2001-1009-1

Translation and Cultural Adaptation of Quality of Life Questionnaires: An Evaluation of Methodology

DIRCILENE da MOTA FALCÃO, ROZANA MESQUITA CICONELLI, and MARCOS BOSI FERRAZ

ABSTRACT. ***Objective.*** To evaluate the methodological steps currently proposed in translation and cultural adaptation of quality of life questionnaires.

Methods. Fifty patients with rheumatoid arthritis were invited to participate. Two versions each of the Stanford Health Assessment Questionnaire (HAQ), the MOS 36 Item Short-Form Health Survey (SF-36), and the Arthritis Impact Measurement Scales 2 (AIMS-2) were administered: version 1 was a literal translation of the questionnaire; version 2 resulted from a process of translation and cultural adaptation following internationally accepted guidelines. For each patient we applied 2 questionnaires before and after consultation. The questionnaire, the order of administration, and the version were randomly assigned. The interviews were performed by a single interviewer. Several clinical and laboratory outcome measures were assessed simultaneously. Descriptive statistical analysis was used to characterize the data. Spearman and intraclass correlation coefficients were used to evaluate reliability and validity of each version of each questionnaire.

Results. Patients' mean (SD) age was 47 (12) years and mean (SD) disease duration was 10 (7) years. The differences of the mean in the physical and affective AIMS-2 components between the literal and adapted versions (0.21 and 0.11, respectively) were similar to the differences in the intraobserver application of the same version of culturally adapted AIMS-2 components (0.03 and 0.20) (the component scores range from 0 to 10). The same results were observed when considering other components of AIMS and SF-36, as well as HAQ scores. Version 1 and 2 presented a similar clinically and statistically significant correlation with clinical and laboratory measures used in the validation process of the questionnaires.

Conclusion. The complex methodologies proposed in the translation and validation of the questionnaires should be carefully reevaluated. The simplification of this methodology should be studied. (J Rheumatol 2003;30:379–85)

Key Indexing Terms:
QUALITY OF LIFE STANFORD HEALTH ASSESSMENT QUESTIONNAIRE
SHORT-FORM HEALTH SURVEY (SF-36) ARTHRITIS IMPACT MEASUREMENT SCALES
CROSS CULTURAL ADAPTATION

In addition to traditional measures such as the number of painful or swollen joints, clinical assessment of patients with rheumatoid arthritis (RA) currently includes the evaluation of other aspects of quality of life[1-4]. Quality of life assessment has been basically conducted through the use of questionnaires. Most of them have been compiled in the English language, to be used with the English-speaking population. They were developed within a specific cultural scenario[5]. With the increase in the number of questionnaires, the process of translation and cultural adaptation must be considered when using these instruments in other countries, cultures, and languages. The translation should not be exclusively linguistic, but it must be culturally adapted to maintain the same measurement properties. So, to be applied, the questionnaire must be previously translated and also culturally adapted to the target population. This process follows internationally accepted guidelines, which make it costly and time-consuming[5,6]. We evaluated the need for the proposed and internationally accepted methodological steps in the translation and cultural adaptation of quality of life questionnaires.

MATERIALS AND METHODS

Patients. Fifty patients with RA diagnosed according to the American College of Rheumatology (ACR) classification criteria[7] were selected and invited to participate in our study. The patients were consecutively selected from the outpatient clinic of the Rheumatology Division of the Escola Paulista de Medicina–UNIFESP, from June 1997 to April 1998. These patients were not accustomed to answering questionnaires and they were not previously exposed to any of the versions.

Patients with other diseases that precluded them from providing reliable data (e.g., dementia, sequelae of neurological disease) were excluded from the study.

Interviews using questionnaires. Two versions of the Stanford Health Assessment Questionnaire (HAQ)[8], the Arthritis Impact Measurement Scales 2 (AIMS-2)[9], and the MOS 36 Item Short-Form Health Survey (SF-36)[10] were used. The version corresponding to the first translation of the original

From the Division of Rheumatology, Department of Medicine, University of São Paulo-UNIFESP, São Paulo, Brazil.

D.M. Falcão, MD, Postgraduate Student; R.M. Ciconelli, MD, Assistant Physician; M.B. Ferraz, MD, Associate Professor.

Address reprint requests to Dr. M.B. Ferraz, Disciplina de Reumatologia, Departamento de Medicina, Universidade Federal de São Paulo-UNIFESP, Rua Botucatu 740, São Paulo, Brazil CEP 04023-062. E-mail: marcos.ferraz@fleury.com.br

Submitted November 23, 2001; revision accepted August 6, 2002.

questionnaire, which was obtained from previous validation studies of the HAQ[11], AIMS-2[12], and SF-36[13], is here called the literal version (LV). A second version, resulting from the process of translation and cultural adaptation of the questionnaire following internationally accepted criteria [which includes initial translation, evaluation of the initial translation (back-translation), and evaluation of cultural equivalencies (pretesting)], is here called the adapted version (AV). This provided us with 6 different questionnaires, the LV and the AV of the HAQ, the AIMS-2, and the SF-36, respectively.

Fifty-five patients were interviewed and required to answer 2 different questionnaires, each in the 2 versions described above.

Two random selections were made: first, it was predefined by means of a raffle which 2 out of the 3 questionnaires should be administered to each patient, and second, which version (LV or AV) of each of the selected questionnaires would be completed immediately before and after the clinic visit. Thus 4 versions were completed, 2 before and 2 after the routine clinic visit. The interviews were conducted on the same day and by the same observer. The interviewer simply read the questions, and neither helped to answer nor explained the meaning of the sentence.

Assessment of patient using clinical and laboratory measures. Clinical assessment of the patient was performed for comparative purposes using the conventional measures for evaluation of RA, for example, morning stiffness, numerical rating scale for pain, functional class, global evaluation of disease activity by the patient and by the physician, and painful and swollen joint counts. All these variables were evaluated at the time of the patient's clinic visit.

Morning stiffness was expressed in minutes; the evaluation of pain was performed using a numerical rating scale for pain with scores ranging from 0 to 10 (0 = no pain and 10 = extreme pain), patients being instructed to choose the score that best defined their pain during the last week; functional class was determined according to the ACR 1992 criteria[14]. Global evaluation of disease activity by the patient and the physician was performed using a numerical rating scale with scores ranging from 0 to 10 (0 = no activity and 10 = extreme activity). Joint count was performed according to ACR recommendation: 68 painful joints and 66 swollen joints were evaluated (coxofemoral joints were excluded)[15]. Laboratory assessment consisted of analysis of the erythrocyte sedimentation rate (ESR) and rheumatoid factor (latex agglutination test).

Statistical analysis. Descriptive statistical analysis was used to characterize patients' demographic, clinical, and laboratory features. Spearman's correlation coefficient was used to assess correlation between the different components of the different versions of the same questionnaire and also between the questionnaires and clinical data. Correlation between the scores of LV and AV of the questionnaires, when applied by the same observer, was compared to the intraobserver reproducibility observed in the translation, cultural adaptation, and validation studies of these questionnaires. Spearman's coefficient correlation was used. Intraclass correlation coefficients were used to evaluate the degree of agreement between the LV and AV of the questionnaires.

RESULTS

Table 1 shows the patients' clinical and sociodemographic characteristics. Ninety-two percent were female, with mean age (SD) of 47.56 (11.27) years. Eighty-four percent were equally distributed in functional classes I and II. Seventy-four percent of the patients were literate, although the great majority had formally studied for a period of less than 5 years. Twenty-seven patients did not show comorbidities; hypertension and dyspeptic syndromes were the most frequently observed comorbidities.

Table 2 shows the clinical and laboratory findings of the patients. The mean numbers (SD) of painful and swollen joints were 4.7 (9.9) and 1.86 (2.75), respectively. The mean (SD) score for pain assessment by the visual analog scale (range 0–10) was 4.24 (3.01).

Table 1. Clinical and sociodemographic characteristics of the 50 patients with RA.

Sex	
Female (%)	46 (92)
Male (%)	4 (8)
Age, yrs, mean (SD)	47.56 (11.27)
Functional Class*, n (%)	
I	21 (42)
II	21 (42)
III	8 (16)
IV	0
Duration of disease, yrs, mean (SD)	10.62 (7.70)
Race, n (%)	
White	26 (52)
Non–white	24 (48)
Schooling, n (%)	
Literate	37 (74)
Illiterate	13 (26)
Family's monthly income, 33 patients	
1 MW	1
1 to 4 MW	17
> 6 MW	15
Comorbidity, n	
Hypertension	12
Dyspeptic syndrome	8
Other	3
None	27
Latex ≥ 1/80, n (%), 43 patients	22 (51)

*According to the ACR[14] criteria. MW: minimum wage.

Table 3 shows the scores obtained for each component of the LV and the AV of the questionnaires. The scores obtained in both versions of the HAQ were very similar, close to 1. Most mean values for each component of the SF-36 (also similar between versions) were around 40 and 60, the lowest mean values being observed in relation to the physical aspects, and the highest mean values in relation to social aspects and mental health. Values reported for the AIMS-2 components also showed similarity between versions; no clinically relevant differences were observed between LV and AV values.

Table 4 shows the correlation between the same components of the different versions (LV compared to AV) of the same questionnaires, verified by Spearman correlation coefficient, intraclass correlation coefficient, and 95% confidence interval. Statistically significant correlation was found in all of the components. For the HAQ, statistically significant values are those > 0.460 ($p < 0.01$); for the AIMS-2, values > 0.451 ($p < 0.01$), and for the SF-36, values > 0.433 ($p < 0.01$).

Table 5 shows another comparison between that result and the correlation coefficient for intraobserver reproducibility (AV compared to AV) for the same components in previous validation studies. The statistically significant values reported in these studies were similar to those we observed in the comparison of the same components in different versions. The dif-

Table 2. Clinical and laboratory variables of the 50 patients with RA.

Variable	Mean	SD	Minimum	Maximum
Morning stiffness, min	32.7	66.7	0	360
Number of painful joints (0 to 68)	4.7	9.9	0	67
Number of swollen joints (0 to 66)	1.86	2.75	0	13
Physician's global assessment of disease activity (numerical scale ranging from 0 to 10)	3.22	2.64	0	8
Patient's global assessment of disease activity (numerical scale ranging from 0 to 10)	4.24	3.17	0	10
Pain assessment (numerical scale ranging from 0 to 10)	4.24	3.01	0	10
ESR	43.58	24.2	2	100

Table 3. Scores for the literal and the adapted versions of the HAQ, SF–36, and AIMS–2 obtained through interviews with 50 patients with RA.

Questionnaire	Mean (SD)	Mean (SD)
HAQ	1.01 (0.68)	1.07 (0.66)
SF–36		
Physical functioning	48.14 (25.61)	48.43 (28.07)
Role–physical	33.57 (44.12)	35.57 (45.53)
Bodily pain	51.66 (25.16)	51.80 (23.06)
General health perceptions	53.63 (19.31)	52.71 (19.24)
Vitality	54.71 (22.78)	48.86 (24.68)
Social functioning	67.50 (27.31)	65.36 (27.97)
Role–emotional	56.14 (43.36)	60.89 (46.09)
Mental health	64.23 (24.19)	63.09 (23.84)
AIMS2		
Physical	2.63 (1.73)	2.41 (1.70)
Emotional	4.44 (2.14)	4.33 (2.00)
Symptoms	4.83 (2.53)	4.83 (2.72)
Social interaction	3.80 (2.08)	3.82 (2.09)
Work	2.70 (2.46)	3.17 (2.49)

LV: literal version; AV: adapted version.

ference observed between the mean scores of the LV and AV of the HAQ, the AIMS-2, and the SF-36, applied on the same day, was similar to the difference observed when the adapted version of each questionnaire was applied on different days.

Tables 6, 7, and 8 show the Spearman correlation coefficients between the LV and the AV of the HAQ, AIMS-2, and SF-36, respectively, and the clinical variables used. We observed that most clinical variables correlated with the questionnaires. The number of swollen joints and the ESR did not show significant correlation with any questionnaire. For the AIMS-2, the social interaction component was the one that least correlated with the clinical measures. For the SF-36, the component that least correlated was the emotional aspect.

DISCUSSION

One challenge faced by the rheumatology community has been to establish measures for the evaluation of rheumatologic diseases in general, and RA in particular. These measures may quantify objective signs such as number of painful or swollen joints, symptoms such as duration of morning stiffness or the intensity of pain, and even the physical function or global health status as perceived by the patient or by the physician[4]. In 1993, the ACR and an international committee initiative — the Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT)[16] — recommended the use of a standardized set of outcome measures in RA trials. The suggested outcome measures included not only the assessment of the frequency and severity of the disease, but also an evaluation of quality of life.

Most instruments developed for assessment of quality of life are based on questionnaires that, although widely available, have a great limitation for physicians and researchers: the great majority of these tools have been produced and validated in English-speaking countries[5,6]. Although valid in their countries of origin, these measures are not directly applicable elsewhere due to cultural differences among the nations[5,6]. In order to be appropriately used in a new context, they must be submitted to a complex process that includes translation of the instrument into a different language, followed by a detailed process of cultural adaptation and validation of its measurement properties[5,6,17]. This process has been heterogeneously described in the literature, and there has been a tendency to use more complex methodologies in which different numbers of steps and qualification components have been proposed by researchers and methodologists.

In 1993, Guillemin, *et al*[5] conducted an extensive revision of the literature and verified that most of the studies conducted for the translation and validation of quality of life questionnaires did not follow the literature guidelines. This study assessed a total of 712 references, and of these, only 17 papers were selected for revision. All the rest were excluded for failing to provide a clear description of the method used for the translation and validation process. The authors then proposed a methodologically standardized guideline that included the following steps: translation, back-translation, committee review, pretesting, and weight of scores.

However, although this methodology has been internationally accepted, it should be noted that not all the proposed methodological steps have been applied in a uniform way when translating and adapting quality of life questionnaires to culturally different scenarios. There are variations in the num-

Table 4. Spearman's correlation coefficient between versions of the HAQ, the SF–36, and the AIMS–2 when different versions (LV compared to AV) are applied, compared to the intraobserver reproducibility of the adapted version (AV compared to AV) and the intraclass correlation coefficient (ICC) between the LV and the AV of the HAQ, SF–36, and AIMS–2 questionnaires.

Questionnaire	LV/AV Spearman Correlation	AV/AV* Spearman Coefficient	LV/AV ICC	LV/AV 95% CI
HAQ	0.90	0.91	0.91	0.82–0.89
SF–36				
Physical functioning	0.75	0.80	0.79	0.62–0.89
Role–physical	0.92	0.63	0.94	0.89–0.97
Bodily pain	0.72	0.54	0.76	0.57–0.87
General health perception	0.66	0.84	0.63	0.38–0.80
Vitality	0.85	0.65	0.81	0.64–0.90
Social functioning	0.81	0.75	0.79	0.63–0.89
Role–emotional	0.78	0.44	0.77	0.59–0.88
Mental health	0.67	0.69	0.77	0.60–0.88
AIMS–2				
Physical	0.94	0.94	0.87	0.75–0.93
Emotional	0.80	0.95	0.81	0.65–0.90
Symptoms	0.90	0.96	0.82	0.67–0.91
Social interaction	0.81	0.95	0.92	0.84–0.96
Work	0.64	0.97	0.76	0.54–0.88

*Results from previous validation studies.

Table 5. Difference in the mean scores of the literal version (LV) and the adapted version (AV) of the HAQ, the SF–36, and the AIMS–2 applied on the same day, compared to the difference between the adapted version of the same questionnaires applied on different days.

Questionnaire	Difference Between the Mean Scores of Both Versions LV/AV*	AV/AV**
HAQ (0–3)	0.06	0.05
SF–36 (0–100)		
Physical functioning	0.29	1.7
Role–physical	2.0	2.0
Bodily pain	0.14	3.4
General health perception	0.92	2.8
Vitality	5.85	1.1
Social functioning	2.15	1.3
Role–emotional	4.75	2.0
Mental health	1.14	1.5
AIMS–2 (0–10)		
Physical	0.21	0.03
Emotional	0.11	0.20
Symptoms	0	0.16
Social interaction	0.1	0.29
Work	0.47	0.35

*Data are from the present study. **Data are from previous translation and validation studies[11–13].

ber of translators and back-translators required, the definition of the characteristics of these translators, the importance of translators knowing or not knowing the objective of the translation being requested and, finally, discrepancies regarding the need to specify the types of equivalence used, or the adequacy of the weight given to the scores.

Table 6. The relationship between the LV and AV of the HAQ and the various clinical variables, as assessed by the Spearman correlation coefficient.

Clinical Variable	Version of the HAQ LV	AV
No. swollen joints	0.11	0.20
No. painful joints	0.38*	0.41*
Patient global assessment	0.44*	0.45*
Physician global assessment	0.51**	0.46**
Numerical rating scale for pain	0.52**	0.49**
Morning stiffness	0.54**	0.58**
ESR	0.34	0.20

*$p < 0.05$. **$p < 0.01$.

Another issue is the circumstances in which the process of translation and cultural adaptation should be conducted. The analysis of a large country with an enormous amount of cultural variations leads us to believe that it is unlikely that language will be the only obstacle when trying to use the same translated instrument[18].

There are reports such as that of González, *et al*[19] in 1995, whose objective was to produce Spanish versions of the quality of life questionnaires most frequently used in rheumatology, such as the HAQ. According to the authors, Hispanics represent about 8.8% of the United States population, living as an heterogeneous group originating from different Latin American countries with varying lengths of stay in the US. All these factors make the oral and written Spanish language used by these patients distinctly different according to the region

Table 7. Spearman correlation coefficient between various clinical and laboratory variables and the literal and the adapted versions of the AIMS-2.

Variable	Literal Version					Adapted Version				
	Physical	Emotional	Symptoms	Social Interaction	Work	Physical	Emotional	Symptoms	Social Interaction	Work
No. swollen joints	0.50**	0.14	0.60**	0.10	0.62**	0.42*	0.42*	0.66**	0.01	0.58**
No. painful joints	0.55**	0.27	0.72**	0.20	0.62**	0.62**	0.49**	0.71**	0.10	0.58**
Patient global assessment	0.72**	0.61**	0.87**	0.62**	0.56**	0.73**	0.79**	0.89**	0.43*	0.82**
Physician global assessment	0.63**	0.44*	0.84**	0.50**	0.41*	0.54**	0.60**	0.83**	0.24	0.76**
Numerical rating scale for pain	0.72**	0.66**	0.85**	0.64**	0.57**	0.72**	0.77**	0.87**	0.45*	0.83**
Morning stiffness	0.51**	0.37	0.79**	0.67**	0.28	0.49**	0.53**	0.80**	0.56**	0.65**
ESR	0.51*	0.54**	0.41*	0.57**	0.25	0.41*	0.43*	0.37	0.26	0.51*

*p < 0.05. **p < 0.01.

Table 8. Spearman correlation coefficient between the literal and the adapted versions of the SF–36 and each of the clinical and laboratory variables considered.

	Literal Version								Adapted Version							
	PF	RF	BP	GHP	VIT	SF	RE	MH	PF	RF	BP	GHP	VIT	SF	RE	MH
No. swollen joints	–0.26	–0.69**	–0.35	–0.04	–0.13	–0.03	–0.04	–0.07	–0.37*	–0.58**	–0.49**	–0.14	–0.22	–0.19	–0.03	–0.03
No. painful joints	–0.25	–0.56**	–0.50**	–0.27	–0.31	–0.07	–0.14	–0.11	–0.46**	–0.43**	–0.73**	–0.28	–0.39*	–0.23	–0.02	–0.33
Patient global assessment	–0.42*	–0.62**	–0.39*	–0.43**	–0.47**	–0.34	–0.17	–0.25	–0.41*	–0.47**	–0.64**	–0.29	–0.59**	–0.37*	–0.12	–0.58**
Physician global assessment	–0.43**	–0.70**	–0.51**	–0.26	–0.36*	–0.24	–0.26	–0.12	–0.57**	–0.60**	–0.62**	–0.22	–0.47**	–0.35	–0.22	–0.43**
Numerical rating scale for pain	–0.55**	–0.57**	–0.51**	–0.38*	–0.52**	–0.39*	–0.27	–0.33	–0.53**	–0.45**	–0.72**	–0.42*	–0.60**	–0.50**	–0.10	–0.66**
Morning stiffness	–0.15	–0.44**	–0.32	–0.44**	–0.43**	–0.20	–0.15	–0.20	–0.29	–0.28	–0.60**	–0.43**	–0.50**	–0.21	0.04	–0.49**
ESR	–0.13	–0.51*	–0.20	–0.03	0.15	0.11	0.11	0.13	–0.29	–0.52*	–0.35	–0.32	–0.06	–0.09	0.10	–0.03

*p < 0.05. ** p < 0.01. PF: physical functioning; RP: role-physical; BP: bodily pain; GHP: general health perceptions; VIT: vitality; SF: social functioning; RE: role emotional, MH: mental health.

they come from, sometimes using the same words for totally different concepts. The authors conclude that regional or national variables are common to all languages.

It may be possible that some of the methodological steps proposed today might not add much to the process of translation and cultural adaptation of the questionnaires. Considering that, the translation and cultural adaptation process should be more simple and flexible to allow a considerable reduction in its costs. With this objective in mind it would be worthwhile to consider simplifying the methodological process.

The questionnaires used in this study have already been translated and adapted to our cultural scenario, following the internationally accepted criteria proposed by Guillemin, *et al*[5], for a population of patients from the same outpatient clinic in which the current study was developed.

The population included in our study is very similar to the population described in the studies dealing with the translation and validation of questionnaires. We noted, as shown in Table 3, that mean values for the HAQ and all the components of the SF-36 and AIMS-2 showed great similarity of scores, with only slight variations; this demonstrates that the different versions of the same questionnaire applied to that group of patients had a comparable capacity to detect a particular health status. The greatest variations were observed for the components "vitality" and "role emotional" of the SF-36 questionnaire, and also the component "work" of the AIMS-2.

In regard to the SF-36 questionnaire, we also observed considerable variation in one of the modified components when comparing the original version to the adapted version, and this component was vitality. In spite of this finding, the physical functioning component, which was also modified, failed to show the same behavior.

Despite the modifications in questions number 7, 11, 15, 19, and 20, the HAQ showed no variation in the mean results obtained. Ten patients did not answer question number 11 of the literal version, 8 patients answered promptly — in spite of not having a bathtub at home, and only one patient did not answer question number 7. Since, according to the method used to count the scores, a single question may remain unanswered without interfering in the accountability of the results, no great variation was observed in the results obtained. Patients did not encounter difficulties in understanding the remaining questions.

The modified questions in the SF-36 questionnaire were

numbers 3 and 9. Only 3 patients mentioned that they had never played bowling or golf (question number 3), while all others answered naturally to the question; maybe because in addition to these activities, which are not very common among the Brazilian population, other more familiar activities were included, such as pushing a table or vacuum cleaning, thus allowing the patient to answer the question by considering either of the activities presented.

Question number 9, which refers to vitality, represented a higher degree of difficulty for the patients since it required quantification of the response, which could be all the time, most of the time, a considerable part of the time, sometimes, seldom, or never, instead of requiring a specific answer to a question. The same fact was observed regardless of the version applied.

With regard to the AIMS-2, although questions number 3, 6, 27, 41, 43, and 51 were modified, a total of 5 patients had difficulty in understanding a single question, number 3, claiming they had not completely understood the wording of the question.

In the study conducted by Abello-Banfi, *et al*[20] related to the translation and validation of the AIMS into Spanish, some of the questions left unanswered by patients were those requiring quantification of the answers; for example: "how long were you able to relax without difficulty during the last month," "how long did you feel calm and peaceful during the last month," or "how long did you feel relaxed and free from stress during the last month."

These results lead us to believe that the modifications in the original instruments did not substantially alter the results obtained when compared to the literally translated version.

Table 4 shows the correlation between the literal version and the adapted version using the Spearman correlation coefficient, compared to the coefficient of correlation of the intraobserver reproducibility retained from previous translation and validation studies. It is important to point out that the Spearman correlation coefficient represents the relatedness between 2 measures. The coefficient itself is not sufficient to make sure that the level of measure is similar; however, the intraclass correlation coefficient could indicate that.

In a 1995 study, Perneger, *et al*[21] demonstrated that a quick process of translation and cultural adaptation may represent an alternative to the production of cultural adaptations when resources are scarce. The SF-36 questionnaire was translated and adapted using this process, and was then compared to the officially accepted version that was translated and validated according to the guidelines developed by the International Quality of Life Assessment (IQOLA) Project[22], and the results showed no major differences. In 1995, Sullivan, *et al*[23] produced a Swedish version of the SF-36, the first phase of which was conducted according to norms different from those traditionally accepted. When the questionnaire was applied, there were no significant differences in the properties tested versus the other versions of the IQOLA.

In 1997, in different studies conducted by Goycochea, *et al*[24] and Arguedas, *et al*[25], methodological studies for the translation and cultural adaptation of the Childhood Health Assessment Questionnaire (CHAQ) into Spanish were developed in Mexico and Costa Rica with various methodological steps that did not totally comply with international guidelines. Despite the common language and the geographic proximity, only one of the 26 questions included in the questionnaire had the same written wording; others were only similar and the great majority differed in the equivalence of vocabulary, syntax, or experience.

Thus, in the absence of convincing evidence in the literature in support of this lengthy and costly methodology for the translation and cultural adaptation of questionnaires in scenarios with limited research resources, we propose the simplification of the methodology used to translate and adapt quality of life questionnaires.

The basic process is still maintained, since the steps of translation, back-translation, cultural adaptation, and evaluation of the measurement properties remain, yet the complexity of each of these steps were reduced as follows:

Translation: must be performed by a physician or health professional (translator 1) who must be fluent in both languages, the original language and the target language of the questionnaire. He/she must have a thorough knowledge of the languages in question and concepts related to quality of life and the cultural adaptation of the questionnaires. In case of questionnaires developed for a specific disease, a clear understanding of the disease in question is also desirable. Thus, version 1 of the process will be generated.

Back-translation and revision of version 1: must be performed by a native teacher of the language of the original questionnaire (translator 2) in order that the translation is as close as possible to the original language. The version resulting from the back-translation will be reviewed by 2 translators (1 and 2) and, if necessary, and after having reached a consensus, alterations to version 1 should be implemented. Special attention must be given to verb tenses and colloquial expressions. Thus, version 2 of the process will be generated.

Cultural adaptation: a group composed of 2 physicians or health professionals familiar with quality of life questionnaires and their applications, one knowledgeable patient with formal instruction, and a native teacher of the language used in the original questionnaire will be responsible for this process. This group, after a clear presentation of the types and examples of the cultural equivalents (semantic, conceptual, experiential, and idiomatic) by the project coordinator, will perform a detailed revision of the questionnaire, analyzing item by item, and will substitute culturally equivalent items for the potentially problematic ones. If necessary, the weight of the scores will also be reviewed in this phase. Thus, the final version (version 3) of the questionnaire in the target language will be established.

Evaluation of the measurement properties of the translated version (version 3): assessment of reproducibility and validity, and responsiveness if applicable.

The simplified methodology proposed in this study needs to be evaluated in prospective studies.

REFERENCES

1. Buchbinder R, Bombardier C, Yeung M, Tugwell P. Which outcome measures should be used in rheumatoid arthritis clinical trials? Arthritis Rheum 1995;38:1568-80.
2. Pincus T, Brooks RH, Callahan LF. Prediction of long-term mortality in patients with rheumatoid arthritis according to simple questionnaire and joint count measures. Ann Intern Med 1994;120:26-34.
3. Pincus T, Callahan LF, Brooks RH. Self-reported questionnaire scores in rheumatoid arthritis compared with traditional physical, radiographic and laboratory measures. Ann Intern Med 1987;110:259-66.
4. Boers M, Tugwell P. The validity of pooled outcome measures (indices) in rheumatoid arthritis clinical trials. J Rheumatol 1993;20:568-74.
5. 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.
6. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health status measures. Scand J Rheumatol 1995;24:61-3.
7. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.
8. Fries JF, Spitz PW, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137-45.
9. Meenan RF, Gertman PM, Mason JH. The Arthritis Impact Measurement Scales. Arthritis Rheum 1980;23:146-52.
10. Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): I: Conceptual framework and item selection. Med Care 1992;30:473-83.
11. Ferraz MB, Oliveira LM, Araujo PMP, Atra E, Tugwell P. Cross-cultural reliability of the physical ability dimension of the Health Assessment Questionnaire. J Rheumatol 1990;17:813-7.
12. Brandão L. Qualidade de vida em artrite reumatóide: validação de uma versão do Arthritis Impact Measurement Scales II para a língua portuguesa (Brasil - AIMS-II) [thesis]. São Paulo: Escola Paulista de Medicina-UNIFESP; 1995.
13. Ciconelli RM. Tradução para o português e validação do questionário genérico de avaliação de qualidade de vida Medical Outcomes 36-Item Short-form Health Survey (SF-36) (Brasil-SF36). Rev Bras Reumatologia 1999;39:143-50.
14. Hochberg MC, Chang RW, Dwosh I, Lindsen S, Pincus T, Wolfe F. The American College of Rheumatology 1991 revised criteria for the classification of global functional status in rheumatoid arthritis. Arthritis Rheum 1992;35:498-502.
15. Felson DT, Anderson JJ, Bombardier C, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. Arthritis Rheum 1993;36:729-40.
16. Fried BJ, Boers M, Baker PRA, for the OMERACT Committee. A method for achieving consensus on RA outcome measures: the OMERACT conference process. J Rheumatol 1993;20:548-51.
17. Secherest L, Bay TL, Zaidi SMH. Problems of translation in cross-cultural research. J Cross-cultural Psychol 1972;3:41-56.
18. Ferraz MB. Cross-cultural adaptation of questionnaires: what is it and when should it be performed? J Rheumatol 1997;24:2066-8.
19. González VM, Stewart A, Ritter PL, Lorig K. Translation and validation of arthritis outcomes into Spanish. Arthritis Rheum 1995;38:1429-46.
20. Abello-Banfi M, Cardiel MH, Ruiz-Mercado R, Alarcón-Segovia D. Quality of life in rheumatoid arthritis: validation of a Spanish version of the Arthritis Impact Measurement Scales (Spanish-AIMS). J Rheumatol 1994;21:1250-5.
21. Perneger TV, Leplège A, Etter JF, Rougemont A. Validation of a French-language version of the MOS 36-Item Short Form Health Survey in young healthy adults. J Clin Epidemiol 1995;48:1051-60.
22. Ware JE, Gandek B, and the IQOLA Project Group. The SF-36 Health Survey: development and use in mental health research and the IQOLA project. Int J Mental Health 1994;23:49-73.
23. Sullivan M, Karlsson J, Ware J. The Swedish SF-36 Health survey. I. Evaluation of data quality, scaling assumptions, reliability and construct validity across general populations in Sweden. Soc Sci Med 1995;41:1349-58.
24. Goycochea-Robles MV, Garduño-Espinosa J, Vilchis-Guizar E, Ortiz-Alvárez O, Burgos-Vargas R. Validation of a Spanish version of the Childhood Health Assessment Questionnaire. J Rheumatol 1997;24:2242-5.
25. Arguedas O, Andersson Gare BA, Fasth A, Porras O. Development of a Costa Rican version of the Childhood Health Assessment Questionnaire. J Rheumatol 1997;24:2233-41.