Practical aspects for minimizing errors in the cross-cultural adaptation and validation of quality of life questionnaires

A. Lauffer*, L. Solé, S. Bernstein, M.H. Lopes, C.F. Francisconi

*Servicio de Gastroenterología, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil
Gastroenterología Diagnóstica y Terapéutica, Buenos Aires, Argentina

Received 8 Mar 2012; accepted 23 January 2013

Abstract
Background: The development and validation of questionnaires for evaluating quality of life (QoL) has become an important area of research. However, there is a proliferation of non-validated measuring instruments in the health setting that do not contribute to advances in scientific knowledge.
Aims: To present, through the analysis of available validated questionnaires, a checklist of the practical aspects of how to carry out the cross-cultural adaptation of QoL questionnaires (generic, or disease-specific) so that no step is overlooked in the evaluation process, and thus help prevent the elaboration of insufficient or incomplete validations.
Methods: We have consulted basic textbooks and Pubmed databases using the following key-words quality of life, questionnaires, and gastroenterology, confined to «validation studies» in English, Spanish, and Portuguese, and with no time limit, for the purpose of analyzing the translation and validation of the questionnaires available through the Mapi Institute and PROQOLID websites.
Results: A checklist is presented to aid in the planning and carrying out of the cross-cultural adaptation of QoL questionnaires, in conjunction with a glossary of key terms in the area of knowledge. The acronym DSTAC was used, which refers to each of the 5 stages involved in the recommended procedure. In addition, we provide a table of the QoL instruments that have been validated into Spanish.
Conclusions: This article provides information on how to adapt QoL questionnaires from a cross-cultural perspective, as well as to minimize common errors.

© 2012 Asociación Mexicana de Gastroenterología. Published by Masson Doyma México S.A. All rights reserved.
PALABRAS CLAVE
Calidad de vida; Cuestionarios; Gastroenterología; Transculturalidad; Vali-
dación; Clinimetría

Cómo minimizar errores al realizar la adaptación transcultural y la validación de los cuestionarios sobre calidad de vida: aspectos prácticos

Resumen
Antecedente: El desarrollo y la validación de los cuestionarios para evaluar calidad de vida (QoL) se han transformado en un área importante de la investigación. Sin embargo, la literatura revela una proliferación de instrumentos de medición no validados en el escenario de la salud, los cuales, no contribuyen al perfeccionamiento del conocimiento científico.

Objetivo: Presentar mediante el análisis de los cuestionarios validados disponibles, una lista de verificación sobre los aspectos prácticos de cómo llevar a cabo la adaptación transcultural de los cuestionarios de calidad de vida (genéricos, o específicos de una enfermedad) para que no sea olvidado ningún paso en el proceso de evaluación, evitándose de esta manera, validaciones sin suficientes o incompleteas.

Metodología: Hemos consultado los libros de texto básicos, la base de datos de Pubmed, usando las siguientes palabras claves: calidad de vida, cuestionarios y gastroenterología, con límites en «estudios de validación», tanto en inglés, como en español y portugués, sin límite de tiempo, con el objetivo de analizar la traducción y validación de los cuestionarios disponibles en Mapi Research y en el sitio PROQOLID.

Resultados: Se presenta una lista de verificación, para ayudar en la planificación y ejecución de la adaptación transcultural de los cuestionarios de calidad de vida, acompañado de un glosario de términos clave en esta área del conocimiento. Se utilizó el acrónimo DETAC que son 5 pasos, cada uno en referencia a una fase del procedimiento a seguir.

Conclusiones: Este artículo brinda información acerca de cómo debe realizarse la adaptación cultural de cuestionarios sobre calidad de vida y cómo minimizar errores comunes.

© 2012 Asociación Mexicana de Gastroenterología. Publicado por Masson Doyma México S.A. Todos los derechos reservados.

Introduction
Cross-cultural research has been carried out for years in the social sciences and its importance has been recognized in the health sciences, especially in the development of the concept of health-related quality of life (HRQOL).

In phase III clinical trials for the development of new drugs, quality of life measures have been almost systematically incorporated as one of the aims to be evaluated so that these drugs can be considered adequate for their proposed clinical application. The concept of quality-adjusted life years (QALY) is relatively new and is used to evaluate cost-benefit studies for the development of strategies in the area of health. On the other hand, it is often seen that these clinical trials use non-validated instruments for measuring quality of life, limiting the interpretation of their study data.

Physicians and researchers that do not have adequate instruments for measuring quality of life in their languages have 2 options: to develop a new instrument or to modify one that has been previously validated into another language, known as the process of cross-cultural adaptation.

In the health setting, there is a proliferation of instruments that are neither valid nor accurate, and therefore do not contribute to the advancement of scientific knowledge.

Given this reality, the aim of the present article is to review the practical aspects of how to carry out the cross-cultural adaptation process of quality of life questionnaires, both general and HRQOL, developing a checklist that attempts to ensure that no step in this process is left out, thus avoiding the elaboration of insufficient or incomplete validations of these tools. A glossary of terms is also presented to aid in understanding the terminology used in this area of knowledge.

Methodology
We have consulted basic textbooks and the Pubmed database, using the following keywords: quality of life, questionnaires, and gastroenterology, limiting them to “validation studies” in English, Spanish, and Portuguese, with no time restrictions, for the purpose of finding and analyzing validated general or disease-specific questionnaires in the field of gastroenterology.

The articles that resulted from the search were selected according to the study aim and analyzed with respect to the methodology used for their translation and cross-cultural validation. In addition, the questionnaires were analyzed on the available PROQOLID and Mapi Research Institute websites. After analyzing the translations and validations, common errors that impede the use of these questionnaires in clinical trials were identified; they are necessary mainly in the follow-up and evaluation of patients with functional...
digestive disorders. In regard to elaborating the checklist, the guidelines of the Mapi Research Institute were followed.

The initial process of questionnaire selection and study design

Quality of life can be evaluated by means of 2 basic approaches: through general questionnaires or through disease-specific questionnaires. When the study aim is to evaluate quality of life in general, the generic questionnaire can be chosen. When this aim is solely to evaluate the frequency, intensity, and duration of the symptoms of a given disease, then a one-dimensional disease-specific questionnaire should be selected. If the study aim is to evaluate HRQOL, the best choice is a multidimensional disease-specific questionnaire that examines aspects beyond the physical symptoms, such as the effect on the patient’s social and emotional life and/or the impact on daily activities.

Scale selection should be determined by the content and context of its use, given that there is no single evaluation tool; because every instrument has its advantages and disadvantages, the one that best adapts to the desired aim must be found. The questionnaire that is to be translated and validated into another language should be one that is practical to apply and that has the capacity to be generalized, so it can be used in different populations without losing its basic characteristics.

The translation process

The first step in a formal translation process implies that 2 or more translators work either separately or together to produce a concordant version of the questionnaire. Another approach is the back-translation or inverse translation method in which at least 2 persons competent in the source language of the process concurred.

Questionnaire translation recommendations from the Mapi Research Institute and Guillemín et al. have been widely used in validation studies and employ steps that are similar to those recommended by other authors. The process of translation and inverse translation (back-translation), as well as the work of a review committee, should mainly focus on the semantic equivalence evaluation. But there are other equivalence aspects that must also be evaluated, such as idiomatic equivalence and experimental or cultural equivalence that involve slang expressions, sayings, or words that are particular to a given culture. Some translators are not aware of the strict requisites involved in the translation of cross-cultural studies, and thus waste time making a literal translation, rather than paying proper attention to cultural meanings.

The need to make adjustments to the measuring instruments is not limited to situations of different countries and/or languages; local and regional adjustments are also required. It is difficult to decide whether the translated text is in accordance with the cultural characteristics of the population on which the version will be used. When choosing terminology, how much is gained with the cultural approximation and how much is lost in terms of generalization and comparison possibilities, has to be taken into account. Linguistic changes are produced in the same population with the passage of time, sometimes making transitory adjustments necessary.

It is advantageous to use words that are applicable to large geographic areas and cultural regions; experience has shown that an instrument is rarely used only in the country or cultural region in which it was created or for which it was adapted.

Applying the questionnaire: practical aspects

Important methodological problems are related to the decision of whether a healthy control group is necessary. One alternative is to use blood donors, because this group is naturally regarded as one of healthy volunteers. Even though it may not be the ideal control group, from a methodological viewpoint, it is the best that has been used in various studies. Blood donors are loyal to their centers and are considered healthy, and they tend to care about others. The subjects should always answer the questionnaires before their medical consultation or treatment in order to avoid measurement bias, except when response to the intervention is being analyzed; this is the test that evaluates the sensitivity of the instrument for detecting individual symptom variation.

Care should be taken that the clinical examination, diagnosis, test results, or medical rulings do not influence the answers to the questionnaire.

Linguistic validation of the questionnaires: practical aspects

The validation process of a questionnaire must also follow well-defined stages so that its usefulness and safety in clinical research is confirmed. This is produced through the gauging of clinical measurement properties known as “psychometric properties”. Different psychometric requisites should be covered in the process of linguistic validation, such as reliability, validity, and response to the intervention. Reliability can be evaluated by internal consistency, reproducibility, and discriminant validity. Validity can be evaluated by content value, criterion value, and construct value. Response to the intervention is regarded as a separate and distinct property from the psychometric properties of validity and reliability in quality of life questionnaires because it evaluates questions related to the sensitivity of the instrument.

The majority of researchers evaluate and/or publish response to the intervention results separately or after the validity and reliability results.

Statistical analysis and result presentation: practical aspects

To advance the statistical analysis, the score of each question and answer (item) of the questionnaire should be tabulated, as opposed to the result of the sum of the domains or the total score.

The score of added dimensions derived from the individual items should be simply expressed; for example, as a percentage in relation to the maximum score of the results. This manner of expression enables direct interpretation, even when the reader is not familiar with the instrument.
With this scenario in mind, the present article attempted to analyze information as to how the cultural adaptation of quality of life questionnaires should be carried out and how to reduce the possibility of committing the common errors that are seen in this field of investigation.

A checklist with the acronym DSTAC was proposed to help in the planning and implementation of these types of studies, along with a glossary of key terms in this area of knowledge.

**Final considerations**

The development and validation of instruments for evaluating quality of life and their specific components have become an important area in medical research. However, in order to demonstrate their measurement properties, these instruments must be evaluated and re-evaluated in different situations, in different research centers, and by many researchers in different populations.

In our analysis, we have found the following problems: questionnaires whose aims are not in accordance with the themes and questions; questionnaires originally developed for a given disease, but validated and used for others, thus laying open the risk for bias; questionnaires that evaluate long periods of time before the medical history, which is conducive to memory bias in the patients; reproducibility tests carried out in short periods of time that can also lead to patient memory bias; questionnaires validated specifically for one disease, when in reality their design is multidimensional or vice versa; scoring systems that are too complex, that make interpretation and statistical analysis difficult; questionnaires based on outdated diagnostic criteria; questionnaires designed for specific health systems that often do not state their aims in their titles; translations without back-translations and/or incomplete validations (psychometric properties).

Financial disclosure

No financial support was received in relation to this article.

Conflict of interests

The authors declare that there is no conflict of interest.

Appendix 1. DSTAC protocol for the cross-cultural adaptation of quality of life questionnaires

The aim of this protocol is to aid in the planning and culmination of cross-culturally adapted and validated generic or disease-specific quality of life questionnaires. A glossary of principal terms used in this area of research is listed at the end of the protocol.

The protocol is divided into 5 stages, each stage referring to a research phase, as follows:

D Define the aims, instruments, and tests of the study
S Schematize the study protocol
T Translate the questionnaire
A Apply the questionnaires
C Consolidate, analyze, and present the study data
## Stage D. Define the aims, instruments, and tests

<table>
<thead>
<tr>
<th>Steps</th>
<th>Problem</th>
<th>Suggestion/orientation</th>
</tr>
</thead>
</table>
| Step 1 | What kind of questionnaire is being validated? | - Generic  
- One-dimensional disease-specific  
- Multidimensional disease-specific |
| Step 2 | Is there another questionnaire for this purpose available in the target language to be validated?  
See subitem A | Look for the questionnaires available in the literature and determine whether they have already been adapted to the target language |
| Subitem A | If you find a validated questionnaire in the literature in the target language, go to Step 4. If not, go to step 3 |
| Step 3 | Does the questionnaire to be adapted cover the study aims? | Carry out a conceptual analysis of the items and dimensions of the questionnaire to see whether it covers the study aims and the sample characteristics |
| Step 4 | Is the evaluated period of time adequate? | The evaluated period of time should not be longer than 6 months (due to memory bias) and should be consistent with the evaluated disease |
| Step 5 | How has the questionnaire been developed?  
See Subitem B | Determine whether this questionnaire has been adequately developed and validated |
| Subitem B | If the questionnaire found has not been properly validated, is it possible to revalidate it (see step 6) or should the search be reinitiated (return to step 3)? |
| Step 6 | Does the present questionnaire have a copyright? | If it does, contact the authors |
| Step 7 | What kind of scoring mechanism is used in the questionnaire? | Examine the syntax of the questionnaire |

The following steps are also related to stage E (step 15) and should be planned ahead of time with the statistician

| Step 8 | What are the psychometric tests intended to be carried out? | Plan the psychometric tests. Continue with steps 9 through 13 |
| Step 9 | Is the evaluation of content validity an aim? | This should be done after obtaining the translated version, but before applying it to the samples. Determine beforehand how possible differences of opinion between the evaluators will be resolved |
| Step 10 | Is the evaluation of the criterion validity an aim?  
See subitem C | Carefully choose the questionnaire to be used because it will be the gold standard for the score comparison: pay attention to the syntax for later interpretation of the questionnaire and for the possible correlation difficulties between the evaluated domain scores and what is being validated |
| Subitem C | Get in touch with the authors of the questionnaire being studied and become familiar with its syntax (score calculation manual) |

Steps 11 through 14 concern the psychometric tests in regard to the validation study aims

| Step 11 | Does the questionnaire evaluate discriminant validity? | Choose the sample for the control group carefully, plan the sample calculation, and carry out the necessary matching |
| Step 12 | Does the questionnaire evaluate reproducibility? | Take into account that there should be no diagnostic or therapeutic intervention in the sample in regard to the measuring range |
| Step 13 | Is it an aim of the questionnaire to evaluate internal coherency? | Analyze the sample size calculation so that the results are not affected; it is possible to include a control sample in the calculation if discriminant validity data are being collected |
Step 14  Does the questionnaire intend to evaluate the response to the intervention?  This is important in disease-specific questionnaire validation because it will give the study greater credibility. It should be accommodated into a clinical trial, or a longitudinal study should be developed for this purpose. Review step 12

Stage S. Schematize the study

<table>
<thead>
<tr>
<th>Steps</th>
<th>Problem</th>
<th>Suggestion/orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 15</td>
<td>What will the study protocol be like? Make sure the study protocol includes the following: Selection and origin of the population studied Questionnaire placement Keep in mind the number of questionnaires to be used (whether they are included in other study protocols in the same study population) The amount of time available for applying the questionnaires</td>
<td></td>
</tr>
</tbody>
</table>

Step 16  Consult the statistician before beginning the translation process  - Planning: Calculate the sample size based on the psychometric tests to be carried out, the statistical analysis of the scoring methodology, and access to the syntax of the questionnaire(s) to be used - Evaluate the instrument’s type of scale and scoring

The syntax for many instruments is available and should be used to guarantee coherence between studies

Step 17  What is the sample population? The samples should be accessible and representative of the community in general

Step 18  What is the case and control group inclusion and exclusion criterion (if it is necessary from the methodological standpoint)? - The inclusion and exclusion criteria should be carefully planned out and well-defined to avoid including subjects with concomitant diseases that also affect quality of life (measuring bias) - Carry out the appropriate matching according to the methodological necessities

Stage T. Translate the questionnaire

<table>
<thead>
<tr>
<th>Steps</th>
<th>Problem</th>
<th>Suggestion/orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 19</td>
<td>How will the translation of the questionnaire be carried out? Establish a methodology</td>
<td></td>
</tr>
</tbody>
</table>

If the questionnaire has a company copyright, the methodology will most likely have to be proposed by the company. If this is not the case, the methodology described in step 20 is recommended

Step 20  Translation  Two qualified translators should separately translate the questionnaire from the source language into the target language from 2 independent texts. The translators’ native language should be the source language of the questionnaire, they should know the study aim, and they should not be in contact with one another during the translation. Note: the translators should know the study aim of the translation, but not of the back-translation

Step 21  Consensus Version  The researchers analyze the 2 translations to create a consensus version

Step 22  Back-translation  The consensus version should be back-translated to the source language by 2 translators that are fluent in the target language and whose native tongue is the source language. This should be done separately using 2 independent texts and the translators should not be in contact with one another during the back-translation. They should also not know the study aim
Step 23  | Quality control  | The back-translations are cross-analyzed and discussed in relation to the source version to evaluate the reliability of the translation.

Step 24  | Evaluation of the consensus version to elaborate the final text  | The semantic, idiomatic, and cultural equivalency of the final version of the questionnaire is analyzed by a multidisciplinary committee in order to elaborate the final text.

Step 25  | A pilot test will be carried out on representatives of the population for test validation  | At least 5 persons that are representative of the sample will be tested before the study is initiated in order to evaluate the comprehension of the themes, the guidelines/instructions of how to complete the questionnaire/form, and the format of the questionnaire. Make the necessary adjustments.

**Stage A. Apply the questionnaires**

<table>
<thead>
<tr>
<th>Steps</th>
<th>Problem</th>
<th>Suggestion/orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 26</td>
<td>Is the questionnaire applied during an interview?</td>
<td>In these cases, the questionnaire should be given to 50% of the sample by a second interviewer in order to evaluate inter and intraobserver concordance. The interviewers should be trained to apply the questionnaires.</td>
</tr>
<tr>
<td>Step 27</td>
<td>Is the questionnaire self-administered?  Subitem D</td>
<td>In these cases the same researcher should distribute the questionnaires to all the subjects.</td>
</tr>
<tr>
<td>Subitem D</td>
<td>Pay special attention to steps 28 through 30</td>
<td></td>
</tr>
<tr>
<td>Step 28</td>
<td>Do the subjects understand the questions asked in the questionnaire?</td>
<td>If they do not, the researcher must not interfere with the subject’s response. The researcher should read the question to the subject out loud and in the same tone of voice. If the doubt persists the researcher should move on to the next question.</td>
</tr>
<tr>
<td>Step 29</td>
<td>The handing in of the questionnaire by the subject to the researcher</td>
<td>The questionnaires should be checked when they are handed in to the researcher and they may be returned to the subjects to answer or complete the answer to any question left blank.</td>
</tr>
<tr>
<td>Step 30</td>
<td>And if the questionnaire is still returned incomplete?</td>
<td>Incomplete questionnaires will be excluded.</td>
</tr>
</tbody>
</table>

**Stage C. Consolidate the data**

<table>
<thead>
<tr>
<th>Steps</th>
<th>Problem</th>
<th>Suggestion/orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 31</td>
<td>How are the data tabulated before the statistical analysis?</td>
<td>Tabulate the results of each item of the questionnaires and do not add the domains together or the total scores.</td>
</tr>
<tr>
<td>Step 32</td>
<td>Statistical analysis</td>
<td>The professional that analyzes the data should present the results in an easily accessible manner so that readers who are not familiar with these types of data can understand them.</td>
</tr>
<tr>
<td>Step 33</td>
<td>The statistical analysis, if the response to the intervention was evaluated</td>
<td>The instrument’s sensitivity and specificity should be calculated.</td>
</tr>
</tbody>
</table>
Appendix 2. Glossary of common terms

Cross-cultural adaptation: the production of an equivalent instrument adapted to another culture.7

Quality of life (QOL): A subjective parameter by which direct questioning is a simple and appropriate way to gather information about how a patient feels and lives,16 improving the comprehension of the disease and treatment impact11; the perception of the individual as to his or her position in life in the context of the culture and value system in which he or she lives in relation to that individual’s objectives, expectations, standards, and concerns9,12; it is the existing relation between how the individual actually lives and the desired life pattern, which represents the comparative parameter within the community itself of those persons that feel they have or do not have quality of life.17

Health-related quality of life (HRQOL): the elements of quality of life directly related to individual health. It examines how the individual feels and perceives the manifestations of disease in daily life; it is a disease severity marker.7,17,33

Generic questionnaires: these instruments attempt to measure the important aspects of quality of life, making a health profile analysis.7

Disease-specific questionnaires: instruments that attempt to measure HRQOL, focusing on the health aspects that are within the field of interest. The questionnaire can be specific for a disease, population, or for a health-related problem or function.7

One-dimensional disease-specific questionnaires: they deal with HRQOL measurements that evaluate the frequency, intensity, and duration of disease symptoms.7

Multidimensional disease-specific questionnaires: these instruments deal with measurements for evaluating domains that go beyond symptoms, such as the impact on the patient’s social and emotional life and/or daily or routine activities.7

Internal consistency: a measurement used to evaluate accuracy through examining the coherence among the items and the homogeneity of the instrument; it is the most widely used measurement for estimating accuracy.7,28,29

Construct: these are abstractions, such as anxiety, pain, and fear – theoretical constructions whose aim is to organize and give meaning to our environment. Their formation is based on the relations between the measured variables that, in turn, are incalculable, and it is believed that they are responsible for the relation between the measured variables.7,28

Culture: values, beliefs, norms, and practices of a particular group of persons that direct their thoughts, decisions, and actions in a standardized way.1

Dimensions / domains: used synonymously, they signify the fragmentation of the global concept of quality of life into various components that represent the elements of the questionnaire, simplifying its evaluation; the fractioning of the overall concept of quality of life into various components, which are the dimensions, and taking the simplest evaluations.17,27

Concept and item equivalence: the exploration of the different items and areas covered by the source instrument for determining if they are relevant and pertinent to the culture to which they are to be adapted.9,16

Cultural equivalence: the evaluation of the use of terms, so that they are coherent with the experiences lived by the study population within its cultural context.2,15,16

Idiomatic equivalence: the translation of idiomatic and colloquial expressions that can rarely be translated and in these cases have to be substituted with equivalent expressions.2,15,16

Semantic equivalence: the evaluation of the equivalence of the grammar, vocabulary, and words of the source instrument with the one being adapted; it implies the capacity to transfer the meaning of the concepts contained in the source instrument to the produced version, having a similar effect on the subjects in the setting of the two cultures that participate in this process.2,7,15,16

Ethnicity: it is regarded as a measurement of cultural heritage.1

Reliability: it is the certainty and confidence that the score represents the true score, which is a question of stability during the time of result repetition; the capacity of the test to be repeated several times and produce the same result; it provides information on the stability of the construct and whether it can be generalized.7

Disease-specific instruments: they focus on measuring the HRQOL, and are centered on specific aspects of health status for the area of interest.7

Generic instruments: they evaluate all the important aspects of quality of life, making a health profile analysis.7

Multidimensional instruments: they evaluate social and emotional aspects, the impact of disease on daily life, as well as the severity, frequency, and duration of symptoms at a given period of time.10

One-dimensional instruments: they evaluate the frequency, severity, and duration of symptoms at a given period of time.10

Cross-cultural research: a term that is frequently applied to prevalence and incidence studies of different diseases or of psychosocial variables in different countries or distinct ethnic or social groups.1,7

Psychometrics: this refers to the discipline related to psychological or mental tests and to any quantitative analysis of the psychological traits or attitudes of an individual, as well as to his or her mental processes11; it is a branch of statistics that studies and measures psychological phenomena through tests that analyze the accuracy and validity of the questionnaire.3,2,21

Race: it refers to the phenotype, for example, the color of the skin.1

Reproducibility: also called test-retest, it is the test that attempts to show the accuracy of data obtained through the application of the instrument at different points in time; the obtention of similar data in repeated evaluations (temporary data repetition); the correlation between the points obtained by the same person on different occasions, usually within a 15-day interval, in the search for similar results to demonstrate data accuracy.14

Response to the intervention: a test that evaluates the sensitivity of the instrument for detecting variations in an individual’s symptoms.4,35

Back-translation: to translate a document from the source language into the target language and then back to the source language.1,2,7,36
Validation: a test that evaluates whether the instrument actually measures what it is supposed to and if it maintains the measurement value when the hypotheses about the relation between the scale score and a particular criterion are confirmed.\(^7\)

Discriminant validity: a test that evaluates the capacity of the instrument to distinguish between extremes, such as patient groups and control groups; it evaluates specificity, as well as the discrepancy between the questionnaire results and the variables, that theoretically, should not correlate.\(^37\)

Content validity: a test through which the instrument is analyzed by reviewers that are experts in the construct’s area of research, that did not participate in the experiment, for the purpose of evaluating whether the translation and the instrument content are representative of the condition intended to be analyzed (semantic equivalence and construct validity)\(^37\); it is an interview-based qualitative evaluation.\(^7\)

Construct validity: a test that analyzes whether the instrument is representative of the theoretic referential of the construct being evaluated\(^37\) through statistical processes for supporting the hypothesis of the structure of the questionnaire’s scale.\(^33\)

Criterion validity (concurrent): a test that analyzes the quality of the construct of the instrument that is being evaluated in relation to the established and widely accepted criterion standard.\(^34\)

Validity of outcome measures: the psychometric validation of a measure based on the symptoms incorporating various components such as content validity, construct validity, reliability, response capacity, and criterion validity. The participation of the patients in developing their result measurement is stressed by the Food and Drug Administration. This can be fostered by structured interview sessions, focal groups, and quality research methods. The measuring of the result must have an effective measurement range so that the instrument can detect changes in the results during the clinical trial.\(^38\)
<table>
<thead>
<tr>
<th>Author and year. Name and acronym</th>
<th>Questionnaire characteristics</th>
<th>Questionnaire aims</th>
<th>Comments on the instrument and original validation</th>
<th>Validation into Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>El-Omar et al., 1996</td>
<td>Multidimensional</td>
<td>Evaluates the impact of dyspepsia symptoms on the life of the patients Records symptom frequency, the effect on routine activities, time out of work, frequency of medical consultation, request for clinical evaluations, and the use of over-the-counter and prescription medications.</td>
<td>Original language: English (Scotland) Developed in Scotland at a health system organization with a high environmental impact The items are in line with the aims of the questionnaire Depending on the population, it does not enable adequate symptom evaluation, given that 5 of the 7 items ask questions about the use of health services Validation was conducted on patients with different gastric problems, such as responsiveness in patients with duodenal ulcer. There is only a one-week reproducibility interval The questions referred to the last 6 months, which could be a cause of memory bias</td>
<td>Carried out by Monés et al., 2001 No reference was made to the translation and back-translation method, only comments on the interpretability in relation to cross-cultural adaptation Internal consistency and responsiveness were included It does not seem adequate for the Spanish-speaking population</td>
</tr>
<tr>
<td>Rabeneck, 2001</td>
<td>Multidimensional</td>
<td>To evaluate quality of life in functional dyspepsia</td>
<td>Original language: English (USA) Validity and responsiveness were evaluated One of the aims of the questionnaire is not contemplated in the title There is low test-retest accuracy due to a long test interval (4 weeks) It evaluates symptom severity but not frequency It was revalidated in 2002 because it had not been completely validated when it was developed</td>
<td>Validated by the Mapi Research Institute</td>
</tr>
</tbody>
</table>

Table of questionnaires validated into Spanish

**Table of questionnaires validated into Spanish**

<table>
<thead>
<tr>
<th>Author and year. Name and acronym</th>
<th>Questionnaire characteristics</th>
<th>Questionnaire aims</th>
<th>Comments on the instrument and original validation</th>
<th>Validation into Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>El-Omar et al., 1996</td>
<td>Multidimensional</td>
<td>Evaluates the impact of dyspepsia symptoms on the life of the patients Records symptom frequency, the effect on routine activities, time out of work, frequency of medical consultation, request for clinical evaluations, and the use of over-the-counter and prescription medications.</td>
<td>Original language: English (Scotland) Developed in Scotland at a health system organization with a high environmental impact The items are in line with the aims of the questionnaire Depending on the population, it does not enable adequate symptom evaluation, given that 5 of the 7 items ask questions about the use of health services Validation was conducted on patients with different gastric problems, such as responsiveness in patients with duodenal ulcer. There is only a one-week reproducibility interval The questions referred to the last 6 months, which could be a cause of memory bias</td>
<td>Carried out by Monés et al., 2001 No reference was made to the translation and back-translation method, only comments on the interpretability in relation to cross-cultural adaptation Internal consistency and responsiveness were included It does not seem adequate for the Spanish-speaking population</td>
</tr>
<tr>
<td>Rabeneck, 2001</td>
<td>Multidimensional</td>
<td>To evaluate quality of life in functional dyspepsia</td>
<td>Original language: English (USA) Validity and responsiveness were evaluated One of the aims of the questionnaire is not contemplated in the title There is low test-retest accuracy due to a long test interval (4 weeks) It evaluates symptom severity but not frequency It was revalidated in 2002 because it had not been completely validated when it was developed</td>
<td>Validated by the Mapi Research Institute</td>
</tr>
<tr>
<td>Author and year</td>
<td>Name and acronym</td>
<td>Questionnaire characteristics</td>
<td>Questionnaire aims</td>
<td>Comments on the instrument and original validation</td>
</tr>
<tr>
<td>----------------</td>
<td>------------------</td>
<td>-----------------------------</td>
<td>-------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Fraser et al., 2007</td>
<td>Leeds Dyspepsia Questionnaire Short-form LDQ-SF</td>
<td>Multidimensional</td>
<td>To validate a short form (SF) of the LDQ dyspepsia questionnaire and to evaluate responsiveness of the evaluation</td>
<td>It has just a 2-day test-retest reliability interval. The questionnaire is not multidimensional. Even though it is regarded as disease-specific for functional dyspepsia, the evaluated symptoms were not included: retrosternal burning sensation, pain or discomfort in the upper abdomen, regurgitation, and nausea. Responsiveness was not evaluated.</td>
</tr>
<tr>
<td>Goldman et al., 2002</td>
<td>Spanish Language Dyspepsia Questionnaire SLDQ</td>
<td>Disease-specific</td>
<td>To develop a questionnaire in Spanish for evaluating the symptoms of dyspepsia</td>
<td>Validated by the Mapi Research Institute. The questionnaire was developed from the revision and shortening of the previously validated LDQ that evaluates the frequency and severity of symptoms, increasing the accuracy of the evaluation. It is regarded as disease-specific for functional dyspepsia, multidimensional. It was validated by the Mapi Research Institute. It is completed by the researchers. Acceptability, interpretability, validity, reliability, internal consistency, and responsiveness were evaluated.</td>
</tr>
<tr>
<td>Author and year. Name and acronym</td>
<td>Questionnaire characteristics</td>
<td>Questionnaire aims</td>
<td>Comments on the instrument and original validation</td>
<td>Validation into Spanish</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>------------------------------</td>
<td>------------------</td>
<td>---------------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Raymond et al., “Reflux Qual” 1999</td>
<td>Multidimensional Disease-specific Self-administered Records 37 articles or items occurring over the last 4 weeks organized into 7 dimensions (daily life, discomfort, well-being, physical aspects, anxiety, sleep, eating habits) with the 5-point Likert scale</td>
<td>Evaluates quality of life in patients with reflux</td>
<td>Original language: French (France) Well-conducted validation (instrument acceptability, validity, and accuracy) A short form of the Reflux Qual (RQSF) was developed in 2005 by Amouretti et al. with 8 items, adequately validated for daily use in clinical practice The short form questions are oriented towards quality of life in general, not specifically GERD-directed, and could be applicable to other upper GI disturbances</td>
<td>Validated by the Mapi Research Institute Short-form still not available in Spanish</td>
</tr>
<tr>
<td>Wiklund Ingela QOLRAD-Heartburn QoLRAD-Dyspepsia QOLRAD-NSAID: questionnaire for patients with Gastrointestinal Symptoms</td>
<td>Multidimensional Disease-specific Self-administered Records 25 articles or items occurring over the previous week, combined into 5 dimensions (emotional stress, sleep disorders, vitality, eating and drinking problems, physical and social functioning) Graduated 7-point Likert scale (low values indicate a greater impact on daily performance)</td>
<td>To evaluate digestive tract diseases, pathologic conditions, and signs and symptoms</td>
<td>Original language: English Initially, only the QOLRAD existed for evaluating dyspepsia and reflux, but the majority of questions were directed towards gastroesophageal reflux. QOLRADs for heartburn and dyspepsia and dyspepsia and NSAIDs have recently been developed The psychometric validation methods were questioned in some revision articles</td>
<td>Appropriately translated and validated by the author in a 6-country study, but not using the Mapi Research Institute reference methodology</td>
</tr>
<tr>
<td>Rentzet al., 2004 Patient Assessment of Gastrointestinal Disorders Symptom Severity Index PAGI-SYM</td>
<td>One-dimensional Disease-specific Self-administered Records 20 articles or items occurring over the past 2 weeks</td>
<td>To measure symptom severity in patients with upper gastrointestinal disorders: gastroesophageal reflux, dyspepsia, and gastroparesis</td>
<td>Original language: English (USA) Well-conducted validation; the construct appears to support the evaluation of subjects with reflux, dyspepsia, and gastroparesis A one-dimensional questionnaire for evaluating 3 gastrointestinal disorders. It is nonspecific for a particular disorder It evaluates the severity, but not the frequency, which diminishes the accuracy of the instrument</td>
<td>Validated by the Mapi Research Institute</td>
</tr>
<tr>
<td>Author and year</td>
<td>Name and acronym</td>
<td>Questionnaire characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>De La Loge et al., 2004</td>
<td>PAGI-QOL</td>
<td>Multidimensional, Disease-specific, Self-administered, Records 30 articles or items occurring over the past 2 weeks, combining them into 5 dimensions (daily activities, dressing, dietary and eating habits, ... wellbeing) with a graduated 5-point Likert scale (low values indicate a more severe impact on daily functioning)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talley et al., 1999</td>
<td>NDI</td>
<td>Multidimensional, Disease-specific, Self-administered, Records 42 or 55 articles or items occurring over the past 2 weeks, combining with daily activities, eating/drinking, knowledge, control, study/work</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nepean Dyspepsia Index</td>
<td>NDITalley et al., 2001</td>
<td>Multidimensional, Disease-specific, Self-administered, Records 42 or 55 articles or items occurring over the past 2 weeks, combining with daily activities, eating/drinking, knowledge, control, study/work</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDI-SF (short form)</td>
<td>NDITalley et al., 2001</td>
<td>Multidimensional, Disease-specific, Self-administered, Records 42 or 55 articles or items occurring over the past 2 weeks, combining with daily activities, eating/drinking, knowledge, control, study/work</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Questionnaire aims**

- To measure quality of life and the specific symptoms for: dyspepsia, gastroesophageal reflux, and gastroparesis
- To measure the deterioration of the capacity of a subject to participate and enjoy relevant aspects of his or her life
- To evaluate the frequency, intensity, and inconvenience or discomfort in 15 upper gastrointestinal tract symptoms

**Comments on the instrument and original validation**

- Adequate multicenter validation
- Despite having jointly evaluated reflux and gastroparesis, it has a positive approach to functional dyspepsia within the spectrum of the upper digestive functional disorders
- The reliability, discriminant validity, and responsiveness of the list of symptoms have been established. The NDI register of the 42 items is too long. The symptom list validity has not been evaluated in comparison with other instruments
- The short form does not include one of the symptom components of the symptom articles or items are separated from quality of life
<table>
<thead>
<tr>
<th>Author and year. Name and acronym</th>
<th>Questionnaire characteristics</th>
<th>Questionnaire aims</th>
<th>Comments on the instrument and original validation</th>
<th>Validation into Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chassany et al., 1999</td>
<td>Multidimensional</td>
<td>To specifically measure the physical and psychological effects and the constant impacts of dyspepsia and irritable bowel syndrome</td>
<td>Original language: French (France) Well-conducted validation (validity, reliability, reproducibility, acceptability, responsiveness). Adequately translated into English and German, using the translation and back-translation method. It only evaluates quality of life, without assessing the frequency or severity of symptoms. Long questionnaires tire the patient and he or she may answer quickly in order to finish it sooner.</td>
<td>Validated by the Mapi Research Institute.</td>
</tr>
<tr>
<td>Patrick et al., 1998</td>
<td>Multidimensional</td>
<td>To evaluate the impact of Irritable Bowel Syndrome and its treatment</td>
<td>Original language: English (USA) Well-conducted validation. Questions directed towards quality of life in IBS. It only evaluates quality of life without taking into account the specific symptoms of IBS.</td>
<td>Schmulson et al., 2007 According to the abstract, the linguistic validation followed the standard guidelines for Mexican Spanish and further psychometric validation. Validated by the Mapi Research Institute</td>
</tr>
<tr>
<td>Revicki et al., 2004</td>
<td>One-dimensional</td>
<td>To evaluate the severity of symptoms reported by patients presenting with gastroparesis that are enrolled in clinical trials</td>
<td>Original language: English (USA) Rapid application. It evaluates the severity of symptoms without evaluating the frequency.</td>
<td>Distributed by the Mapi Research Trust.</td>
</tr>
<tr>
<td>Author and year</td>
<td>Name and acronym</td>
<td>Questionnaire characteristics</td>
<td>Questionnaire aims</td>
<td>Comments on the instrument and original validation</td>
</tr>
<tr>
<td>----------------</td>
<td>------------------</td>
<td>------------------------------</td>
<td>-------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>Svedlund et al., 1988</td>
<td>GSRS</td>
<td>One-dimensional Disease-specific</td>
<td>Interviewer-administered</td>
<td>The questionnaire records 15 articles or items occurring over the last month or week, and combines them into 5 symptom groups that describe reflux, abdominal pain, indigestion, diarrhea, and constipation. Graduated 7-point Likert scale (1 is the absence of bothersome symptoms and 7 is for very bothersome symptoms). To measure a wide range of gastrointestinal symptoms (digestive tract diseases, tumors, peptic ulcer, irritable bowel syndrome, gastric cancer); it does not include colorectal, small bowel, and hepatobiliary tumors. Due to the fact that it evaluates different disorders, not all are done adequately; for example, reflux is examined with only 2 well-directed questions. The interviewer-administered format can be a negative point.</td>
</tr>
<tr>
<td>Kulich et al., 2005</td>
<td>QoL-PEI</td>
<td>Multidimensional Disease-specific</td>
<td>Self-administered</td>
<td>Its results are acceptable. Internal consistency and validity are adequate for most of the questions. It is sometimes confused with a multidimensional instrument.</td>
</tr>
<tr>
<td>Francis et al., 1997</td>
<td>IBSSS</td>
<td>One-dimensional Disease-specific</td>
<td>Self-administered</td>
<td>Records 9 questions on events occurring over the last 10 days that are organized with categorical response options in relation to a 0-100 severity score and a maximum sum of 500 points. To evaluate IBS severity. Original language: English (UK). It was validated for severity, reproducibility, and responsiveness; it evaluates the frequency, intensity, and duration of symptoms and their impact on daily life. Reproducibility was carried out in a 6 to 24-h interval. The psychometric properties of validity were not evaluated. A small number of patients are tested. It evaluates the severity but not the frequency of the symptoms, thus reducing instrument accuracy.</td>
</tr>
<tr>
<td>Ruiz et al., 2001</td>
<td>IBSSS</td>
<td>One-dimensional Disease-specific</td>
<td>Self-administered</td>
<td>It was adapted and validated as multidimensional, which does not seem appropriate. The adaptation and validation process includes the translation and back-translation, dimensionality, accuracy, and validity (of the content, of the convergent, predictive, and discriminative construct). The translation into Spanish was apparently well done. In regard to validation, they followed the same format as that used in the original validation.</td>
</tr>
<tr>
<td>Almansa et al., 2011</td>
<td>DRHSD</td>
<td>Multidimensional Disease-specific</td>
<td>Self-administered</td>
<td>The translation into Spanish was apparently well done. It was adapted and validated as a one-dimensional instrument. The method used for cross-cultural adaptation is not mentioned in the abstract. The questionnaire has internal consistency and validity but not reproducibility or responsiveness. Its results are acceptable. It evaluates dyspepsia in relation to quality of life. The questionnaire is essentially one-dimensional. It was adapted and validated as multidimensional, which does not seem appropriate. Due to the fact that it evaluates different disorders, not all are done adequately; for example, reflux is examined with only 2 well-directed questions. The interviewer-administered format can be a negative point. It is sometimes confused with a multidimensional instrument.</td>
</tr>
<tr>
<td>Cook et al., 1999</td>
<td>DRQ</td>
<td>Multidimensional Disease-specific</td>
<td>Self-administered</td>
<td>Records 15 articles or items occurring over the last month, and combines them into 5 symptom groups that describe reflux, abdominal pain, indigestion, diarrhea, and constipation. Graduated 7-point Likert scale (1 is the absence of bothersome symptoms and 7 is for very bothersome symptoms). To measure a wide range of gastrointestinal symptoms (digestive tract diseases, tumors, peptic ulcer, irritable bowel syndrome, gastric cancer); it does not include colorectal, small bowel, and hepatobiliary tumors. Due to the fact that it evaluates different disorders, not all are done adequately; for example, reflux is examined with only 2 well-directed questions. The interviewer-administered format can be a negative point. It is sometimes confused with a multidimensional instrument.</td>
</tr>
<tr>
<td>Nordin et al., 1984</td>
<td>GSRS</td>
<td>One-dimensional Disease-specific</td>
<td>Interviewer-administered</td>
<td>Records 15 articles or items occurring over the last month, and combines them into 5 symptom groups that describe reflux, abdominal pain, indigestion, diarrhea, and constipation. Graduated 7-point Likert scale (1 is the absence of bothersome symptoms and 7 is for very bothersome symptoms). To measure a wide range of gastrointestinal symptoms (digestive tract diseases, tumors, peptic ulcer, irritable bowel syndrome, gastric cancer); it does not include colorectal, small bowel, and hepatobiliary tumors. Due to the fact that it evaluates different disorders, not all are done adequately; for example, reflux is examined with only 2 well-directed questions. The interviewer-administered format can be a negative point. It is sometimes confused with a multidimensional instrument.</td>
</tr>
<tr>
<td>Svedlund et al., 1988</td>
<td>GISSS</td>
<td>One-dimensional Disease-specific</td>
<td>Self-administered</td>
<td>Records 15 articles or items occurring over the last month, and combines them into 5 symptom groups that describe reflux, abdominal pain, indigestion, diarrhea, and constipation. Graduated 7-point Likert scale (1 is the absence of bothersome symptoms and 7 is for very bothersome symptoms). To measure a wide range of gastrointestinal symptoms (digestive tract diseases, tumors, peptic ulcer, irritable bowel syndrome, gastric cancer); it does not include colorectal, small bowel, and hepatobiliary tumors. Due to the fact that it evaluates different disorders, not all are done adequately; for example, reflux is examined with only 2 well-directed questions. The interviewer-administered format can be a negative point. It is sometimes confused with a multidimensional instrument.</td>
</tr>
<tr>
<td>Author and year. Name and acronym</td>
<td>Questionnaire characteristics</td>
<td>Questionnaire aims</td>
<td>Comments on the instrument and original validation</td>
<td>Validation into Spanish</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>------------------------------</td>
<td>------------------</td>
<td>-----------------------------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Paré et al., 2003 Gastrointestinal short form questionnaire GSFQ</td>
<td>Multidimensional Disease-specific Self-administered 0-4 ordinal scale (frequency in time: never - always) 6 questions</td>
<td>To evaluate symptoms and quality of life in patients with reflux disease</td>
<td>Original language: English (Canada) Well-designed questionnaire, well-planned validation with internal consistency, reproducibility on days 7 and 28 after the initial evaluation, divergent and convergent validity, and responsiveness The scale has a simple scoring system, but it can be easily manipulated by the patient</td>
<td>Ruiz Díaz et al., 2009 A well done translation, adaptation, and validation</td>
</tr>
<tr>
<td>Shaw et al., 2001 Reflux Disease Questionnaire RDQ</td>
<td>One-dimensional Disease-specific Self-administered Records events occurring over the last 4 weeks The response options were evaluated with a Likert scale of 0-5 points for frequency (from not present to daily) and severity (from not present to severe) 12 articles or items</td>
<td>To evaluate the frequency and intensity of heartburn, regurgitation, and dyspepsia complaints (4 articles or items each) and to facilitate primary attention diagnosis of reflux disease</td>
<td>Original language: English It was developed in 2001 by Shaw et al. as a diagnostic questionnaire and was also validated for dyspeptic symptoms. Since 2008 it has been validated to evaluate symptom frequency and severity (which increased instrument accuracy) A questionnaire developed with one specific aim (diagnosis) was then validated for another (frequency and severity) It is regarded as ideal for evaluating reflux in GERD, but it does not assess dyspepsia, whose subscale was not adequately validated</td>
<td>Nuevo et al., 2009 It evaluated viability, accuracy, validity, and sensitivity to changes in a multicenter study It is reproducible for 16 days in good samples of patients with reflux disease</td>
</tr>
<tr>
<td>Eypach et al., 1995 Gastrointestinal Quality of Life Index GIQLI</td>
<td>Multidimensional Disease-specific Self-administered Records events occurring over 2 weeks 36 articles or items organized into 5 dimensions (symptoms, physical, psychological and social aspects, and disease-specific items 5-point Likert scale (high scores represent better quality of life [QOL])</td>
<td>To evaluate quality of life, specifically for patients with gastrointestinal disease It was developed to distinguish healthy patients from patients with gastrointestinal diseases, but it cannot identify the disorder</td>
<td>Original languages: English and German Tests for reliability, validity, reproducibility, and responsiveness were carried out More than half of the questions are related to symptom frequency as opposed to quality of life How the questionnaire was cross-culturally adapted to the second language was not mentioned</td>
<td>Quintana et al., 2001 Validity, accuracy, and responsiveness of the GIQLI were studied The abstract does not state how the cross-cultural adaptation was carried out</td>
</tr>
</tbody>
</table>

* The translations listed below may not have gone through a complete process of linguistic validation and could require further work to be of adequate use in a study. This translation list is constantly changing, so please confirm the translation status with Mapi Research Trust / the developers / copyright holders / distributors.
Practical aspects for minimizing errors in the cross-cultural adaptation and validation of quality of life questionnaires

References


