J Clin Epidemiol Vol. 46, No. 12, pp. 1417–1432, 1993
Printed in Great Britain.

0895-4356/93 $6.00 + 0.00

CROSS-CULTURAL ADAPTATION OF HEALTH-RELATED QUALITY OF LIFE MEASURES: LITERATURE REVIEW AND PROPOSED GUIDELINES

FRANCIS GUILLEMIN,[1]* CLAIRE BOMBARDIER[2,3] and DORCAS BEATON[3]
[1]School of Public Health, University of Nancy, France, [2]Clinical Epidemiology Division, Wellesley Hospital, University of Toronto and [3]Ontario Worker Compensation Institute, Toronto, Canada

(Received in revised form 11 June 1993)

Abstract—Clinicians and researchers without a suitable health-related quality of life (HRQOL) measure in their own language have two choices: (1) to develop a new measure, or (2) to modify a measure previously validated in another language, known as a cross-cultural adaptation process. We propose a set of standardized guidelines for this process based on previous research in psychology and sociology and on published methodological frameworks. These guidelines include recommendations for obtaining semantic, idiomatic, experiential and conceptual equivalence in translation by using back-translation techniques and committee review, pre-testing techniques and re-examining the weights of scores. We applied these guidelines to 17 cross-cultural adaptation of HRQOL measures identified through a comprehensive literature review. The reporting standards varied across studies but agreement between raters in their ratings of the studies was substantial to almost perfect (weighted $\kappa = 0.66$–0.93) suggesting that the guidelines are easy to apply. Further research is necessary in order to delineate essential versus optional steps in the adaptation process.

Quality of life Health status index Cross-cultural comparison Culture Validity Guidelines

RATIONALE

A large body of research has recently been devoted to the development of health-related quality of life (HRQOL) measures. In 1991 alone, over 160 different measures were used in the published literature [1]. Such techniques are increasingly used in clinical trials [2, 3] to determine the impact of medical intervention on quality of life (QOL), and by public health researchers [4] to assess the outcome of health care services. With a few exceptions [5, 6] all the measures so far developed are in the English language and are intended for use in English-speaking countries. There is nonetheless a need for measures specifically designed to be used in non English-speaking countries and also among immigrant populations, since cultural groups vary in disease expression and in their use of various health care systems. This need has become more acute with the growing number of large multicentre multicountry trials.

In order to meet that need, two options are available: (1) to develop a new measure, and (2) to use a measure previously developed in another language. The first option, the generation of a new HRQOL measure is a time-consuming process in which the bulk of the effort is devoted to the conceptualization of the measure and the selection and reduction of its items. In the second option, if the transposition of a measure from its original cultural context

*All correspondence should be addressed to: Dr Francis Guillemin, Ecole de Santé Publique, Faculté de Médecine BP 184, 54 505 Vandoeuvre-les-Nancy Cedex, France.

is done by simple translation it is unlikely to be successful because of language and cultural differences [7]. Furthermore, the perception of QOL and the ways in which health problems are expressed vary from culture to culture [8]. To be successful this option requires a systematic approach to the translation and to the cross-cultural adaptation process of HRQOL measures. A recent effort was made by Hunt and the European Group for Health Measurement and Quality of Life Assessment through a cross-cultural adaptation of the Nottingham Health Profile (NHP) to several European countries using a systematic method [9]. This effort of standardization needs to be expanded.

In this paper, we propose a set of standardized guidelines for the cross-cultural adaptation of HRQOL measures based on previous research in psychology and sociology [10–18] and on published methodological frameworks for HRQOL validity [19, 20]. We review the published literature on the cross-cultural adaptation of HRQOL instruments and evaluate the practicality of our proposed guidelines to this literature.

METHODS

Literature search strategy

Relevant papers reflecting the methods used for cross-cultural adaptation were identified from three databases: **Medline** (1966–1992), **Health Planning and Administration** (1975–1992) and **Embase** (Excerpta Medica) (1990–1992). The search strategy used was to identify articles with "quality of life", "health status", "health status indicator", "functional status", "questionnaires" and "interviews" as main subject headings (exploded) or text words in titles and abstracts. This was matched with "cross-cultural", "cross-cultural comparison", "translation" and "languages" as main subject headings (exploded) or text words. Papers published between January 1966 and October 1992 were considered, without language restriction. All references obtained were entered into Reference Manager computer software [21] to check for duplication.

Development of the guidelines

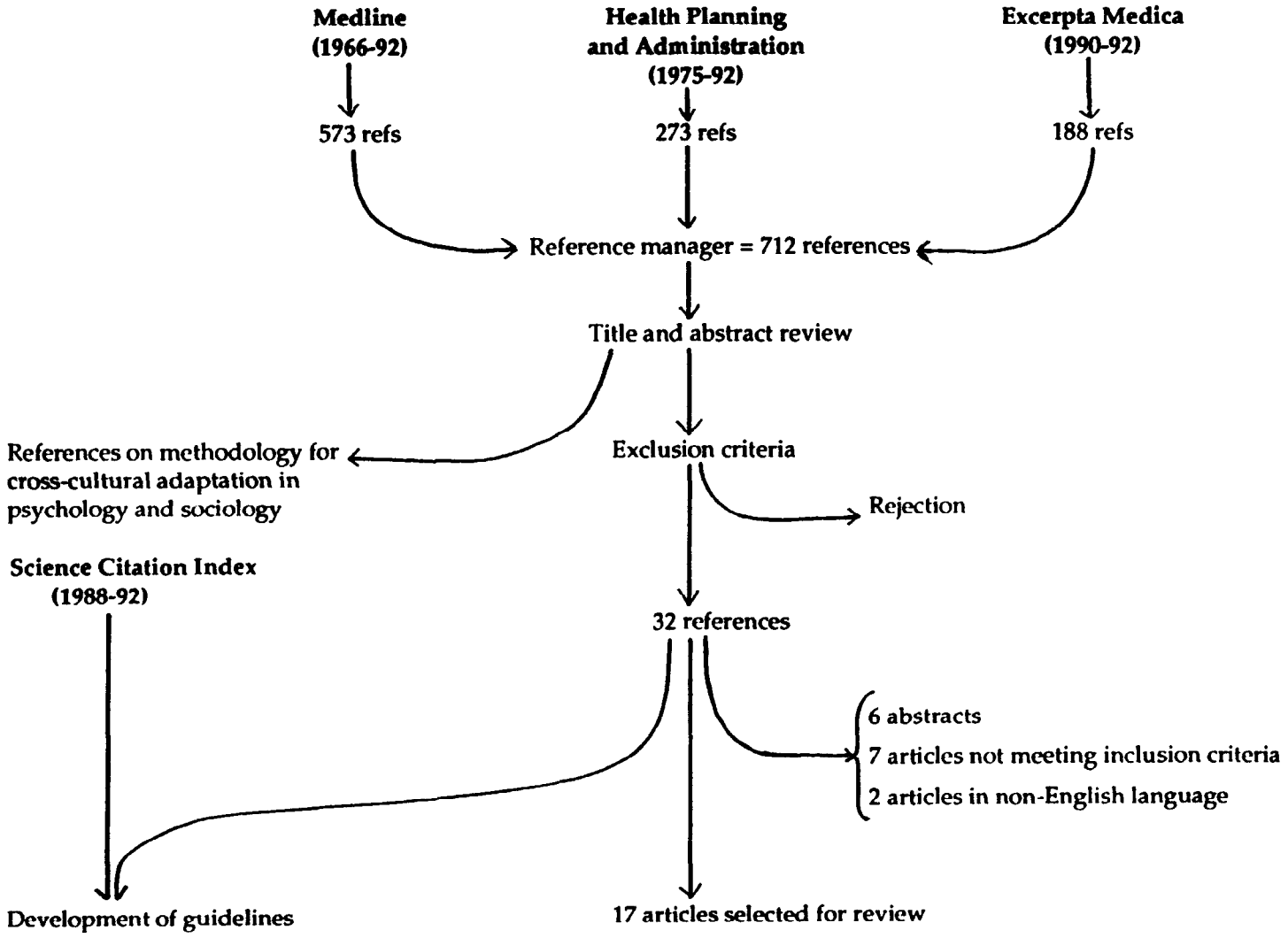
The literature review identified several publications in the field of psychology and sociology addressing the methodology of cross-cultural adaptation. For example, the General Health Questionnaire has been translated into at least 36 languages [10] and the State Trait Anxiety Inventory into at least 21 languages [11]. Although these papers were excluded from our formal evaluation because they addressed mainly mental health, their work suggests methodological approaches, developed to overcome the inadequacy of simple translation, which may be useful in the cross-cultural adaptation of HRQOL measures. We have developed guidelines and a scoring method which can be applied in a standardized manner to evaluate the quality of cross-cultural adaptations of HRQOL measures. This system was based on both empirical and theoretical findings extracted from the literature. The empirical basis was derived from a systematic review of the published work on cross-cultural adaptation. Theoretical foundations were gained from guidelines on the methodology of assessing the validity of HRQOL measures [19, 20].

Selection of articles for review

Our literature search identified 712 references (Table 1). Their titles and abstracts were reviewed by one author (FG) for relevance to the study. Papers were included if they contained a description of the methodological process used to adapt a HRQOL measure from a source to a target culture. Papers were excluded if they presented only results of cross-cultural comparisons, or simply mentioned the use of HRQOL measures in different countries or in international trials without describing the translation and adaptation process. Papers concerning instruments to measure only pain, symptoms or mental status/disorders as well as utilities were also excluded.

During the selection process, any reference to methodology quoted in these papers were used to conduct a supplementary search in the **Science Citation Index** (1980–1992) for additional material on cross-cultural adaptation of HRQOL instruments. From 1966 to 1992, only 32 papers met the inclusion and exclusion criteria for review. Six of these publications were in abstracts form [22–27] and were not included in the present review since not enough information was available to assess the quality of the adaptation process. Seven other papers [28–34] were rejected during the review process as they were judged by subsequent raters (CB, DB) not to have met the original selection criteria, i.e. either not dealing with a quality of life measure or not with the cross-cultural adaptation of such

Table 1. Selection of articles in the databases



measure or not containing a description of the method used for cross-cultural adaptation. Of the remaining 19 articles, only 2 were excluded [35, 36] by restricting our review to English language papers. Thus, the review of 17 studies was completed [37–53].

Application of the guidelines to assess quality of studies

The proposed guidelines include 5 sections: (1) translations and (2) back-translations by qualified people, (3) committee review of those translations and back-translations, (4) pretesting for equivalence using adequate techniques (with bilingual or monolingual individuals), and (5) reexamination of the weighting of scores, if relevant. If an instrument was adapted from one culture to another using a similar language (from American English, U.S.A., to British English, U.K.), the steps of translation and back-translation obviously were not required and therefore were not assessed.

The quality of each study was assessed by two of the investigators (DB, CB) blinded to authors' names, journal titles, and city or area of the study. Each investigator was provided with an operational definition of the evaluation criteria for each section and was asked to rate each section as "good", "moderate", "poor" quality or "not done" using a standard data extraction form (Appendix). Agreement between the two judges in these ratings was assessed using the weighted kappa statistic for categorical judgement [54]. For each section, a mean score across studies was calculated as the mean of quality ratings assigned for that section with the following values: good = 3; moderate = 2; poor = 1; if the section was rated as "not done" the study was not included in this calculation.

An overall score was also calculated for each study by adding the scores across all sections for that study and dividing by the number of sections. For this calculation a value of 0 was assigned to sections rated as "not done". If translation and rescoring were not considered relevant for a particular study, these sections were omitted from the calculation of the overall score for that study. Agreement between judges on these overall scores was assessed using the intra-class correlation coefficient for continuous data [55]. Recognizing that the concepts and techniques involved in adapting measures for people using similar language in another culture, using another language in another country and using another language in the same country (immigrants) differ, agreement between the judges was also considered separately for each subgroup of studies.

RESULTS

Settings for Cross-cultural Adaptation

Our literature review identified several settings for the cross-cultural adaptation process. A range of situations may be encountered depending on similarities and differences between the cultures and languages of the populations concerned. An instrument originally developed in the English language in the U.S.A. can readily be used by a majority of the American population (Table 2: example 1).

Immigrants using the same language may encounter particular problems in expressing themselves with regard to health and HRQOL. Therefore, particular attention should be paid to the adaptation of cross-cultural HRQOL measures to such populations. Immigrants to the U.S.A., for instance Hispanics, will judge their health and related QOL according to their cultural origin and their degree of assimilation into the host culture. They may have been settled for long enough to have mastered the English language sufficiently well to answer the original instrument, and still refer to their Spanish culture in assessing their situation (Table 2: example 2).

An instrument used in a country other than that in which it was developed may require adaptation if the populations concerned have another culture with similar language. For instance British English should be used in Great Britain rather than American English. There are sufficiently meaningful differences between the British and American cultures to necessitate modification of some items and validation of the measure in its new setting [37–39] (Table 2: example 3).

Recently settled immigrants with a low degree of acculturation may require an instrument that is cross-culturally adapted to their Spanish (native) language and culture, but appropriate to the American situation (Table 2: example 4).

Under most circumstances, instruments require adaptation for use in a different country with both a different culture and a different language. For instance, the American measure would need to be modified for use in the French language in France or in Canada (Table 2:

Table 2. Settings (examples 2 to 5) for cross-cultural adaptation of a HRQOL measure originally developed in English in the U.S.A. (example 1)

Examples	Culture of the target population		Language of the measure		Country of utilization	Translation required	Adaptation required
1	Same culture	American	Similar language	English	Same country	U.S.A.	—	—
2	Other culture	Established immigrants (Hispanics)	Similar language	English	Same country	U.S.A.	—	✓
3	Other culture	British	Similar language	British English	Other country	U.K.	—	✓
4	Other culture	Recent immigrants (Hispanics)	Other language	Spanish	Same country	U.S.A.	✓	✓
5	Other culture	French	Other language	French	Other country	France	✓	✓

example 5). The degree of adaptation required depends on similarities in language structure (there are fewer differences between most of the European languages than there are between European and Arabic or Asian languages) and in culture [12].

Cross-cultural adaptation has two components: the translation of the HRQOL 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 regard to idiom, and to cultural context and lifestyle. Translation and adaptation are required for examples 4 and 5 while only cross-cultural adaptation is necessary in examples 2 and 3. The quality of the adapted measure is then assessed with regard to its sensibility. The elements of sensibility, as defined by Feinstein, which need to be considered include the purpose of the measure, its comprehensibility, its content and face validity, its replicability and the suitability of the scales [19].

Guidelines for Cross-cultural Adaptation

The following guidelines concentrate on the points that must be addressed in order to preserve the sensibility of the tool in the target culture. Table 3 summarizes the steps that are essential in order to ensure the quality of the procedure.

1. Translation

Produce several translations. Translations are of higher quality when undertaken by at least two independent translators. This allows for the detection of errors and divergent interpretations of ambiguous items in the original. The quality will be even higher if each translation is undertaken by teams rather than single individuals, who are more likely to introduce personal idiosyncrasies.

Use qualified translators. The qualifications and characteristics of the translators are also important. Highly educated individuals may not be culturally representative of the target population [13]. Translators should preferably translate into their mother tongue [40]. Some of them should be aware of the objectives underlying the material to be translated and the concepts involved so as to offer a more reliable restitution of the intended measurement [56, 41]. Other translators who are unaware of these objectives and concepts may usefully elicit unexpected meanings from the original tool.

Table 3. Guidelines to preserve equivalence in cross-cultural adaptation of HRQOL measures.*

1. *Translation*
   Produce several translations
   Use qualified translators

2. *Back-translation*
   Produce as many back-translations as translations
   Use appropriate back-translators

3. *Committee review*
   Constitute a committee to compare source and final versions
   Membership of the committee should be multidisciplinary
   Use structured techniques to resolve discrepancies
   Modify instructions or format, modify/reject inappropriate items, generate new items
   Ensure that the translation is fully comprehensible
   Verify cross-cultural equivalence of source and final versions

4. *Pre-testing*
   Check for equivalence in source and final versions using a pre-test technique
   Either use a probe technique
   Or submit the source and final versions to bilingual lay people
   Immigrants: Choose the language of administration or use a dual-format measure

5. *Weighting of scores*
   Consider adapting the weights of scores to the cultural context

*It is not always possible to follow all the steps described here due to the design of the measure (e.g. no weighting score to be examined).

2. *Back-translation*

Produce as many back-translations as translations. Back-translation, translating back from the final language into the source language, has been shown to help improve the quality of the final version [7, 14]. Each first translations should be back-translated independently from each other. Misunderstandings in the first translation may be amplified in the back-translation, and thereby revealed. Failure to adapt to the cultural target context and ambiguity in the source version can also be uncovered.

Use appropriate back-translators. Back-translation is of better quality if those who do it are fluent in the idioms and colloquial forms of the source language, i.e. the result of their back-translation. Thus, they should also translate into their mother tongue. Unlike some of the first translators, back-translators should preferably not be aware of the intent and concepts underlying the material. Back-translators without *a priori* knowledge of the intent of the original instrument are free of biases and expectations and their back-translation may reveal unexpected meanings or interpretations in the final version.

3. *Committee review*

Constitute a committee to compare source and final versions. A committee should be constituted in order to produce a final version of the modified measure based on the various translations and back-translations obtained as described above. Part of that committee's role should also review the introduction and instruction to the questionnaire as well as review the scaling of responses to each question (i.e. the translation should maintain equivalence of steps in Likert-type scales).

Membership in the committee should be multidisciplinary. To use the analogy of the development of a new health status measure, the committee should consist of individuals expert in the disease(s) explored, and in the intent of the measure and the concepts to be explored. Bilingual members are of particular value for such committee [15].

In case of a cross-cultural adaptation for an immigrant population, individuals representative of the target group are likely to be available. Their input is likely to result in a measure better adapted in terms of idioms and colloquialisms than that which would be produced by highly educated people [9, 42]. A scale referring to the ability to speak, write, read and understand both languages has been developed [16] and can be useful in selecting these bilingual committee members.

Use structured techniques to resolve discrepancies. The committee may resolve problems by considering the material it has now collected. It may further decide to repeat the translation–back-translation process. A decentring technique [17] has also been proposed as a way

of improving cross-cultural adaptation. This technique considers the source and final versions equally important. Both are open to modification during the translation procedure. In other words, the measure is not considered to centre on one of the languages. Decentring is best conducted in close collaboration with the authors. If problem items are found, the authors may provide a working version of the instrument or items, maintaining the concept of the questions, but avoiding colloquialisms. Searching for a common way to express a concept in both languages is the best way to ensure that the final version maintains content validity. It is unusual for authors to be available, and this process may need to be conducted by committee members.

Modify instructions or format, modify or reject inappropriate items, generate new items. The committee must ensure that the introduction to the research tool and the instructions for filling in the questionnaire are carefully translated in order to preserve the replicability of the measure [19]. The redundancy principle, i.e. repeating the same instruction in a different manner, may help to reduce comprehension errors [12].

The review committee is also likely to modify or eliminate irrelevant, inadequate and ambiguous items and may generate substitutes better fitting the cultural target situation while maintaining the general concept of the deleted items.

Ensure that the translation is fully comprehensible. Guidelines about how to produce translations comprehensible to a majority of people have suggested using language which can be understood by 10 to 12-year old children [12]. Recommendations include: short sentences with key words in each item as simple as possible [11]; the active rather than the passive voice; repeated nouns instead of pronouns; and specific rather than general terms. Authors should avoid using: metaphors and colloquialisms; the subjective mode; adverbs and prepositions telling "where" and "when", possessive forms; words indicating vagueness; and sentences containing two different verbs that suggest different actions.

Verify cross-cultural equivalence of source and final versions. Several taxonomies of cross-cultural equivalence have been proposed in the psychiatry literature [12, 13, 60, 61]. The ultimate parity is equivalence of HRQOL concepts within the cultures concerned. Translators aiming for conceptual equivalence should consider the following:

- *semantic equivalence* is equivalence in the meaning of words, and achieving it may present problems with vocabulary and grammar. For example, vocabulary problems may be encountered in the question "are you able to bend ?" which can refer to several parts of the body, such as the arm, back or knees, and might have been intended—and translated—only to explore the ability to flex (arm), bend over (back) or squat (knees). Furthermore, some words, such as "happy", have several subtly different meanings depending on the context.

  Grammatical alterations are sometimes necessary in the construction of sentences. For example, languages without the gerund form may be more difficult to adapt [13]: activities couched in terms such as dancing, singing or eating (gerund form of to dance, sing and eat) may not be readily translatable.
- *idiomatic equivalence.* Since idioms and colloquialisms are rarely translatable, equivalent expressions have to be found or items have to be substituted. This is more likely to be necessary in the emotional and social dimensions. For example, "Do you feel downhearted and blue?" or "Do you feel at home?" are untranslatable idioms for which equivalents must be found. The item "I am feeling on edge" in the NHP was translated into "I have my nerves outside my skin" in Italian [9], "I feel nervous, tense" in French [43] and "I am afraid" in Arabic [60].
- *experiential equivalence.* The situations evoked or depicted in the source version should fit the target cultural context. This may result in the modification of an item. For example, in the Brazilian version of the HAQ, "using public transportations" was substituted for "using a private car", since most of the people in Brazil have no car [44]. "I have forgotten what it is like to enjoy myself" [60] and "How many hours a week do you have leisure activities?" do not refer to usual experiences in a number of cultures and equivalent feelings (enjoy) or activities (leisure) must be found or the items discarded.
- *conceptual equivalence* refers to the validity of the concept explored and the events experienced by people in the target culture, since items might be equivalent in semantic meaning but not conceptually equivalent.

  For example, "cousin" and "brother" may mean more than simply second or first-degree relative of the same generation. In many cul-

tures in developing countries they have a wider meaning within the social network. "I have pain in my head" may translate perfectly into another language semantically, but have a totally different conceptual meaning for the target culture [60].

4. Pre-testing

Check for equivalence in source and final versions using a pre-test technique. In pre-testing, a sample population replies to the questionnaire in order to check for errors and deviations in the translation. Two techniques are available: a probe technique and appraisal by bilingual individuals. Both allow for the checking of face validity, i.e. the confirmation that questions are acceptable without arousing reluctance or hesitation. If the final version does not achieve a satisfactory level of equivalence, further revision can be performed by the committee.

Either use a probe technique. The answer to an item might appear adequate, yet be consistently misunderstood. In order to determine whether a questionnaire is being understood correctly, it can be administered to a group of patients as follows. After each answer (or a random sample of answers), the patient is asked the probe question: "What do you mean ?" and is encouraged to elucidate his or her understanding of the item in an open-ended manner [18]. This ensures that the final item is understood as having a meaning equivalent to that of the source item.

Or submit the source and final versions to bilingual lay people. The source and the final versions of a measure can be administered to a group of bilingual individuals in order to detect possible discrepancies. This method can also help pinpoint any inadequacy of the final version with the cultural context. They are asked to rate the equivalence of each item between the source and final versions. Those items with low level of equivalence or rated discrepantly by different people can still be revised at this stage [11, 12]. Administration of a questionnaire to bilingual lay people is not practical in every setting but may be possible with immigrants.

The case of adaptation for immigrants requires two additional considerations:

Choose the language of measurement administration for immigrants. During questionnaire administration, some immigrant respondents express a preference for their native language. Several methods have been described for the choice of language of administration; it can be the decision of the respondent her/himself, or of the interviewer or research assistant, or it can be based on a measure of acculturation. Several such measures have been proposed [57–59] and include language proficiency and preference, country of birth and origin, location of education, ethnic identification, contact with homeland, and ethnicity of children's friends, combined to form an acculturation score. Even if not used to determine the interview language, the score can be a useful covariable when investigating several cultural groups.

Use a dual-format measure for immigrants. On the other hand, immigrants may switch language during an interview and responses should be recorded in the language used (on a two-language form). With regard to self-administered questionnaires, the best option is to present the material in a dual-language format, either on two separate pages [62] or item by item [40]. The measure of HRQOL by proxies is not recommended, even when respondents are illiterate, since individual subjective appreciation is not reliably assessed by other raters [63]. In this situation, interview must be preferred to self-administered questionnaire. Interview may also be appropriate when questionnaires are not appropriate to the respondent's culture [9].

5. Weighting scores

Consider adapting the weights of scores to the cultural context. A scoring method using weights is provided with the source versions of some instruments (Sickness Impact Profile [64], NHP [65]) in order to combine the information in an index or in several indices (profile). However, the weighting may not apply to the new cultural situation. It can be reexamined either by judgement or using a mathematical approach. Using judgement, the cross-cultural validity of the weighting of items is reexamined by experts, who may be health care professionals, patients or lay people. Several techniques are available to elicit culture-adapted weights from expert opinion. With a mathematical approach, data obtained from a sample of patients are analysed by various statistical techniques for scalability (Gutmann analysis) or dimensionality (factor analysis) in order to work out the best way of aggregating the information in one index or several indices.

Application of the Guidelines

Study characteristics

A description of the content of the 17 studies analysed is given in Table 4. Some instruments have been adapted to another language by different research teams addressing different aspects of the adaptation. For instance, a Chicano version of the Sickness Impact Profile (SIP, originated in English [64]) produced by one team [50] has been further examined for score weights by another team [42].

The methods for adaptation reported in the 17 studies were heterogenous. In the adaptation of the instruments from one culture in another with the same language (three studies), i.e. from American to British English, rewording of items was never carried out by a specified committee. Pretesting of the measure was conducted in two studies using either a probe technique [37] or a comparison of the original and final versions [38]. The weight of scores were reassessed in one study [39]. In the adaptation of instruments from one culture to another which uses a different language (8 studies), the translation techniques used varied from one literal translation to three translations performed independently by one to two translators with varying degrees of qualification. Back-translation was described in six studies. Only one study with multiple translations specified using as many back-translations as translations [41]. When a committee for review was constituted (5 studies), its composition varied from two authors to 12 people including physicians, other health professionals, patients, non-patients and bilingual people. Pretesting of the measure was conducted in four studies. The weighting of scores, examined in two of three studies where it was applicable [43, 47] involved the use of large samples of patients and non-patients as experts. Weights were derived by expert elicitation using the Thurstone method, as in the original instrument.

Six papers addressed important cultural groups of immigrants in the host country: Asians in the U.K. and Hispanics in the U.S.A. The translation techniques used varied considerably, from one to five translations, each involving one to three translators in one to four back-translations. A review committee of 2 to 15 carefully selected bilingual people was constituted in five studies. A pre-test of the new versions was conducted in three studies by 12 to 31 bilingual individuals. Weighting of score was reassessed in one of three relevant studies (SIP in Spanish [42]) using expert evaluation by a group of health care consumers, as in the original instrument.

Quality of the studies

For each of the individual guidelines, the mean quality scores on a scale of 1 = poor, 2 = moderate and 3 = good, ranged between 1.9 and 2.4 (Table 5). Thus, when a guideline was addressed in a study, raters judged the methods to be of moderate quality. According to Landis [66], there was substantial to almost perfect agreement between the two judges in these ratings (weighted κ 0.66–0.93).

Since not all of the relevant guidelines were considered in each study, the overall scores were much lower. In calculating the overall score when a guideline rated as "not done" is assigned a value of 0, the mean quality score overall studies was 1.3. It was higher in immigrant populations (1.6) than in adaptation to another country using the same or another language (0.8 and 1.3 respectively). The agreement between judges for the overall quality score for the 17 studies was high (ICC = 0.92). It was higher in studies where the target was another country using the same language (ICC = 1) or adaptation to another language in another country (ICC = 0.96) than in studies for adaptation to immigrant populations (ICC = 0.87).

COMMENTS

Cross-cultural adaptation must be clearly distinguished from cross-cultural comparison since the two processes rely on different research hypotheses [67]. Adaptation is oriented towards measuring a similar phenomenon in different cultures; it is essentially the production of an equivalent instrument adapted to another culture. Cross-cultural comparison refers to the comparative study of a phenomenon across cultures in order to identify differences attributable to culture. It is possible only after the measurement tool has been adapted and is equivalent in both cultures. Thus, the cross-cultural adaptation of a measure is a prerequisite for the investigation of cross-cultural differences.

The articles reviewed were mainly describing the cross-cultural adaptation of English-language instruments into European languages in European countries. Although our search strategy included all languages, 17 of the 19

Table 4. Reviewed articles with description of the methodology used for the cross-cultural adaptation of HRQOL measures

Original instrument	Language	Country	Translation	Back-translation	Committee	Pre-testing	Weighting score
Adaptation in similar language, other country							
AIMS, 1980 [70]	AIMS British, 1990 [37]	U.K.	N/A*	N/A	—	30 RA patients Probe technique: questionnaire + interview	N/A
HAQ, 1980 [71]	HAQ British, 1986 [38]	U.K.	N/A	N/A	—	27 RA patients Check understandability of the original: 22% pbs 33 RA patients Check understability of modified HAQ: 6% pbs	N/A
SIP, 1981 [64]	SIP British, 1985 [39]	U.K.	N/A	N/A	—	—	135 consumers (stratified by age/sex) 54 health prof: 46 nurses + 8 physicians Thurstone method
Adaptation in another language, other country							
AIMS, 1980 [70]	AIMS French, 1990 [45]	Canada	1 translation 1 translator aware of objective and disease	—	2 bilingual researchers	—	N/A
HAQ, 1980 [71]	HAQ Swedish, 1988 [46]	Sweden	—	—	—	64 RA patients Check relevance	N/A
—	HAQ Portugese, 1990 [44]	Brazil	1 translation by 2 translators aware of objectives and intent	1 back-translation by 2 translators	2 Rheumatologists 2 English professors 2 Physiotherapist 1 Occupational therapist 5 RA patients	31 RA patients with translation 22 RA patients with substitute version Probe technique	N/A
—	HAQ French, 1992 [41]	France	3 translations (2 translators in each team, aware of objectives and disease)	3 back-translations (1 English-cultured)	Authors + 2 bilinguals	—	N/A

QOLS, 1978 [72]	QOLS Swedish, 1992 [48]	Sweden	1 translation by 2 translators aware of objective and disease (authors)	1 back-translation by 1 translator	—	In a pilot study	N/A
NHP, 1980 [65]	NHP Swedish, 1987 [47]	Sweden	—	—	—	—	259 patients + relatives Thurstone method
—	NHP French, 1990 [43]	France	1 translation	1 back-translation	Bilinguals	—	355 Hosp patients 270 non patients Thurstone method
—	NPH Spanish + Catalan, 1990 [49]	Spain	2 translations (1 translator) by experts in NHP	1 back-translation undetailed	2 experts 10 unskilled workers	Cardiology + rheumatology patients Check acceptability	—
Adaptation in another language, same country (immigrants)							
AIMS, 1980 [70]	AIMS Spanish, 1989 [40]	U.S.A.	2 translations: 1 idiomatic + 1 literal (3 translators each) 3 previous translations	2 back-translations (1) 2 previous back-translation	15 selected Hispanic American Selection of best version Iteration of back-translation	12 selected bilinguals Check comprehension and readability	N/A
—	AIMS Spanish, 1989 [51]	U.S.A.	1 translation (1)	1 back-translation (1)	2 translators?	—	N/A
SIP, 1981 [64]	SIP Spanish, 1980 [50]	U.S.A.	several translations	—	—	—	—
—	SIP Spanish, 1984 [52]	U.S.A.	4 translations: 3 by 1 translator and 1 by 2 translators	4 back-translations, each by 4 bilingual translators	3 English and 3 bilinguals	31 bilingual comparison to SIP English overall r −0.95 (0.65–0.95)	N/A
—	SIP Spanish, 1992 [42]	U.S.A.	2 tranlations by 3 translators each: 1 literal + 1 idiomatic	1 back-translation by other bilinguals?	2 reviewers for 2 previous + 2 new translations	12 bilinguals choose the best version	29 health care consumers valuing statements
NHP, 1980 [65]	NHP Urdu, 1989 [53]	U.K.	2 translations (1 translator each)	1 back-translation (1 translator)	3 authors	—	—

*N/A, not applicable.

Table 5. Mean scores and agreement between two judges in assessment of quality of cross-cultural adaptation studies

	Number of studies	Mean score* [range]	Intra-class correlation coefficient	Weighted kappa
Individual guidelines (5 sections)				
Translation	12	2.1 [1–3]		0.70
Back-translation	10	1.9 [1–3]		0.66
Committee	10	2.2 [1–3]		0.86
Pre-testing	9	2.1 [1–3]		0.88
Weighting scores	6	2.4 [2–3]		0.93
Overall guidelines				
Similar language, other country	3	0.8 [0.5–1]	1	
Other language, other country	8	1.3 [0.4–2.1]	0.96	
Other language, same country (immigrants)	6	1.6 [0.4–2.6]	0.87	
All articles	17	1.3 [0.4–2.6]	0.92	

*Mean score across studies calculated as the mean quality ratings assigned for each section with the following values:
Individual guidelines: good = 3, moderate = 2, poor = 1;
Overall guidelines: good = 3, moderate = 2, poor = 1, 0 = not done.

eligible articles were papers published in the English language. It may be that the use of self-administered instruments to measure QOL is an English cultural phenomena. Alternatively, non-English papers may have been missed since papers published only in national journals may not be included in the three medical databases we searched. These papers would have been overlooked unless they were cited in another article included in a database [35]. Finally, much of this research may not be published because it is not the main topic of the research but only a preliminary step toward an application of a QOL measure. However, the increasing number of publications appearing in recent years reflects the growing importance and interest attached to the methodology of cross-cultural adaptation.

Our review of the literature indicates a lack of standardized approach to the cross-cultural adaptation of HRQOL instruments. The methodologies vary and often the authors do not give the readers essential information to understand the strength of the translation. Interestingly, citation searches (**Science Citation Index**) using methodology papers found in the psychology and sociology literature for adaptation of tools in these fields failed to produce additional references relating to HRQOL tools. This suggests that many researchers in QOL may not be aware or do not quote this methodological work developed in the psychology and sociology literature. Based on our review of the methods of cross-cultural adaptation in the field of psychology and sociology, and our review of the HRQOL measures, we propose a set of guidelines which includes 5 essential steps for the translation and cross-cultural adaptation of HRQOL measures. The agreement between the judges using the proposed guidelines appeared substantial to almost perfect. The results were consistent within and across the sections of these guidelines and in different cross-cultural adaptation settings. This indicates that they are appropriate, easy to interpret and suggests that they can be used further with satisfactory reliability.

The quality of the methodology employed for the adaptation of HRQOL measures was rated between 1.9 and 2.4 (on a 1–3 scale) in each section. However, in each study assessed, some aspect of the adaptation process is likely to have been underreported (even if it was carried out) perhaps because it was not at the time regarded as important. Hence, a score of 0 ("not done") might have underrated a work that authors merely did not made explicit. For this reason we are not reporting on individual paper scores. For instance, the replicability of the adapted instrument, i.e. the clarity of the presentation and the thoroughness of the directions provided for its use [19] was addressed in none of the papers, although it might well have been considered by authors.

Whether reliability, validity and sensitivity to change should also be considered in the cross-cultural adaptation process is a matter of controversy. Overall the 17 studies, the reliability of the final version was assessed in 6 studies, the construct validity in 9 studies and the responsiveness in 1 study. On the one hand, one may think that the full achievement of cross-cultural equivalence conveys the equivalence of the original measurement properties. But on the other hand, one may argue that because of the adaptation process, the modified instrument has unknown reliability, validity or sensitivity to

change in the new culture. Since this question is not clear, we did not address this methodology.

Overall, the guidelines used here cannot yet be taken as firm recommendations. Very little research has been done in this field to delineate what is essential from what is supplementary in the process of cross-cultural adaptation. It is important to stress that our guidelines address only the quality of the "process" of adaptation. The quality of the final product (the adapted version) can only be judged by a qualified committee. A single investigator or clinician can only judge the process used in the translation and cross-cultural adaptation since a single individual cannot fulfil the linguistic and other committee qualifications. Our literature review suggests that each step adds quality to the final version in terms of equivalence of concepts explored between source and final instruments, but this benefit has to be weighed against the feasibility of the process.

Important factors in determining the feasibility of a particular technique are the constraints of time and resources. Bilingual people are often in short supply. Variations in how bilingualism is defined may mean they are even more rare*. This problem can be more difficult to deal with when an original instrument is developed in a language other than English. In some such situations, for example in international studies involving several countries in Europe, investigators may opt to translate the instrument into English first, as a common communication language, and then adapt it into the other language as required. If a shortage of bilingual people impedes the pre-testing with bilingual respondents, a probe technique with monolingual respondents can be used.

The preservation of the sensibility of an instrument is time-consuming but unavoidable. The only step that can possibly be shortened is the examination of the weighting of scores, if any, by simply accepting the weights of the original instrument. It should be borne in mind, however, that the validity of the final score may be diminished [43].

Further research is required to establish a method to quantify the equivalence of source and final instruments across cultures, and to identify essential versus optional steps in the adaptation process. Research is also needed to determine adaptation needs for HRQOL measures where items are selected by the patient (Patient Elicitation Technique [68], SEIQoL [69]) rather that the questionnaires addressed in this paper.

In conclusion, the adaptation of a preexisting measure to the cultural context of a target population, as described above, has several advantages:

- it provides a common measure for the investigation of HRQOL within different cultural contexts;
- it offers a standard measure for use in international studies, many of which are now being conducted;
- it allows comparisons between national/cultural groups relying on a standard measure designed and adapted to measure the phenomenon cross-culturally;
- it allows the inclusion of immigrants avoiding the frequent bias of representing only the dominant culture of the country;
- it is less costly and time-consuming than generating a new measure. Nevertheless, it should be borne in mind that the cross-cultural adaptation of HRQOL also requires careful attention, involves numerous people and is time-consuming.

Acknowledgements—The authors are grateful to Dr Stephen Stansfeld and Dr James Wright for fruitful comments on earlier drafts of this manuscript. F. Guillemin was supported by a Bourse Lavoisier from the French Ministry of Foreign Affairs and by the Fondation pour la Recherche Therapeutique.

*Some authors have considered as bilingual people (with one language) who have lived at least one year in another country (with another language) [42]. Other define people as bilingual only when they have been reared in two cultural and language contexts, keeping in touch with both [40]. Although the former definition probably does not allow for an understanding and mastery of idioms and colloquialisms, the latter situation may well be too rare to be useful.

REFERENCES

1. Spilker B, Simpson RL, Tilson HH. Quality of life bibliography and indexes: 1991 update. **J Clin Res Pharmacoepidemiol** 1992; 6: 205–266.
2. Aaronson NK, Meyerowitz BE, Bard M *et al.* Quality of life research in oncology. Past achievements and future priorities. **Cancer** 1991; 67: 839–843.
3. Fitzpatrick R, Fletcher A, Gore S, Jones D, Spiegelhalter D, Cox D. Quality of life measures in health care. I: Applications and issues in assessment. **Br Med J** 1992; 305: 1074–1077.
4. Advances in Health Status Assessment: Proceedings of a conference. **Med Care** 1992; 30: MS1–MS293.
5. Gerin P, Dazord A, Cialdella P, Leizorovicz A, Boissel JP. Le questionnaire "Profil de la qualité de la vie subjective". Premiers éléments de validation. **Therapie** 1991; 46: 131–138.
6. Kempen GI, Suurmeijer TP. The development of a hierarchical polychotomous ADL–IADL scale for

noninstitutionalized elders. **Gerontologist** 1990; 30: 497–502.
7. Berkanovic E. The effect of inadequate language translation on Hispanics' responses to health surveys. **Am J Public Health** 1980; 70: 1273–276.
8. Kleinman A, Eisenberg L, Good B. Culture, illness and care: clinical lessons from anthropologic and cross-cultural research. **Ann Intern Med** 1978; 88: 251–258.
9. Hunt SM, Alonso J, Bucquet D, Niero M, Wiklund I, McKenna S. Cross-cultural adaptation of health measures. **Health Policy** 1991; 19: 33–44.
10. Goldberg D, Williams P. **A User's Guide to the General Health Questionnaire**. Windsor: NFER-NELSON; 1990.
11. Spielberger CD, Sharma S. Cross-cultural measurement of anxiety. In: Spielberger CD, Diaz-Guerrero R, Eds. **Cross-cultural Anxiety**. Washington: Hemisphere Publ. Corp; 1976: 13–25.
12. Brislin RW, Lonner WJ, Thorndike RM. Questionnaire wording and translation. In: **Cross-cultural Research Methods**. New York: John Wiley; 1973: 32–58.
13. Secherst L, Fay TL, Zaidi SMH. Problems of translation in cross-cultural research. **J Cross-cultural Psychol** 1972; 3: 41–56.
14. Brislin RW. Back-translation for cross-cultural research. **J Cross-cultural Psychol** 1970; 1: 185–216.
15. Deutscher I. Asking questions cross-culturally: some problems of linguistic comparability. In: Warwick DP, Oskerson S, Eds. **Comparative research methods**. New Jersey: Prentice Hall; 1973: 163–186.
16. Gonzalez-Reigosa F. The anxiety-arousing effect of taboo words in bilinguals. In: Spielberger CD, Diaz-Guerrero R, Eds. **Cross-cultural Anxiety**. Washington: Hemisphere Publ Corp; 1976: 89–105.
17. Werner O, Campbell DT. Translating, working through interpreters, and the problem of decentering. In: Narroll R, Cohen R, Eds. **A Handbook of Method in Cultural Anthropology**. New York: American Museum of Natural History Press; 1970: 398–420.
18. Schuman H. The random probe: A technique for evaluating the quality of closed questions. **Am Sociol Rev** 1966; 31: 218–222.
19. Feinstein AR. The theory and evaluation of sensibility. In: Feinstein AR, Ed. **Clinimetrics**. New Haven: Yale University Press; 1987: 141–166.
20. Bombardier C, Tugwell PX. Methodological consideration in functional assessment. **J Rheumatol** 1987; 14 (Suppl 15): 6–10.
21. **Reference Manager, version 4**. Carlsbad, CA: Research Information System, Inc.; 1989.
22. Perez ER, Mackenzie CR, Ryan C. Development of a Spanish version of the Modified Health Assessment Questionnaire. **Arthritis Rheum** 1990; 33 (Suppl 9): S100 (Abstract).
23. Cardiel MH, Abello-Banfi M, Ruiz-Mercado R, Alarcon-Segovia D. Quality of life in rheumatoid arthritis: validation of a Spanish version of the disability index of the Health Assessment Questionnaire. **Arthritis Rheum** 1991; 34 (Suppl. 9): S183 (Abstract).
24. Raynauld JP, Singh G, Shiroky JB, Haraoui B, Choquette D, Blier C, Fortin P, Pelletier JP, Gascon J, Goulet JR, Neville C, Fries JF. A French-Canadian version of the Health Assessment Questionnaire. **Arthritis Rheum** 1992; 35 (Suppl. 9): S125 (Abstract).
25. Archenholtz B, Bjelle A. Reliability and validity of a Swedish version of the Arthritis Impact Measurement Scales. **Scand J Rheumatol** 1990; 20: 219 (Abstract).
26. Abello-Banfi M, Cardiel MH, Ruiz-Mercado R, Alarcon-Segovia D. Quality of life in rheumatoid arthritis: validation of a Spanish version of the Arthritis Impact Measurement Scales. **Arthritis Rheum** 1991; 34 (Suppl. 9): S183 (Abstract).
27. Esteva M, Gonzalez N, Ruiz M. Reliability and validity of a Spanish version of the Sickness Impact Profile. **Arthritis Rheum** 1992; 35 (Suppl. 9): S219 (Abstract).
28. Chang RH, Dodder RA. The modified purpose in life scale: a cross national validity study. **Int J Aging Hum Devel** 1983–84; 18: 207–217.
29. Kaplan RM, Bush JW, Berry CC. The reliability, stability and generalizability of a health status index. **Proc Social Statistics Section**. American Statistical Association; 1978: 704–709.
30. Brazier JE, Harper R, Jones NMB, O'Cathain A, Thomas KJ, Usherwood T, Westlake L. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. **Br Med J** 1992; 305: 160–164.
31. Siegert CEH, Vleming LJ, Vandenbroucke JP, Cats A. Measurement of disability in Dutch rheumatoid arthritis patients. **Clin Rheumatol** 1984; 3: 305–309.
32. Van der Heidje DMFM, Van Riel PLCM, Van de Putte LBA. Sensitivity of a Dutch Health Assessment Questionnaire in a trial comparing hydroxychloroquine vs sulphasalazine. **Scand J Rheumatol** 1990; 19: 407–412.
33. Cavalieri F, Salaffi F, Ferracioli GF. Relationship between physical impairment, psychosocial variables and pain in rheumatoid disability. An analysis of their relative impact. **Clin Exp Rheumatol** 1991; 9: 47–50.
34. Wiklund I, Romanus B, Hunt SM. Self-assessed disability in patients with arthrosis of the hip joint. Reliability of the Swedish version of the Nottingham Health Profile. **Int Disabil Studies** 1988; 10: 159–163.
35. Ferracioli GF, Cavalieri F, Salaffi F, Mercadanti M. Compromissione funzionale, invalidita ed handicap nell'artrite reumatoide. Studio prospettico mediante questionari orientari. **Reumatismo** 1988; 40: 45–56.
36. Jacobs HM, Luttik A, Touw-Otten FW, de Melker RA. The Sickness Impact Profile (SIP); results of an evaluation study of the Dutch version. **Nederlands Tijdschift Voor Geneeskunde** 1990; 134: 1950–1954.
37. Hill J, Bird HA, Lawton CW, Wright V. The Arthritis Impact Measurement Scales: An anglicized version to assess the outcome of British patients with rheumatoid arthritis. **Br J Rheumatol** 1990; 29: 193–196.
38. Kirwan JR, Reeback JS. Stanford Health Assessment Questionnaire modified to assess disability in British patients with RA. **Br J Rheumatol** 1986; 25: 206–209.
39. Patrick DL, Sittampalam Y, Sommerville SM, Carter WB, Bergner M. A cross-cultural comparison of health status values. **Am J Public Health** 1985; 75: 1402–1408.
40. Hendricson WD, Russel IJ, Jacobson JM, Rogan H, Bishop GD, Castillo R. Development and initial validation of a dual language English–Spanish format for the Arthritis Impact Measurement Scales. **Arthritis Rheum** 1989; 32: 1153–1159.
41. Guillemin F, Briançon S, Pourel J. Validity and discriminant ability of a French version of the Health Assessment Questionnaire in early RA. **Disabil Rehabil** 1992; 14: 71–77.
42. Deyo RA. Pitfalls in measuring the health status of Mexican Americans: comparative validity of the English and Spanish Sickness Impact Profile. **Am J Public Health** 1984; 74: 569–573.
43. Bucquet D, Condon S, Ritchie K. The French version of the Nottingham Health Profile. A comparison of items weights with those of the source version. **Soc Sci Med** 1990; 30: 829–835.
44. Bosi-Ferraz M, Magalhaes Oliveira L, Araujo PMP, Atra E, Tugwell P. Cross-cultural reliability of the physical ability dimension of the Health Assessment Questionnaire. **J Rheumatol** 1990; 17: 813–817.

45. Sampalis JS, Pouchot J, Beaudet F, Carette S, Gutkowski A, Harth M, Myhal D, Senecal JL, Yeadon C, Williams JI, Esdaile JM. Arthritis Impact Measurement Scales: reliability of a French version and validity in adult Still's disease. **J Rheumatol** 1990; 17: 1657–1661.
46. Ekdahl C, Eberhardt K, Andersson ST, Svensson B. Assessing disability in patients with rheumatoid arthritis. **Scand J Rheumatol** 1988; 17: 263–271.
47. Hunt SM, Wiklund I. Cross-cultural variation in the weighting of health statements: a comparison of English and Swedish valuations. **Health Policy** 1987; 8: 227–235.
48. Burckhardt CS, Archenholtz B, Bjelle A. Measuring the quality of life of women with rheumatoid arthritis or systemic lupus erythematosus: a Swedish version of the Quality of Life Scale (QOLS). **Scand J Rheumatol** 1992; 21: 190–195.
49. Alonso J, Anto JM, Moreno C. Spanish version of the Nottingham Health Profile: translation and preliminary validity. **Am J Public Health** 1990; 80: 704–708.
50. Gilson BS, Erickson D, Chavez CT, Bobbitt RA, Bergner M, Carter W. A Chicano version of the Sickness Impact Profile. **Cult Med Psychiat** 1980; 4: 137–150.
51. Coulton CJ, Hyduk CM, Chow JC. An assessment of the Arthritis Impact Measurement Scales in 3 ethnic groups. **J Rheumatol** 1989; 16: 1110–1115.
52. Hendricson WD, Russel IJ, Prihoda TJ, Jacobson JM, Rogan A, Bishop GD. An approach to developing a valid Spanish language translation of a health status questionnaire. **Med Care** 1989; 27: 959–966.
53. Ahmad WIU, Kernohan EEM, Baker MR. Cross-cultural use of socio-medical indicators with British Asians. **Health Policy** 1989; 13: 95–102.
54. Fleiss JL. **The Design and Analysis of Clinical Experiments**. New York: Wiley; 1986.
55. Dunn G. **Desin and Analysis of Reliability Studies**. New York: Oxford University Press; 1989.
56. Del Greco L, Walop W, Eastridge L. Questionnaire development. 3. Translation. **Can Med Assoc J** 1987; 136: 817–818.
57. Chesney AP, Chavira JA, Hall RP, Gary HE. Barriers to medical care of Mexican-Americans: the role of social class, acculturation and social isolation. **Med Care** 1982; 20: 883–891.
58. Schribner R, Dwyer JH. Acculturation and low birthweight among Latinos in the Hispanic HANES. **Am J Public Health** 1989; 79: 1263–1267.
59. Marks G, Solis J, Richardson JL, Collins LM, Birba L, Hisserich JC. Health behavior of elderly Hispanic women: does cultural assimilation make a difference? **Am J Public Health** 1987; 7: 1315–1319.
60. Hunt SM. Cross-cultural issues in the use of socio-medical indicators. **Health Policy** 1986; 6: 149–158.
61. Flaherty JA, Gaviria FM, Pathak D, Mitchell T, Wintrob R, Richman JA, Birz S. Developing instruments for cross-cultural psychiatric research. **J Nerv Ment Dis** 1988; 176: 257–263.
62. Karno M, Burnam A, Escobar JI, Hough RL, Eaton WW. Development of the Spanish-language version of the national institute of mental health diagnostic interview schedule. **Arch Gen Psychiat** 1983; 40: 1183–1188.
63. Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: a review. **J Clin Epidemiol** 1992; 45: 743–760.
64. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: Development and final revision of a health status measure. **Med Care** 1981; 19: 787–805.
65. Hunt SM, McKenna SP, McEwen J. A quantitative approach to perceived health status: a validation study. **J Epidemiol Commun Health** 1980; 34: 281–286.
66. Landis JR, Koch GG. The measurement of observer agreement for categorical data. **Biometrics** 1977; 33: 159–174.
67. Campos SS, Johnson TM. Cultural considerations. In: Spilker B, Ed. **Quality of Life in Clinical Trials**. New York: Raven Press; 1990: 163–170.
68. Tugwell PX, Bombardier C, Gent M, Bennett KJ, Bensen WG, Carette S, Chalmers A, Esdaile JM, Klinkhoff AV, Kraag GR, Ludwin D, Roberts R. Low-dose cyclosporine versus placebo in patients with rheumatoid arthritis. **Lancet** 1990; 335: 1051–1055.
69. O'Boyle CA, McGee H, Hickey A, O' Malley K, Joyce CRB. Individual quality of life in patients undergoing hip replacement. **Lancet** 1992; 339: 1088–1091.
70. Meenan RF, Gertman PM, Mason JH, Dunaif R. Measuring health status in arthritis: The Arthritis Impact Measurement Scales. **Arthritis Rheum** 1980; 23: 146–152.
71. Fries JF, Spitz PW, Young DY. Measurement of health outcome in arthritis. **Arthritis Rheum** 1980; 23: 137–145.
72. Flanagan JD. A research approach to improving our quality of life. **Am Psychol** 1978; 33: 138–147.

(Appendix overleaf)

APPENDIX

ABSTRACTION FORM FOR THE APPRAISAL OF CROSS-CULTURAL ADAPTATION OF HRQOL INSTRUMENTS

Original instrument: _____

Population addressed by the target version:

Language: _____

Culture: _____

Country of origin: _____

Country of residence: _____

SCORE

1. Translation technique:

Number of translations: ☐

Number of translators in each translation: ☐

Were they translating into their mother tongue: Yes No

Were they aware of the concepts: Yes No

Were they aware of the target condition(s): Yes No ☐

2. Back-translation technique

Number of back-translations: ☐

Number of translators in each back-translation: ☐

Were they translating into their mother tongue: Yes No

Were they aware of the concepts: Yes No

Were they aware of the target condition(s): Yes No ☐

3. Committee approach

Committee review done: Yes No

Composition of the committee: _____

_____

_____

Translation/back-translation process iterated: Yes No ☐

4. Pre-testing

With bilinguals: Yes No

With monolinguals: Yes No

using a probe technique: Yes No

using another technique: Yes No

specify:

Sample composition: _____

_____

_____

Sample size: ☐ ☐

5. Weighting scores adaptation SCORE

Was the weighting of scores examined: Yes No Not applicable

with patients: Yes No

with other experts: Yes No

What method was used: _____

Which were the results: _____ ☐