

Effects of Method of Translation of Patient-Reported Health Outcome Questionnaires: A Randomized Study of the Translation of the Rheumatoid Arthritis Quality of Life (RAQoL) Instrument for Sweden

Peter Hagell, PhD,¹ Per-Johan Hedin, MD,² David M. Meads, MSc,³ Lennart Nyberg, PhD,⁴ Stephen P. McKenna, PhD³

¹Department of Health Sciences, Lund University, Lund, Sweden; ²Department of Rheumatology, Central Hospital, Falun, Sweden; ³Galen Research Ltd., Manchester, UK; ⁴Centre for Languages and Literature, Lund University, Lund, Sweden

ABSTRACT

Aims: To compare two versions of a questionnaire translated using forward-backward (FB) translation and dual-panel (DP) methodologies regarding preference of wording and psychometric properties.

Methods: The Rheumatoid Arthritis Quality of Life instrument was adapted into Swedish by two independent groups using FB and DP methodologies, respectively. Seven out of thirty resulting items were identical. Nonidentical items were evaluated regarding preference of wording by 23 bilingual Swedes, 50 people with rheumatoid arthritis (RA), and 2 lay panels (n = 11). Psychometric performance was assessed from a postal survey of 200 people with RA randomly assigned to complete one version first and the other 2 weeks later.

Results: Preference did not differ among the 23 bilinguals ($P = 0.196$), whereas patients and lay people preferred DP over FB item versions ($P < 0.0001$). Postal survey response rates were 74% (FB) and 75% (DP).

There were more missing item responses in the FB than the DP version (6.9% vs. 5.6%; $P < 0.0001$). Floor/ceiling effects were small (FB, 6.1/0%; DP, 4.4/0.7%) and reliability was 0.92 for both versions. Construct validity was similar for both versions. Differential item functioning by version was detected for five items but cancelled out and did not affect estimated person measures.

Conclusions: The DP approach showed advantages over FB translation in terms of preference by the target population and by lay people, whereas there were no obvious psychometric differences. This suggests advantages of DP over FB translation from the patients' perspective, and does not support the commonly held view that FB translation is the "gold standard."

Keywords: outcomes research, patient-reported outcomes, quality of life, questionnaire development, translation.

Introduction

There is general agreement that the process of translating questionnaires and rating scales from one language into another needs to be systematic and that the new language version should be evaluated with regard to wording and psychometric properties before it can be used with confidence [1]. A number of different and in part related procedures have been suggested for translation and cross-cultural adaptation of patient-reported health outcome questionnaires [2–6]. The most commonly used approach within the health sciences is forward-backward (FB) translation [7]. Essentially, with this approach one or several forward translations into the target language are produced by independent translators, followed by back-translation into the source language by (an)other translator(s). Differences in the forward- and back-translated versions are typically reconciled after each step. Alternative translation methods include the dual-panel (DP) approach, in which a consensus translation is produced by a panel of bilingual people native to the target language together with a representative of the developers of the adapted instrument. This is followed by review of the first translation by a second panel consisting of monolingual people of average or below average educational levels to ensure acceptability of wording and ease of completion.

Although the FB approach has been recommended by a number of authors [2,3,6,8], such recommendations are not based on empirical evidence and studies comparing different methods are sparse [7]. Perneger et al. [9] compared the psychometric properties of two French language versions of the SF-36 health survey using 946 young adults in a managed care plan in Geneva, Switzerland. A version produced by synthesizing three forward translations was compared with the official translation produced by an iterative FB translation process administered to the same sample 1 year after the first version. Results showed no systematic differences in psychometric performance between the two versions. Preference or ease of use among respondents was not assessed. In another study [10], the performance of forward translations only and FB translated versions of the SF-36, Health Assessment Questionnaire (HAQ) and the Arthritis Impact Measurement Scales 2 was compared among 50 people with rheumatoid arthritis (RA). Questionnaires were randomized by order and version, and interviewer administered before and after a medical consultation. Scores were compared descriptively and correlated with clinical variables. No major differences were observed, although the strengths of correlations with clinical variables often differed between the two versions. No assessments of preference among respondents or psychometric performance of the two versions were reported.

Given the sparse evidence-base for selecting between alternative methods of translation, there is a need for comparative studies before recommendations of any one specific method over another can be made [11,12]. This article reports the results of a randomized prospective study comparing two versions of a

Address correspondence to: Peter Hagell, Department of Health Sciences, Lund University, P.O. Box 157, SE-221 00 Lund, Sweden. E-mail: peter.hagell@med.lu.se
10.1111/j.1524-4733.2009.00677.x

questionnaire translated using either FB or DP methodologies regarding preference of wording and psychometric performance.

Methods

Instrument Translations

The Rheumatoid Arthritis Quality of Life instrument (RAQoL) is an RA-specific needs-based quality of life (QoL) questionnaire [13]. It consists of 30 items (statements), each with “Yes”/“No” response options scored 1 and 0, respectively. The total score ranges from 0 to 30 with high scores indicating poor QoL.

The RAQoL underwent translation and adaptation into Swedish by two independent groups using different methodologies [14,15]. The official Swedish version was produced by means of the DP approach [15]. First, a panel of six bilingual Swedes working together with one of the developers of the RAQoL (to ensure conceptual equivalence of the translation) produced a first draft version of the Swedish RAQoL. This version was then reviewed and revised by a second panel consisting of six monolingual Swedish lay people not suffering from RA. This was followed by face-to-face field-test interviews with 15 people with RA to assess wording and face validity. The interviewees reported no problems with the questionnaire and no changes were made following the field test [15]. The alternative Swedish RAQoL version was translated into Swedish by two independent authorized translators [14]. The two forward translations were then combined into one version by the authors, taking conceptual problems into consideration. This version was back-translated into English by a third authorized translator. Finally, it was assessed whether this Swedish version was easily understood by 10 people with RA. Patients found the RAQoL easy to understand and no changes to the questionnaire were reported [14].

It is important to note that the two research teams were completely independent. Furthermore, the decision to compare the two different Swedish versions was taken after both adaptations were complete. The two translation procedures resulted in seven identical Swedish RAQoL items, and both versions were considered conceptually equivalent (as judged by P. H., P. J. H., and L. N.).

Participants and Procedures

The two RAQoL versions were compared qualitatively and quantitatively. Qualitative evaluations were undertaken with three specific groups of people:

1. Two lay panels consisting of Swedes of average educational achievement.
2. Swedes who were bilingual in English and Swedish (advanced level students at the Department of English, Lund University, Sweden).
3. Swedish RA patients recruited consecutively from a Swedish rheumatology outpatient clinic.

The bilingual assessors had access to both Swedish RAQoL versions as well as the original UK version. The lay individuals and patients only had access to the two Swedish versions. All evaluators were instructed to consider their preference of wording of the 23 nonidentical item pairs with regard to ease of answering, appropriateness, ease, and ambiguity of language. Based on these considerations, they indicated their preferred version of each item. Items were presented in a neutral two-column table format with the two versions (FB and DP) appearing randomly in either of the two columns. Lay panels conducted the evaluation as a group exercise (as an additional task following the review of other

Swedish questionnaire translations), while bilinguals and patients provided individual evaluations.

Quantitative analyses were conducted by means of a repeated postal survey with 200 RA patients randomly selected from a Swedish rheumatology clinic. Those who had participated in the qualitative evaluation (see above) were excluded before patient selection. Patients were randomized to complete either the DP or the FB version first and the other version 2 weeks later. Both RAQoL versions appeared with identical layout in the respective questionnaire packages. In addition to the RAQoL, the survey included the HAQ, the Nottingham Health Profile (NHP), and demographic and RA-related questions. Only the RAQoL was included in both mailings. The NHP is a generic health status questionnaire that consists of 38 items representing six sections (Emotional Reactions, Sleep, Energy, Pain, Physical Mobility, and Social Isolation) [16,17]. NHP section scores are computed as a percentage score ranging between 0–100 (100 = worse). Embedded in the questionnaire is the NHP index of distress, a measure of illness-related distress (score range, 0–24; 24 = greater distress) [18,19]. The HAQ is a patient-reported rating scale that covers eight areas of daily activities [20,21]. The highest scores from each area are added together and divided by eight to derive the final HAQ score, which can range from 0 to 3 (3 = worse).

All patients provided written informed consent and the study was approved by the local ethics committee.

Analyses

Preference data comparing the two RAQoL versions were analyzed by means of chi-square tests. Comparisons of the overall preference were conducted by cross-tabulating data from all 23 nonidentical items (preference \times item version) for each of the three groups. A similar comparison was also made at the individual item level for the preference data from the bilinguals and people with RA, and the number of instances where preference for an item differed significantly between the two questionnaire versions was recorded. For the lay panels, who conducted the assessment as a group exercise, the number of instances when both panels preferred the same item version was recorded instead.

Postal survey RAQoL data were analyzed separately for the two questionnaire versions with respect to descriptive statistics, data quality (percentage of missing item responses; should be <10% [22]), floor and ceiling effects (should be <15% [23]), internal consistency reliability (Cronbach’s coefficient alpha; should be >0.7 and preferably >0.8 [24]), and construct validity. Construct validity was assessed by comparing (Kruskal–Wallis and Mann–Whitney *U*-tests) RAQoL scores between people perceiving their general health as excellent or good versus fair or poor; across levels of perceived RA severity (rated as mild, moderate or severe); and between people with and without a current flare-up of their arthritis. Fair/poor general health, more severe perceived disease, and current flare-up were hypothesized to be associated with higher RAQoL scores. Scores from the two RAQoL versions were also correlated (Spearman correlations) with each other as well as with HAQ and NHP scores. Correlations with the HAQ and NHP were compared with those in the original RAQoL report [13], as well as in previous reports on the two Swedish RAQoL versions [14,15].

In addition, the two questionnaire versions were analyzed regarding overall and item level fit to the Rasch model [25]. Overall fit was assessed by the chi-square based item-trait interaction statistic and item level fit was analyzed by analysis of variance (ANOVA) of the residuals (differences between

observed and expected responses) between people with different levels of QoL according to their RAQoL scores [26]. Additionally, Rasch analysis was used to identify any differential item functioning (DIF) by translation method. DIF is an aspect of fit to the Rasch model and occurs when people at comparable levels on the measured variable respond systematically differently to items, either in a uniform (responses differ uniformly regardless of people's location on the variable) or nonuniform (differences in responses vary across the variable) manner [26]. DIF analyses were conducted by means of a two-way ANOVA of the differences between observed and expected responses to DP and FB items across five QoL levels according to the RAQoL. Items with significant F -values ($P < 0.05$ following Bonferroni correction) were considered to have DIF by item version [26]. The practical significance of any observed DIF was assessed by testing whether DIF influenced the estimated person locations (logit measures). First, DIF was adjusted for by splitting items one by one (starting with the item displaying most DIF) into version specific items, until no significant DIF remained. The person locations obtained after adjustment for DIF were then compared to those estimated from the non-DIF-adjusted scale. Before doing so, items without significant DIF ($P > 0.05$ without Bonferroni adjustment) in the non-DIF-adjusted scale were anchored by their item locations from the DIF-adjusted scale to assure that the two sets of person estimates were on the same metric. The two sets of person locations were then plotted and correlated to assess the influence of DIF on people's estimated QoL measures. Finally, to explore further any differences in the two questionnaire versions, items in the DIF adjusted scale that displayed signs of misfit were examined and deleted one by one until no misfitting items remained. If this process identified any of the version specific (split) items as misfitting, this was interpreted as a psychometric disadvantage relative to the other version.

All analyses were conducted using SPSS 14 for windows (SPSS Inc., Chicago, IL) and RUMM 2020 (Rumm Laboratory Pty Ltd., Perth, Australia). The alpha level of significance was set to 0.05 (two-tailed, following Bonferroni correction).

Results

Preference of Wording

Qualitative evaluations were completed by 11 lay people (73% women; mean age, 48.7 years), 23 bilingual Swedes (78% women; mean age, 30.1 years), and 50 people with RA (78% women; mean age, 59.7). There was one case of missing preference data among the bilinguals (no preference indicated for one item by one participant) and 24 instances among people with RA (1–4 missing responses for 14 items). The reason(s) for these instances of missing data are unknown. Data were complete for the lay people. Overall preference for the 23 nonidentical RAQoL items (Fig. 1) did not differ among the 23 bilingual assessors (χ^2 , 1.674; $P = 0.196$). In contrast, lay people (χ^2 , 14.087) and patients (χ^2 , 17.059) preferred the DP over the FB item versions ($P < 0.0001$). Figure 2 shows the number of items for each questionnaire version that was significantly ($P < 0.05$) more often preferred over its comparator item. For example, significantly more patients preferred the DP over the FB versions of 10 items, with the opposite observed for five items. In the remaining instances, not depicted in Figure 2, there was no significant difference in the number of people preferring one version over the other. Ten DP items were preferred to their FB comparators and one FB item was preferred to its DP comparator by the lay panels; the remaining 12-item pairs were preferred by one panel each.

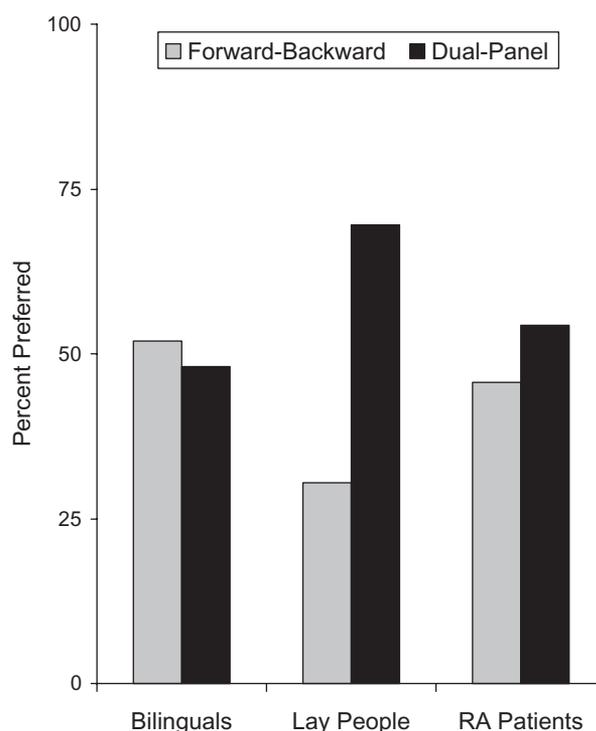


Figure 1 Overall preference for the dual-panel and forward-backward item versions of 23 nonidentical Swedish Rheumatoid Arthritis Quality of Life instrument (RAQoL) items of 23 bilinguals, two panels of lay people ($n = 11$) and 50 people with rheumatoid arthritis (RA). Bars show the percentage of instances that either version was preferred over the other within the respective groups. There was no difference in preference among bilingual assessors ($P = 0.196$) but lay people and patients preferred the dual-panel over the forward-backward item versions ($P < 0.0001$).

Psychometric Properties

A total of 175 out of 200 patients responded to the postal survey (88%), of whom 157 (79%) consented and 18 did not. Twenty-five patients chose not to respond. Of the 157 respondents, 142 responded to the second mailing after a mean of 17 days. Response rates for the two RAQoL versions were 74% (FB) and 75% (DP). RAQoL scores did not differ between responses from the first (median (q1–q3), 11 (4–17)) and second (11 (4–15)) mailing ($P = 0.528$; Wilcoxon's signed-rank test). Respondent characteristics are shown in Table 1.

Data quality was acceptable for both versions but better for the DP than the FB version (Table 2). Among the 23 nonequal RAQoL items, the proportion of missing item responses was larger for the FB than the DP versions of 21 items and equal for two item pairs. Total scores, floor/ceiling effects, and internal consistency did not differ between the two RAQoL versions and both were able to discriminate between respondents according to perceived health, RA severity, and whether or not they had a current flare-up of their arthritis (Table 2). Spearman and intra-class correlations between the two RAQoL versions were 0.87 and 0.88 (95% CI 0.83–0.91), respectively. Correlations between RAQoL scores and scores on the HAQ and NHP were very similar for the two versions (Table 3). With the possible exception of Social Isolation scores (NHP), these correlations were also similar to those reported previously for the original UK and the two Swedish RAQoL versions (Table 3).

Both questionnaire versions showed overall misfit to the Rasch model (FB: χ^2 , 128.4 (df, 90), $P = 0.004$; DP: χ^2 , 160.9 (df,

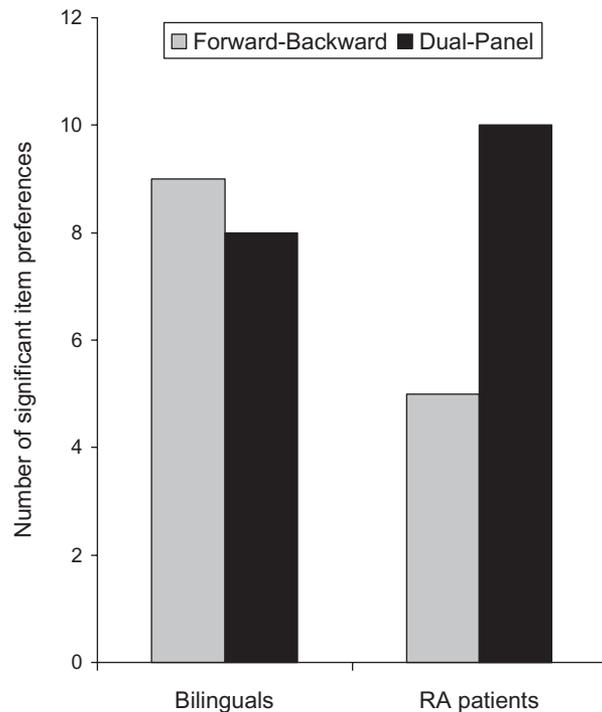


Figure 2 Number of items for each questionnaire version that was significantly ($P < 0.05$) more often preferred over its comparator item by 23 bilinguals and 50 people with rheumatoid arthritis (RA).

90), $P < 0.001$). At the item level, there was one misfitting item in the FB version ($P < 0.001$) and two misfitting items in the DP version ($P \leq 0.002$). Examination of DIF between the two versions displayed significant signs of uniform DIF for five items (Table 4) of which all had different wording in the two versions. These items continued to display DIF at each step during the stepwise process of splitting one item at the time and no additional DIF was detected during this process. The non-DIF-adjusted scale was then anchored on the DIF adjusted locations of 13 items without any signs of DIF (including four items with identical wording in the two questionnaire versions). Plots of estimated person measures derived from the scale with and without adjustment for DIF were virtually identical (mean difference, 0.006 logits) with Pearson and intraclass correlations of 1.0 (Fig. 3). Examination of item level fit in the DIF adjusted scale found five misfitting items, none of which had displayed

Table 1 Postal survey respondent characteristics (n = 157)

Male/Female, n (%)	46 (30%)/111 (70%)
Age (years), mean (SD)	62.3 (11.8)
RA duration (years), mean (SD)	16.2 (13.4)
HAQ, median (q1–q3)	0.88 (0.25–1.25)
NHP emotional reactions, median (q1–q3)	11.1 (0–33.3)
NHP Sleep, median (q1–q3)	20.0 (0–60.0)
NHP Energy, median (q1–q3)	33.3 (0–66.7)
NHP Pain, median (q1–q3)	25.0 (0–62.5)
NHP Physical Mobility, median (q1–q3)	25.0 (0–50.0)
NHP Social Isolation, median (q1–q3)	0 (0–20.0)
NHPD, median (q1–q3)	4 (1–8)

HAQ, Health Assessment Questionnaire; NHP, Nottingham Health Profile; NHPD, NHP index of distress; RA, rheumatoid arthritis; SD, standard deviation.

DIF. Stepwise deletion of these items did not result in any additional misfit.

Discussion

This is the first study comparing two commonly used methods of translating cross-cultural questionnaires. Results demonstrate that the DP methodology appears to be able to produce item wording that is perceived to be more acceptable by patients and lay people than FB translation. Nevertheless, the two methods do not seem to result in any major differences in psychometric performance of the resulting questionnaire versions.

The DP approach appears to have yielded wording that was preferred to that produced by FB translation. This was evidenced by preference data from representatives of the target patient population, as well as from lay people. Similarly, survey data quality was also better for the DP compared with the FB version of the questionnaire. Although data quality can be considered acceptable for both questionnaire versions [22], the almost consistent pattern of more missing item responses for the FB version may be seen as an indication of poorer respondent perceived quality of the translation [27]. It has been argued that there are a number of pitfalls associated with the use of FB translations that render the approach doubtful as a reliable method for quality control of the target questionnaire [5,28]. Because bilinguals tend to be better educated in general and linguistically more sophisticated than the general population, relying on bilingual people only in the translation process may tend to produce translations that differ somewhat from everyday language [1]. This, in turn, may affect how the resulting wording is perceived by respondents. In this respect, the DP approach of using lay people to assess the translation linguistically may be advantageous, because their task is to review the translation for ease and clarity of language [5]. Some support for this hypothesis was found in the present study as there was no difference in preferences for the two sets of RAQoL item versions among bilingual people, whereas patients and lay people generally favored DP item versions.

Table 2 Descriptive and psychometric postal survey RAQoL data (n = 157)

	RAQoL versions	
	Forward-backward	Dual-panel
Total score, mean (SD)	10.2 (7.4)	11.3 (7.5)*
Total score, median (q1–q3; range)	10.0 (4–16; 0–29)	12.0 (4–17; 0–30)†
Data quality (% missing item responses), mean (range per item)	6.9 (6.4–8.9)	5.6 (4.5–7.0)‡
Floor/ceiling effects (%)	6.1/0	4.4/0.7
Coefficient alpha	0.92	0.92
Perceived general health§		
Excellent/good (n = 46/43)¶	3 (1–6)	3 (1–6)
Fair/poor (n = 85/94)¶	12 (9–17)¶	14 (11–18)¶
Perceived RA severity§		
Mild (n = 40/34)¶	3 (1–8)	3 (1–7)
Moderate (n = 67/73)¶	15 (12–6)	15 (10–7)
Severe (n = 24/30)¶	18.5 (14–22)#	18 (14–23)#
Current symptom flare-up§		
No (n = 92/95)¶	10 (3–15)	7 (2–14)
Yes (n = 38/42)¶	13 (10–20)¶	13.5 (10–18)**

* $P = 0.080$ (paired t-test).

† $P = 0.151$ (Wilcoxon signed-rank test).

‡ $P < 0.001$ (Wilcoxon signed-rank test).

§Data are median (q1–q3).

¶n-values are for forward-backward/dual-panel RAQoL versions.

$P < 0.001$ (Mann–Whitney U-test).

** $P < 0.001$ (Kruskal–Wallis test).

*** $P = 0.003$ (Mann–Whitney U-test).

RAQoL, Rheumatoid Arthritis Quality of Life instrument; SD, standard deviation.

Table 3 Spearman correlations between the two RAQoL versions and scores on the HAQ and NHP

	Present study		Original UK RAQoL [13]	Previous studies	
	RAQoL-DP	RAQoL-FB		Original Swedish RAQoL-FB [14]	Original Swedish RAQoL-DP [15]
HAQ	0.76	0.78	—	0.67	0.79
NHP					
Emotional reactions	0.59	0.58	0.66	0.68	0.56
Sleep	0.59	0.52	0.59	0.38	0.30
Energy level	0.68	0.67	0.80	0.71	0.70
Pain	0.67	0.66	0.77	0.72	0.68
Physical mobility	0.76	0.75	0.87	0.77	0.74
Social isolation	0.49	0.42	0.62	0.48	0.53
NHPD	0.77	0.75	—	—	—

DP, dual-panel; FB, forward-backward; HAQ, the Health Assessment Questionnaire; NHP, Nottingham Health Profile; NHPD, NHP index of distress; RAQoL, the Rheumatoid Arthritis Quality of Life instrument.

Table 4 RAQoL items with uniform DIF by questionnaire version (DP vs. FB)^{*†}

Item no. (abridged content)	F-ratio [‡]	P-value	DIF direction [§]
6 (difficult walking to shops)	25.93	<0.001	FB > DP
17 (unable to join in activities with family/friends)	14.69	<0.001	FB > DP
23 (condition is always on my mind)	110.12	<0.001	DP > FB
27 (difficult taking care of people I am close to)	14.24	<0.001	DP > FB
28 (unable to control my condition)	34.64	<0.001	FB > DP

^{*}Performed with the sample divided into five class intervals according to person locations on the latent trait.

[†]Nonuniform DIF was not detected.

[‡]Two-way analyses of variance of deviations from model expectation along the latent trait between the two RAQoL versions.

[§]Direction of observed DIF; DP > FB indicates higher probability of item endorsement for the DP compared to the FB version, and vice versa.

DIF, differential item functioning; DP, dual panel; FB, forward-backward; RAQoL, Rheumatoid Arthritis Quality of Life instrument.

Differences in preference may also reflect an age difference (the bilingual evaluators were about 30 years old, whereas the lay people and patients were approximately 50 and 60 years old, respectively). Nevertheless, patients with chronic disease also tend to be older. Indeed, the DP approach requires lay panels to include people with a wide range of ages [5].

Many ideas and concepts can be expressed in different ways. When choices are available it is preferable to use wording that is perceived to be most acceptable, easy, and unambiguous by as many people as possible. To achieve this, the DP approach uses a lay panel that has the final say in the selection of alternative possible translations suggested by the bilingual panel and the ability to produce new forms of wording [5]. This may have been reflected in the results of this study.

It may be considered whether the relatively simple approach taken to assess wording in this study (i.e., asking participants to indicate their preferred version based on certain considerations) is an optimal one. However, this decision was made to keep participant burden at a minimum, particularly for the people with RA. Nevertheless, future studies could consider using a more detailed protocol.

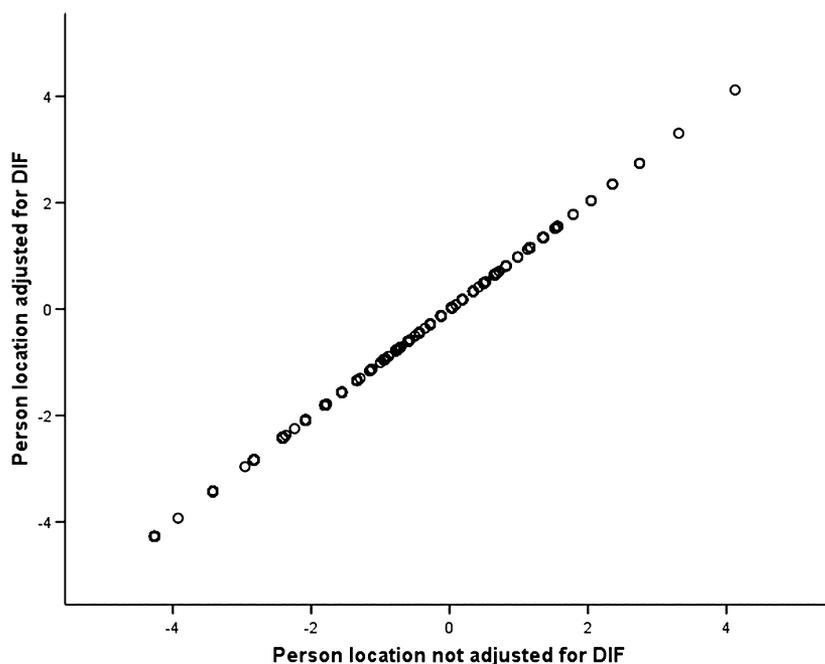


Figure 3 Impact of differential item functioning (DIF) by questionnaire version (dual-panel vs. forward-backward translation) on person measures, as Rasch analyzed from postal survey data from 157 people with rheumatoid arthritis (RA). The scatterplot depicts locations (logit measures) of each person estimated from the Rheumatoid Arthritis Quality of Life instrument (RAQoL) after adjustment for DIF by means of item split (y-axis) compared to those obtained from the original items not adjusted for DIF but anchored by DIF-free item calibrations from the DIF-adjusted scale (x-axis).

No clear psychometric advantages were found for one translation method over the other. This is in accordance with observations from previous studies comparing questionnaire versions translated according to different methods [9,10]. Although comparing different translation protocols, all three studies compared questionnaire versions that had been translated using an FB approach with versions that had been produced without the involvement of back-translation. This study therefore adds further doubt about the value added by back-translation.

There was DIF by questionnaire version on five RAQoL items. Nevertheless, the observed DIF cancelled out (i.e., DIF favoring one version of some items was balanced out by DIF favoring the other version of other items) and did not have any influence on the estimated person measures. This is an important observation as it suggests that choice of translation methodology would not affect the measures derived from the resulting questionnaire. Nevertheless, the degree to which this is the case is an empirical question and more studies are warranted before these observations can be generalized beyond the current questionnaire versions.

The present study concerned the translation of a questionnaire developed in the UK for use in Sweden. British and Swedish cultures can be considered to be relatively similar, which eases the task of achieving conceptual equivalence. Where a scale developed in Western Europe or North America is adapted for use in Africa or Asia the task is more challenging. Acceptability to potential patients then becomes even more important. In such circumstances it is possible that the advantages of the DP method might also be evident psychometrically.

The conduct of the present study was prompted by two factors. First, the need for more empirical comparisons of the influence of choice of translation protocol on outcome measures when adapted for use in a new language. Secondly, the unintended situation that presented itself when the RAQoL was being independently adapted for use in Sweden by two different groups. The advantage of this is that there was minimal investigator bias involved in the production of the two versions. That is, both are probably reflective of typical resulting target language questionnaire versions since the two groups were unaware of each other and of the upcoming conduct of this comparative study. Nevertheless, this may also pose limitations since the two approaches differed regarding aspects that go beyond the translation process itself. For example, whereas the DP version was pretested among 15 people with RA [15], the FB approach only used 10 people [14]—although FB protocols recommend pretesting with as few as five people [6]. It is possible that issues may have been identified in the FB version at this stage if more people had been included.

This is the first study designed to compare the preference in wording and measurement properties of a questionnaire translated for use in a new language according to different methodologies. In the case of the Swedish RAQoL versions, it was found that the DP approach showed advantages over FB translation in terms of preference by the target population and by lay people, whereas no obvious psychometric differences between the versions were found. This suggests potential advantages of the DP over the FB method from the patients' perspective. These advantages may result in higher item response rates, which would have important implications for the quality of data collected in clinical studies. Importantly, our observations do not support the commonly held view that FB translation is the "gold standard." Additional head-to-head comparisons using other scales, languages and target groups are required to allow fully generalizable conclusions to be drawn.

The authors want to thank all participating patients, bilingual and lay people for their cooperation.

Source of financial support: The study was supported by the Swedish Research Council, the Skane County Council Research and Development Foundation, and the Faculty of Medicine, Lund University.

References

- 1 Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use* (4th ed.). New York: Oxford University Press Inc., 2008.
- 2 Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186–91.
- 3 Bonomi AE, Cella DF, Hahn EA, et al. Multilingual translation of the Functional Assessment of Cancer Therapy (FACT) quality of life measurement system. *Qual Life Res* 1996;5:309–20.
- 4 Hunt SM, Alonso J, Bucquet D, et al. Cross-cultural adaptation of health measures. European Group for Health Management and Quality of Life Assessment. *Health Policy* 1991;19:33–44.
- 5 Swaine-Verdier A, Doward LC, Hagell P, et al. Adapting quality of life instruments. *Value Health* 2004;7(Suppl. 1):S27–30.
- 6 Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation Process for Patient-Reported Outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. *Value Health* 2005; 8:94–104.
- 7 Acquadro C, Conway K, Hareendran A, et al. Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. *Value Health* 2008;11:509–21.
- 8 Maneesriwongul W, Dixon JK. Instrument translation process: a methods review. *J Adv Nurs* 2004;48:175–86.
- 9 Perneger TV, Leplege A, Etter JF. Cross-cultural adaptation of a psychometric instrument: two methods compared. *J Clin Epidemiol* 1999;52:1037–46.
- 10 da Mota Falcao D, Ciconelli RM, Ferraz MB. Translation and cultural adaptation of quality of life questionnaires: an evaluation of methodology. *J Rheumatol* 2003;30:379–85.
- 11 Lenderking WR. Comments on the ISPOR Task Force Report on Translation and Adaptation of Outcomes Measures: guidelines and the need for more research. *Value Health* 2005;8:92–3.
- 12 McKenna SP, Doward LC. The translation and cultural adaptation of patient-reported outcome measures. *Value Health* 2005; 8:89–91.
- 13 de Jong Z, van der Heijde D, McKenna SP, et al. The reliability and construct validity of the RAQoL: a rheumatoid arthritis-specific quality of life instrument. *Br J Rheumatol* 1997;36:878–83.
- 14 Eberhardt K, Duckberg S, Larsson BM, et al. Measuring health related quality of life in patients with rheumatoid arthritis—reliability, validity, and responsiveness of a Swedish version of RAQoL. *Scand J Rheumatol* 2002;31:6–12.
- 15 Hedin PJ, McKenna SP, Meads DM. The Rheumatoid Arthritis Quality of Life (RAQoL) for Sweden: adaptation and validation. *Scand J Rheumatol* 2006;35:117–23.
- 16 Hunt SM, McKenna SP, McEwen J, et al. A quantitative approach to perceived health status: a validation study. *J Epidemiol Community Health* 1980;34:281–6.
- 17 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 Stud* 1988;10:159–63.
- 18 McKenna SP, Hunt SM, Tennant A. The development of a patient-completed index of distress from the Nottingham Health Profile: a new measure for use in cost-utility studies. *Br J Med Econ* 1993;6:13–24.
- 19 Wann-Hansson C, Klewsgard R, Hagell P. Cross-diagnostic validity of the Nottingham Health Profile Index of Distress (NHPD). *Health Qual Life Outcomes* 2008;6:47.
- 20 Ekdahl C, Eberhardt K, Andersson SI, et al. Assessing disability in patients with rheumatoid arthritis. Use of a Swedish version of the

- Stanford Health Assessment Questionnaire. *Scand J Rheumatol* 1988;17:263–71.
- 21 Fries JF, Spitz P, Kraines RG, et al. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137–45.
 - 22 Saris-Baglama RN, Dewey CJ, Chisholm GB, et al. SF Health Outcomes™ Scoring Software User's Guide. Lincoln: Quality-Metric, Inc., 2004.
 - 23 McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995;4:293–307.
 - 24 Nunnally JC, Bernstein IH. *Psychometric Theory* (3rd ed.). New York: McGraw-Hill, Inc., 1994.
 - 25 Andrich D. *Rasch Models for Measurement*. Beverly Hills, CA: Sage Publications, Inc, 1988.
 - 26 Andrich D, Sheridan B, Luo G. *Interpreting RUMM*. Perth: RUMM Laboratory Pty Ltd., 2004–2005.
 - 27 Ware JE Jr, Gandek B. Methods for testing data quality, scaling assumptions, and reliability: the IQOLA Project approach. *International Quality of Life Assessment. J Clin Epidemiol* 1998;51: 945–52.
 - 28 Leplege A, Verdier A. The adaptation of health status measures: methodological aspects of the translation procedure. In: Shumaker S, Berzon R, eds. *International Use and Performance of Health-Related Quality of Life Instruments*. Oxford: Oxford Rapid Communication, 1995.