See related commentary on pg 2299

ORIGINAL ARTICLE

Cross-Cultural Inequivalence of Dermatology-Specific Health-Related Quality of Life Instruments in Psoriasis Patients

Tamar Nijsten[1], David M. Meads[2], John de Korte[3], Francesca Sampogna[5], Joel M. Gelfand[4], Katia Ongenae[6], Andrea W. Evers[7] and Matthias Augustin[8]

The dermatology life questionnaire index (DLQI) and the Skindex are the most commonly used dermatology-specific health-related quality of life (HRQOL) instruments. Although these tools are used in international surveys and clinical trials, the cross-cultural equivalence of their items has not been documented. We used differential item functioning (DIF), which is part of the Rasch model, to assess the impact of cultural background on the items of the DLQI and Skindex-29 and-17. The data of the 450 psoriasis patients, who attended in- and outpatient dermatology centers, was collected retrospectively from five European and one US center. The DLQI and Skindex-29 scales did not fit the Rasch model ($P<0.0008$) and 10/10 of the DLQI and 19/29 of the Skindex-29 items displayed significant DIF. Although the psychosocial scale of the Skindex-17 fitted the Rasch model, half or more of the items of the psychosocial (6/12) and the symptom scale (4/5) showed significant DIF across countries. These findings suggest that psoriasis patients from different countries respond differently to a substantial proportion of DLQI and Skindex items despite having the same level of underlying HRQOL impairment. Therefore, these instruments should not be used in their current form in international studies.

Journal of Investigative Dermatology (2007) **127,** 2315–2322; doi:10.1038/sj.jid.5700875; published online 10 May 2007

INTRODUCTION

Patient reported outcomes such as health-related quality of life (HRQOL) are increasingly important in the assessment of disease severity and evaluating the effect of interventions. In dermatology, HRQOL impairment is pivotal because most skin diseases are chronic and intermittent but not life threatening or physically disabling. Moreover, several studies have shown that clinical disease severity does not necessarily correlate well with patients' perspectives of disease severity and HRQOL impairment (Sampogna *et al.*, 2004).

[1]*Department of Dermatology, Erasmus MC, Rotterdam, The Netherlands;* [2]*Galen Research, Manchester, UK;* [3]*Department of Dermatology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands;* [4]*Health Services Research Unit, Istituto Dermopatico dell'Immacolata IDI-IRCCS, Rome, Italy;* [5]*Department of Dermatology and Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, Pennsylvania, USA;* [6]*Department of Dermatology, University Hospital of Ghent, Ghent, Belgium;* [7]*Department of Medical Psychology, St. Radboud University, Nijmegen, The Netherlands and* [8]*Klinik fur Dermatologie und Venerologie, Universitats-Klinikum Hamburg-Eppendorf, Hamburg, Germany*

Correspondence: Dr Tamar Nijsten, Department of Dermatology, Erasmus MC, PO Box 2040, 3000 CA Rotterdam, The Netherlands.
E-mail: t.nijsten@erasmusmc.nl

Abbreviations: ANOVA, analysis of variance; DIF, differential item functioning; DLQI, dermatology life questionnaire index; HRQOL health-related quality of life; ICC, item characteristic curve

Received 12 December 2006; revised 9 February 2007; accepted 12 March 2007; published online 10 May 2007

The dermatology life questionnaire index (DLQI) (Finlay and Khan, 1994) and Skindex-29 (Chren *et al.*, 1997) are the most commonly used dermatology-specific HRQOL instruments. Most large dermatological clinical trials are multinational and include multiple sites located in 10 or more countries and use the change in DLQI score as an end point. In contrast to widely used HRQOL tools such as the Short Form-36 and -12, WHOQOL-100, EuroQoL-5D (Wagner *et al.*, 1998; Gandek *et al.*, 1998; Skevington, 2002; Prieto *et al.*, 2003), no cross-cultural equivalence studies of the DLQI and Skindex-29 have been performed to test whether items' responses are independent of cultural background or not. A recent validation study of the Skindex-17 (Nijsten *et al.*, 2006a) in psoriasis patients from three European countries demonstrated that nationality did affect how patients, with the same amount of psychosocial impairment, responded to some items (Both *et al.*, submitted) confirming that it is important to test for cultural differences in dermatology-related HRQOL tools as well before they are used in multinational studies (Anderson *et al.*, 1993, 1996; Bullinger *et al.*, 1993a; McHorney and Fleishman, 2006).

The objective of this analysis was to use differential item functioning (DIF) (Angoff, 1993; McHorney and Fleishman, 2006), which is set within the framework of the Rasch unidimensional measurement, to study the effect of cultural background (e.g., nationality) on retrospectively collected DLQI and Skindex data from five European countries and the

US. It was not the intention of this study to adjust these instruments for potential cultural differences or to differentiate between a suboptimal translational process and true measurement unequivalence.

RESULTS

Study population

For each country, 75 psoriasis patients were included, except for the Skindex-29 in the US ($n=72$). Table 1 presents the origin of the study patients, the nationality of the psoriasis patients, and the mean scores of the DLQI and the Skindex-29. Patients from Germany and Ireland reported greater impact of psoriasis on their lives than those from other countries. Of the included patients, the mean score of the DLQI was 7.1 (SD 6.4, interquartile range: 1.0–13.0) and 32.9 (SD 23.8, interquartile range: 13.0–48.0) for functioning, 46.7 (SD 24.9, interquartile range: 28.0–63.0) for emotions, and 54.5 (SD 23.3, interquartile range: 39.0–71.0) for symptom's scale of the Skindex-29. The age and gender distribution between countries was similar (male percentage ranged between 44.0 and 54.5% and the mean age ranged between 43.8 and 47.1 years, respectively), except that German psoriasis patients were younger (mean age of 38.6 and 37.2 years).

Dermatology life quality index

Overall, the DLQI significantly misfitted the Rasch model (mean item interaction 0.00, SD 0.45; mean person interaction –1.29, SD 1.28; item–trait interaction, $P<0.0001$) but had an excellent person separation index of 0.88. After stratification by country, the DLQI misfitted the Rasch model in half the countries (Table S1). After combining all countries, half of the items demonstrated significant individual misfit (items 1, 3, 5, 8, and 10), three items had reversed thresholds but no DIF was seen across gender and age. All items had uniform DIF and four items also had nonuniform DIF across cultures (Table S2). The item characteristic curve (ICC) of item 1 shows that Belgian, Dutch, and Italian patients reported more psoriasis-related skin symptoms than expected and than those from US, UK, and Germany (Figure 1). Italians and Belgians experienced significantly less often ''problems with loved ones'' (item 8) and ''sexual difficulties'' (item 9) because of their psoriasis compared with remaining patients with similar levels of disability measured by the DLQI.

For the 10 DLQI items, psoriasis patients from specific countries were not persistently scoring higher than others (data not shown). However, the ICC of Italians, Belgians, and often the Dutch were closely aligned in majority of the items (8/10) both below and above the expected curve as were those of the UK and US. On average, the German ICC was either more in the middle or close to those of the English-speaking patients.

Skindex-29

Overall, the items of the Skindex-29 fitted the Rasch model (item fit mean 0.00, SD 0.52) but the persons did not (person fit mean –0.47, SD 0.94), which resulted in a significant item–trait interaction statistic (Table S3). The person separation index was excellent (0.95). Of the 29 items, seven showed significant item misfit, nine threshold

Table 1. Countries of residence of patients included for the DLQI and Skindex-29 scales and the mean scores[1]

	DLQI		Skindex-29			
Country of residence	Mean score (SD, 25th–75th percentile)	Reference	Functioning mean score (SD, 25th and 75th percentile)	Emotions mean score (SD, 25th and 75th percentile)	Symptoms mean score (SD, 25th and 75th percentile)	Reference
Belgium	6.9 (5.6, 3.0–11.0)	Ongenae *et al.*, 2005[2]	—[3]	—[3]	—[3]	—[3]
Germany	17.7 (6.6, 13.0–23.0)	Augustin *et al.*, 1999[2]	33.2 (29.0, 17.0 and 48.0)	46.8 (25.1, 25.0 and 68.0)	53.8 (24.0, 36.0 and 71.0)	Augustin *et al.*, 2004[2]
Ireland	—[3]	—[3]	47.6 (20.7, 31.0 and 55.5)	62.0 (23.1, 48.0 and 79.5)	75.3 (16.6, 64.0 and 89.0)	de Korte *et al.*, 2005[4]
Italy	7.4 (5.1, 4.0–10.0)	Sampogna *et al.*, 2004[2]	35.8 (23.9, 17.0 and 50.0)	47.4 (22.9, 28.0 and 65.0)	47.7 (21.4, 32.0 and 57.0)	Sampogna *et al.*, 2004[2]
The Netherlands	8.0 (6.3, 3.0–11.0)	Evers *et al.*, 2005[2]	21.7 (18.4, 8.0 and 29.0)	33.4 (17.5, 20.0 and 45.0)	53.2 (18.5, 43.0 and 64.0)	de Korte *et al.*, 2005[4]
UK	11.7 (5.9, 6.0–16.0)	McKenna *et al.*, 2003[2]	23.6 (21.7, 4.0 and 40.0)	40.4 (23.2, 23.0 and 58.0)	53.7 (21.4, 39.0 and 68.0)	de Korte *et al.*, 2005[4]
USA[1]	13.6 (5.0, 12.0–17.0)	McKenna *et al.*, 2005[2]	33.0 (26.3, 10.0 and 54.0)	50.6 (27.7, 30.0 and 73.0)	44.1 (21.7, 29.0 and 57.0)	Joel Gelfand[2,5]

DLQI, dermatology life questionnaire index; SD, standard deviation.
[1]Each country contributed 75 randomly selected psoriasis patients to the analyses, except US, which contributed 72 patients for the Skindex. The Skindex-17 was derived from the Skindex-29 instrument. Range of DLQI is 0–30 and Skindex-29 is 0–100.
[2]Patient recruitment at university centers.
[3]No data was obtained for these countries.
[4]Patient recruitment at treatment centers.
[5]Unpublished data.

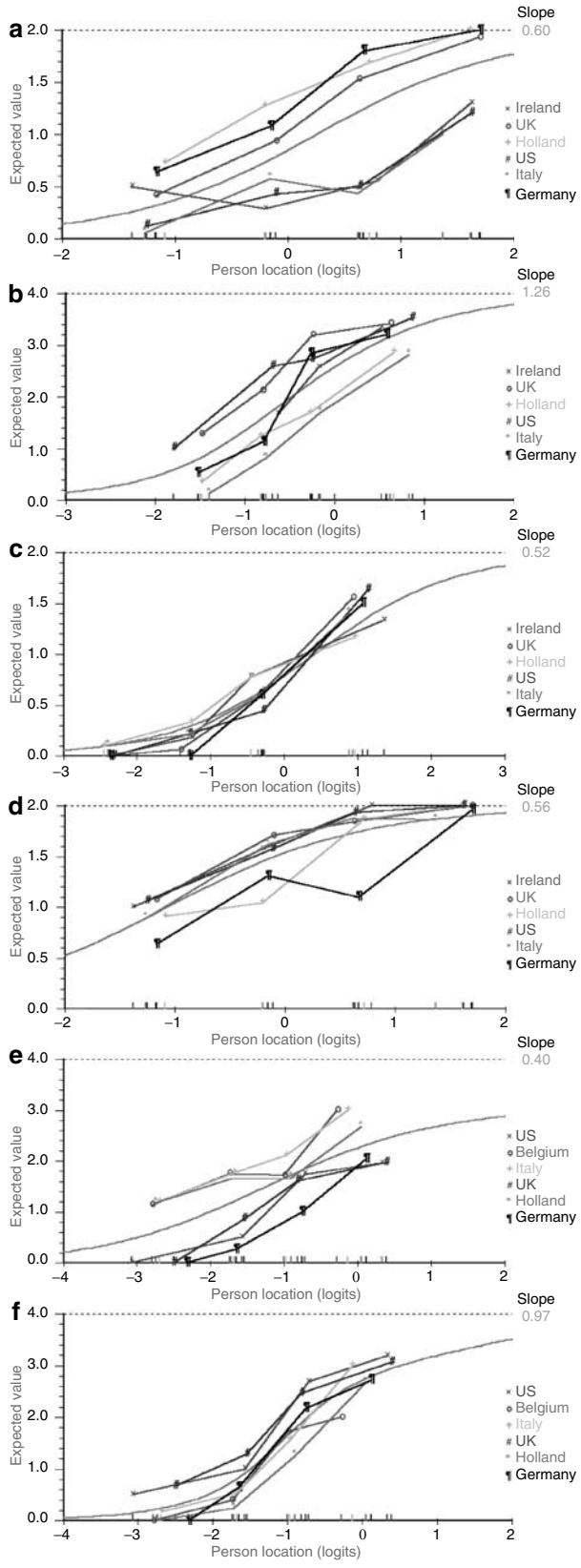

Figure 1. ICC plotted against country. (**a**) Bothered by water (Skindex-17), (**b**) frustrated (Skindex-29), (**c**) desire to be with people (Skindex-17), (**d**) itches (Skindex-17), (**e**) itchy, sore, painful, or stinging (DLQI), and (**f**) work around the house (DLQI).

disorder, and two items had significant DIF for age. Although no nonuniform DIF was detected across countries (except item 16), more than a third of the 29 items showed significant uniform DIF (Table S4).

Psoriasis patients from specific countries were not persistently scoring higher than others, except US, British, and Italian patients who scored 10 or more times more extreme (higher or lower) than expected, whereas the patients of other countries scored more often in accordance with the expected. Also, especially the US and to a lesser extent British and Italian patients scored substantially higher on the emotion items. In contrast to the DLQI, no clear ICC associations were seen between countries.

If analyzed separately, the functioning scale did not fit the overall model, four items showed significant misfit, one item had threshold disorder, and two items had DIF for age and gender (Table S3). However, across countries, 7/12 items showed significant differences (Table S5). For example, people from the UK reported significantly more often that psoriasis affected their social life (item 5) but despite that they stayed at home significantly less frequently (item 8) compared with Italian patients with similar amounts of functioning impairment.

The emotions scale of the Skindex-29 misfitted the Rasch model (item mean interaction 0.00, SD 0.56; person mean interaction –034, SD 1.30; item trait interaction, $P<0.0001$; person separation index 0.91). Only item 9 misfitted, two items had reversed thresholds, and one demonstrated DIF for age and gender. In contrast, significant uniform DIF across country was seen in 7/10 items (Table S5). No nonuniform DIF was present. Despite similar levels of emotional impairment, the Germans and Italians were more angry and depressed, the Dutch more ashamed and annoyed, the Italians and Americans more often embarrassed, and the British, Irish, and American patients were more frequently frustrated compared with the other patients from different countries.

Of the symptom scale, the item and person mean interaction fitted the model well (0.00, SD 0.62 and –0.07, SD 1.25, respectively) but with a significant item–trait interaction ($P=0.0008$). The person separation index was 0.87. Only item 18 showed significant misfit, items 7 and 16 had disordered thresholds and one item with DIF was detected across age and gender. All items had uniform DIF across cultures (Table S5). Psoriasis patients from the UK, Netherlands, and Germany were significantly more often ''bothered by water'' (item 16) than those from Philadelphia (US), and Italy. Italians reported more often symptoms such as burning and stinging, irritation, bleeding, and that they had a sensitive skin compared with patients from most other countries.

Skindex-17

The 12 items of the psychosocial scale of the Skindex-17 fitted the Rasch model relatively well (item mean interaction 0.00, SD 0.61; person mean interaction 0.51, SD 0.79; item–trait interaction, $P=0.07$) (Table S4). The person separation index was excellent (0.91). None of the items had individual misfit, threshold order, or DIF across gender

and age, except item 6, which showed significant DIF across age levels (patients less than 30 years had significantly felt more often depressed than older patients). Items 4, 5, 8, 20, and 22 demonstrated different responses than expected across cultures (the last four items all ask about direct interactions with others) (Table S5). Despite different scoring categories, items 5 (social life), 8 (stay home), 20 (embarrassed), and 22 (frustrated) show patterns comparable with that seen in the functioning and emotions subscales of the Skindex-29 (Figure 1). Despite a changed scoring system and underlying construct, the observed ICC patterns of the items of the Skindex-17's psychosocial scale and Skindex-29 were comparable.

Of the Skindex-17, the mean item and person interaction of the symptom scale was good (0.00, SD 0.94 and 0.15, SD 1.45) but the item–trait interaction was significant ($P=0.001$) and the person separation index was 0.78 (Table S2). Item 18 (irritated) showed significant individual misfit but none of the five items had threshold disorder or DIF across gender and age. All items showed uniform DIF across cultures, item 16 showed nonuniform DIF also (Table S5). As seen in the symptom subscale of the Skindex-29, Italians were more frequently bothered by psoriasis symptoms and the German patients reported significant fewer itches despite similar levels of underlying construct (Figure 1). Also, patients from Manchester (UK) scored lower on most symptom-related items than excepted compared with patients from other nations. The observed ICC patterns of the symptom items of the Skindex-17 and those of the Skindex-29 showed less variation but were comparable.

Sensitivity analyses

Restricting the DIF analyzes across cultures to native English-speaking psoriasis patients (Ireland, UK, and US) changed the results substantially for Skindex-29. Instead of 16/29 displaying significant DIF, only six had significant DIF across these three countries (items 3, 8, 16, 20, 26, and 29). The same restrictions improved the psychosocial scale (only two (items 1 and 21) instead of 6/12 displayed significant DIF) and the symptom scale of the Skindex-17 (not all items, only items 16 and 18 remained significant) as well. Restricting the analyses to the non-native English-speaking patients (Dutch, Germans, and Italians) did not substantially reduce the number of items with significant DIF across culture (data not shown) suggesting that the improvement by restricting it to English-speaking patients is not only because of reduced sample size and number of categories. For the DLQI, no significant differences were detected between British and American psoriasis patients. However, among the non-native English-speaking patients, nationality had a significant effect on 7/10 items. Limiting the DIF analyzes to northern European countries (Germany, Ireland, The Netherlands, and UK) decreased the number of items with DIF across nations to nine (items 4, 6, 8, 12, 15, 16, 20, 22, and 25) for the Sklindex-29. Of the Skindex-17 psychosocial scale, 6/12 items (items 4, 5, 6, 8, 20, and 22) displayed significant DIF and all items, except item 26, displayed DIF in the symptom scale. Including only northern European countries in the DLQI analyses showed that seven items displayed uniform and five nonuniform DIF. To test whether the overall fit of the Skindex-17 was owing to reduced number of items, we randomly selected 12 items form the emotions and functioning scale and five from the symptom scale of the Skindex-29 and demonstrated that both item sets significantly misfitted the Rasch model (item–trait interaction = 159.8, df = 72, and $P<0.0001$ and item–trait interaction = 66.5, df = 30, and $P<0.0001$, respectively).

DISCUSSION

Unidimensionality

The DLQI and the Skindex-29 are the most commonly used dermatology-specific HRQOL instruments. However, in psoriasis patients, they significantly misfit the Rasch model, in part owing to individual item misfit, threshold disorder, and DIF, confirming other findings suggesting that these instruments fail to fulfill fundamental measurement requirements such logical ordering of scoring categories, unidimensionality, and lack of impact of external factors on its items (Nijsten *et al.*, 2006a, b). In contrast, the psychosocial and, to a lesser extent, the symptom scale of the Skindex-17 behaved well in this multinational sample of psoriasis patients. This is probably because of the fact that the Skindex-17 was developed using the Rasch models.

Cross-cultural differences

In addition to suboptimal psychometric properties, the majority of the items of the DLQI and Skindex dysfunctioned between psoriasis patients from different cultural backgrounds. Although it is not uncommon to spot many DIF items in HRQOL instruments and the effect of the cross-cultural DIF in this analysis needs to studied (Hambleton, 2006; McHorney and Fleishman, 2006), the observed cross-cultural inequivalence may result in true measurement bias where multinational study data is combined (Anderson *et al.*, 1993, 1996; Bullinger *et al.*, 1993a, b). Also, more qualitative research into what causes DIF is needed to minimize its effect (McHorney and Fleishman, 2006). Therefore, the Skindex and DLQI should not be used in multinational (psoriasis) studies in their current form and, if use continues, attempts should be made to adjust cross-cultural DIF. All items of the DLQI and about half of the Skindex-29 and -17 showed DIF across the seven countries suggesting that despite similar amounts of underlying HRQOL impairment patients' nationality influenced scoring of these items. Similar items of the DLQI and Skindex such as symptom-related items confirmed the findings in the separate instruments. Interestingly, it seemed that the responses of patients from Latin-oriented countries (Belgium and Italy) clustered as well as those of English-speaking patients (UK and US). This was confirmed by sensitivity analysis restricted to the native English-speaking patients suggesting that the relevance, meaning, attitude, and expression differ for several items across nations. Completing a HRQOL questionnaire can be considered a ''social encounter'' and, as such, is governed by social rules and norms, which are culturally dependent. Because all questionnaires were either the original or translated version using guidelines (Guillemin *et al.*, 1993; Wild *et al.*, 2005), it is

unlikely that the cultural differences are because of translational issues only. However, it is interesting that lesser differences were detected among native English-speaking psoriasis patients, which may be owing to cultural equivalence, the use of the original versions of the tools, and/or both. Emotions such as embarrassment, self-consciousness, frustration, shame, irritation, humiliation, annoyance, and anger may be adequately translated but still have different meanings and interpretations across cultures (e.g., false cognates (meaning of words is not same in different languages) or homographs (words are spelled alike but have different meanings) may play a role in this issue). It seems that patients from some countries such as Italy and Belgium report less difficulties showing physical intimacy, which is reflected in several items, than those from other countries arguing that there are psychological differences such as identification, coping, and negation between patients from different cultures. This is in accordance with US studies showing that it is less stigmatizing for Hispanics to admit symptoms of psychological distress and for blacks to exhibit greater social desirability, resulting in under-reporting (McHorney and Fleishman, 2006). Although German psoriasis patients had high overall impairment levels, they reported significantly less symptoms, Italians reported a high frequency of most types of symptoms, and cultural inequivalence was detected in all symptom (sub)scales suggesting that the interpretation of skin symptoms varies substantially between countries. Another explanation of the observed inequivalence may be that scoring categories are interpreted differently and/or response patterns such as an extreme response style differ between cultures, but collapsing the five response categories of the Skindex to three categories, as is done in the Skindex-17, did not substantially reduce the observed differences.

In the light of the increasing importance of HRQOL evaluation in clinical trials and public health studies of chronic non-life-threatening diseases such as most skin diseases and the often multinational nature of these studies, we need a dermatology-specific instrument that is in line with the current psychometric standards and can be used across cultures. Despite the enormous research effort in the field of HRQOL in dermatology, the overwhelming majority of these studies are descriptive and very few methodologically oriented. Of the DLQI and the Skindex-29, which were published 15 years or more ago and are widely used, crucial instrument characteristics such as minimal clinical importance differences and cultural equivalence have yet not been reported. For now, the DLQI is considered the standard HRQOL measurement in dermatology by many and has introduced HRQOL to mainstream dermatology, but it is a first-generation instrument. The creation and refinement of the Skindex-29 lifted HRQOL in dermatology to another level making it a second-generation instrument. Third-generation HRQOL instruments are developed according to guidelines, use item response theory models to assure fundamental measurement requirements, and are tested in patients that differ in demographic, including cultural background, and disease characteristics. The Skindex-17 was an attempt to upgrade the Skindex to a third-generation instrument but despite the fact that it fitted the Rasch model reasonably well, responses by psoriasis patients to half of its items were influenced by patients' nationality.

The need of a dermatology-specific HRQOL instrument

Are we in need of a new dermatology-specific HRQOL instrument? A psychometrically good and widely accepted dermatology-specific HRQOL instrument is pivotal because of comparative reasons and because several studies have shown that they provide additional information to generic HRQOL tools. However, more studies are needed to clarify which generic tool (e.g., Short Form-36 and-12, EuroQol-5D, Nottingham Health Profile, Sickness Impact Profile) is preferable in all or specific skin disease and how they behave across cultures in dermatologic patients. To come to a third-generation HRQOL instrument that is widely accepted, existing instruments have to be adjusted and refined or a new instrument has to be developed. Although it is not the objective of this study, DIF misfit can be adjusted for by altering scoring categories for these items and/or using item-split techniques for uniform DIF (Angoff, 1993; Tennant *et al.*, 2004). These approaches may be useful if sufficient items measure a unidimensional construct (e.g., fit the Rasch model) and a minority of items have uniform DIF to be adjusted for a limited number of external factors such as age, gender, cultural background, and skin disease. These basic assumptions make it extremely complicated to adjust the DLQI or Skindex-29 because large number of their items demonstrate some kind of misfit. In addition to the observation that quantifying skin-related symptoms depends on diagnosis (Nijsten *et al.*, 2006a), the measurement of symptoms is greatly influenced by cultural background suggesting that it should not be compared in international patient groups with different skin diseases. Therefore, adjustment of the five items of the symptom scale of the Skindex-17 is very complex and may not be worthwhile. Of the tools and scales tested in this study, the psychosocial scale of the Skindex-17 may be the only one eligible to be adjusted for because of the good overall fit, a smaller number of items with individual misfit, threshold disorder, and DIF.

Strengths and limitations

This is the first cultural equivalence study for the DLQI and Skindex in a large international sample using state-of-the-art psychometric techniques. The importance of cross-cultural comparison studies is illustrated by the fact that some very well-established HRQOL instruments such as the Short Form-36 and -12 have been tested for it and it was included in the design of the development study of the EuroQOL-5D and WHOQOL-100. Moreover, large clinical trials are almost always multicenter studies with patients from multiple countries and include HRQOL (most often the DLQI in dermatology) as a primary or secondary outcome. However, this study suggest that the current dermatology-specific instruments do not meet the new standards in psychometrics, except may be the psychosocial scale of the Skindex-17 (McHorney, 1997).

Although there are more traditional psychometric approaches to test for cross-cultural differences (Gandek *et al.*,

1998), Rasch analysis is among the most commonly used statistical methods in this type of studies (Prieto *et al.*, 2003; Roorda *et al.*, 2004; Tennant *et al.*, 2004). In addition to the theoretical advantages of this technique, Rasch analysis was especially interesting to use because it assumes sample-free measurement. This implied that we were able to collect patients retrospectively from different countries, clinical settings, health-care systems, and demographic characteristics, with varying degrees of psoriasis severity without compromising the study findings. For example, differences in clinical psoriasis severity among patients from six countries should have no substantial impact on the study findings because DIF analysis did not directly compare psoriasis patients from the included nations but compared those patients of different nations who have the same level of underlying construct (i.e., HRQOL impairment). Because sample size may affect the results of the analysis of variance (ANOVA) analysis, which is used for DIF, (Wright and Tennant (1996); Hambleton, 2006), we choose *a priori* to create equally sized patient groups from several countries to have about 500 eligible patients. The fit of each country to the model should be interpreted with caution because of the relatively small sample sizes. In contrast to the results of each country separately, none of the analyses of the total study population showed an overall fit to the model (this may also be explained by the observed cultural differences). Also, a Rasch analysis of a larger group of British psoriasis patients, from which we randomly selected 75 observations, did show significant misfit in contrast to our findings (McKenna *et al.*, 2004). However, a sample size of 75 is large enough to have 99% confidence that the parameter estimates are within 1 logit of the stable value (Linacre, 1994). The majority of the study patients had moderate to severe psoriasis from university and/or psoriasis treatment centers, which is described in the references and confirmed by the high mean DLQI and Skindex-29 scores. Because the sensitivity especially of the DLQI at the lower ends of the HRQOL spectrum is likely to be low, it makes the study results more robust but decreases their generalizibility. Additional prospective studies with a greater variety of skin diseases and cultural backgrounds are needed to confirm our findings and study the meaningfulness of DIF (e.g., comparing results of unadjusted pooled scores with transformed scores to eradicate DIF) (McHorney and Fleishman, 2006).

Conclusion

The importance of HRQOL measurement in dermatology warrants more methodological studies to test existing instruments and refine them, if possible, or create a new tool that complies with its purposes. In the attempt to create a tool that measures HRQOL impact in dermatology in a scientific way, cultural differences should be taken into account.

MATERIALS AND METHODS

Study population

We searched Pubmed to identify papers that included DLQI and/or Skindex-29 data in psoriasis patients and peers interested in HRQOL and psoriasis. Our aim was to have about 75 anonymous patients with predominantly plaque-type psoriasis from more than five countries along with their age and gender. Researchers with eligible data were contacted and all but one provided their available data. If collaborators provided a larger sample, a random sample of 75 patients was generated using random number-generator method to conform to equal group size requirements for the ANOVA of the residuals (Tabachnick and Fidell, 2001) because 75 was the smallest group of patients from one country. Patients were excluded if two or more items were not scored for the DLQI (Finlay and Khan, 1994) and/or if more than 25% of the Skindex-29 items were not recorded (Chren *et al.*, 1997). No additional information on type of psoriasis, presence of arthritis, and extent of disease was collected.

In brief, the DLQI consists of 10 items with four response categories (range 0–30) (Finlay and Khan, 1994). The Skindex-29, has 29 items (five response categories) divided over three subscales (emotions, functioning, and symptoms) (Chren *et al.*, 1997). Although a composite score of the Skindex-29 has not been formally studied and has little face validity (Chren, personal communication), several studies have used it. Therefore, an overall score of the Skindex-29 (converted to range between 0 and 100) was also analyzed. The Skindex-17 is two-dimensional and the score of its two scales (12 psychosocial items (range 0–24) and five symptom items (range 0–10)) were derived from the Skindex-29 data after rescoring the items appropriately (Nijsten *et al.*, 2006a). For each of the three instruments, a higher score indicates more HRQOL impairment.

Statistical analysis

In general, the Rasch model is a unidimensional model developed in the field of education, which asserts that the easier the item the more likely it will be passed, and the more able the person, the more likely they will pass an item compared with a less able person. The probability that a person will affirm an item is a logistic function of the difference between the person's ability (or level of disease severity) and the difficulty of the item (or the level of disease severity represented in the item) and only a function of that difference. This approach can also contribute to the issue of cross-cultural validity and has been used in several studies (Prieto *et al.*, 2003; Roorda *et al.*, 2004; Tennant *et al.*, 2004). Responses to the scale's items should only depend on the amount of underlying construct (i.e., dermatology-specific HRQOL impairment) and not on external factors such as cultural background (Angoff, 1993; McHorney and Fleishman, 2006) (e.g., the score of any given DLQI item should not be significantly different between UK and US patients who both have a mean DLQI score of 10). Under the requirement that the ability under consideration is unidimensional, if the item measures the same ability across groups then, except for random variations, the same proportion is found irrespective of the nature of the group for whom a function is plotted. Items that do not yield the same item response function for two or more groups display DIF and are violating the requirement of unidimensionality. If the location of items along the measurement construct should remain the same across cultures, there is no significant DIF. Bonferroni corrections were applied to the DIF statistics because of multiple testing for each of the three instruments separately.

The statistical test used for detecting DIF is an ANOVA of the person–item deviation residuals with person factors (e.g., country of

residence) and class intervals (e.g., number of groups along the trait and the number depends on the sample size with about 30–50 persons per group) as factors (Angoff, 1993), which is a well accepted method to study DIF (Tennant *et al.*, 2004). Two types of DIF can be identified. With uniform DIF there is a constant difference between groups in the probability in affirming a category across the trait (ANOVA main effect) resulting in parallel ICCs. With nonuniform DIF, the difference varies across trait (ANOVA interaction effect) and results in a nonparallel and/or crossing ICC. The importance of this difference is that uniform DIF can be corrected by assigning different scores or weighing the scores for the person factor, whereas nonuniform DIF cannot be corrected in existing instruments. In addition to DIF across countries, overall fit to the Rasch model, individual item fit, threshold order, and DIF across gender and age were studied (Nijsten *et al.*, 2006a). The Rasch analysis was performed with RUMM2020 (RUMM Laboratory Pvt Ltd, Perth, Australia).

CONFLICT OF INTEREST

The authors state no conflict of interest.

ACKNOWLEDGMENTS

This work was done in Rotterdam, The Netherlands.

SUPPLEMENTARY MATERIAL

Table S1. Overall fit to the Rasch model, number of significant individual item misfits, threshold disorder, and DIF for the DLQI for each country.

Table S2. DIF across countries for the DLQI.

Table S3. Overall fit to the Rasch model, number of significant individual item misfits, threshold disorder, and DIF for the Skindex-29 and its scales for each country.

Table S4. Overall fit to the Rasch model, number of significant individual item misfits, threshold disorder, and DIF for the two scales of the Skindex-17 for each country.

Table S5. DIF across countries for the Skindex-29 and Skindex-17.

REFERENCES

Anderson RT, Aaronson NK, Bullinger M, McBee WL (1996) A review of the progress towards developing health-related quality-of-life instruments for international clinical studies and outcomes research. *Pharmacoeconomics* 10:336–55

Anderson RT, Aaronson NK, Wilkin D (1993) Critical review of the international assessments of health-related quality of life. *Qual Life Res* 2:369–95

Angoff WH (1993) Perspectives on differential item functioning methodology. In: *Differential Item Functioning*. (Holland PW, Wainer H, eds) Hillsdale, NJ: Lawrence Erlbaum

Augustin M, Wenninger K, Amon U, Schroth MJ, Kuster W, Chren M *et al.* (2004) German adaptation of the Skindex-29 questionnaire on quality of life in dermatology: validation and clinical results. *Dermatology* 209:14–20

Augustin M, Zschocke I, Lange S, Seidenglanz K, Amon U (1999) Quality of life in skin diseases: methodological and practical comparison of different quality of life questionnaires in psoriasis and atopic dermatitis. *Hautarzt* 50:715–22

Bullinger M, Alonso J, Apolone G *et al.* (1993b) Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. International Quality of Life Assessment. *J Clin Epidemiol* 51:913–23

Bullinger M, Anderson R, Cella D, Aaronson N (1993a) Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Qual Life Res* 2:451–9

Chren MM, Lasek RJ, Flocke SA, Zyzanski SJ (1997) Improved discriminative and evaluative capability of a refined version of Skindex, a quality-of-life instrument for patients with skin diseases. *Arch Dermatol* 133: 1433–40

de Korte J, Van Onselen J, Kownacki S, Sprangers MA, Bos JD (2005) Quality of care in patients with psoriasis: an initial clinical study of an international disease management programme. *J Eur Acad Dermatol Venereol* 19:35–41

Evers AWM, Lu Y, Duller P, van der Valk PGM, Kraaimaat F, van de Kerkhof PCM (2005) Common burden in chronic skin disease? Contributors to psychological distress in adults with psoriasis and atopic dermatitis. *Br J Dermatol* 152:1275–81

Finlay AY, Khan GK (1994) Dermatology life quality index (DLQI) – a simple practical measure for routine clinical use. *Clin Exp Dermatol* 19: 210–216

Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE *et al.* (1998) Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol* 51:1171–8

Guillemin F, Bombardier C, Beaton D (1993) Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 46:1417–32

Hambleton RK (2006) Good practices for identifying differential item functioning. *Med Care* 44:S182–8

Linacre JM (1994) Sample size and item calibration stability. 7:4 p.328 http://www.rasch.org/rmt/rmt74m.htm accessed on 9 February, 2007

McHorney CA (1997) Generic health measurement: past accomplishments and a measurement paradigm for the 21st century. *Ann Intern Med* 127:743–50

McHorney CA, Fleishman JA (2006) Assessing and understanding measurement equivalence in health outcome measures: issues for further quantitative and qualitative inquiry. *Med Care* 44:S205–10

McKenna SP, Cook SA, Whalley D, Doward LC, Richards HL, Griffiths CE *et al.* (2003) Development of the PSORIQoL, a psoriasis-specific measure of quality of life designed for use in clinical practice and trials. *Br J Dermatol* 149:323–31

McKenna SP, Lebwohl M, Kahler KN (2005) Development of the US PSORIQoL: a psoriasis-specific measure of quality of life. *Int J Dermatol* 44:462–9

McKenna SP, Meads DM, Doward LC (2004) Scaling properties of the dermatology life quality index (DLQI). *Value Health* 7:750–1

Nijsten T, Meads DM, McKenna SP (2006b) Dimensionality of the dermatology life quality index (DLQI): a commentary. *Acta Derm Venereol* 86:284–5

Nijsten TE, Sampogna F, Chren MM, Abeni DD (2006a) Testing and reducing skindex-29 using Rasch analysis: Skindex-17. *J Invest Dermatol* 126:1244–50

Ongenae K, Van Geel N, De Schepper S, Naeyaert JM (2005) Effect of vitiligo on self-reported health-related quality of life. *Br J Dermatol* 152: 1165–72

Prieto L, Novick D, Sacristan JA, Edgell ET, Alonso J SOHO Study Group (2003) A Rasch model analysis to test the cross-cultural validity of the EuroQoL-5D in the Schizophrenia Outpatient Health Outcomes Study. *Acta Psychiatr Scand Suppl* 416:24–9

Roorda LD, Jones CA, Waltz M, Lankhorst GJ, Bouter LM, Vanlder Eijken JW *et al.* (2004) Satisfactory cross cultural equivalence of the Dutch WOMAC in patients with hip osteoarthritis waiting for arthroplasty. *Ann Rheum Dis* 63:36–42

Sampogna F, Sera F, Abeni D, IDI Multipurpose Psoriasis Research on Vital Experiences (IMPROVE) Investigators (2004) Measures of clinical severity, quality of life, and psychological distress in patients with psoriasis: a cluster analysis. *J Invest Dermatol* 122:602–7

Skevington SM (2002) Advancing cross-cultural research on quality of life: observations drawn from the WHOQOL development. World Health Organization Quality of Life Assessment. *Qual Life Res* 11:135–44

Tabachnick BG, Fidell LS 2001 *Using multivariate statistics*, 4th ed. Boston: Allyn and Bacon

Tennant A, Penta M, Tesio L *et al.* (2004) Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. *Med Care* 42:I37–48

Wagner AK, Gandek B, Aaronson NK, Acquadro C, Alonso J, Apolone G *et al.* (1998) Cross-cultural comparisons of the content of SF-36 translations across 10 countries: results from the IQOLA Project. International quality of life assessment. *J Clin Epidemiol* 51:925–32

Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A *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 translation and cultural adaptation. *Value Health* 8:94–104

Wright BD, Tennant A (1996) Sample size again. *Rasch Measurement Transactions*. 9:4 p468 http://www.rasch.org/rmt/rmt94h.htm accessed on 9 February, 2007