Qual Life Res (2013) 22:1483–1490
DOI 10.1007/s11136-012-0276-6

An evaluation of the response category translations of the EORTC QLQ-C30 questionnaire

Neil W. Scott · Josephine A. Etta · Neil K. Aaronson · Andrew Bottomley · Peter M. Fayers · Mogens Groenvold · Michael Koller · Dagmara Kuliś · Debbi Marais · Morten A. Petersen · Mirjam A. G. Sprangers

Accepted: 20 September 2012 / Published online: 2 October 2012

Abstract
Purpose The aim of this study was to evaluate the translated response categories used in eight translations of the European Organisation for Research and Treatment of Cancer QLQ-C30 questionnaire, which is used in many international clinical trials. Twenty-eight of the 30 items in the questionnaire have the same four response categories: "Not at all", "A little", "Quite a bit" and "Very Much".
Methods People with knowledge of both English and one of eight European languages were asked to complete an Internet survey. The strength (intensity) of the translated categories was assessed using two separate approaches: one using a verbal response scale and the other a visual analogue scale (VAS).
Results Two hundred and seventy-nine people participated in the survey. Most translations were rated similarly to English. The largest differences were for the German translation of "Quite a bit", which was rated 16.3 points lower than the corresponding English category on a 0–100 VAS.
Conclusions Most of the translated categories were found to be similar to the English versions and should continue to be used. We recommend that three translated categories should be considered for revision. Similar surveys could be used to assess the categories used in other translated quality of life instruments.

This study is conducted on behalf of the EORTC Quality of Life Group.

N. W. Scott (✉) · P. M. Fayers
Medical Statistics Team, Division of Applied Health Sciences, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK
e-mail: n.w.scott@abdn.ac.uk

J. A. Etta
Division of Applied Health Sciences, University of Aberdeen, Aberdeen, UK

N. K. Aaronson
Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, The Netherlands

A. Bottomley · D. Kuliś
Quality of Life Department, European Organisation for Research and Treatment of Cancer Headquarters, Brussels, Belgium

P. M. Fayers
Department of Cancer Research and Molecular Medicine, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway

M. Groenvold · M. A. Petersen
Department of Palliative Medicine, Bispebjerg Hospital, Copenhagen, Denmark

M. Groenvold
Institute of Public Health, University of Copenhagen, Copenhagen, Denmark

M. Koller
Centre for Clinical Studies, University Hospital Regensburg, Regensburg, Germany

D. Marais
Public Health Nutrition, Division of Applied Health Sciences, University of Aberdeen, Aberdeen, UK

M. A. G. Sprangers
Department of Medical Psychology, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands

Keywords EORTC QLQ-C30 · Translations · Quality of life

Background

The European Organisation for Research and Treatment of Cancer (EORTC) QLQ-C30 is a questionnaire for assessing the quality of life of cancer patients [1, 2]. It has been widely used in many multicentre clinical trials and other national and international studies [3, 4]. The current version of the questionnaire (version 3.0) comprises 30 items, of which 28 use the same four response categories which measure the strength or intensity of agreement with a particular statement: "Not at all", "A little", "Quite a bit" and "Very much" [5].

As the EORTC QLQ-C30 is often used in cancer clinical trials and other studies involving more than one country, it is important that the various translated versions of the questionnaire are equivalent. The original instrument was developed in English with parallel development in other languages using input from a multidisciplinary international group [the EORTC Quality of Life Group (QLG)]. It has now been translated into over 80 language versions, and the EORTC QLG has developed a rigorous translation procedure using forward and backward translations in line with published guidelines for translations of patient reported outcomes [6–9].

The equivalence of the EORTC QLQ-C30 item translations was previously assessed using differential item functioning (DIF) analyses as well as a survey of bilingual individuals [10–12], but the translations of the four response categories have not yet been formally evaluated. As the same response categories are used for nearly all the items in the questionnaire, lack of translation equivalence in one of these categories could have a much larger impact on the cross-cultural validity of the instrument than poor equivalence for any given translation of an item. Such differences have the potential to introduce bias to the results of international clinical trials using the questionnaire, although it may not be possible to find category translations that are exactly equivalent to the English version.

Although it is becoming more common to use DIF analyses to assess the translations of health status questionnaires, there has been little research into response category translations in quality of life instruments [13]. Surveys have been used to assess the strength of various response categories in the development of the Short Form 36 (SF-36) and the United Kingdom World Health Organisation Quality of Life (UK WHOQOL) questionnaires [13, 14].

The aim of this study was to evaluate the wording of the response categories used in eight translated versions of the EORTC QLQ-C30 and to examine how closely equivalent these are to the English version.

A secondary aim was to assess whether the four response categories of the EORTC QLQ-C30 are considered to be equally spaced—this would be desirable given that each questionnaire subscale is derived by summing item scores [5].

Methods

Eight European languages were selected for this study: Danish, Dutch, French, German, Italian, Norwegian, Spanish and Swedish. These are among the most frequently used translations that have been used in many international studies using the EORTC QLQ-C30 and are languages commonly spoken by members of the EORTC QLG. These languages were also among those evaluated in our previous DIF studies focusing on the item translations [10, 11, 15].

Eight separate Internet surveys were conducted, one for each translation. The target population consisted of bilingual people defining themselves as having "good knowledge" of both English and one of the studied languages. A varied recruitment strategy was used. A message explaining the purpose of the study and with links to the eight online surveys was sent via e-mail to staff at the University of Aberdeen, members of the EORTC QLG and to personal contacts of the research team. University of Aberdeen students were also recruited using the "Message of the Day" service seen by students logging into the University system. Messages were posted targeting speakers of each of the studied languages and including a link to the relevant online survey. Student associations catering for different national groups were also asked to distribute the message. Recipients were asked to forward the e-mail to any person who they thought might be able to complete the survey (the "snowball" method). The EORTC QLG comprises researchers from all of Europe, many of whom were able to distribute the survey to English-speaking colleagues.

The survey was first pilot tested using both face-to-face and electronic administration. Around ten bilingual people were involved with pilot testing, and several monolingual people also gave advice about the format of the survey. Modifications to help improve understanding were made to the survey instructions based on the feedback received. The main study period was from April to May 2011. As participation was voluntary and the responses were anonymous, containing no identifying information, ethics approval was not required.

The survey measured the equivalence of the translations in two different ways. The first part of the survey used a seven-point verbal response scale (VRS) (Likert scale) to compare the response categories in the translated and

English categories. It comprised four questions with, for example, the following format: "Compared with the English phrase 'Not at all' the German phrase 'Überhaupt nicht' is: very much stronger/moderately stronger/slightly stronger/no difference/slightly weaker/moderately weaker/very much weaker". The respondents were then asked about the translations of the other three phrases: "a little", "quite a bit" and "very much" using the same seven-point scale.

The second part of the survey used 0–100 visual analogue scale (VAS) ratings to measure the strength (intensity) of these four response categories, in both the studied language and English. After giving examples of EORTC QLQ-C30 items in order to provide context for their answers, respondents were asked, "Please state how strong you think the following phrases are by choosing a point on a scale of 0–100 using the boxes below". A picture of a graded 0–100 scale was displayed to illustrate this concept (Fig. 1) but respondents had to type their responses into designated boxes. Respondents were also asked to assess where the three category thresholds would lie on the same 0–100 scale; the first of these was defined as the boundary below which "Not at all" would be chosen and above which "A little" would be chosen. Similarly, ratings were obtained for the boundaries between "A little" and "Quite a bit" and between "Quite a bit" and "Very much".

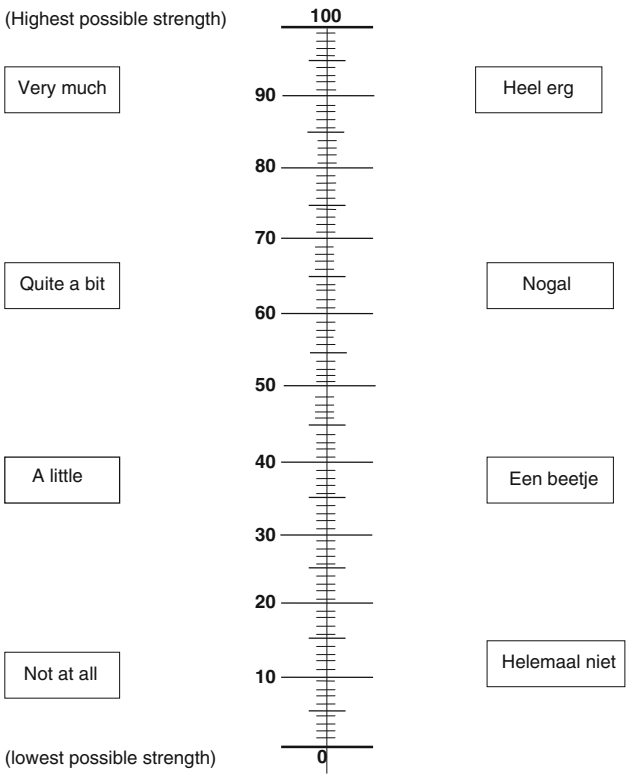

Fig. 1 An example of the graded VAS shown to respondents (English/Dutch)

Respondents were encouraged to write comments about each translated category and to make general comments about the survey. Suggestions for improved category translations are reported in the text if made by more than one respondent.

The results of Part 1 of the survey were analysed using the one-sample Wilcoxon signed rank test—this tested the null hypothesis that the median VRS rating was zero (no difference). For Part 2, the distributions of the differences between the English and translated VAS ratings were assessed and, as most of these were judged to be roughly normally distributed, paired t tests were used to compare means for each of the English and translated categories and thresholds. Because of the large number of tests performed, interpretation of the results used a combination of statistical significance and the magnitude of the observed differences. Translations with $p < 0.001$ and a difference of over five points on the 0–100 VAS were noted as candidates for possible revision. p values less than 0.05 are also indicated in the tables.

Results

Two hundred and seventy-nine people completed one of the eight surveys: the numbers of respondents ranged from 26 for Swedish to 49 for Dutch (Table 1). All 279 respondents completed the VRS ratings and were included in the first part of the survey, but only 212 (76 %) were included in the second part (VAS ratings). Some respondents did not complete this part, and the answers of others had to be excluded as they had clearly misunderstood this part of the survey. For example, several respondents had clearly used the ratings to indicate the equivalence of the English and translated categories, instead of the strength of an individual category on a 0–100 scale.

As the full results of Part 1 of the survey (VRS ratings) are too extensive to present here, two summaries are given instead (Tables 2, 3). Table 2 gives mean (SD) ratings for each translated category assuming the following codes to measure the translated version compared with the English translation: very much stronger (3), moderately stronger (2), slightly stronger (1), no difference (0), slightly weaker (−1), moderately weaker (−2), very much weaker (−3). Table 3 shows the proportion of ratings that were considered moderately or very much stronger or weaker, that is, ignoring differences considered slightly stronger or weaker.

The results for the second part of the survey (VAS ratings) are shown in Table 4. Combined ratings for English are given first using responses from all available raters, and then, separate results for both English and the studied language are presented for each of the eight surveys. Mean [standard deviation (SD)] ratings for both the four response

Table 1 Number of respondents included in each survey

Survey	Number of respondents *N* (%)		Translations evaluated			
	Included in Part 1 (VRS ratings)	Included in Part 2 (VAS ratings)	Not at all	A little	Quite a bit	Very much
Danish	36	24 (67)	Slet ikke	Lidt	En del	Meget
Dutch	49	41 (84)	Helemaal niet	Een beetje	Nogal	Heel erg
French	31	22 (71)	Pas du tout	Un peu	Assez	Beaucoup
German	47	39 (83)	Überhaupt nicht	Wenig	Mäßig	Sehr
Italian	28	17 (61)	No	Un po'	Parecchio	Moltissimo
Norwegian	27	25 (93)	Ikke i de hele tatt	Litt	En del	Svært mye
Spanish	35	28 (80)	En absoluto	Un poco	Bastante	Mucho
Swedish	26	16 (62)	Inte alls	Lite	En hel del	Mycket
Total	**279**	**212 (76)**				

Table 2 Mean (SD) VRS ratings comparing each translated category with English

Translation	Max *N*	Not at all	A little	Quite a bit	Very much
Danish	36	0.0 (0.9)	−0.1 (0.6)	−0.4 (1.1)*	−0.3 (1.3)
Dutch	49	0.2 (0.5)*	0.0 (0.5)	0.1 (1.1)	0.5 (0.8)*
French	31	0.1 (0.8)	0.1 (0.6)	−0.5 (1.4)	−0.4 (1.0)*
German	46	0.2 (0.6)	−0.2 (1.0)	−1.1 (1.2)**	−0.2 (0.7)
Italian	28	−0.7 (1.8)	0.1 (0.5)	0.8 (1.3)*	0.6 (1.3)*
Norwegian	27	0.1 (0.6)	−0.1 (0.6)	−0.5 (1.0)*	0.4 (0.9)*
Spanish	35	0.1 (1.4)	0.1 (0.8)	0.5 (1.4)	−0.3 (1.2)
Swedish	26	0.1 (1.0)	0.1 (0.8)	0.5 (1.1)*	−0.6 (1.4)*

The following coding was used: very much stronger (3), moderately stronger (2), slightly stronger (1), no difference (0), slightly weaker (−1), moderately weaker (−2), very much weaker (−3) than the English category

* $p < 0.05$; ** $p < 0.001$ (Wilcoxon signed rank test)

Table 3 Number (%) of VRS responses considered moderately or very much stronger/weaker than English

Translation	Max *N*	Not at all		A little		Quite a bit		Very much	
		Stronger	Weaker	Stronger	Weaker	Stronger	Weaker	Stronger	Weaker
Danish	36	3 (8)	2 (6)	0 (0)	2 (6)	1 (3)	5 (14)	5 (14)	3 (8)
Dutch	49	2 (4)	0 (0)	1 (2)	0 (0)	4 (8)	5 (10)	4 (8)	1 (2)
French	31	2 (6)	0 (0)	1 (3)	0 (0)	3 (10)	7 (23)	1 (3)	4 (13)
German	47	1 (2)	1 (2)	1 (2)	2 (5)	2 (5)	17 (40)	1 (2)	1 (2)
Italian	28	5 (19)	8 (31)	0 (0)	0 (0)	9 (32)	1 (4)	8 (29)	1 (4)
Norwegian	27	0 (0)	1 (4)	0 (0)	1 (4)	0 (0)	4 (15)	5 (19)	0 (0)
Spanish	35	5 (15)	4 (12)	2 (6)	0 (0)	7 (20)	2 (6)	3 (9)	4 (11)
Swedish	26	2 (8)	0 (0)	1 (4)	1 (4)	4 (15)	1 (4)	2 (8)	7 (27)

The columns labelled "Stronger" and "Weaker" include only those VRS ratings considered moderately or very much stronger/weaker but not those considered slightly stronger/weaker

Table 4 Mean (SD) VAS ratings of response categories

Translation	Max *N*	Not at all	Threshold 1	A little	Threshold 2	Quite a bit	Threshold 3	Very much
English (all respondents)	210	2.3 (4.2)	11.3 (8.3)	24.9 (8.3)	41.0 (11.5)	59.3 (11.5)	72.4 (11.1)	87.2 (8.6)
Danish	24	0.7 (1.7)	9.7 (6.5)	21.2 (7.7)	32.4 (11.1)*	50.7 (13.9)**	62.1 (13.1)*	81.6 (9.3)**
English	24	0.4 (1.4)	9.9 (8.2)	21.8 (7.9)	37.0 (10.3)	58.8 (14.9)	71.3 (10.5)	88.0 (11.4)
Dutch	41	1.5 (3.4)	11.0 (6.9)	24.0 (8.8)	47.5 (10.5)*	61.1 (11.6)	76.8 (8.9)*	90.3 (7.5)*
English	41	1.5 (3.2)	10.8 (6.9)	24.1 (8.8)	43.9 (10.7)	58.4 (10.8)	73.4 (9.5)	87.1 (9.0)
French	22	3.2 (5.9)	10.8 (7.4)	22.8 (7.9)	39.0 (11.9)	53.6 (6.6)	68.6 (8.7)	83.6 (7.9)
English	22	3.2 (4.8)	10.8 (7.5)	22.5 (9.9)	41.0 (12.2)	54.7 (13.8)	71.3 (12.7)	86.2 (7.9)
German	39	1.0 (3.0)	7.2 (5.5)	21.3 (7.9)	32.0 (8.5)*	45.0 (10.8)**	66.2 (10.7)*	85.3 (10.1)*
English	39	1.4 (3.1)	8.5 (6.4)	23.4 (8.2)	36.5 (9.5)	61.3 (12.0)	72.4 (8.7)	88.1 (8.9)
Italian	17	8.5 (3.8)	18.3 (8.0)*	30.9 (8.2)**	45.2 (13.3)*	64.1 (13.0)*	79.2 (15.3)*	93.8 (5.4)*
English	17	3.2 (5.0)	13.1 (9.2)	28.5 (7.9)	40.3 (13.5)	57.9 (11.9)	68.8 (10.5)	87.9 (7.3)
Norwegian	25	1.2 (2.9)	14.5 (11.9)	26.9 (5.7)	38.8 (13.9)	51.2 (11.2)*	72.1 (16.3)	88.4 (6.6)*
English	25	2.1 (3.9)	12.2 (10.5)	26.7 (6.4)	40.0 (14.3)	58.9 (8.9)	71.8 (14.9)	85.2 (7.1)
Spanish	28	2.8 (6.5)	13.9 (10.2)	27.5 (7.2)	48.7 (14.4)*	66.9 (11.7)	72.7 (12.4)	83.2 (10.6)*
English	28	4.2 (6.2)	15.6 (10.3)	27.5 (7.4)	43.9 (12.0)	62.7 (8.9)	72.0 (13.1)	87.9 (8.8)
Swedish	16	3.9 (4.1)	11.7 (6.5)	25.6 (10.9)	46.4 (10.3)	62.0 (10.1)	74.4 (8.1)*	82.9 (9.7)
English	16	4.1 (4.9)	12.6 (6.7)	27.2 (8.4)	45.7 (7.6)	59.4 (9.1)	77.5 (8.5)	87.0 (7.7)

0 represents the lowest possible strength, 100 the highest possible strength. Thresholds 1–3 indicate the boundaries between "Not at all" and "A little", "A little" and "Quite a bit", and "Quite a bit" and "Very much", respectively

* $p < 0.05$; ** $p < 0.001$ (paired t test between translated version and English)

categories and the three thresholds between categories are shown.

The following sections discuss the results for each translation in more detail.

Danish

There was a slight suggestion from the VRS ratings that the translations of "Quite a bit" ("En del") and "Very much" ("Meget") were weaker than English (Table 2), although more people (5/36 vs. 3/36) rated the latter category as moderately/very much stronger than very moderately/very much weaker (Table 3). Mean differences in the VAS ratings were 8.1 and 6.4 for "Quite a bit" and "Very much", respectively, again suggesting that these translated labels were weaker than English (Table 4).

There were six suggestions of "En hel del" as an alternative translation of "Quite a bit" and five suggestions of "Rigtig meget" for "Very much", although one of these respondents stated that "Meget" was still a good translation.

Dutch

There were relatively small differences between the English and Dutch translated categories. The results for "Quite a bit" were the most variable, but respondents did not agree whether this was stronger or weaker than English (Table 3). Two people suggested "Een klein beetje" for the translation of "A little".

French

Ratings for the English and French translations were similar, and there was no evidence of differences for any of the category translations or thresholds.

German

The translation of "Quite a bit" ("Mäßig") had the largest difference of any category. The mean VRS rating was −1.1 indicating the German translation was weaker than English with 17/42 respondents (40 %) rating this as moderately or very much weaker (Table 3). There was a large difference of 16.3 points (45.0 vs. 61.3) in the VAS ratings for this category, and the threshold ratings on either side of this also seemed to be affected (Table 4). The other German translations seemed to be reasonably similar to English.

"Ein bisschen" and "Ein wenig" were both suggested several times as alternative translations of "A little", "Ziemlich" and "Ziemlich viel" were suggested for "Quite a bit," and

"Sehr viel" was suggested instead of "Sehr" for the translation of "Very much".

Italian

The most variable category examined was the translation of "Not at all" ("No"). Only one person thought there would be no difference between English and Italian, but the responses were inconsistent as to whether this would be stronger (7/26) or weaker (18/26) (data not shown). This was also reflected in higher VAS ratings for this category, but there was no evidence of a difference between English and Italian using the paired *t* test (Table 4). The translations of "Quite a bit" ("Parecchio") and "Very much" ("Moltissimo") were also considered stronger than English in the VAS ratings.

Norwegian

Overall, the results for Norwegian were similar to English, but there was a seven-point difference in the ratings for "Quite a bit" ("En del") (Table 4). Although all but four rated it as "slightly weaker" (data not shown), 16/27 people rated this category as weaker than English. Three people suggested that "En god del" might be closer to "Quite a bit" than the existing translation.

Spanish

The VRS results showed a spread of ratings with opinion divided as to whether categories were stronger or weaker than English. There were moderate differences in VAS ratings for two of the categories: "Bastante" was rated as four points stronger than the English "Quite a bit" and "Mucho" four points weaker than "Very much" (Table 4). Four people suggested the translation "Muchísimo" for "Very much".

Swedish

Differences between the English and Swedish categories were not large, but there was a 4.1 point difference for the translation of "Very much" ("Mycket") (Table 4). Seven people mentioned that this category could be translated as "Väldigt mycket".

Spacing of categories

Using the combined VAS results from all eight surveys (Max $N = 210$), the four English response labels had mean ratings of 2.3, 24.9, 59.3 and 87.2, respectively, on the 0–100 scale (Table 4). This implies that these four labels can be considered to be reasonably equally spaced, as assumed by the sum scoring used to derive the EORTC QLQ-C30 subscales. The results for the German response categories were less evenly spaced: 1.0, 21.3, 45.0 and 85.3, respectively. Mean ratings for the three English category thresholds were 11.3, 41.0 and 72.4.

Discussion

Our study found that most translations were rated similarly to the original English version. This is encouraging as all these versions of the EORTC QLQ-C30 have been used for many years in clinical trials of cancer treatments.

We chose to collect the information on agreement using two separate approaches (using both a VRS and a VAS). There was reasonable agreement between Parts 1 and 2 of the survey. The Internet survey generally worked well and enabled collection of a reasonable sample size in a short period of time. A disadvantage of electronic administration was that not all participants appeared to understand the instructions to Part 2 and the use of the word "strength" in the instructions to this part may have been confusing; comprehension might have been improved with more pilot testing. Some respondents also reported that special characters (diacritics) were not always displayed correctly on their computer screen.

The nature of the "snowball" sampling meant that we could not guarantee similar numbers for each translation. We did not prespecify a sample size but aimed to obtain results for 30 respondents for each language. Although final numbers were lower, especially for Part 2 of the survey, we judged that sufficient results were available for all eight languages for presentation in this paper.

No demographic information was collected about the respondents, so it is not possible to distinguish between native speakers of English and native speakers of the target language. We were therefore unable to assess respondents' level of familiarity with each language—in retrospect, it would have been interesting to include questions on fluency in both English and the target language. We also do not know the age or socioeconomic status of the respondents, but it is likely that many may have been from an academic or clinical background and therefore be more likely to be familiar with this type of questionnaire than members of the general population. We do not, however, believe that there would be major differences by age or socioeconomic status in how common phrases such as "Very much" are understood and rated.

It could be argued that the "Not at all" category implied a strength of zero on the 0–100 VAS, but some respondents rated this higher. Most translations of this category were rated similarly to English. Although the ratings for Italian were variable, perhaps reflecting the broader meaning of

"No" in Italian compared with the English "Not at all", there was no consensus on whether this was stronger or weaker than English.

The category "A little" was usually rated similarly to English with a maximum difference of 2.54 points on the 0–100 VAS (Table 4). In fact, there was more variation (over six points) in how the English label "A little" was rated with speakers of Danish, French and German rating this relatively low and speakers of Italian, Spanish and Swedish rating this higher.

The largest differences were for the translations of "Quite a bit", suggesting that this phrase was the hardest to render accurately in other languages. Differences of over three points in the VAS ratings were found for Danish, Dutch, Italian, Norwegian and Spanish, but the most extreme results were for the German translation. This had a mean VRS rating of −1.1 and a VAS rating that was rated 16.3 points weaker than English. The translation "Quite a bit" appeared to be the most difficult phrase to translate as this seems to be an expression with no direct equivalent in other languages. A similar finding was found with the category "A good bit of the time" which was dropped during the development of the SF-36 questionnaire [13]. The ratings for "Very much" were the second most variable.

The ratings boundaries are important as they determine the cut-offs when respondents choose between categories, but it appeared to be quite difficult for some of those completing our survey to understand this concept and rate these thresholds and there tended to be greater variability between raters for the threshold ratings compared with the categories themselves. There were relatively few between-group differences between the ratings for the first two thresholds, but Danish, German and Italian were substantially different to English for the threshold between "Quite a bit" and "Very much".

Three category translations were associated with a p value less than 0.001 using the paired t test and also had VAS differences of over five points and therefore meet the criteria for possible investigation or revision. In each case, alternative translations were suggested by survey respondents and these could be tested in a subsequent study.

There was strong evidence that the German translation of "Quite a bit" ("Mäßig") was considerably weaker than the English version. Concerns had been raised about this particular translated category, usually translated into English as "Moderately", at past QLG meetings. We will be undertaking further work to explore alternative translations of this category. Interestingly, a previous study evaluating the intensity of 18 German response categories on a 0–10 scale found few options that would correspond well with "Quite a bit" with a gap of 1.3 points in this area of the scale between "Einigermaßen" (5.4) and "Ziemlich" (6.7) [16]. Two further studies reported in the same article showed mean ratings of 6.0 and 6.4 for "Ziemlich", which would be closer to the ratings that we obtained for "Quite a bit".

Two Danish translations, "En del"/"Quite a bit" and "Meget"/"Very much", were also associated with moderately large VAS differences (8.1 and 6.4 points, respectively), although the evidence from the VRS ratings was less clear. As these are adjacent response categories, it is possible that ratings for one may have influenced the other.

When considering the three Scandinavian languages (Danish, Norwegian and Swedish), an interesting pattern emerges. With no available phrase exactly matching the strength of the English categories, when translating "Quite a bit" and "Very much", the original translators had a choice of possible options. For example, for "Quite a bit", all three languages have the same choice of using "En del" or a stronger translation with an intensifier ("En hel del"). The latter translation was chosen for Swedish, and this was judged to be stronger than English using both VRS and VAS ratings, whereas "En del", chosen for Danish and Norwegian, was considered weaker than English (Table 5). For "Very much", all three Nordic languages used a cognate word ("Meget", "Mye" or "Mycket"), but again the only translation to qualify this with an intensifier (Norwegian: "Svært mye") was judged to be stronger than English. Overall, these results form a natural experiment that highlights the difficulty of finding an exact match for the English categories and the dilemma of choosing between a translation that is too strong and one that is too

Table 5 A comparison of the three Scandinavian languages [mean VRS score (−3 to 3), mean VAS rating (0–100)]

English	Quite a bit		Very much	
	Weaker translation (without intensifier)	Stronger translation (with intensifier)	Weaker translation (without intensifier)	Stronger translation (with intensifier)
Danish	**En del (−0.4, 50.7)**	*En hel del*	**Meget (−0.3, 81.6)**	*Rigtig meget*
Norwegian	**En del (−0.5, 51.2)**	*En hel del*	*Meget/Mye*	**Svært mye (0.4, 88.4)**
Swedish	*En del*	**En hel del (0.5, 62.0)**	**Mycket (−0.6, 82.9)**	*Väldigt mycket*

Translations in bold are currently used in the EORTC QLQ-C30—for these mean VRS scores (Table 2) and mean VAS ratings (Table 4) are displayed in brackets. Translations in italics are alternative response categories that could have been chosen

weak. In this case, it might be better to choose categories that are consistently stronger or weaker than English.

Although the rating for "Quite a bit" was lower than might be expected, there was some support for the equidistance of categories used in the EORTC QLQ-C30. It is questionable as to whether this matters as there is some evidence that those completing VRSs in questionnaires assume that categories are equally spaced and choose their responses regardless of the labels used [17, 18]. Other research has shown the importance of the wording of response categories in surveys as respondents tend to assume that middle categories represent typical behaviour [19]; this is important for the EORTC QLQ-C30 as data from both cancer patients and the general population has shown variation in the distributions of the 30 items with some showing marked floor effects [20]. Our current study also used a VRS and made an assumption of equidistance when calculating means in Table 2. Unlike previous studies [13, 14], we did not treat the extreme categories as anchor points as we were interested in examining the equivalence of all four response choices.

We believe that this type of survey evaluation is relatively easy to conduct and could be used more often in health outcomes and quality of life research to evaluate the equivalence of translated response labels. Although the best time to conduct this type of study would be during the development of a questionnaire and its translations, it is still useful to evaluate existing translations to determine whether they should be revised.

References

1. Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., et al. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute, 85*, 365–376.
2. Aaronson, N. K., Cull, A. M., Kaasa, S., & Sprangers, M. A. G. (1996). The European Organization for Research and Treatment of Cancer (EORTC) modular approach to quality of life assessment in oncology: An update. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (pp. 179–189). Philadelphia: Lippincott-Raven.
3. Fayers, P., Bottomley, A., & EORTC Quality of Life Group, and EORTC Quality of Life Unit. (2002). Quality of life research within the EORTC-the EORTC QLQ-C30. *European Journal of Cancer, 38*, S125–S133.
4. Garratt, A., Schmidt, L., Mackintosh, A., & Fitzpatrick, R. (2002). Quality of life measurement: Bibliographic study of patient assessed health outcome measures. *BMJ, 324*, 1417.
5. Fayers, P., Aaronson, N., Bjordal, K., Groenvold, M., Curran, D., & Bottomley, A. (2001). *EORTC QLQ-C30 scoring manual*. Brussels: European Organization for Research and Treatment of Cancer.
6. Cull, A., Sprangers, M., Bjordal, K., Aaronson, N., West, K., & Bottomley, A. (2002). *EORTC Quality of Life Group translation procedure*. Brussels: European Organization for Research and Treatment of Cancer.
7. Koller, M., Aaronson, N. K., Blazeby, J., Bottomley, A., Dewolf, L., Fayers, P., et al. (2007). Translation procedures for standardised quality of life questionnaires: The European Organisation for Research and Treatment of Cancer (EORTC) approach. *European Journal of Cancer, 43*(12), 1810–1820.
8. Kulis, D., Arnott, M., Greimel, E. R., Bottomley, A., & Koller, M. (2011). Trends in translation requests and arising issues regarding cultural adaptation. *Expert Review of Pharmacoeconomics & Outcomes Research, 11*, 307–314.
9. Wild, D., Grove, A., Martin, M., et al. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translating adaptation. *Value Health, 2*, 94–104.
10. Scott, N. W., Fayers, P. M., Bottomley, A., Aaronson, N. K., de Graeff, A., Groenvold, M., et al. (2006). Comparing translations of the EORTC QLQ-C30 using differential item functioning analyses. *Quality of Life Research, 15*, 1103–1115.
11. Scott, N. W., Fayers, P. M., Aaronson, N. K., Bottomley, A., de Graeff, A., Groenvold, M., et al. (2007). The use of differential item functioning analyses to identify cultural differences in responses to the EORTC QLQ-C30. *Quality of Life Research, 16*, 115–129.
12. Scott, N. W., Fayers, P. M., Aaronson, N. K., Bottomley, A., de Graeff, A., Groenvold, M., et al. (2010). Interpretation of differential item functioning (DIF) analyses using external review. *Expert Reviews in Pharmacoeconomics and Outcomes Research, 10*, 253–258.
13. Keller, S. D., Ware, J. E., Jr, Gandek, B., Aaronson, N. K., Alonso, J., Apolone, G., et al. (1998). Testing the equivalence of translations of widely used response choice labels: Results from the IQOLA project. *Journal of Clinical Epidemiology, 51*, 933–944.
14. Skevington, S. M., & Tucker, C. (1999). Designing response scales for cross-cultural use in health care: Data from the development of the UK WHOQOL. *British Journal of Medical Psychology, 72*, 51–61.
15. Scott, N. W., Fayers, P. M., Aaronson, N. K., Bottomley, A., de Graeff, A., Groenvold, M., et al. (2009). Differential item functioning (DIF) in the EORTC QLQ-C30: A comparison of baseline, on-treatment and off-treatment data. *Quality of Life Research, 18*, 381–388.
16. Rohrmann, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie, 9*, 222–245.
17. Finn, R. H. (1972). Effects of some variations in rating scale characteristics on the means and reliabilities of ratings. *Educational and Psychological Measurement, 32*, 255–265.
18. Spector, P. E. (1980). Ratings of equal and unequal response choice intervals. *Journal of Social Psychology, 112*, 115–119.
19. Schwarz, N. (1990). What respondents learn from scales: The informative functions of response alternatives. *International Journal of Public Opinion Research, 2*, 274–285.
20. Scott, N. W., Fayers, P. M., Aaronson, N K., Bottomley, A., de Graeff, A., Groenvold, M. et al. on behalf of the EORTC Quality of Life Group. (2008). *Reference values manual*. Brussels: European Organization for Research and Treatment of Cancer.