Validating an Amharic version of the 36-Item Short Form Health Survey (SF-36) in individuals with leprosy in Ethiopia

Saba Maria Lambert[a,b], Shimelis Nigusse[c], Digafe Alembo[c], Stephen Walker[a], Peter Nicholls[d] & Diana Nancy Lockwood[a]

[a]*London School of Hygiene and Tropical Medicine, Clinical Research, Keppel Street, WC1E 7HT, UK*
[b]*ALERT, Leprosy Clinic, Jimma Road, Ethiopia*
[c]*All Africa Leprosy Rehabilitation and Training (ALERT) Center, Addis Ababa, Dermatology, Ethiopia*
[d]*School of Health Sciences, University of Southampton, SO17 1BJ, UK*

Summary
Background Health-related quality of life (HRQoL) has now become an indispensable outcome measure in many randomized clinical trials and other studies. It provides the patient's voice in measuring health improvement or decline and assessing treatment effectiveness. A validated Amharic version of HRQoL assessment tool was needed for leprosy clinical trials in Ethiopia. The SF-36 was chosen but a validated Amharic version was not available. We describe how this was developed.
Methods The SF-36 was translated from English into Amharic and evaluated for content acceptability in a patient focus group. Back translation was performed. Validity and reliability of the Amharic SF-36 in leprosy affected individuals was tested with 100 patients with leprosy attending the leprosy clinic at ALERT hospital and compared to the Amharic version of the WHOQOL-BREF.
Results Amharic versions of both the WHOQOL-BREF and the SF-36 had good reliability and validity amongst leprosy affected individuals. Internal consistency reliability estimates for each domain/scale exceeded 0.70. The Amharic SF-36 had better convergent and discriminant validity than WHOQOL-BREF in this group of patients. Good known-group validity was seen in both WHOQOL-BREF and SF-36 in leprosy affected patients. The Amharic SF-36 had good inter-rater reliability with seven out of 8 domains scoring above 0.8 in intra-class correlation.
Conclusion This Amharic version of the SF-36 is a valid instrument to measure HRQoL in studies and clinical settings involving leprosy affected individuals in Ethiopia.

Keywords: Leprosy, quality of life, HRQOL, Amharic, SF-36, WHOQOL-BREF, Ethiopia

Correspondence to: Dr. Saba Maria Lambert, London School of Hygiene and Tropical Medicine, Clinical Research, Keppel Street, WC1E 7HT, UK (Tel.: 447921266473; e-mail: sabalambert@hotmail.com)

Background

Health-related quality of life (HRQoL) as a patient-centred outcome measure is increasingly being included in clinical trials but is rarely used in comparative clinical trials in leprosy.

Leprosy is a chronic granulomatous infection principally affecting the skin and peripheral nerves caused by the obligate intracellular organism *Mycobacterium leprae*.[1] In 2020, the World Health Organization (WHO) reported 202,185 new cases globally.[2] Multi-drug therapy (MDT) cures the infection, but the damage done to nerves by *M. leprae* and immunological reactions leads to disability and visible deformity mainly affecting eyes, hands and feet. Patients with leprosy are often stigmatised and experience economic hardship,[3] both because of stigma[4] and poor health. The treatment of leprosy reactions is currently inadequate, with prednisolone being the main drug used with limited success and high rates of adverse effects.[5] Identifying better agents for treating leprosy reactions and understanding the effects of these treatments on HRQoL is a priority.[6]

Due to the international nature of clinical research, the need for cross-culturally valid patient reported outcomes questionnaires has grown considerably. For cross-cultural clinical research, the ultimate goal is to pool data across languages in order to evaluate the effect of treatment on an outcome measured by the same questionnaire. To achieve this objective, for each language, the concepts assessed by each item should be as semantically close as possible, the aggregation of items should result in the same constructs and the metric scales should be similar.

When a questionnaire has already been developed and used in one culture, the sequential approach based on a thorough translation is essential for controlling potential bias at the level of each item. Many translation guidelines have been published and most describe a forward-backward translation by a qualified team, followed by pilot testing with patients.[7,8] This does rely on the assumption that the constructs of a questionnaire, and their content, are relevant and equivalent across cultures.

At the start of this study, a few cross-sectional studies of HRQoL in leprosy patients had been published. The three main tools used were the World Health Organisation Quality of Life Bref (WHOQOL-BREF),[9–12] the Dermatology Life Quality Index (DLQI)[13,14] and the Short Form-36 Health Survey SF-36.[15,16]

The WHOQOL-BREF is a shortened version of the original WHOQOL-100, with the following domains: physical health, psychological health, level of independence, social relationships, spiritual and environmental conditions.[17] The validated Amharic WHOQOL-BREF had been used in an assessment of 749 women displaced by conflict,[18] and of 346 individuals affected by podoconiosis.[19]

The DLQI consists of ten questions, designed to assess the effect of dermatological conditions on HRQoL in adults. It covers several dimensions of life quality, including pain, embarrassment, interference with activities, and social and sexual relationships with questions aimed at how the patient/individual has felt in the last week. An Amharic version of the DLQI was used to study podoconiosis in Ethiopia.[20] The questions in the DLQI do not capture information on the impact of nerve damage and disability caused by leprosy.

The SF-36 was developed as a way of measuring the outcome of healthcare delivery in the United States.[21] The SF-36 comprises 36 items assessing eight health concepts using multi-item scales and is administered using a past four weeks reporting interval. It tests physical functioning (10 items), role limitations caused by physical health problems (4 items), role limitations caused by emotional problems (3 items), social functioning (2 items), emotional well-being (5 items), energy/fatigue (4 items), pain (2 items), general health perception (5 items) and perceived change in health during the last 12 months. Two summary scores

can be calculated: a mental health component summary score (MCS) and a physical health component summary score (PCS).

Four studies reported using the SF-36 in Ethiopia. A study published in the Ethiopian Medical Journal[22] evaluated the SF-36 to obtain normative data in a general health survey in order to establish general population norms and to describe the effects of socio-demographic factors on SF 36. It concluded that the Amharic SF-36 had acceptable psychometric properties and construct validity. The translation system used was not reported. It was later used to assess HRQoL in 271 individuals with schizophrenia,[23] and 315 individuals with bipolar disease[24] and 420 people living with HIV and on anti-retroviral therapy.[25]

Unfortunately, despite many attempts, we were not able to obtain a copy of this Amharic version of the SF-36, and the developers of SF-36 did not hold an Amharic version.

Although cross validation of item selection and scoring of SF-36 has been done,[26] this has often been done on patients living in developed countries with similar standards of living. At first glance, the face validity of some SF-36 items appear questionable for patients in low-income settings, such as questions about "playing golf", "bowling", "pushing a vacuum cleaner" and "climbing several flights of stairs" in a country where only urban buildings have several floors. This observation pointed to a need to explore the construct validity of the SF-36 before adopting it for use with leprosy patients in our study in Ethiopia.

Validation of a translated questionnaire can be done by comparing its reliability and validity with a validated QOL tool in that language. Previous comparisons between SF-36 and WHOQOL-BREF have been successfully done in patients with HIV, showing that there are good correlations between the corresponding domains/scales of the two instruments.[27] Validation of the Amharic SF-36 in our study was done by comparison with an already validated Amharic WHOQOL-BREF.[18,28] Another measure of validity for SF-36 in leprosy patients was to assess known-group validity by comparing SF-36 scores with symptom frequency and symptom severity in leprosy patients. The decision to select SF36 for our clinical trial if it proved superior to WHOQOL BREF was taken.

Methodology

STUDY SETTING AND POPULATION

The study took place at ALERT hospital, a tertiary referral centre for leprosy, in Addis Ababa, Ethiopia. Adult individuals diagnosed with leprosy attending the hospital for clinical care were approached to participate in the study.

STUDY AIM

To validate an Amharic version of SF36, useable in future clinical studies we hypothesized that if both instruments captured the health related QOL of leprosy patients, then:

(1)   The corresponding domain/scale of both instruments should be positively correlated, i.e. the physical, psychological, and social domains of the WHOQOL-BREF should be significantly correlated with PF, MH and SF scales of the SF-36 respectively;
(2)   The physical and psychological domains of the WHOQOL-BREF should have weak associations with MCS and PCS of the SF-36, respectively;
(3)   The domain/scale score of both instruments should be positively correlated with self-perceived health status (question 2 in both instruments);
(4)   The domain/scale score of both instruments should be inversely correlated with the number and intensity of leprosy related symptoms.

SAMPLE SIZE

After review of published literature, a minimum sample size of 30 participants was the usual size in validation studies, especially in low prevalence conditions.[29,30] We decided on 50 participants for the comparison of SF-36 to WHOQOL-BREF and 50 participants to assess inter-rater reliability, giving a total of 100 participants for the relationship between SF-36 scores and leprosy symptoms.

ETHICAL APPROVAL

Ethical approval was obtained as part of the larger clinical trial from the Ethics Committee of the London School of Hygiene and Tropical Medicine (5376), the ALERT and AHRI Ethical Review Committee (AA/ht/248/09), and the National Ethics Review Committee of Ethiopia (RDHE/34-90/2009). Written informed consent was obtained in Amharic. Data were anonymised and stored in a password protected Access database.

INSTRUMENT TRANSLATION AND ADAPTATION

The SF-36 questionnaire was translated by two native Amharic speakers fluent in English. The translators, two doctors, a social worker and a nurse reviewed the translation to ensure the translation replicated the original as closely as possible but was appropriate for the socio-economic and cultural setting. For example, "pushing a vacuum cleaner, bowling, or playing golf" were removed, leaving only "moving a table" as an example for moderate activity. The two previous reports on the use of Amharic SF-36 mentioned that "climbing stairs" was replaced by "walking up a hill" in their translation, but we felt comfortable using "climbing stairs" in an urban setting, with "walking up a hill" as a second option. Distance in miles and yards were changed to kilometres and metres which are more commonly used in Ethiopia. Following this, our Amharic version was then discussed in a focus group of two doctors, two nurses, an occupational therapist and seven patients. The patients were of various ages and leprosy experience – two were newly diagnosed patients, three patients had leprosy reactions and two long-term patients had attended for management of neuropathic ulcers. After minor changes in wording, a final version was agreed and back translated into English by an independent translator. The new English translation was then compared with the original SF-36 for conceptual equivalence and found to be satisfactory by two English language native speakers. The final Amharic SF-36 version (Supplementary Appendix 2) was the one tested for validity and reliability.

VALIDITY AND RELIABILITY OF AMHARIC SF-36 IN LEPROSY PATIENTS

Amharic speaking patients with leprosy were interviewed after obtaining informed consent, by a nurse or social worker trained in questionnaire administration.

Demographic and clinical data were collected using a standard form (Supplementary Appendix 1). Disability grading was assessed using the Eye Hand Foot (EHF) score recommended by the WHO which has a range of 0–12. The higher the score the greater the disability.[31]

Participants were alternatively allocated to one of the two groups. Group A were interviewed by the same interviewer, on the same day with two different questionnaires: Amharic WHOQOL-BREF and Amharic SF-36. Individuals in Group B were interviewed separately by two interviewers, on the same day with the Amharic SF-36. Each interviewer was blinded to the other's interview results.

Statistical analysis

Reverse score items were adjusted in SF-36 for questions SF02, GH02, GH04, VT03, VT04, MH01, MH02, MH04 and in WHOQOL-BREF for questions 3, 4 and 26. The scoring systems recommended by the tool developers were followed for both the WHOQOL-BREF[17] and the SF-36.[21] Data were then analysed in the following aspects:

(1) Descriptive statistics
(2) Tests of scaling assumptions (multi-trait scaling methods)
(3) Reliability (Cronbach's α for internal consistency reliability)
(4) Convergent and discriminant validity (correlations between scores of the two instruments)
(5) Known-groups validity (correlations between scores and symptoms)
(6) Inter-rater reliability (correlation between two interviewers per domain/scale and global score).

Data were analysed using Statistical Package for the Social Sciences (SPSS) for Windows, version 20. Quality Metric Health Outcomes Scoring Software 4.0 was used to derive SF-36 scores.

Results

BASELINE CHARACTERISTICS

The characteristics of the 100 participants are summarized in Table 1.

There was a 1:3 ratio of female to males in the group of 100 participants interviewed. Although only 2% had received tertiary education, 58% had been to school and were literate. 27% of participants interviewed were rural residents. Most participants (81%) attending the clinic were being treated for a reaction and were on steroids; only 31% were acutely unwell on the day of the interview. A total of 41% were on MDT.

The EHF scores were categorized into "0 = No disability" (24%), "1–4 = Moderate disability" (55%) and "5–12 = Severe disability" (21%). A high percentage (79%) of participants reported experiencing more than four leprosy related symptoms.

All the participants were interviewed; none self-completed the questionnaires. This ensured 100% completion rate. The approximate completion time was around 20 min for each questionnaire.

DESCRIPTIVE STATISTICS FOR THE WHOQOL-BREF VS. SF-36 COMPARISON

Each of the 50 Group A participants interviewed had their scores analysed by domains for both questionnaires. The score distribution is shown in Table 2.

The physical and environmental domains of WHOQOL-BREF and six out of the 8 scales of the SF-36 were positively skewed, indicating distributions with more participants scoring lower than the mean group score.

All four domains of the WHOQOL-BREF had trivial floor and ceiling effects. Ceiling effect is measured by the proportion of people getting the highest possible score, whilst floor effects reflect the proportion of people receiving the lowest possible score. The highest ceiling effect was noted in the physical functioning (PF) scale of SF-36 (34%) indicating that one third of participants were able to perform physical activities without limitations. Noteworthy ceiling effects were observed for the role-disability scales (24% for role physical (RP) and 22% for role emotional (RE)) in the SF-36, indicating that almost one quarter of individuals affected

Table 1. Characteristics of 100 patients with leprosy enrolled in this study

Characteristics of patient group		Total group $n = 100$ (%)	Group A: WHOQOL-BREF vs. SF-36 comparison group	Group B: SF-36 inter-rater reliability group
Number of patients		100	50	50
Age (years) Median		35	33.5	37.4
Female: male ratio		1:3	2:3	1:2
Education level	None	42%	17 (34%)	25 (50%)
	Primary	30%	17 (34%)	13 (26%)
	Secondary	26%	15 (30%)	11 (22%)
	Tertiary	2%	1 (2%)	1 (2%)
Literacy	No	42%	18 (36%)	24 (48%)
	Yes	58%	32 (64%)	26 (52%)
Lives:	Alone	16%	5 (10%)	11 (22%)
	With others	84%	45 (90%)	39 (78%)
Residence: Rural/Urban	Rural	27%	10 (20%)	17 (34%)
	Urban	73%	40 (80%)	33 (66%)
Duration of leprosy symptoms (years: mean range)		2.9 (0–14)	2.8 (0–14)	3 (0–10)
On MDT		41%	22 (44%)	19 (38%)
On steroids for reactions		81%	42 (94%)	39 (78%)
Type of leprosy reaction	ENL	20%	9 (18%)	11 (22%)
	T1R	61%	33 (66%)	28 (56%)
Health status at today's attendance	Sick	31%	19 (38%)	12 (24%)
	Stable	69%	31 (62%)	38 (76%)
Hospital admission	Never	66%	33 (66%)	33 (66%)
	Past	27%	14 (28%)	13 (26%)
	Presently	7%	3 (6%)	4 (8%)
Disability grading (Total EHF score)	None = 0	24%	15 (30%)	9 (18%)
	Moderate = 1–4	55%	18 (36%)	37 (74%)
	Severe = 5–12	21%	17 (34%)	4 (8%)
Number of positive symptoms:	None	5%	5 (10%)	0 (0%)
	1–3 symptoms	16%	9 (18%)	7 (14%)
	4–7 symptoms	79%	36 (72%)	43 (86%)
Severity of symptoms (higher or lower than group mean of reported severity of symptoms)	None	4%	4 (8%)	0
	Moderate (lower than group mean)	50%	23 (46%)	27 (54%)
	Severe (higher than group mean)	46%	23 (46%)	23 (46%)
Self-perceived health status (GH1 from SF-36):	Poor	7%	3 (6%)	4 (8%)
	Fair	40%	20 (40%)	20 (40%)
	Good	33%	15 (30%)	(36%)
	Very good	16%	8 (16%)	8 (16%)
	Excellent	4%	4 (8%)	0

by leprosy did not feel that their physical health or emotional problems resulted in difficulties with work or other activities. A modest ceiling effect was observed for social functioning (SF) with 20% of participants able to perform social activities without interference.

Table 2. Score distribution of the WHOQOL-BREF and SF-36 (*n* = 50)

	Number of items	Mean	Standard deviation	Minimum	Percentile 25	Median	Percentile 75	Maximum	Percent scoring at the floor (%)	Percent scoring at the ceiling (%)
WHOQOL-BREF										
PHYS	7	51.2	16.0	10.7	39.3	50.0	64.3	75.0	2.00	8.00
PSYCH	6	53.7	16.4	12.5	41.7	58.3	66.7	80.0	2.00	2.00
SOCIAL	3	46.1	22.6	0.0	33.3	50.0	66.7	75.0	8.00	16.00
ENVIR	8	40.9	14.9	12.5	31.3	40.6	50.0	71.9	2.00	2.00
SF-36										
PF	10	71.5	33.9	0.0	50.0	90.0	100.0	100.0	6.00	34.00
RP	4	61.5	29.1	0.0	43.8	50.0	93.8	100.0	4.00	24.00
BP	2	50.9	30.3	0.0	32.0	42.0	74.0	100.0	8.00	18.00
GH	5	50.0	24.8	5.0	27.0	45.0	67.0	100.0	2.00	4.00
VT	4	55.0	22.5	12.5	50.0	50.0	75.0	100.0	2.00	6.00
SF	2	48.0	33.9	0.0	25.0	50.0	62.5	100.0	16.00	20.00
RE	3	60.2	28.6	0.0	50.0	50.0	75.0	100.0	4.00	22.00
MH	5	52.4	21.8	0.0	45.0	50.0	60.0	100.0	2.00	4.00
PCS		46.4	10.3	27.9	37.4	46.1	56.8	62.0	2.00	2.00
MCS		38.7	10.9	15.4	32.3	38.4	43.9	62.8	2.00	2.00

Phy – physical domain, Psy – psychological domain, Soc – social domain, Env – environment domain, PF – physical functioning, RP – role physical, BP – bodily pain, GH – general health perceptions, VT – vitality, SF – social functioning, RE – role emotional, MH – mental health, PCS – physical component summary, MCS – mental component summary.

TEST OF SCALING ASSUMPTION FOR THE WHOQOL-BREF VS. SF-36 COMPARISON

To evaluate item internal consistency test and item discriminant validity test for both instruments, multi-trait scaling techniques were used (Table 3). Item internal consistency describes to what extent items belonging to the same scale do correlate one with each other and item discriminant validity shows that items belonging to different scales should not correlate to a great extent.

The range of the item internal consistency for the WHOQOL-BREF was 0.4–0.84 and 0.49–0.97 for SF-36. A perfect success rate, with the criteria of correlations, which equal or exceed 0.40, was observed in the tests of the item internal consistency for both instruments. Results of item discriminant validity and scaling success rates are also shown, with a near perfect success rate achieved in tests of the item discriminant validity for both instruments.

RELIABILITY

Table 4 shows internal consistency for reliability tested by Cronbach's a. The Cronbach's a values for internal consistency (reliability) for all the SF-36 scales were above 0.70 showing good internal reliability of SF-36. The physical, psychological and environmental domains of WHOQOL-BREF also had Cronbach's a values above 0.70. The social domain had a lower Cronbach's a than expected at 0.652. Looking back into the three questions being assessed, it was noted that question 21 dealt with sexual function asking: "How satisfied are you with your sex life?" In the Ethiopian context, discussing your sex life is still fairly taboo and it was theorized that this item was negatively influencing the internal reliability of the social domain.

Table 3. Tests of item internal consistency and discriminant validity of the WHOQOL-BREF and SF-36

	Range of correlations		Internal consistency tests[c]		Discriminant validity test[d]	
	Item-internal consistency[a]	Item-discriminant validity[b]	#Success/total	Success rate (%)	#Success/total	Success rate (%)
WHOQOL-BREF (*n* = 50)						
Phy	0.4–0.84	0.01–0.78	7/7	100	27/28	96
Psy	0.53–0.80	0.06–0.78	6/6	100	24/24	100
Soc	0.70–0.81	0.27–0.53	3/3	100	12/12	100
Env	0.46–0.71	0.27–0.74	8/8	100	30/32	100
SF-36 (*n* = 100)						
PF	0.64–0.97	0.25–0.78	10/10	100	78/80	97.5
RP	0.82–0.94	0.31–0.89	4/4	100	32/32	100
BP	0.79–0.79	0.45–0.74	2/2	100	16/16	100
GH	0.65–0.80	0.27–0.59	5/5	100	39/40	97.5
VT	0.49–0.73	0.14–0.66	4/4	100	32/32	100
SF	0.88–0.88	0.29–0.63	2/2	100	16/16	100
RE	0.93–0.97	0.45–0.85	3/3	100	24/24	100
MH	0.61–0.8	0.27–0.61	5/5	100	40/40	100

[a]Correlation between items and hypothesized scale corrected for overlap. [b]Correlation between items and other scales. [c]Number ≥ 0.40. [d]Number of correlations significantly higher/total number of correlations.

Table 4. Reliability statistics for WHOQOL-BREF and SF-36

QoL questionnaire	Domains	Cronbach's *a*	*N* of items
WHOQOL-BREF	Phy	0.768	7
	Psy	0.738	6
	Soc	0.652	3
	Envir	0.744	8
SF-36	PF	0.968	10
	RP	0.966	4
	BP	0.923	2
	GH	0.877	5
	VT	0.817	4
	SF	0.928	2
	RE	0.975	3
	MH	0.894	4

Table 5. Reliability Statistics for WHOQOL social domain

	Cronbach's *a*	Cronbach's *a* based on standardized items	*N* of items
Qu 20, 21 and 22	0.652	0.685	3
Qu 20 and 22 only	0.851	0.854	2

Re-running the analysis by omitting question 21, greatly improved the Cronbach's *a* from 0.652 to 0.851 (Table 5).

Table 6. Pearson's correlation coefficients between WHOQOL-BREF and SF-36 (*n* = 50)

Name	WHOQOL-BREF						SF-36							
	Q1	GH1	PHY	PSY	SOC	ENV	PF	RP	BP	GH	VT	SF	RE	MH
Q1	1.00													
GH1	0.41**	1.00												
PHY	0.69**	0.43**	1.00											
PSY	0.56**	0.32*	0.76**	1.00										
SOC	0.21	0.27	0.46**	0.48**	1.00									
ENV	0.48**	0.25	0.64**	0.68**	0.50**	1.00								
PF	0.23	0.46**	0.33*	0.32*	0.25	0.20	1.00							
RP	0.34*	0.59**	0.46**	0.35*	0.17	0.30*	0.58**	1.00						
BP	0.36*	0.76**	0.47**	0.35*	0.16	0.21	0.54**	0.70**	1.00					
GH	0.25	0.78**	0.28	0.26	0.10	0.19	0.41**	0.60**	0.75**	1.00				
VT	0.33*	0.70**	0.42**	0.39**	0.05	0.23	0.38**	0.55**	0.66**	0.70**	1.00			
SF	0.16	0.68**	0.38**	0.25	0.15	0.37*	0.39**	0.57**	0.72**	0.59**	0.55**	1.00		
RE	0.44**	0.52**	0.62**	0.50**	0.19	0.30*	0.64*	0.74**	0.71**	0.50**	0.54**	0.57**	1.00	
MH	0.51**	0.67**	0.60**	0.46**	0.08	0.34*	0.46**	0.47**	0.73**	0.62**	0.63**	0.60**	0.65**	1.00
PCS	0.23	0.67**	0.32*	0.29*	0.22	0.20	0.81**	0.83**	0.80**	0.76**	0.59**	0.59**	0.65**	0.49**
MCS	0.45**	0.70**	0.61**	0.47**	0.06	0.35*	0.35*	0.54**	0.75**	0.63**	0.75**	0.77**	0.77**	0.89**

*Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.001 level (2-tailed).

Q1 – overall QOL in WHOQOL-BREF, GH1 – general health in WHOQOL-BREF, Phy – physical domain, Psy – psychological domain, Soc – social domain, Env – environment domain, PF – physical functioning, RP – role physical, BP – bodily pain, GH – general health perceptions, VT – vitality, SF – social functioning, RE – role emotional, MH – mental health, PCS – physical component summary, MCS – mental component summary.

In analysing level of correlation, the following have been assumed: high correlation: 0.5 to 1.0 or −0.5 to 1.0; moderate correlation: 0.3 to 0.5 or −0.3 to 0.5; low correlation: 0.1 to 0.3 or −0.1 to −0.3.

CONVERGENT AND DISCRIMINANT VALIDITY

The correlations for inter-domain/scale of the WHOQOL-BREF and the SF-36 are presented in Table 6. The range of correlations for inter-domain/scales of the WHOQOL-BREF is 0.46–0.76 (all *p* < 0.001), showing a range of moderate (30%) to high (60%) associations among domains. All the inter-scale correlations of the SF-36 showed moderate (14%) to high (76%) associations (*r* range 0.38–0.89, all *p* < 0.001). Within WHOQOL-BREF, the range of correlation between the general QoL question (Q1) and the domains was from 0.21–0.69, whilst between the general health question (GH1) and the domains it was 0.25–0.43. The correlations were between the social domain and general QoL and health questions (*r* = 0.21 and *r* = 0.27 respectively) and the environmental domain and the general health question (*r* = 0.25). Within SF-36, there was good correlation between the general health question (GH) and all the scale (*r* range 0.47–0.76).

Correlation between scores of the WHOQOL-BREF and the SF-36 are also shown in this table. The relationship of the general item, Q1 (overall QoL from WHOQOL-BREF) showed weak to moderate associations with scales of SF-36, including with the general health question GH (*r* = 0.41, *p* < 0.001). The highest association (*r* = 0.51) was between Q1 and MH of the SF-36. This implies that both measured similar concepts. Question GH from SF-36 showed weak to moderate associations with WHOQOL-BREF domains but a high association (*r* = 0.78) with GH1 implying that both measured similar concepts, and that participants were responding to this question consistently with both questionnaires. The hypothesis that domain/scale scores should be positively correlated to self-perceived health status is better supported with the SF-36 in this group of patients.

Looking in more detail at the associations between the SF-36 scales and the WHOQOL-BREF domains, weak associations occurred between the social domain of WHOQOL-BREF

and all SF-36 scale items (*r* range 0.05–0.25), as well as between the environmental domain of WHOQOL-BREF and scale items PF, BP, GH, and VT of the SF-36 (*r* range 0.19–0.23). Moderate associations were seen between the physical domain of WHOQOL-BREF and PF, RP, BP, GH and VT of SF-36 (*r* range 0.28 and 0.47); and the psychological domain of WHOQOL-BREF and PF, BP, RP, VT and MH of SF-36 (*r* range 0.32–0.46). The highest correlations were found between the physical domains of WHOQOL-BREF and RE (*r* = 0.62) and MH (*r* = 0.60) of the SF-36.

The correlation between the physical and psychological domains of the WHOQOL-BREF and PF and MH of the SF-36 were 0.33 and 0.46 respectively, but the association between the social domain and SF scale was low (*r* = 0.15). The first hypothesis that the corresponding domain/scale of both instruments should be positively correlated is partially supported.

Regarding the summary measures of the SF-36, the physical domain of the WHOQOL-BREF has weak association with PCS (*r* = 0.32) and the psychological domain of WHOQOL-BREF has strong association with MCS (*r* = 0.47). A weak association was found between the psychological domains of the WHOQOL-BREF and PCS of SF-36 (*r* = 0.29). This supports the hypothesis that the psychological domain of WHOQOL-BREF should have weak association with PCS, but the correlation between the physical domain and MCS was found to be strong (*r* = 0.61).

Within SF-36, the strongest association were between PCS and PF, RP, BP and GH (0.75–0.83) and between MCS and BP, VT, SF, RE and MH (0.76–0.89). Previous studies have found that, three scales (PF, RP, BP) correlated most highly with the physical component (PCS) measure whilst the mental component (MCS) correlated most highly with the MH, RE, and SF scales.[21,32,33]

Overall, the results of validity examination showed that SF-36 has better convergent and discriminant validity than WHOQOL-BREF in this group of participants. The social domain of WHOQOL-BREF showed particularly poor correlation, which might be related to the small number of questions in this domain or to poor internal validity of this domain (Cronbach's *a* = 0.652).

KNOWN GROUP VALIDITY

Table 7 shows that, in general, leprosy affected individuals with a greater number of symptoms scored significantly lower on the physical domain of WHOQOL-BREF and in three scales of SF-36 (RP, BP, MH) and MCS (all *p* < 0.05). Patients with less severe symptoms scored significantly higher scores on physical, psychological and environmental domains as well almost all the scales of SF-36. This supports the fourth hypothesis and indicated good known-groups validity of both instruments.

INTER-RATER RELIABILITY

One way of performing reliability testing is to use the intra-class correlation coefficient (ICC). It can be defined as, "the proportion of variance of an observation due to between-subject variability in the true scores". The range of the ICC may be between 0.0 and 1.0. The ICC will be high when there is little variation between the scores given to each item by the raters, e.g. if all raters give the same, or similar scores to each of the items. The ICC is an improvement over Pearson's *γ* and Spearman's *ρ*, as it takes into account of the differences in ratings for individual segments, along with the correlation between raters.[34]

In this study intra-class correlation was calculated by using ICC (2), "Two-Way Random" method which works on two assumptions: (1) it models both an effect of rater and of ratee

Table 7. Comparison of the mean scores in different domains of WHOQOL-BREF and SF-36 for leprosy affected individuals with different symptoms severity

		WHOQOL-BREF				SF-36									
		PHYS	PSYCH	SOC-IAL	ENVIR	PF	RP	BP	GH	VT	SF	RE	MH	PCS	MCS
Severity level	None	67.9	68.8	41.7	52.3	76.3	87.5	83.0	69.3	73.4	62.5	87.5	80.0	51.7	51.4
	Low	55.5	56.6	54.3	44.9	75.7	69.6	61.5	56.8	57.1	60.3	63.0	55.2	49.7	40.3
	High	43.7	48.0	38.3	34.6	66.5	48.9	34.7	39.9	49.7	33.2	52.5	44.8	42.2	34.9
p value		<0.001	<0.001		<0.001		<0.001	<0.001	<0.001		<0.05	<0.05	<0.05	<0.05	<0.05
Number of symptoms	None	67.9	68.8	41.7	52.3	76.3	87.5	83.0	69.3	73.4	62.5	87.5	80.0	51.7	51.4
	1–3	56.4	52.9	42.5	40.9	83.0	75.6	57.7	59.4	63.8	63.8	68.3	64.0	50.1	43.9
	4–7	47.8	52.2	47.6	39.5	67.8	54.7	45.4	45.3	50.5	42.0	54.9	46.1	44.8	35.9
p value		<0.05				<0.001	<0.05						<0.001		<0.001

Table 8. Intra-class correlation coefficient

SF-36 domain	Intra-class correlation[a] Average measures	95% confidence interval		*F* test with true value 0			
		Lower bound	Upper bound	Value	df1	df2	Significance
PF	0.830	0.700	0.903	5.873	49	49	<0.001
RP	0.737	0.537	0.851	3.804	49	49	<0.001
BP	0.809	0.664	0.892	5.239	49	49	<0.001
GH	0.987	0.977	0.993	76.179	49	49	<0.001
VT	0.988	0.978	0.993	81.584	49	49	<0.001
SF	0.976	0.957	0.986	41.101	49	49	<0.001
RE	0.805	0.657	0.890	5.140	49	49	<0.001
MH	0.934	0.884	0.963	15.178	49	49	<0.001

Two-way random effects model where both people effects and measures effects are random. [a]Type C intra-class correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance.

(i.e. two effects) and (2) assumes both are drawn randomly from larger populations (i.e. a random effects model). Mean rating was selected, computing first the mean of each of the 8 domains of SF-36 (PF, RP, BP, GH, VT, SF, RE, MH), for each of the 50 participants in both sets of interviews. The measure of consistency was chosen as this is recommended when comparing means and results are summarized in Table 8.

An intra-class correlation of 0.7 is deemed acceptable, above 0.8 is optimal and a score of above 0.9 would be considered excellent inter-rater reliability. Our results show that for four out of the eight domains of SF-36 inter-rater reliability was excellent, three were in the optimal range and one, physical role (RP) was in the acceptable range. The *p*-values, all under 0.001, were statistically significant.

Discussion

This study assessed the reliability and validity of our Amharic version of the SF-36 by comparing it to the Amharic WHOQOL-BREF in measuring HRQoL in individuals with leprosy.

The Amharic translation of SF-36 was translated following standard procedures.[35] Interviewers reported that it was easy to use and that most participants understood the questions.

Descriptive statistics revealed positively skewed score distributions of the WHOQOL-BREF domains and SF-36 scales indicating more patients scored less than the mean group

scores. Taking into consideration that 69% of patients interviewed were attending hospital because they were unwell, this result would be expected, and would be an indication of validity. This was further supported by the high ceiling effect noted in the PF scale of SF-36 (34%) and 24% for RP and 25% for RE, supporting the theory that our patient group would have a large proportion of people with some limitations in physical functioning, and work/social activities.

Comparison between the Amharic SF-36 and the validated Amharic WHOQOL-BREF found that item internal consistency and item discriminant validity were good. Internal consistency reliability estimates for each domain/scale exceeded 0.70, except for the social domain of WHOQOL-BREF where results were heavily skewed by Question 21 relating to satisfaction in sex life. This is traditionally a taboo subject in Ethiopia, and a subject that is often difficult to discuss with healthcare professionals. The strong correlation between all WHOQOL-BREF domains (except the social domain) with the mental component rather than the physical component may reflect the strong mental health component of the WHOQOL-BREF. The Amharic SF-36 may have better reliability than the Amharic WHOQOL-BREF in this group of participants.

The results of validity examination showed that convergent and discriminant validity for SF-36 inter-scale was better than that for WHOQOL-BREF inter-domain, in this group of participants. Correlations between the scores of corresponding domains/scales between the WHOQOL-BREF and SF-36 supported the first hypothesis that the corresponding domain/scale of both instruments should be positively correlated with the exception that the association between the social domain and SF scale was low ($r = 0.15$). The second hypothesis stating that the physical and psychological domains of the WHOQOL-BREF should have weak associations with MCS and PCS of the SF-36, respectively was partially supported as the correlation between the physical domain and MCS was found to be strong ($r = 0.61$), possibly reflecting the strong mental health component of the WHOQOL-BREF. The third hypothesis that domain/scale scores should be positively correlated to self-perceived health status is better supported with the SF-36 in this group of patients. Good known-group validity for both instruments supported the fourth hypothesis as there was a consistent trend of decreasing scores in the WHOQOL-BREF and SF-36 with increasing severity and number of leprosy related symptoms.

The inter-rater reliability was very good with all the scales scoring between the acceptable and excellent range.[36]

At the commencement of our study, there were two published studies assessing quality of life using SF-36 in leprosy patients in a clinical situation. Both studies were based in Brazil. One assessed the quality of life in 107 patients attending a health facility for leprosy treatment[15] and the second study assessed the quality of life in 49 patients on treatment for paucibacillary (PB) leprosy.[16] The second study found that quality of life scores in 63% of patients with PB leprosy were not affected. Most of these patients were diagnosed early with no leprosy reaction or nerve function impairment. The study of Lustosa *et al.*[15] found that patients with reactions, increased disability grades and a perception of stigma had a significantly lower score in all scales of SF-36. Recently, more studies with people affected by leprosy are reporting the use of SF-36 to assess HRQoL. In Bangladesh, 29 patients with erythema nodosum leprosum (ENL), a chronic and often severe complication of leprosy, had significantly worse HRQoL in all 8 domains of SF-36, compared to 46 leprosy patients without ENL.[37] Similar findings in Brazil, discussed the effect of ENL on the HRQoL of leprosy patients as measured by SF-36.[38] In 2019, a Brazilian study compared SF-36 and DLQI in leprosy patients, finding that SF-36

was the better HRQoL assessment tool covering non-dermatological sequelae of leprosy such as body or nerve pain, and disability.[39] The use of SF-36 is reported in African leprosy patients for the first time here.

The Amharic SF-36 scores in our sample of 100 Ethiopian patients with leprosy were much lower compared to the Ethiopian normative data.[22] The difference was more marked in the scales regarding bodily pain and social functioning. This may be explained by the fact that 81% of patients interviewed were on treatment for leprosy reaction, 31% of which felt unwell on the day of the interview and 46% had severe symptoms. The significant relationship between reduced HRQoL and physical pain has been previously described in other studies.[12] The lower social scores in the social functioning of individuals with leprosy may reflect the stigma that exists in leprosy. The scores in both emotional and physical role scales were lower in leprosy affected people indicating difficulties with work or other activities because of physical health and emotional problems.

In our study, strong correlations were found between higher grades of disability (determined by EHF scores) and lower SF-36 scores, in PF, BP, SF, RE, MH, PCS, and MCS. The correlation between higher level of severity of symptoms and lower HRQoL scores was statistically significant in all the scales of SF-36. This was also mostly true for the number of symptoms experienced and for patients who were unwell on the day of the interview. These differences in scores in the Amharic SF-36 between patient categories indicate that the questionnaire has good construct validity.

The study is not without limitations. Self-administered questionnaires are preferable, as interviewer led questionnaire may score higher,[40] but have better completion rates. Using interviewers is essential in populations were literacy rates and levels of education are low, but conversations and explanations may have influenced some of the answers.

Conclusion

Our Amharic version of the SF-36 is valid and reliable. We are confident that this instrument is useful to measure HRQoL in clinical trials involving Ethiopian participants with leprosy.

List of abbreviations

AHRI	Armauer Hansen Research Institute
ALERT	All Africa Leprosy Rehabilitation and Training Center
BP	Bodily pain
DLQI	Dermatology Life Quality Index
EHF	Eye, hand and foot score
ENL	Erythema nodosum leprosum
Env	Environment domain
GH	General health perceptions
HRQoL	Health-related quality of life
MDT	Multi-drug therapy
MH	Mental health
PCS	Physical component summary
PF	Physical functioning
Phy	Physical domain

Psy	Psychological domain
RE	Role emotional
RP	Role physical
SF	Social functioning
SF-36	36-Item Short Form Health Survey
Soc	Social domain
T1R	Type 1 Reaction
VT	Vitality
WHO	World Health Organisation
WHOQOL-BREF	World Health Organisation Quality of Life Bref

Ethics approval and consent to participate

Ethical approval was obtained as part of the larger clinical trial from the Ethics Committee of the London School of Hygiene and Tropical Medicine (5376), the ALERT and AHRI Ethical Review Committee (AA/ht/248/09), and the National Ethics Review Committee of Ethiopia (RDHE/34-90/2009). Written informed consent was obtained in Amharic. Data was anonymised and stored in a password protected Access database.

Non-commercial licence agreement with Quality metric: QM011524.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SML, PGN conceived the study. SML and SDN supervised the study. SML conducted the analysis. SML, SDN, PGN and DNJL interpreted the findings. All authors read and approved the final manuscript.

Supplementary data

Supplemental information for this article can be found online at
https://doi.org/10.47276/lr.94.1.53.

References

1 Lockwood DNJ. Leprosy. In: Burns DA, Breathnach SM, Cox NH, Griffiths CEM (eds), *Rook's Textbook of Dermatology*. 7th edn., Oxford: Blackwell Publishing, 2004; pp. 1215–1235.

2 World Health Organization. WHO Weekly Epidemiological Review (WER), vol. 95(36). 2020. Available from: https://reliefweb.int/sites/reliefweb.int/files/resources/WER9536-eng-fre.pdf [Internet] [cited 2021 June 25].

3 Chandler DJ, Hansen KS, Mahato B, Darlong J, John A, Lockwood DNJ. Household costs of leprosy reactions (ENL) in rural India. *PLoS Negl Trop Dis*, 2015; **9**(1): e0003431. Available from: http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0003431 [Internet] [cited 2015 September 17].

4 Seddon D, Seeley J. Leprosy and stigma. *Asia Pacific Disabil Rehabilitation J*, 2006; **17**(2): 3–6. Available from: http://english.aifo.it/disability/apdrj/apdrj206/guested-leprosy.pdf [Internet] [cited 2015 September 18].

5 Walker SL, Lebas E, Doni SN, Lockwood DNJ, Lambert SM. The mortality associated with erythema nodosum leprosum in ethiopia: a retrospective hospital-based study. *PLoS Negl Trop Dis*, 2014; **8**(3): e2690. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3953021&tool=pmcentrez&rendertype=abstract [Internet] [cited 2014 May 27].

6 Mieras L *et al.* Global partnership on zero leprosy research agenda working group subgroup on disability. Global Partnership on Zero Leprosy (GPZL). 2019. Available from https://zeroleprosy.org/wp-content/uploads/2019/06/GPZL-RWG-Disability.pdf [Internet] [cited 2021 June 25].

7 Meadows K. Criteria for translations of health measurement instruments. *Qual life Res*, 1994; **3**(1): 67–67. Available from: https://www.jstor.org/stable/i386473 [Internet] [cited 2022 June 20].

8 Brazier JE, Sullivan M, Ware JE. Evaluating translations of health status questionnaires. Methods from the IQOLA project. International Quality of Life Assessment. *Int J Technol Assess Health Care*, 1995; **11**(3): 525–551. Available from: https://pubmed.ncbi.nlm.nih.gov/7591551/ [Internet] [cited 2022 June 20].

9 Joseph GA, Rao PS. Impact of leprosy on the quality of life. *Bull World Health Organ*, 1999; **77**(6): 515–517. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2557686&tool=pmcentrez&rendertype=abstract [Internet] [cited 2013 September 1].

10 Tsutsumi A, Izutsu T, Islam AM, Maksuda AN, Kato H, Wakai S. The quality of life, mental health, and perceived stigma of leprosy patients in Bangladesh. *Soc Sci Med*, 2007; **64**(12): 2443–2453. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17382441 [Internet].

11 Mankar MJ, Joshi SM, Velankar DH, Mhatre RK, Nalgundwar AN. A comparative study of the quality of life, knowledge, attitude and belief about leprosy disease among leprosy patients and community members in Shantivan Leprosy Rehabilitation centre, Nere, Maharashtra, India. *J Glob Infect Dis*, 2011; **3**(4): 378–382. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3249995&tool=pmcentrez&rendertype=abstract [Internet] [cited 2014 May 5].

12 Costa MD, Terra F de S, Costa RD, Lyon S, Costa AMDD, Antunes CM de F. Assessment of quality of life of patients with leprosy reactional states treated in a dermatology reference center. *An Bras Dermatol*, 2012; **87**(1): 26–35. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22481648 [Internet] [cited 2013 August 25].

13 An J-G, Ma J-H, Xiao S-X, Xiao S-B, Yang F. Quality of life in patients with lepromatous leprosy in China. *J Eur Acad Dermat Venereol*, 2010; **24**(7): 827–832. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20015055 [Internet] [cited 2013 August 11].

14 Proto RS, Machado Filho CDS, Rehder JRCL, Paixão MP, Angelucci RI. Quality of life in leprosy: a comparative analysis between patients in the Amazon region and patients in Santo André in the ABC region of São Paulo, Brazil. *An Bras Dermatol*, 2010; **85**(6): 939–941. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21308328 [Internet] [cited 2013 August 26].

15 Lustosa AA, Nogueira LT, Pedrosa JIDS, Teles JBM, Campelo V. The impact of leprosy on health-related quality of life. *Rev Soc Bras Med Trop*, 2011; **44**(5): 621–626. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0037-86822011000500019&lng=en&nrm=iso&tlng=en [Internet] [cited 2013 September 2].

16 Bottene IMC, Reis VMS dos. Quality of life of patients with paucibacillary leprosy. *An Bras Dermatol*, 2012; **87**(3): 408–411. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22714756 [Internet] [cited 2014 May 5].

17 WHO. World Health Orgaization. *Division of Mental Health and Prevention of Substance Abuse: WHOQOL User Manual*. Geneva, Switzerland, 2012.

[18] Araya M, Chotai J, Komproe IH, de Jong JTVM. Quality of life after postconflict displacement in Ethiopia: comparing placement in a community setting with that in shelters. *Soc Psychiatry Psychiatr Epidemiol*, 2011; **46**(7): 585–593. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20383488 [Internet] [cited 2013 August 23].

[19] Mousley E, Deribe K, Tamiru A, Davey G. The impact of podoconiosis on quality of life in Northern Ethiopia. *Health Qual Life Outcomes*, 2013; **11**: 122–134. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3726315&tool=pmcentrez&rendertype=abstract [Internet] [cited 2013 August 26].

[20] Henok L, Davey G. Validation of the dermatology life quality index among patients with podoconiosis in southern Ethiopia. *Br J Dermatol*, 2008; **159**(4): 903–906. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18671781 [Internet] [cited 2013 September 1].

[21] Ware JE. *SF-36 Health Survey: Manual and Interpretation Guide*. Boston, MA: The Health Institute, New England Medical Center, 1993; p. 1993.

[22] Kebede D, Alem A, Shibre T, Negash A, Deyassa N, Beyero T. Health related quality of life (SF-36) survey in Butajira, rural Ethiopia: normative data and evaluation of reliability and validity. *Ethiop Med J*, 2004; **42**(4): 289–297. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16122121 [Internet].

[23] Kebede D, Alem A, Shibre T, Negash A, Deyassa N, Beyero T *et al.* Short-term symptomatic and functional outcomes of schizophrenia in Butajira, Ethiopia. *Schizophr Res*, 2005; **78**(2–3): 171–185. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16112845 [Internet] [cited 2014 May 7].

[24] Kebede D, Alem A, Shibire T, Deyassa N, Negash A, Beyero T *et al.* Symptomatic and functional outcome of bipolar disorder in Butajira, Ethiopia. *J Affect Disord*, 2006; **90**(2–3): 239–249.

[25] Abera K, Gedif T, Engidawork E, Gebre-Mariam T. Quality of life of people living with HIV/AIDS and on highly active antiretroviral therapy in Ethiopia. *African J AIDS Res*, 2010; **9**(1): 31–40. Available from: http://www.informaworld.com/openurl?genre=article&doi=10.2989/16085906.2010.484560&magic=crossref%7C%7CD404A21C5BB053405B1A640AFFD44AE3 [Internet] [cited 2013 August 25].

[26] Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE *et al.* 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*, 1998; **51**(11): 1171–1178. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9817135 [Internet] [cited 2013 August 26].

[27] Hsiung P-C, Fang C-T, Chang Y-Y, Chen M-Y, Wang J-D. Comparison of WHOQOL-bREF and SF-36 in patients with HIV infection. *Qual Life Res*, 2005; **14**(1): 141–150. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15789948 [Internet] [cited 2013 August 27].

[28] Araya M, Chotai J, Komproe IH, de Jong JTVM. Effect of trauma on quality of life as mediated by mental distress and moderated by coping and social support among postconflict displaced Ethiopians. *Qual Life Res*, 2007; **16**(6): 915–927. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17440829 [Internet] [cited 2013 August 26].

[29] Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*, 2000; **25**(24): 3186–3191. Available from: https://pubmed.ncbi.nlm.nih.gov/11124735/ [Internet] [cited 2022 June 20].

[30] Price VE, Klaassen RJ, Bolton-Maggs PHB, Grainger JD, Curtis C, Wakefield C *et al.* Measuring disease-specific quality of life in rare populations: a practical approach to cross-cultural translation. *Health Qual Life Outcomes*, 2009; **7**(1): 1–8. Available from: https://link.springer.com/articles/10.1186/1477-7525-7-92 [Internet] [cited 2022 June 20].

[31] Brandsma JW, Van Brakel WH. WHO disability grading: operational definitions. *Lepr Rev*, 2003; **74**(4): 366–373. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14750582 [Internet] [cited 2013 September 1].

[32] Gandek B, Ware JE, Aaronson NK, Alonso J, Apolone G, Bjorner J *et al.* Tests of data quality, scaling assumptions, and reliability of the SF-36 in eleven countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol*, 1998; **51**(11): 1149–1158. Available from: http://www.jclinepi.com/article/S0895-4356(98)00106-1/abstract [Internet] [cited 2014 May 11].

[33] McHorney CA, Ware JE, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care*, 1993; **31**(3): 247–263. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8450681 [Internet] [cited 2014 April 28].

[34] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*, 1979; **86**(2): 420–428. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18839484 [Internet] [cited 2013 August 7].

[35] Acquadro C, Conway K, Hareendran A, Aaronson N. Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. *Value Heal*, 2008; **11**(3): 509–521. Available from: https://pubmed.ncbi.nlm.nih.gov/18179659/ [Internet] [cited 2021 January 24].

[36] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*, 1977; **33**(1): 159–174.

37 Bowers B, Butlin RC, Alam K, Lockwood DNJ, Walker SL. Health-related quality of life amongst people affected by Erythema Nodosum Leprosum in Bangladesh: a cross sectional study. *Lepr Rev*, 2017; **88**: 488–498. Available from: https://www.lepra.org.uk/platforms/lepra/files/lr/Dec17/Lep488-498.pdf [Internet] [cited 2019 September 26].

38 Sales A, Illarramendi X, Walker SL, Lockwood DNJ, Sarno EN, Nery JA. The impact of Erythema Nodosum Leprosum on health related quality of life in Rio de Janeiro. *Lepr Rev*, 2017; **88**: 499–509. Available from: https://www.lepra.org.uk/platforms/lepra/files/lr/Dec17/Lep499-509.pdf [Internet] [cited 2019 September 26].

39 Guimenes Albuquerque R, Grüdtner Buratto G, Hirotsu C, Maeda SM, Floriano MC, Levy Andersen M *et al.* Comparison of quality of life evaluated by SF-36 and DLQI in multibacillary and paucibacillary leprosy patients from Sao Paulo, Brazil. *Int J Dermatol*, 2019; **58**(12): 1415–1422. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/ijd.14489 [Internet] [cited 2019 September 26].

40 Açma A, Carrat F, Hejblum G. Comparing SF-36 scores collected through web-based questionnaire self-completions and telephone interviews: An ancillary study of the SENTIPAT multicenter randomized controlled trial. *J Med Internet Res*, 2022; **24**(3): e29009. Available from: https://pubmed.ncbi.nlm.nih.gov/35266869/ [Internet] [cited 2022 June 20].

41 Lambert SM, Alembo DT, Nigusse SD, Yamuah LK, Walker SL, Lockwood DNJ. A randomized controlled double blind trial of ciclosporin versus prednisolone in the management of leprosy patients with new type 1 reaction, in Ethiopia. *PLoS Negl Trop Dis*, 2016; **10**(4): e0004502. Available from: https://pubmed.ncbi.nlm.nih.gov/27046330/ [Internet] [cited 2021 June 25].

42 Lambert SM, Nigusse SD, Alembo DT, Walker SL, Nicholls PG, Idriss MH *et al.* Comparison of efficacy and safety of ciclosporin to prednisolone in the treatment of erythema nodosum leprosum: Two randomised, double blind, controlled pilot studies in Ethiopia. *PLoS Negl Trop Dis*, 2016; **10**(2): e0004149. Available from: https://pubmed.ncbi.nlm.nih.gov/26919207/ [Internet] [cited 2021 June 25].