

# Literature Review of Methods to Translate Health-Related Quality of Life Questionnaires for Use in Multinational Clinical Trials

Catherine Acquadro, MD,<sup>1</sup> Katrin Conway, MA,<sup>2</sup> Asha Hareendran, PhD,<sup>3</sup> Neil Aaronson, PhD,<sup>4</sup> for the European Regulatory Issues and Quality of Life Assessment (ERIQA) Group

<sup>1</sup>Mapi Research Trust, Lyon, France; <sup>2</sup>Mapi Research Institute, Lyon, France; <sup>3</sup>Pfizer Group of Pharmaceuticals, Sandwich, UK; <sup>4</sup>The Netherlands Cancer Institute, Amsterdam, The Netherlands

## ABSTRACT

**Objectives:** We conducted a literature review to respond to regulatory concerns about the quality of translated patient-reported outcome questionnaires. Our main objective was to answer two questions: What do the methods have in common (and how do they differ)? Is there evidence of the superiority of one method over another?

**Methods:** We identified 891 references by searching MEDLINE, Embase, and the Mapi Research Trust's database with "quality-of-life," "questionnaires," "health status indicators" matched with "translating," "translation issues," "cross-cultural research," and "cross-cultural comparison." Articles were included if they proposed, compared or criticized translation methods.

**Results:** Forty-five articles met our inclusion criteria: 23 representing 17 sets of methods, and 22 reviews. Most articles recommend a multistep approach involving a centralized review process. Nevertheless, each group proposes its own sequence of translation events and weights each step

differently. There is evidence demonstrating that a rigorous and a multistep procedure leads to better translations. Nevertheless, there is no empirical evidence in favor of one specific method.

**Conclusions:** We need more empirical research on translation methodologies. Several points emerge from this review. First, producing high-quality translations is labor-intensive. Second, the availability of standardized guidelines and centralized review procedures improves the efficiency of the production of translations. Although we did not find evidence in favor of one method, we strongly advise researchers to adopt a multistep approach. In line with the recent Food and Drug Administration recommendations, we developed a checklist summarizing the steps used for translations, which can be used to evaluate the rigor of the applied methodologies.

**Keywords:** clinical trials, cross-cultural research, health-related quality of life, patient-reported outcomes, PRO questionnaires, review, translation issues.

## Introduction

The increased prevalence of chronic illnesses [1,2], the call for greater patient empowerment [3], and advances in information technology [4–7] have contributed to a rising interest in patients' views about outcomes of treatment. Changes in a patient's self-assessment of the impact of treatment can only be measured by asking the patient. Patrick [8] points out that some outcomes are known only to the patient, such as the evaluation of how satisfied one is with one's life, and these cannot be verified by other evidence. Other outcomes, such as reports of physical function, can be observed by others.

The extent to which these patient-reported outcomes can be used to evaluate and communicate the effect of new drugs and devices is still a subject of

much debate. Although there remains concern about the value and application of patients' self-assessment in the evaluation of new therapies, regulators are showing interest in these outcomes. Both the European Agency for the Evaluation of Medicinal Products (EMA) [9] and the United States Food and Drug Administration (FDA) [10] have recently issued guidance for the use of patients' self-assessed outcomes in the evaluation of medicinal products.

When pharmaceutical companies evaluate new medicines in clinical trials in different countries, as is the case with regulatory agencies, they need assurance that patients' self-assessment of their condition and the effects of treatment are valid across the different countries, irrespective of the patients' language and cultural background. Many questionnaires, mainly developed in English for Anglo-Saxon cultures, have been translated for use in other countries and/or cultures, and European regulators are rightfully concerned about their validity in measuring the same concepts. The EMA clearly voiced these concerns as one of their

*Address correspondence to:* Katrin Conway, Mapi Research Trust, 27 rue de la Villette, 69003 Lyon, France. E-mail: kconway@mapi.fr  
10.1111/j.1524-4733.2007.00292.x

key issues: “Are HRQL instruments internationally validated?” [11] The FDA shares the same concerns in its recent draft guidance on the use of Patient-Reported Outcome (PRO) measures in the evaluation of medicinal products, section IV.D.5. Changed Culture or Language of Application, and provides some recommendations [10]. The guidance states that: “The FDA recommends that sponsors provide evidence that the methods and results of the translation process were adequate to ensure that the validity of the responses is not affected. [ . . . ] Sponsors should consider whether generally accepted standards for translation and cultural adaptation have been used to support the validity of data from a translated/adapted PRO instrument. As reported by Bullinger and colleagues [12], “Culture and nations differ with regard to a more ethnological or political perspective.” Cultural diversity may exist in one nation (e.g., USA) although other nations may be relatively culturally close because of similar origins (e.g., Scandinavian countries). Acquadro and colleagues [13] also state that communication between cultures depends on equivalence in thoughts and situations and not just equivalence in expressions. Nevertheless, similarity in thoughts and situations between cultures may not exist. Sartorius and Kuyken [14] highlighted this issue when they pointed out differences that exist between cultures in their concepts of health and illness, levels of literacy, reading level, concordance between written and spoken versions of language, taboo subjects, and social desirability effects. Thus, equivalence of international versions of Health-Related Quality of Life (HRQL) instruments is key to their use in cross-cultural research. How this equivalence can be reached is still under discussion, and several taxonomies of equivalence have been developed [14–19].

Stewart and Napoles-Springer [19] provide a taxonomy for assessing equivalence that integrates and modifies the approaches of numerous researchers. They describe six levels of equivalence and the basic methods for addressing each type of equivalence. Conceptual equivalence of constructs and items is achieved when the constructs exist, are relevant and are acceptable in all cultures. The items should also represent the definition of the constructs. Semantic equivalence is obtained when items mean the same thing to people from different groups and in the target and source language. Operational equivalence ensures that standardized methods of survey administration are appropriate for the target culture. Psychometric or measurement equivalence is achieved when comparable psychometric properties are observed in the source and target measures. Item equivalence is observed when 1) items are not differentially more difficult (e.g., biased) in the target culture than in the original; 2) item weights reflect comparative importance of items in all groups; and 3) the meaning of and

distance between response choices is similar across cultures. Finally, criterion equivalence is obtained when the interpretation of scores is the same across groups, and when compared with norms for each group. Translated versions should also demonstrate similar associations with any previously set independent criterion established during the validation of the original version.

## Objectives

In response to regulators’ concerns, Mapi Research Institute and the European Regulatory Issues on Quality of Life Assessment (ERIQA) group investigated current methods and guidelines for translating HRQL questionnaires. A literature review was conducted of methods used to achieve conceptual, semantic and operational equivalence [19] (often referred to as linguistic validation). The review was intended to answer the following questions: 1) what do the various methods and guidelines that are used to translate HRQL questionnaires have in common and how do they differ? and 2) is there empirical evidence pointing to one or more methods as being superior to others?

## Methods

We identified articles published between January 1966 and May 2005 relevant to translating HRQL questionnaires into other languages and adapting them to other cultures. We first performed a search of the MEDLINE and Embase databases with the MeSH terms: “quality of life,” “questionnaires,” and “health status indicators” matched with “translating” and “cross-cultural comparison.”

We considered articles without language restrictions and excluded duplicates, and identified 468 references.

In addition, we performed a search of the database of the Information Resources Center of Mapi Research Trust using the search terms “translation issues,” “cross-cultural comparison,” and “cross-cultural research.” This provided us with another 423 references.

We chose to use this latter source of information, because it was a repository of 13,400 articles, 180 books, periodicals, and unpublished reports on health outcomes—compiled and regularly updated since 1995. The database is well indexed and is constantly updated [20]. Most of the published material (articles and book chapters), and unpublished documents have been directly collected from the developers of PRO questionnaires. The database is accessible at [http://www.mapi-research.fr/t\\_03\\_intr.htm/](http://www.mapi-research.fr/t_03_intr.htm/) Access is free for researchers associated with academic institutions.

Two authors (C.A. and K.C.) reviewed the titles and abstracts of the 891 references for relevance to the study. The authors reviewed the articles together and

compared their choices. All discrepancies were discussed and consensus was reached in all cases.

Articles were included if: 1) they proposed a set of guidelines or recommendations; or 2) they reviewed, compared or criticized methods to adapt HRQL questionnaires from a source culture to a target culture.

## Results

Forty-five articles met our inclusion criteria. Twenty-three articles representing 17 guidelines [14–17,21–39]:

- American Association of Orthopedic Surgeons (AAOS) [21,22];
- The European Organisation for Research and Treatment of Cancer (EORTC) Group [23];
- The European Quality of Life Instrument (EUROQOL) Group [24];
- The European Group for Health Measurement and Quality of Life Assessment (Nottingham Health Profile—NHP) [25,26];
- The Functional Assessment of Cancer Therapy (FACT) Group [27];
- Flaherty et al. [15];
- Herdman et al. [16,17];
- The International Quality of Life Assessment (IQOLA) Group [28,29];
- The Johns Hopkins University (Sickness Impact Profile—SIP) [30];
- Mapi Research Institute [31,32];
- Mathias et al. [33];
- The Medical Outcomes Trust (MOT) [34];
- Rahman et al. [35];
- Sperber [36];
- Spielberger and Sharma [37];
- Swaine-Verdier et al. [38];
- The World Health Organization (WHO) [14,39].

And 22 articles reviewing methods [12,13,18,19,40–57].

Among these 45 articles, six discuss the language issues fundamental to translating HRQL questionnaires [13,14,18,25,26,50], and seven explore the key issue of “equivalence” in the international use of HRQL questionnaires [14–19,43].

It is important to note that we found only three studies that empirically compared different translation methods: Falcao et al. [48], Maneesriwongul and Dixon [53], and Perneger et al. [55].

One article written by the ISPOR Task Force for translation and cultural adaptation proposes good practices for the translation and the cultural adaptation process for PROs [57]. Two other articles comment and discuss the principles proposed by this Task Force [51,54].

### Review of Translation Methods Used or Recommended

Table 1 summarizes translation methods that were either used and/or recommended in the articles

reviewed. It is beyond the scope of this article to describe each of the 17 guidelines in detail. We chose to focus on the following three guidelines because they were not developed for a particular instrument, but a wide range of questionnaires. In addition, they presented a novel feature, which distinguished them from the others:

1. The guidelines proposed by Guillemin and Beaton (i.e., the AAOS guidelines) were chosen because they were the first to propose an extensive review of cross-cultural adaptation in 1993, and followed up their research in 2000;
2. The Mapi Research Institute’s approach was chosen because it was the first to introduce a harmonization step when using multiple language versions of a given instrument in the same study.
3. Swaine and Verdier’s article was chosen because the dual-panel approach they propose represents an interesting contrast to the use of back-translation, a step seen as essential by a majority of groups and individuals working in the field.

It should be noted that the first two methods have greatly benefited from the pioneering work performed by the EORTC Group [23] and the IQOLA Group [28,29].

*AAOS guidelines.* Beaton and colleagues [21] propose that cross-cultural adaptation is required for all self-reported measures. They define cross-cultural adaptation as “a process that looks at both language (translation) and cultural adaptation issues in the process of preparing a questionnaire for use in another setting.” They suggest in another article [22] that this process should be “adapted” according to five different situations, defined by the target population (native, established immigrants, new immigrants), the culture, the language and the country of use. The options are summarized in Table 2.

The guidelines described by Beaton and colleagues are currently used by the AAOS Outcomes Committee. The process for translation involves six stages, which are described below.

Stage I: forward translation. They recommend that at least two translations of the questionnaires be made from the original language (source) to the target language. Bilingual translators whose mother tongue is the target language should produce the two independent translations. Translator 1 (T1) should be informed of the concepts being covered by the questionnaire, and should have a medical or clinical background. The idea is to produce a translation providing equivalence from a measurement perspective. Translator 2 (T2) should be “naive,” less influenced by an academic goal. The idea is to produce a translation that reflects the language used by the layman. Each

**Table 1** Summary of content of Translation Guidelines reviewed

Reference	Recruitment criteria	Forward translation	Synthesis	Back-translation	Review	Pretesting	Specificities
AAOS Guidelines [21,22]	Bilingual translators	Two translators work independently	Yes, done by the translators	Two, work independently, have no knowledge of the underlying concepts	Yes, methodologist, health professionals, language professionals and the translators	Yes, n = 30–40 from the target setting, followed by interviews	The method emphasizes clear justification and documentation of each step
EORTC Quality of Life Study Group [23]	Translators are native speakers, otherwise not specified	Two (minimum) translators work independently	Reviewed by coordinator with the translators; difficulties are resolved with the help of a new independent translator	Two independent translators work independently	Reviewed by coordinator with the translators. Difficulties are resolved with the help of a new independent translator	Yes n = 10–15 Individual interviews	The method emphasizes clear justification and documentation of each step
EuroQol Group [24]	The guidelines specify that qualifications be reported	Two native speakers in source, at least one with health related experience, trained by the Project Manager, they work independently	Yes, two translators and Project Manager	Two back-translators, native in English work independently, one literal and one “polished” translation	Yes	Eight subjects, preferably with low level of education, including healthy individuals as well as patients respond to the PRO and are interviewed	Detailed documentation is required
EuroGroup for Health Measurement, translation of the NHP [25,26]	Panel members and other participants briefly described	Translated by 8–12 from target population, consensus version produced	26 bilinguals respond to both original and consensus target version	Yes, for problematic items only, by teachers of English/target language	Yes	Yes, subjects in the target culture responded to and discussed the questionnaire	
FACT Group [27]	Translators are native speakers of the target language, otherwise not reported	Two translators, one living in USA, one living in target country	Yes, by a third translator	One, by a fourth translator	Yes, reviewed by three to four bilingual health professionals from the target country	Yes, 15–16 patients	Rigorous and time-consuming method; development time of a new version of the SF-36 is several years
IQOLA [28,29]	Not reported in these articles	Minimum of two, translators work independently and also rate difficulty	Yes, first by original translators, then by two other translators who work independently	Yes, two translators	Yes, by US IQOLA team	Yes, n up to 50	
Johns Hopkins, SIP [30]		2 translators whose native tongue is the target language, and who live in the target country	Yes, a focus group or two or more independent translators review each item and synthesize	Yes, by at least one native in the source language	4–5 experts review each item in the pooled forward translation, identify troublesome items, and propose alternatives	Lay panel	The testing procedure is time consuming

Mapi Research Institute [31,32]	Detailed requirements for each member of the team are specified in the article; the entire process takes place in the target culture	Two translators; they reside in the target country, and work independently	Yes, in consultation with the instrument's developer	One translator having no knowledge of the source version	Yes, in consultation with the developer	Two parallel phases: Clinician's Review (users) and Cognitive Debriefing with a sample of 5–10 respondents	The method emphasizes clear justification and documentation of each step; International Harmonization is also done when a questionnaire is linguistically validated in more than one language at the same time
Mathias et al. [33]	Certified translators but not necessarily native speakers, all living in the source country	One by a single translator	No	Yes, by a different translator	Yes, by the instrument developer	Yes, 3–5 bilinguals living in the source country respond to both the original and target language versions. No focus group or debriefing step	This method was used to simultaneously translate a questionnaire into 10 European languages and to modify the original (International Harmonization)
Medical Outcomes Trust (MOT) [34]	Not reported	At least two forward translations	Yes	At least one	Yes—expert panels	Lay panel	Presents only criteria for acceptance and not methods to achieve acceptance
Rahman et al. [35]	Characteristics of the translators described in detail	At least two forward translations	Yes through Key Informant Interviews; key persons across the community are interviewed, and asked to comment on each item	Yes, by a third translator	No	Yes, focus group of 8–10 people representing the study population, paired in 4–5 groups Cognitive debriefing	Use of Key Informant Interviews
Sperber [36]	Not specified in this article	Not specified	Not specified	Yes, comparison of original and back-translation by at least 30 raters fluent in source language; each item in the 2 versions ranked in terms of comparability of language and interpretability	Not specified	Not specified	Propose a new validation method of the translation through comparison and rating of source and back-translation

Table 1 continued

Reference	Recruitment criteria	Forward translation	Synthesis	Back-translation	Review	Pretesting	Specificities
Spielberger and Sharma [37]	Yes. Psychologists, psychiatrists, languages experts. Academic grounding in psychometrics, test theory, extensive experience in the field in which translated scale will be used	Number not specified	Called preliminary translation step	Yes	Expert review	Yes, on bilingual subjects	Review of methods used to translate the STA
Swaine-Verdier et al. [38]	Yes, translators should be as "ordinary" as possible; at least one native speaker of the source language	Number not specified	Yes	Not recommended	Project Coordinator	Lay panel of 5-7 people. Inclusion criteria well described	Back-translation not recommended. Dual translation panels approach. Detailed report essential.
WHO [14,39]	Team members described in detail with motivation for the requirements	Yes, two translators work together	Bilingual panel reviews translation and a monolingual panel "tests" the instrument; bilingual panel then modifies the translation	Yes, one translation	Forward and back-translations are administered to a bilingual group; alternatively, a bilingual panel assesses equivalence		Emic + Etic Approach. The common ground between the two cultures is addressed (Etic) and at another stage, issues specific to the culture are targeted separately (Emic). Creation of international core instrument containing Etics that enable comparison between versions

AAOS, American Association of Orthopedic Surgeons; EORTC, European Organisation for Research and Treatment of Cancer; FACT, Functional Assessment of Cancer Therapy; IQOLA, International Quality of Life Assessment Group; NHP, Nottingham Health Profile; PRO, Patient-Reported Outcome; STA, State-Trait Anxiety Inventory; WHO, World Health Organization.

**Table 2** Cross-cultural adaptation of HRQL questionnaires in five different settings [21]

Target population	Culture	Language	Country	Recommended process
Same population	Same	Same	Same	None
Established immigrants in the same country	Different	Same	Same	Cultural adaptation
Another country, but with the same language (e.g., questionnaire developed in France to be adapted in Canadian French)	Different	Same	Different	Cultural adaptation
New immigrants, not speaking the source language, living in the same source country	Different	Different	Same	Translation and cultural adaptation
Different population	Different	Different	Different	Translation and cultural adaptation

HRQL, Health-Related Quality of Life.

translator should produce a written report summarizing all difficulties encountered, choices made or remaining uncertainties.

Stage II: synthesis of the translations. Both translators, T1 and T2, should work together with a recording observer to produce one common consensus translation called T-12. Again, a written report should document all issues addressed and how they were resolved.

Stage III: back-translation. Two translators, totally blind to the original version, translate the T-12 version back into the original language. The objective is to check that the translated version reflects the same item content as the original. Two back-translations are considered as a minimum, and should be produced by two “naive” individuals whose mother tongue is the source language. Each translator should produce a written report.

Stage IV: expert committee. This committee should be composed of methodologists, health professionals, language professionals, and all the translators involved in the process. The original developers of the questionnaire should be in close contact with this committee. Its role is to develop the prefinal version for field-testing. The committee should review all the translations, and reach consensus on any discrepancy. The goal is to achieve four types of equivalence: semantic (i.e., equivalence in meaning of words), idiomatic (i.e., equivalent expressions have to be found or items have to be substituted), experiential (i.e., the situation evoked or depicted in the source version should fit the target cultural context) and conceptual (i.e., is the concept explored valid in the target culture?). Again all decisions should be documented.

Stage V: test of the prefinal version. Ideally, 30 to 40 persons should be tested. Each subject should complete the questionnaire and be interviewed about the meaning of each item. Distribution of responses should be examined to check the proportion of missing items. This stage provides a rough evaluation of content validity.

Stage VI: submission of documentation to the developers or coordinating committee for appraisal of the adaptation process. This is a process to ensure that all steps have been performed and fully documented.

*Mapi Research Institute’s methodology.* Since 1995, Mapi Research Institute [31,32] has proposed methods similar to those described by Beaton et al. and has labeled the process Linguistic Validation. The Institute has, however, added a step: the International Harmonization. It takes place after the test of the prefinal version, whenever the original questionnaire is translated into several languages simultaneously. The aim is to perform further quality control and to ensure greater comparability between source and target versions. In contrast to the other steps of the linguistic validation, this takes place in one country and in the presence of professional translators representing each target language. The harmonization is achieved at a meeting between translators, the coordinating center and the author. It follows a specific pattern:

1. A reminder of the concepts intended to be represented by each item.
2. The analysis of the translation decisions made in each language group of the same origin (Latin, Scandinavian, Asian, etc.) and a review of certain choices to reconcile them with those of the same group.
3. The submission of the proposed changes to the local team for approval and production of the final target version; and the production of a report explaining the methodology followed and the translation choices made.

Great care is taken at this stage not to smooth over each language’s distinctiveness but to respect their conceptual, semantic, and cultural differences, which lead to the production of appropriate, target versions. In contrast to other methodologies [27,33], the Institute’s approach stresses the importance of performing translation work in the target countries as translations made in the source country have often been inadequate either because of gross errors of syntax or mistakes at the conceptual level.

The Institute specifies that in a classical linguistic validation process no less than 15 people are involved for each language, each having a precise role in the production of an appropriate target version. Organization of these steps by a coordinating center, together with close collaboration with the author of the questionnaire, ensures greater coherence of the final versions. The Institute also emphasizes the importance of documenting each step in detail to facilitate access to translation issues and their solutions at a later stage.

*Dual translation panels (Swaine-Verdier et al.).* Swaine-Verdier and colleagues [38] argue that the forward/backward translation method is controversial, and describe an alternative method involving dual translation panels used in the production of all adaptations of needs-based quality of life (QoL) instruments. The needs-based model of QoL postulates that life gains its quality from the ability of the individual to satisfy his or her needs. QoL is high when these needs are fulfilled and low when few needs are satisfied [58,59].

Swaine-Verdier et al. base their arguments on the assumption that it is better to produce quality in the translation, rather than checking it through back-translation.

They make the following recommendations:

- Recruit translators (five to seven) with varied profiles to work as a team in a group meeting.
- Inform the group of the concepts underlying the questionnaire, its development, design, and content.
- Inform them of the translations requirements (i.e., conceptual equivalence, accessibility and acceptability of wording).
- Have them work under the supervision of an experienced coordinator.
- Have the agreed translation assessed by a lay panel working as a focus group.
- The whole procedure should be reported in detail.

The authors emphasize that translation is only the start of the adaptation process. Further steps include:

1. Pilot testing by means of face-to-face interviews with several (15–20) representatives of the target population, to ensure linguistic, face, and content validity.
2. The evaluation of the psychometric properties of the adapted questionnaire.
3. The assessment of differential item functioning (DIF) between the source and the target version; the absence of DIF indicating measurement equivalence.

The basis of this theory lies in the item response function, the S-shaped trace of the proportion of indi-

viduals at the same ability level who answer a given item correctly. Under the assumption that the ability under consideration is unidimensional and that the item measures the same ability, the trace is unique under the conditions of a particular model. Except for random variations the same curve is found, irrespective of the nature of the group for whom a function is plotted. Items that do not yield the same item response function for two or more groups are violating one of the fundamental assumptions of Item Response Theory, that the item and the test in which the item is contained are measuring the same unidimensional trait [60].

#### *Review of Articles Comparing or Discussing Methods*

Perneger et al. [55] compared the IQOLA group's methodology (see Table 1) and a less rigorous method in which neither a back-translation nor an expert review was done. In Perneger's "Geneva" method, translators who knew nothing about the IQOLA French version of the SF-36 translated the original SF-36 into French.

The Geneva method is briefly described below:

- 1 Recruited three translators from the medical/health field, each having a different specialty.
- 2 Produced three independent forward translations.
- 3 Synthesized the three translations into a single version by a panel of experts from various fields of language and health survey.
- 4 Tested the "final" version for acceptability on two sample groups (n = 15 and 35) belonging to the target population.

The authors then compared the psychometric properties of the "official" IQOLA translation to those of their "Geneva" version administered to the same group of respondents. The investigators concluded that, in this particular case, the more complex IQOLA translation procedure did not pay off in measurably better psychometric performance.

Falcao et al. [48] drew similar conclusions, questioning the complexity of the methodologies proposed in the translation of questionnaires. Two versions of the Health Assessment Questionnaire (HAQ), the SF-36 and the Arthritis Impact Measurement Scale (AIMS-2) were administered to a sample of 50 patients with rheumatoid arthritis. Version 1 was a literal translation of the questionnaire. Version 2 resulted from a process of cultural adaptation following internationally accepted guidelines as recommended by Beaton et al. [21] and Guillemin et al. [22]. Questionnaires were administered before and after a medical consultation. The questionnaire, the order of administration and the version were randomly assigned. Versions 1 and 2 yielded a similar clinically and statistically significant correlation with clinical and laboratory measures in the validation process of the questionnaires. Hence, Falcao et al. proposed the simplification of the AAOS method-



ology by maintaining the basic process (i.e., forward/backward translations, review by expert committee, pretest), but reducing the complexity of each step.

Perneger and Falcao do not question the need for rigorous methods and the importance of a multistep approach in the translations of HRQL measures, but they discuss the complexity of some procedures, arguing that lighter approaches are acceptable.

In another article, Leplège [52] questioned the value of the back-translation method proposed by Brislin [44]. Providing examples that back-translations are at least as misleading as they are informative, he concluded that back-translation is “not the infallible quality control tool it is purported to be.” Therefore, assuming that the priority is to obtain high-quality forward translations, Leplège provides suggestions about the translators’ recruitment criteria and training (e.g., linguistically competent, fully briefed, should have prior experience in the field and should be able to comment on their own translation). Along the same lines, Swaine-Verdier et al. [38] advise against back-translation and proposed the use of dual translation panels. Nevertheless, we could not locate any study that compared these methods (back-translation vs. dual-panel translation). Finally, other authors [11,31], aware of the pitfalls of the back-translation, suggest that the back-translator should produce a version as literal as possible.

In a recent article [53], Maneesriwongul and Dixon reviewed different processes of instrument translation and evaluation of translation adequacy in the published nursing research. They examined 47 studies, which they classified into six categories. For studies with forward translations only, they created two categories to distinguish between those with no test of the translation from those with a test. For studies with forward and back-translation, four categories were created to distinguish between those with no test, monolingual test, bilingual test or both kinds of test. They clearly indicated a hierarchy in quality, with category 1 (forward-only translation) indicating a minimal level of effort and a lower quality of translations, and category 6 (back-translation with monolingual and bilingual tests) demonstrating a substantial effort to ensure the validity of the translation. In the light of their findings, they advised that multiple techniques should be used in all cross-cultural research, and regretted the lack of consensus among researchers on how these techniques should be used or combined. They recommended that minimum standards for applying an instrument developed in another language should include back-translation and testing among target language subjects to allow detection and correction and discrepancies of translation, as well as evaluation of clarity and appropriateness with future subjects. Finally, they called for more detailed information about translation processes in reports and articles.

## Discussion

### *Issues of Definition*

During our literature review we encountered a lack of consensus among authors on critical terms used in the articles. We found, for example, different descriptions of the process of “translating” or “culturally adapting” a HRQL questionnaire.

Although the IQOLA project guidelines [28,29] define “translation methods” to include the production of forward and backward translations, use of difficulty and quality ratings, pilot testing, and cross-cultural comparison of the translation work, the guidelines for the translation of a generic questionnaire (the SIP) [30] have proposed that translating HRQL instruments into other languages implies the translation itself, the evaluation of the psychometric properties and weighing of the translation, and field-testing (responsiveness, norms, etc.).

The difference in meaning between the terms “cultural” and “cross-cultural,” as used in cultural/cross-cultural adaptation was also not clear. According to one author [21], the “cultural adaptation” step follows the translation step and these two steps together form the process of “cross-cultural adaptation.” Other authors seem to use these terms synonymously. In early publications (1993) Guillemin et al. [22] state that “cross-cultural adaptation has two components: the translation of HRQL measure and its adaptation, i.e., a combination of the literal translation of individual words and sentences from one language to another and an adaptation with regards to idiom, and to cultural context and lifestyle.” More recent publications [31,32] state that “cultural adaptation” needs to be performed in two steps: translation (linguistic validation) and evaluation of the psychometric properties (psychometric validation) of the HRQL questionnaire.

The translation of HRQL instruments for use in another culture is still a science under development. Because the terminology of a developing science has not had sufficient time to mature, investigators need to define clearly and precisely the terms they use. Herdman et al. [16] emphasize this need when they point out that there are at least 19 different types of “equivalence” mentioned in the literature and “conceptual equivalence” is used in at least 19 different ways.

We would recommend that the scientific community produce a consensus glossary, which would be made available to nonexperts in the field (e.g., regulators, clinicians, etc.). As part of its objectives, the Cochrane Patient-Reported Outcomes Methods Group is currently undertaking such a task [61].

### *Comparison of Translation Methods*

All the articles discuss the subjectivity of translating HRQL questionnaires and provide substantial information about the requirements, qualifications or char-

acteristics of the people involved in the process. Among the guidelines listed in Table 1 some common features were: a multistep and centralized review process, at least one forward translation and some form of pretesting. Other than that, the methods vary considerably. Some groups recruit professionally educated people in their procedure and others recruit people of the same educational level as that of the target population; some groups have the translators work independently and others have them work as a team; some groups include a back-translation step in their procedure and other groups do not; some groups consolidate several forward translations into one version for back-translation and others back-translate each of the forward translations separately; some consolidate the translations using the same translators and others use independent people to do the consolidation. The pretesting step also varies considerably between groups. Some groups involve patients in a focus group; others use monolingual or bilingual panels and some include a cognitive debriefing step in their method. Finally, some groups recommend the international harmonization step when several translations are performed concomitantly. These differences reflect theoretical differences between groups in definitions of equivalence [14–19], approaches to development [12,40,41,47,56], as well as the trade-offs made (e.g., available resources [33,42,49]).

Beaton [21] and Guillemin [22] highlight the importance of thoroughly documenting each step of the process. Criteria for recruiting translators are provided in detail. There is no indication, however, as to whether the forward translators should be residents of the target country. The categorization of the target population (i.e., native, established or new immigrants) is very sensible. In line with other authors [19], it underscores the need to take the immigration status into consideration.

Although there is some evidence that different methods (i.e., “light/simpler” vs. “heavy/complex”) yield similar results [48,55], this has been tested empirically only on a very limited scale, and needs further research to be widely acceptable. Other findings suggest that a one-step approach (i.e., one forward only or a committee approach only) casts a shadow on the validity of the final outcome [53].

Producing only one forward translation involves a total dependence on the translator’s skill and knowledge, and often results in low validity and reliability. Using the committee approach, serious limitations arise if committee members have common views or if pressure is felt to form a consensus. Finally, we found evidence that most rigorous and centralized procedures provide the best outcomes in terms of equivalence [53].

Our findings are similar to those published by the ISPOR Task Force for translation and cultural adaptation, although the methods under review do not overlap [57]. In addition, we agree with McKenna [54]

and Lenderking [51] when they argue that evidence should be collected before asserting that any specific method represents principles of good practice.

Although, as a result of the lack of scientific evidence in favor of one specific method of translation we could not identify a “best bet” method, we strongly advise researchers to adopt a multistep approach to ensure quality [53]. This advice is in line with the current recommendation of the recent FDA draft guidance, to consider generally accepted methods for translations and cultural adaptations [10]. Nevertheless, we argue that the main subject of controversy with the draft guidance lies in the last sentence of the recommendation: “Sponsors should consider whether generally accepted standards for translation and cultural adaptation have been used to support the validity of data from a translated/adapted PRO instrument, including but not restricted to the following: [. . .] The evidence that measurement properties for translated versions are comparable.”

Although reaching measurement equivalence would be ideal, we would argue that there is a need to adopt a more pragmatic approach in the context of clinical trials that emphasizes the need for conceptual equivalence, and allows evidence of measurement equivalence to emerge over time. Cognitive debriefing in the target countries is one of the means to ensure that conceptual equivalence between the source and the target versions is retained, and represents a way of “bridging” between languages and cultures. Another option would be to add to a well-conducted cultural adaptation process a test of between-country heterogeneity as a prerequisite for pooling trial data. This could form the basis for the proper interpretation of international trial results.

## Conclusions

Several points emerge from this review. First, producing high-quality translation is labor-intensive and time-consuming. Second, the availability of standardized guidelines and centralized review procedures can improve the quality of the translations and the efficiency with which those translations are produced. There is some evidence that a rigorous and a multistep processed method with centralized review procedures leads to better translations. In addition, the people involved in the translation process are critical in determining a questionnaire’s performance in a new country or culture. Consequently, to ensure credibility of their methods, investigators need to specifically describe the process used and justify their recruitment criteria.

Although we did not find evidence in favor of one specific method of translation, we strongly advise researchers to adopt a multistep approach as a guarantee of quality. Toward this end, and in line with the FDA recommendations, we have developed a checklist,

which summarizes the major steps commonly included for the translation of HRQL questionnaires for international use, and can be used to evaluate the rigor of the translation methodology employed.

The major focus of this checklist is on whether the translation process is based on a multistep approach, and whether each step is thoroughly documented. We believe that documentation of each step is crucial in that it allows tracking of all the decisions made during the process.

Item 1 requests information about potential contacts with the developer(s) of the HRQL questionnaire to be translated. Was (were) developer(s) aware of and involved in the translation process?

Items 2 to 4 deal with the forward-translation process. The checklist explores the availability of detailed information about the translators involved in the process, the number of forward translations produced and the steps taken to produce a reconciled forward version.

In the absence of a back-translation step (i.e., use of the dual-panel method), Item 5 deals with the approval of the reconciled forward translation as the Target Version for pilot testing.

Items 6 to 8 deal with the back-translation process. The checklist explores the availability of detailed information about the translators involved in the process, the number of back-translations produced, the analysis of the back-translation(s), and how the reconciled forward translation was revised according to the review of the back-translation(s).

In the absence of a back-translation step, Item 9 deals with the approval of the reconciled forward translation as the Target Version for pilot testing.

Item 10 deals with the description of the pilot testing of the target version to verify that respondents from the target population clearly understand, accept and can easily respond to the target language version.

Item 11 is concerned with the use of International Harmonization.

To provide insight into the types of equivalence covered by this checklist, we have linked each step to its corresponding level of equivalence as described by Stewart and Napoles [19]. Based on the definitions of equivalence and the methods to achieve them provided by the authors, Items 1 and 11 of the checklist cover conceptual equivalence, Items 2 to 12 apply to semantic equivalence, and Item 11 is also relevant to operational equivalence. This checklist only explores three types of equivalence, and does not include the other three types of equivalence described by Stewart and Napoles [19] (i.e., psychometric, item, and criterion) where more empirical research is needed.

Source of financial support: the European Regulatory Issues on Quality of Life Assessment (ERIQA) Group (<http://www.eriqa-project.com>).

Supplementary materials for this article can be found at: <http://www.ispor.org/publications/value/ViHsupplementary.asp>

## References

- 1 CDC. Prevalence of disability and associated health conditions—United States, 1991–1992. *MMWR* 1994;43:730–7.
- 2 CDC Chronic Disease Prevention. The Burden of Chronic Diseases and Their Risk Factor—National and State Perspectives. Atlanta, GA: Department of Health and Human Services, 2002.
- 3 Segal L. The importance of patient empowerment in health system reform. *Health Policy* 1998;44:31–44.
- 4 Eysenbach G, Diepgen TL. The role of e-health and consumer health informatics for evidence-based patient choice in the 21st century. *Clin Dermatol* 2001;19:11–17.
- 5 Ball MJ, Lillis J. E-health: transforming the physician/patient relationship. *Int J Med Inf* 2001;61:1–10.
- 6 Coile RC Jr. The digital transformation of health care. *Physician Exec* 2000;26:8–15.
- 7 Coile RC Jr. E-health: reinventing healthcare in the information age. *J Healthc Manag* 2000;45:206–10.
- 8 Patrick D. Concept of health-related quality of life and of patient-reported outcomes. In: Chassany O, Caulin C, eds. *Health-Related Quality of Life and Patient-Reported Outcomes: Scientific and Useful Criteria*. Paris: Springer, 2002.
- 9 EMEA/CHMP/EWP/139391/2004. Reflection Paper on the Regulatory Guidance for the Use of Health-Related Quality of Life (HRQL) Measures in the Evaluation of Medicinal Products. London: EMEA, 2006.
- 10 Draft Guidance for Industry. Patient-Reported Outcomes Measures: Use in Medical Product Development to Support Labeling Claims. Docket 2006D-0044, Rockville, MD: FDA Federal Register, 2006.
- 11 Chassany O, Sagnier P, Marquis P, et al., for the European Regulatory Issues on Quality of Life Assessment (ERIQA) Group. Patient-reported outcomes: the example of Health-Related Quality of Life – a European guidance document for the improved integration of Health-Related Quality of Life assessment in the drug approval process. *DIA J* 2001;36:209–38.
- 12 Bullinger M, Anderson R, Cella D, Aaronson N. Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Qual Life Res* 1993;2:451–9.
- 13 Acquadro C, Jambon B, Ellis D, Marquis P. Language and translation issues. In: Spilker B, ed. *Quality of Life and Pharmacoeconomics in Clinical Trials* (2nd ed.). Philadelphia, PA: Lippincott-Raven Publishers, 1996.
- 14 Sartorius N, Kuyken W. Translation of health status instruments. In: Orley J, Kuyken W, eds. *Quality of Life Assessment: International Perspectives*. Berlin Heidelberg: Springer Verlag, 1994.
- 15 Flaherty JA, Gaviria FM, Pathak D, et al. Developing instruments for cross-cultural psychiatric research. *J Nerv Ment Dis* 1988;176:257–63.

- 16 Herdman M, Fox-Rushby J, Badia X. Equivalence and the translation and adaptation of health-related quality of life questionnaires. *Qual Life Res* 1997;6:237-47.
- 17 Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQL instruments: the universalist approach. *Qual Life Res* 1998;7:323-35.
- 18 Hui CH, Triandis HC. Measurement in cross-cultural psychology. A review and comparison of strategies. *J Cross Cult Psychol* 1985;16:131-52.
- 19 Stewart AL, Napoles-Springer A. Health-Related Quality of Life assessments in diverse population groups in the United States. *Med Care* 2000;38:102-24.
- 20 Emery MP. The Information Resources Centre of Mapi Research Institute. *Quality of Life Newsletter* 1999;23:9.
- 21 Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186-91.
- 22 Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of Health-Related Quality of Life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993;46:1417-32.
- 23 Cull A, Sprangers M, Bjordal K, on behalf of the EORTC Quality of Life Group. EORTC Quality of Life Group translation procedure. Brussels: EORTC Publications, 2002.
- 24 The EuroQol Group. Draft guidelines for cultural adaptations of EQ-5D. Rotterdam: EuroQol Group, 2000.
- 25 European Group for Health Measurement and Quality of Life Assessment. Cross-cultural adaptation of health measures. *Health Policy* 1991;19:33-44.
- 26 Hunt SM. Cross-cultural issues in the use of socio-medical indicators. *Health Policy* 1986;6:149-58.
- 27 Bonomi AE, Cella DF, Hahn EA, et al. Multilingual translations of the Functional Assessment of Cancer Therapy (FACT) quality of life measurement system. *Qual Life Res* 1996;5:309-20.
- 28 Bullinger M, Alonso A, Apolone G, et al. Translating health status questionnaires and evaluating their quality: the IQOLA approach. *J Clin Epidemiol* 1998;51:913-23.
- 29 Ware JE, Keller SD, Gandek B, et al. Evaluating Translations of Health Status Questionnaires. Methods from the IQOLA Project. *Int J Technol Assess Health Care* 1995;11:525-51.
- 30 The Johns Hopkins University. Guidelines for Translation of the Sickness Impact Profile. Baltimore, MD: The Johns Hopkins University, January 1996.
- 31 Acquadro C, Conway K, Giroudet C, Mear I. Linguistic Validation Manual for Patient-Reported Outcomes (PRO) Instruments. Lyon: Mapi Research Institute, 2004.
- 32 Mear I. Difficulties of international clinical trials: cultural adaptation of quality of life questionnaires. In: Chassany O, Caubin C, eds. *Health-Related Quality of Life and Patient-Reported Outcomes: Scientific and Useful Criteria*. Paris: Springer, 2002.
- 33 Mathias SD, Fifer S, Patrick DL. Rapid translation of quality of life measures for international clinical trials: avoiding errors in the minimalist approach. *Qual Life Res* 1994;3:403-12.
- 34 Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality of life instruments: attributes and review. *Qual Life Res* 2002;11:193-205.
- 35 Rahman A, Iqbal Z, Waheed W, Hussain N. Translation and cultural adaptation of health questionnaires. *J Pak Med Assoc* 2003;53:142-7.
- 36 Sperber AD. Translation and validation of study instruments for cross-cultural research. *Gastroenterology* 2004;126(Suppl. 1):S124-8.
- 37 Spielberger CD, Sharma S. Cross-cultural measurement of anxiety. In: Spielberger CD, Diaz-Guerrero R, eds. *Cross-Cultural Anxiety*. Washington: Hemisphere Pub, 1976.
- 38 Swaine-Verdier A, Doward LC, Hagell P, et al. Adapting quality of life instruments. *Value Health* 2004;7(Suppl. 1):S27-30.
- 39 Skevington SM. Advancing cross-cultural research on quality of life: observations drawn from the WHOQOL development. *Qual Life Res* 2002; 11:135-44.
- 40 Anderson RT, Aaronson NK, Wilkin D. Critical review of the international assessments of health-related quality of life. *Qual Life Res* 1993;2:369-95.
- 41 Anderson RT, Aaronson NK, Bullinger M, McBee WL. A review of the progress towards developing health related quality of life instruments for international clinical studies and outcomes research. *Pharmacoeconomics* 1996;10:336-55.
- 42 Berzon R, Shumaker S. Evaluating health-related quality of life measures for cross-national research. *DIA J* 1994;28:63-7.
- 43 Bowden A, Fox-Rushby JA. A systematic and critical review of the process of translation and adaptation of generic health-related quality of life measures in Africa, Asia, Eastern Europe, the Middle East, South America. *Soc Sci Med* 2003;57:1289-306.
- 44 Brislin RW. Backtranslation for cross-cultural research. *J Cross Cult psychol* 1970;1:185-216.
- 45 Brislin RW. Questionnaire wording and translation. In: Lonner WJ, Thorndike RM, eds. *Cultural Research Methods*. New York: John Wiley, 1973.
- 46 Brislin RW. The wording and translation of research instruments. In: Lonner WJ, Berry JW, eds. *Field Methods in Cross-Cultural Methods*. Beverly Hills, CA: Sage, 1986.
- 47 Bullinger M, Power MJ, Aaronson NK, et al. Creating and evaluating cross-cultural instruments. In: Spilker B, ed. *Quality of Life and Pharmacoeconomics in Clinical Trials* (2nd ed.). Philadelphia, PA: Lippincott-Raven Publishers, 1996.
- 48 Falcao Dda M, Ciconelli RM, Ferraz MB. Translation and cultural adaptation of quality of life questionnaires: an evaluation of methodology. *J Rheumatol* 2003;30:379-85.
- 49 Guyatt GH. The philosophy of health-related quality of life translation. *Qual Life Res* 1993;2:461-5.

- 50 Hunt SM, Bhopal R. Self report in clinical and epidemiological studies with non-English speakers: the challenge of language and culture. *J Epidemiol Community Health* 2004;58:618–22.
- 51 Lenderking WR. Comments on the ISPOR Task Force Report on Translation and Adaptation of Outcomes Measures: guidelines and the need for more research. *Value Health* 2005;8:92–3.
- 52 Leplège A, Verdier A. The adaptation of health status measures: methodological aspects of the translation procedure. In: Shumaker SA, Berzon R, eds. *The International Assessment of Health-Related Quality of Life: Theory, Translation, Measurement & Analysis*. New York: Rapid Communications Oxford, 1995.
- 53 Maneesriwongul W, Dixon JK. Instrument translation process: a methods review. *J Adv Nurs* 2004;48:175–86.
- 54 McKenna SP, Doward LC. The translation and cultural adaptation of patient-reported outcome measures. *Value Health* 2005;8:89–91.
- 55 Perneger TV, Leplège A, Etter JF. Cross-cultural adaptation of a psychometric instrument: two methods compared. *J Clin Epidemiol* 1999;5:1037–46.
- 56 Schmidt S, Bullinger M. Current issues in cross-cultural quality of life instrument development. *Arch Phys Med Rehabil* 2003;84(Suppl. 2):S29–34.
- 57 Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. *Value Health* 2005;8:94–104.
- 58 McKenna SP. A new theoretical approach to the measurement of quality of life. *Drug Inf J* 1994;28:13–18.
- 59 McKenna SP, Doward LC. The needs-based approach to quality of life assessment. *Value Health* 2004;7(Suppl. 1):S1–3.
- 60 Angoff WH. Perspectives on differential item functioning methodology. In: Holland PW, Wainer H, eds. *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum, 1993.
- 61 Acquadro C, Joyce D, Lapalus L, Patrick DL. Cochrane Quality of Life Methods Group. *Cochrane Methods Group Newsletter* 2002;6:24–5.