An approach to validation of a multi-dimensional tool

Gunawardena NS, Seneviratne R de A, Atauda TA

Abstract

Introduction

Short Form Health Survey-36 (SF-36) is a multi-dimensional measure of general health which has been used extensively in research on many population groups.

Objective

To validate SF-36 among lower limb amputee soldiers and a matched group of males, using triangulation.

Methodology

SF-36 was assessed among amputee soldiers (135) and matched healthy males (135) for judgmental, convergent-discriminant and construct validity. **Judgmental validity** assessed appropriateness of translation of conceptual definitions and cultural suitability. Multi-trait Multi-method Matrix technique assessed **convergent-discriminant validity** against another accepted measure of general health, the Nottingham Health Profile (NHP). **Construct validity** was assessed by checking whether: SF-36 distinguished expected differences between and within groups; and eight dimensions of SF-36 emerged from confirmatory factor analyses of data of populations studied. Cronbach's alpha assessed reliability.

Results

Judgmental validity was established. Correlations of similar dimensions of SF-36 and NHP were strong while dissimilar dimensions were weak confirming convergent-discriminant validity. Expected differences between and within groups were seen for scores of SF-36. Five factors among amputee soldiers and six factors among the comparison group were derived from confirmatory factor analyses, which were similar to dimensions of SF-36. Cronbach's alpha for all dimensions exceeded 0.8 for both groups.

Conclusion

Triangulation proved that SF-36 was equally valid for both groups. This approach can be adopted to validate multi-dimensional instruments for cross-cultural research where criterion validity cannot be assessed.

Key words

Validity, multi-dimensional tool

Introduction

Short Form Health survey 36 (SF-36), a multi-dimensional measure of general health status covering all the aspects of health was used in a study to assess the general health status of amputee soldiers and age matched males. It consists of 36 questions or items which forms multi-item scales to measure the eight dimensions of health.

The dimensions of health measured by SF-36 are: physical functioning (PF); role limitation due to physical health problems (RP); bodily pain (P); social functioning (SF); general mental health (MH); role limitation due to emotional problems (RE); vitality, energy or fatigue (V); general health perceptions (GH).

A specimen form of SF-36 and the numbers of questions related to each of its dimensions are annexed (Annex I and Annex II).

Validity of SF-36 has not been assessed in Sri Lanka, though it has been found to be valid in many other countries such as the United States of America, the United Kingdom and France etc. (1).

A 'gold standard' measure for general health cannot be defined and therefore establishing criterion validity by calculating sensitivity and specificity is not possible for this measure. In a situation where criterion validity is not possible, triangulation, i.e. the use of several complementary validation methods, provides the most accurate assessment of the instrument's properties (2).

This study describes the methodology and results for validating a self-administered version of SF-36 using triangulation. The techniques used were assessing judgmental, convergent-discriminant, and construct validity. Reliability was assessed by appraising internal consistency.

Methodology

Study populations

Validity of SF-36 was assessed in the following two study populations.

Male soldiers in Sri Lanka Army who have undergone amputation of one or more lower limbs at transtibial or transfemoral level due to an injury at war and not residing in the study areas

A group of males from general population resident outside the study area, who were within the age limits of 20-50 years

Sample size

The sample size was calculated based on the assessment of convergent and discriminant validity which is a form of validity measured in this study.

The smallest expected correlation coefficient was used for the calculation of the highest required sample size (3).

$N = [(Z_\alpha + Z_\beta) / C]^2 + 3$

$C = 0.5 \times \ln[(1+r) / (1-r)]$

C= correction factor

r= expected correlation coefficient

$\alpha = 0.05$

$\beta = 0.01$

N= total number of subjects required

1. Lecturer, 2. Professor 3. SeniorLecturer, Department of Community Medicine, Faculty of Medicine, University of Colombo

By reviewing the available literature on assessing validity of SF-36 it was decided to use 0.25 as r the smallest correlation coefficient that should be expected for the proposed validation study. Estimated sample size was 135 for each group.

Identification of study units

A convenient sample of eligible amputee soldiers employed in Army headquarters and residing in a 'Ranaviru' village in Colombo district were selected.

The other study population of healthy males in the age groups of 20-50 years was selected from Pita Kotte MOH area. Each of 12 Public Health Midwives (PHMM) areas were used as clusters. From each cluster 12-13 eligible males were selected. The study was conducted on Sundays and public holidays to ensure the presence of employed males in the houses.

Data collection

The instrument SF-36 was translated into Sinhala by consensus of a panel of people fluent in English and Sinhala ensuring that the questions would retain the conceptual meaning. Some words were modified to be culturally suitable. Clear simple language was used to cater to respondents' varying levels of education. It was pre-tested for clarity and acceptance. This translated and pre-tested instrument was used for validation.

Both SF-36 and NHP were incorporated one after the other into a self-administered questionnaire. Questions were also asked on socio-demographic information, recent use of health services and existing chronic illnesses. Special care was taken to ensure smooth flow of questions. An assurance of confidentiality and the purpose of the study were given in simple language on the first page to minimize non-response.

The selected study groups were requested to fill the questionnaire. PHMM of the area and the two soldiers employed in Army headquarters, who distributed the questionnaires were instructed to ensure that questionnaires were to be filled by the respondents themselves. Study investigators were asked to make sure that the questions were answered in a suitable place without disturbances. They were asked to collect the filled forms preferably immediately or within the same day.

Types of validity assessed

a. Judgmental validity

Face, content and consensual validity were the types of validity assessed by judgment (4). Face validity was assessed by appraising the relevance of the tool to the subject under study. Content validity was assessed by checking whether or not all aspects of the measure were covered and consensual validity was determined by assessing the agreement of the experts on whether or not the conceptual definition has been translated appropriately into operational terms in the tool. A panel comprising multidisciplinary experts in the fields of Community Medicine, Clinical Medicine and Psychological Medicine assessed the judgmental validity.

b. Convergent and discriminant validity

Convergent and discriminant validity is assessing the validity of the tool against another known measure (4).

Convergent validity is the degree to which dimensions in the two measures that are related theoretically are interrelated in reality for which the evidence is strong correlation between related dimensions. Discriminant validity is the degree to which dimensions that are not related theoretically are, in fact, not interrelated in reality for producing weak correlations. Both types of validity can be examined simultaneously by Multi-trait Multi-method Matrix technique (MTMM) (4,5).

In this study, another measure of general health status, The Nottingham Health Profile (NHP) which has been validated and used in Sri Lanka (6), was used for this purpose. NHP is also a measure of general health status on a multi-item multi-dimension scale, some of which are comparable with the dimensions of SF-36.

c. Construct validity

Construct validity is assessing whether the tool being validated is able to produce results expected by the researcher (1,4).

Table 1 : Distribution of some of the socio-demographic characteristics of amputee soldiers and comparison group

Socio-demographic characteristics	Amputee soldiers n=135	%	Comparison group n=135	%	Significance
Age (years)					
20-29	74	54.8	70	51.9	χ^2=4.26
30-39	57	42.2	53	39.3	df=2
40-49	4	2.9	12	8.8	p=0.12
Marital status					
Married	85	63	89	65.9	χ^{2*}=1.0
Unmarried	47	34.8	41	30.4	df=2
Separated	3	2.2	5	3.7	p=0.61
Highest educational level achieved					
Grade 6-9	47	34.8	41	30.4	χ^2=7.37
O/L passed	77	57	68	50.3	df=3
A/L passed	6	4.4	17	12.6	p=0.06
Diploma/degree	5	3.7	9	6.7	

* Yates corrected

This was assessed in two ways.

Firstly, this was assessed by checking whether the results of the validation study confirmed the hypotheses of health between and within the group that was examined. It was hypothesized that the health status of amputees especially in dimensions related to physical health would be lower than the comparison group. Within the group of amputees physical health status of above knee amputees was hypothesized to be lower than below knee amputees. Health status of those with a history of chronic illness within the comparison group was hypothesized to be lower than those who had no chronic illness.

Secondly, construct validity was assessed by performing a confirmatory factor analysis on the results of the validation study. Confirmatory factor analysis is a technique of psychometric validation that re-confirms the agreement between hypothetical dimensions that are included in a multi-dimensional tool and the scale designed to measure these dimensions.

If SF-36 is a valid measure of the dimensions within it, these dimensions should emerge from a factor analysis of the data of the population that it is being validated for and items relating to a particular dimension should be grouped together within a single factor (1,7).

d. Reliability- Internal consistency

The stability or consistency of information is the extent to which similar information is gathered when measured more than once. If reliability is low the measure cannot have high validity. Therefore validation of a tool should include information on reliability as well.

In a composite scale reliability can be measured by appraising internal consistency. Internal consistency measures the extent to which similar questions produce consistent answers. This is the extent to which items within a dimension are correlated with each other. If the items on a scale are measures of the same attribute, the extent to which they give the same results is a function of their reliability (4).

Cronbach's alpha which is a measure of the overall correlation between items within a dimension was used to check whether it exceeded the acceptable Nunnally's criterion of 0.7 (8).

Results

Characteristics of the samples

Distribution of some of the socio-demographic characteristics of amputee soldiers and comparison group is shown in Table 1.

The groups were comparable in basic socio-demographic characteristics.

Response rate was 100% among the amputee soldiers and 90% among the comparison group. All respondents from both groups were able to read and write. The rate of completion for each dimension exceeded 90% and for most of the dimensions, the rate of completion was 100%.

Convergent and discriminant validity

Table 2 : MTMM of correlation coefficients for SF-36 and NHP for amputee soldiers

	SF-36					NHP				
	Physical Functioning	Social functioning	pain	Vitality	Mental health	Physical mobility	Social isolation	Pain	Energy	Emotional reactions
SF-36										
Physical functioning										
Social functioning	.331									
Pain	.448	.402								
Vitality	.412	.385	.338							
Mental Health	.295	.401	.303	.415						
NHP										
Physical mobility	**-.689**	-.306	-.437	-.290	-.219					
Social isolation	-.313	**-.567**	-.373	-.316	-.292	.376				
Pain	-.413	-.235	**-.74**	-.304	-.187	.410	.363			
Energy	-.281	-.308	-.282	**-.576**	-.323	.292	.243	.406		
Emotional reactions	-.312	-.293	-.341	-.239	**-.785**	.265	.312	.318	.305	

correlation coefficients are negative because the two scales run in the opposite direction

Table 3 : MTMM of correlation coefficients for SF-36 and NHP for comparison group

	SF-36					NHP				
	Physical Functioning	Social functioning	pain	Vitality	Mental health	Physical mobility	Social isolation	Pain	Energy	Emotional reactions
SF-36										
Physical functioning										
Social functioning	.293									
Pain	.385	.401								
Vitality	.382	.306	.353							
Mental Health	.259	.413	.365	.411						
NHP										
Physical mobility	**-.673**	-.294	-.402	-.271	-.223					
Social isolation	-.364	**-.565**	-.371	-.303	-.293	.363				
Pain	-.396	-.288	**-.626**	-.316	-.203	.327	.360			
Energy	-.295	-.326	-.225	**-.596**	-.399	.293	.236	.401		
Emotional reactions	.345	-.234	-.306	-.338	**-.724**	.251	.284	.304	.319	

correlation coefficients are negative because the two scales run in the opposite direction

Multi-trait multi-method matrix of correlation coefficients for SF-36 and Nottingham Health Profile for amputee soldiers and comparison group are shown in Table 2 and 3 respectively.

The following are the basic principles to interpret MTMM

- Correlations between measures of similar dimension measured using different scales (in bold typing) should be high indicating convergent validity.
- Correlations between dissimilar dimensions of the two scales should be the

lowest in the matrix, confirming the measure being validated is

showing discriminant validity

In the MMTM for amputee soldiers, correlations between similar dimensions were high (r=0.567-0.785) and correlations between dissimilar dimensions were low (r=0.187-0.448) (Table 2).

Correlations between similar dimensions were high (r=0.565-0.724) and correlations between dissimilar dimensions were low (r=0.203-0.413) in the MMTM for the comparison group (Table 3).

Construct validity

Descriptive analysis

Mean scores (SD) for SF-36 amputee soldiers and comparison group are shown in Table 4.

Table 4 shows that scores of all dimensions are lower for amputee soldiers when compared to the comparison group. In the dimensions of PF, SF RP, RE, P and GH and these differences were statistically significant ($p<0.001$). Differences in the dimensions of MH and V were not statistically significant ($p>0.05$).

Tables 5 and 6 show mean scores on dimensions of SF-36 in relation to selected socio-demographic variables and use of health services among amputee soldiers and the comparison group, respectively.

Among the amputee soldiers and comparison group,

Table 4 : Mean scores (SD) SF-36 for amputee soldiers and comparison group

Group	PF	SF	RP	RE	P	MH	V	GH
Amputee soldiers	68.5*** (11.14)	74.87*** (11.35)	66.38*** (9.44)	64.92*** (10.77)	63.4*** (8.55)	69.82** (13.26)	73.19* (12.01)	51.73*** (8.13)
Comparison group	80.18*** (10.57)	79.52*** (9.92)	80.42*** (13.07)	80.74*** (10.56)	78.65*** (9.78)	70.33** (12.84)	75.27* (10.18)	59.78*** (7.64)

*** $p< 0.001$ ** $p> 0.05$ * $0.1 > p> 0.05$

Table 5- Mean scores on dimensions of SF-36 in relation to some socio-demographic variables and use of health services in amputee soldiers

Constructs	SF-36 scale									
	n	%	PF	SF	RP	RE	P	MH	V	GH
Age (years)										
20-29	74	54.8	68.06	81.28	65.14	63.89	66.67	70.08	78.89	50.98
30-39	57	42.2	70.82	74.23	67.51	69.09	69.78	69.71	74.55	52.73
40-49	4	2.9	60	70.03	65.85	65.13	66.67	65.33	65.12	53.33
Recent use of health services										
Yes	36	26.67	66.38	73.43	65.27	67.67	68.77	66.32	67.65	50.56
No	99	73.3	71.9	72.02	69.36	61.01	64.61	71.05	72.45	54.73
Level of amputation										
Above knee	6	4.4	44.33	60	50	56.67	44.53	68.41	67.23	44.45
Below knee	129	95.6	70.52	78.67	69.04	68.53	69.6	69.87	71.36	52.04
Long standing illness										
Yes	20	14.8	69.65	74.34	66.44	66.67	63.18	62.63	70.53	44.21
No	115	85.2	68.63	82.66	66.05	66.78	68.78	69.33	71.31	51.31

those who reported a longstanding illness had scored less for all dimensions of SF-36 when compared to those who did not report any long standing illness (Tables 5 and 6).

Above knee amputees had scored less for all dimensions of SF-36 when compared to below knee amputees and the differences were more in the dimensions related to physical health (Table 5).

Confirmatory factor analysis

Principal component analyses were performed on both study and comparison groups. Factorability of the data was assessed and was confirmed. Factor coefficients of individual questions for amputee soldiers and for the comparison group are shown in Tables 7 and 8 respectively.

Factor analysis for amputee soldiers had identified five relevant factors with Eigenvalues ranging from 12.15 to 1.66. Each item had a loading of greater than 0.4 only in the factor that it belonged to. The dimensions of health identified by factor analysis matched the three dimensions; physical functioning, role limitation due to physical problems and general health perceptions postulated by the author of the tool. Dimensions of mental health and vitality had combined to make up a common dimension while bodily pain, social functioning and role limitation due to emotional problems had combined to make up a single dimension (Table 7).

Factor analysis for the comparison group identified six relevant factors. Eigenvalues ranged from 7.61 to 1.46. Each item had a loading greater than 0.4. The dimensions of health identified by factor analysis matched the four dimensions; physical functioning, role limitation due to physical problems, role limitation due to emotional problems and general health perceptions postulated by the author of the tool. Dimensions of Mental Health and vitality had combined to make up a common dimension while bodily pain and social functioning had combined to make up a single dimension (Table 8).

Internal consistency- Cronbach's alpha

Table 6 : Mean scores on dimensions of SF-36 in relation to some socio-demographic variables and use of health services in comparison group

Constructs	SF-36 scale									
	n	%	PF	SF	RP	RE	P	MH	V	GH
Age (years)										
20-29	70	51.9	70.73	81.41	97.32	83.33	74.38	74.34	76.61	59.29
30-39	53	39.3	84.14	77.32	80.13	83.88	71.28	68.91	78.23	61.33
40-49	12	8.8	75.62	73.81	59.52	66.67	74.41	65.14	65.24	54.18
Recent use of health services										
Yes	29	21.5	74.25	80.68	76.96	83,59	78.76	66.43	71.11	54.82
No	106	78.5	81.83	77.02	84.08	70.24	78.24	70.49	75.58	61.13
Long standing illness										
Yes	28	20.7	70.25	76.73	56.25	72.62	78.37	68.43	65.36	47.14
No	107	79.25	82.74	81.11	87.72	83.27	75.53	64.24	68.43	63.38

Table 7 : Internal consistency (Cronbach's alpha) of SF-36 among amputees and comparison group

Dimensions of SF-36	Amputee soldiers	Comparison group
Physical functioning	.8834	.899
Social functioning	.8437	.8662
Role limitations (physical problems)	.9114	.9487
Role limitations (emotional problems)	.8391	.8231
Pain	.9565	.9369
Mental health	.9015	.9413
Vitality	.8491	.8925
General health perception	.8331	.8240

Internal consistency of SF-36 among amputees and comparison group is shown in Table 9.

Cronbach's alphas were high for all dimensions of SF-36 among amputees (0.83-0.96) and among the comparison group (0.82-0.95).

Discussion

This study validated SF-36 using triangulation, It has been shown that the use of several complementary validation methods, provide the most accurate assessment of the instrument's properties (2).

SF-36 was validated among both study groups that it was intended to be used.

It would be advisable to validate a tool among all the groups that it would be used on to exclude the possibility of differential validity. If the validity of the tool is not uniform in two groups the results obtained by using the instrument would produce deceptive information about association between them (4).

Acceptable response rates of 100% among the amputee soldiers and 90% among the comparison group were shown. Acceptable response rates were 72%-83% in other studies of validation of SF-36 (7,9,10).

Judgmental validity of the scale was established by the team of experts who confirmed that the conceptual definition has appropriately been translated into operational terms.

Basic principles of interpretation of MTMM were fulfilled in both amputee soldiers and the comparative group to satisfy expected relations for convergent and discriminant validity. In a validation study of SF-36 in the U.S.A. which used the NHP and the MTMM method, the expected relations for convergent and discriminant validity were mostly satisfied (7).

The distribution of scores conformed to what was expected as evidence of construct validity. All dimensions of health especially physical functioning and role limitation due to physical health was hypothesized to be lower among amputee soldiers when compared to the comparison group.

Regarding the hypothesis that those who used health services recently and those who are suffering from long standing illnesses would score lower in all dimensions of health measured, the predicted pattern was seen in both amputee soldiers and the comparison group for most of the dimensions.

The health of above knee amputees was expected to be lower than below knee amputees, across all dimensions. The predicted pattern was shown.

These were taken as evidence for construct validity in terms of distinguishing between groups and within group expected health differences.

Confirmation of hypothesized patterns of scores confirmed construct validity in other studies of validation of SF-36 (7,9,10).

Five factors among amputees and six among the comparison group were derived from confirmatory factor analyses, which were similar to dimensions of SF-36. Factors were selected depending on the eigenvalues. The factors were considered relevant only if its eigenvalue exceeded 1.0 (11,12,13).

The results of confirmatory factor analysis of data of amputee soldiers and comparison group was taken as evidence of construct validity of the tool in both amputee soldiers and the comparison group.

Another validation study of SF-36 identified five relevant factors with eigenvalues ranging from 12.8 to 1.3 (7). They concluded that this was evidence for validity of SF-36 as precise correspondence between factors and scales is rare in factor analysis this was evidence for validity of SF-36.

Cronbach's exceeded Nunnally's criterion of 0.7 for all dimensions in both amputee and comparison group, confirming the reliability.

Other studies of validation of SF-36 have confirmed the reliability with high values for Cronbach's alpha which exceeded the Nunnally's criterion (7,9,10).

Conclusion of the validation study

Triangulation of several methods have proved that SF-36 is an acceptable and valid measure to be used to measure the general health status of amputee soldiers and a comparison group. Validity is uniform in both groups.

Acknowledgments

We sincerely thank IDA/WB project for providing funds.

We sincerely thank the Commander of Sri Lanka Army, staff at Adjutant Office and the staff at Directorate of Rehabilitation for their corporation towards success of this study. Special words of

thanks are due to Col. Sriyani Warusawithna for her valuable guidance.

References

1. McDowell I, Newell C. *Measuring health: a guide to rating scales and questionnaires.* 2nd ed. Oxford: University Press Oxford 1996.
2. Etter JF, Perneger TV. Validating a satisfaction questionnaire using multiple approaches: a case study. *Social Sciences Medicine* 1997; 45 (6): 879-885.
3. Hulley SB, Cummings SR. *Designing clinical research; an epidemiological approach.* Baltimore, USA: Williams and Wilkins 1988.
4. Abramson JH, Abramson ZH. *Survey methods in Community Medicine.* 5th ed. Edinburgh: Churchill Livingstone 1999.
5. Trochim WMK. Measurement validity types [online]. Available from: http://trochim.human.cornell.edu/kb/measval.html 2001.
6. Puvimanasinghe JPA. *Prevalence and disability due to rheumatoid aArthritis in the Kalutara district and the development and testing of a health education package for such patients.* Thesis (MD Community Medicine). Post Graduate institute of medicine. University of Colombo 1998.
7. Brazier JE, Harper R, Jones NMB, O'cathain A, Thomas KJ, Usherwood T et al. Validating the SF 36 health survey questionnaire : a new outcome measure for outcome care. *British Medical Journal* 1992;,305: 160-164.
8. Bland JM, Altman DG. Statistical notes; Cronbach's alpha . *British Medical Journal* 1997; 314: 572.
9. Jenkinson C, Coulter A, Wright I. Short form 36 health survey questionnaire: normative data for adults of working age. *British Medical Journal* 1993; 306:1437-1440.
10. Garratt AM, Ruta DA, Abdalla MI, BuckinghamJK, Russell IT. The SF- 36 health survey questionnaire: an outcome measure suitable for routine use within the National Health Service. *British Medical Journal* 1993; 306: 1440-1444.
11. Armitage P, Berry G. *Statistical methods in medical research.* 3rd ed. Oxford: Blackwell Scientific publications 1994.
12. Jaeger RM. *Statistics- a spectator sport.* Newbury Park: Sage 1994.
13. McKeigen ID, Larson LN. Development and validation of an instrument to measure patient satisfaction with pharmacy services. *Medical Care* 1989; 27 (5): 523-536.

Which of the following statements are true or false?

Mark your answer with a X.

1.	Validity of SF-36 would not be affected even if reliability was found to be low in this study.	T	F
2.	Cronbach's ∝ measured overall correlation between, different dimensions of SF-36 in this validation study.	T	F
3.	Reliability is a measure of consistency of results when measured more than once using the tool.	T	F
4.	Cronbach's ∝ could be used in this study to assess reliability of SF-36 as responses were on a composite scale.	T	F
5.	Other selected tests of significance that could be used to assess the degree of reliability in this study is kappa coefficient.	T	F
6.	If the validity has been established in another country it is not necessary to validate the tool again.	T	F
7.	The instrument should be translated into the language that it is expected to be used later, prior to validation.	T	F
8.	To assess construct validity, study population for validation should consist of different subgroups of population that the tool is expected to be used.	T	F
9.	Content validity is an assessment of the extent of inclusion of the content domain under study.	T	F
10.	Response rate in a validation study is not relevant to assessment of validity.	T	F
11.	In the absence of a gold standard to assess criterion validity, triangulation is appropriate to assess validity.	T	F
12.	In this study, Nottingham Health Profile was used to assess convergent and discriminant validity.	T	F
13.	Faced, content and consensual validity are assessed by using statistical methods.	T	F
14.	Construct validity is an assessment of whether hypothesized differences in the study are confirmed by the results.	T	F
15.	Accepted measure of Cronbach's ∝ is Nunnally's criterion.	T	F
16.	Following establishment of validity, this study instrument can be used on populations with any physical disability.	T	F

See page 26 for results.

Answers for CME article questions on page 25

Please give yourself 1 mark if the answer is correct and 0 if the answer is incorrect.

Q No:	Correct answer	Score
1	F	
2	F	
3	T	
4	T	
5	T	
6	F	
7	T	
8	T	

Q No:	Correct answer	Score
9	T	
10	F	
11	T	
12	T	
13	F	
14	T	
15	T	
16	F	

Total

Grading: Score 0-4 = Very poor
5-7 = Poor
8-13 = Good
14-16 = Excellent