results. To determine reliability the size of random measurement error must be assessed. If this is low the instrument provides a consistent measurement of the universe assessed. A number of authors describe this criterion as fidelity, others refer to the precision of the instrument. The most common method used to address this problem are internal consistency reliability and stability (test-retest) reliability. Different items in a dimension must be homogeneous as they relate to the same concept even if they are worded differently. This “internal consistency” is tested by the Cronbach Alpha, values of which range from 0 to 1. Coefficients above 0.70 are generally regarded to be acceptable for psychometric measurements. Intra-class correlation coefficients are used to confirm that the indicator is reproducible, i.e. to ensure that the answers to the same questionnaire remain unchanged in stable patients.

✓ **Construct validity**

An instrument is said to be valid if it measures what it truly purports to measure. This assumes both the absence of random error and systematic bias. Reliability is, therefore, a prerequisite, but is not sufficient for validity. For perfect validity, there must be no consistent error. In the absence of an undisputed reference standard, the validity of a measurement scale is obtained first by
confirming that its factorial structure remains stable on different population samples (structural validity), second, by testing if the results obtained using the scale fit the expected relationship across group of individuals or clinical data available (known group differences), and finally, by comparing its results at a given point of time, and longitudinally either to other validated quality of life scales assessing the same domain or to clinical indicators (convergent validity). Convergent validity is fulfilled when the scales scores for related concept produces a Spearman correlation coefficient of more than 0.4.

✓ **Responsiveness**

The sensitivity of an instrument is its capacity to detect clinically significant changes even if they are of low amplitude. Changes in scores must be able to be produced in patients whose state of health is deteriorating (or improving). An inadequately sensitive indicator might miss differences between two treatments, as it may be unable to detect subtle changes in the patients’ corresponding clinical states. An indicator is sensitive when it detects all changes in a given variable over and above the imprecision due to measurement error Different statistical methods have been described to assess sensitivity in detecting change. The least contentious appears to be the standardised response mean (SRM) and the effect size (ES) as described by Liang and Guyatt. SRM is the ratio of the mean change to the standard deviation of the change. The ES is identical to the SRM but uses the standard deviation of the scores on D0 as the denominator.

**CONCLUSION**

Quality of life is for the physician a mean to rise above too biological an approach,. Beyond organic disease, body spirit must be examined. The physician tries to achieve the best possible management for his patient who entrusts him with his most precious possession : his life. The objective is to control every aspect of the disease which explains the physician desire to remain
within an objective and multidimensional framework. The scientific collection of the data leaves little place to the evaluation of individual’s preferences. For the patient it is important to express the specificity of his complaints. The evaluation instrument has to reflect the patient’s central value and should integrate patient’ preferences in its structure and for the choice of its components. For society the goal of measurement is not to assess "the importance which each of us attaches to our lives," but to produce an overall morbidity indicator through which the effects of actions influencing health may be judged. The tool to be used requires a precise initial definition of users' needs. The choice of an indicator depends on the answers to the four following questions: Does the user require an indicator producing discriminative or evaluative results? Does he wish to assess the overall quality of life or specific facets of it? Which opinion is to be used: that of the doctor, that of the patient, or that of the population? Only too often, the available instruments are used blind without clearly addressing these questions.
BIBLIOGRAPHY


