ORIGINAL ARTICLE

Validity and Reliability of the SF-36: The Malaysian Context

S Sararaks, MPH*, A B Azman, PhD.*, L L Low, MA (Med. Anthropology), B Rugayah, MPH***, A M Aziah, MRCP****, L N Hooi, FRCP*****, M Abdul Razak, FCCP******, M R Norhaya, M.Sc (Respiratory Medicine)*******, K B Lim, MRCP****, A A Azian, MPH*, S Geeta, MBBS********

*Public Health Specialist, Institute for Health Systems Research, 50590 Kuala Lumpur, **Research Officer, Institute for Health Systems Research, 50590 Kuala Lumpur, *** Public Health Specialist, Head, Evidence Based Medicine Unit, Clinical Research Centre, Kuala Lumpur Hospital, ****Consultant Physician, Kuala Lumpur Hospital, *****Consultant Chest Physician, Public Specialist Centre, Penang, ******Consultant Chest Physician, Penang Hospital, *******Consultant Respiratory Physician, Kuala Terengganu Hospital, ********Medical Officer, Klang Hospital

Summary

Results of construct validity and reliability of the SF-36 are described, based on data from a multi-centre study on asthmatics and a population based survey. Questionnaire refinement was carried out between the two studies. Quality of data was good, with all items having less than 0.5% missing values. Floor and/or ceiling effects were observed for REE, REP, PF and SF. For scaling assumptions, correlations between each items and its hypothesized scale were all above 0.50, except for one item in PF, and for both items in SF. Item discriminant validity was an issue for items in VT, SF and MH scales. Cronbach's αs for all scales exceeded the recommended 0.70 level, except for SF. Only one latent dimension was identified in principal component analysis, and only 52-53% of variance accounted for. As expected, PF shows high correlations with the physical component while MH was highly correlated with the mental component. Contrasting findings in the loadings of other scales were observed in the asthma data. Age, disease severity and presence of self-reported handicap/disability significantly affect PF, while MH demonstrates no obvious pattern with declining age. In essence, the Malay version of SF-36 could be used in Malaysia, with its generally acceptable internal consistency and validity. The caveat is in the call for additional domains of importance to Malaysians that is not covered by the instrument, and in the caution to be employed when using and construing the instrument.

Key Words: Validity, Construct validity, Reliability, Internal consistency, Health-related quality of life, SF-36

Introduction

"Health-related quality of life (HRQoL)" is defined as an individual's satisfaction or happiness with domains of life as far as they affect or are affected by health. It is distinguished from Quality of Life (QoL) in that HRQoL is concerned primarily with those factors that fall under the purview of health care providers and health care systems [1], and in health care, sometimes preferred to that of QoL [2, 3].

Measurement of HRQOL is through the use of questionnaire. Amid those instruments which are widely used in health care, the SF-36 (Short form – 36) is one of the most extensively used, widely translated and tested instrument worldwide. It is a generic outcome measure of sickness [4] based on 36-items selected to represent eight health concepts (physical, social and role functioning, mental health, health perceptions, energy fatigue, pain and general health)[4]. The SF-36 is reported to be a sensitive measure for outcome of care

This article was accepted: 26 September 2004

Corresponding Author: Sondi Sararaks, Institute for Health Systems Research, c/o Institute of Public Health, Jalan Bangsar, 50590 Kuala Lumpur

for numerous diseases[4,5] and also sensitive to changes in health in general populations[6,7]. The developers have also methodically documented the validity and reliability of this instrument in other countries[8-12].

Currently there are no published information on the validity and reliability of SF-36 in the Malaysian context. Though there are extensive publications elsewhere on its validity and reliability, given the cultural, social and ethnic differences between Eastern and Western cultures, the SF-36 is postulated to function differently in Malaysia. Therefore, if the SF-36 is to be used in the study of health outcomes such as HRQoL, information on its psychometric properties in the Malaysian context will be of utmost importance.

Here, we explore the construct validity and internal consistency of this instrument in asthmatics and in the general population in Malaysia.

Materials and Methods

The UK version of SF-36 was first translated into the Malay (Bahasa Melayu) language by a group of researchers from Universiti Sains Malaysia[a]. We had subsequently adapted and modified this translated version of the SF-36, changing some Malay words perceived to be a dialect, too localised or vernacular. This modified version of the translated SF-36 was then used in a two-phased study between 1999 and 2000.

In Phase I, the Malay version of the SF-36 was administered to asthmatics at six government hospitals from different regions of the country. Further details on this study are described elsewhere[13].

In addition, using quota sampling to cover the main ethnic groups[b], we also carried out cognitive debriefing in the form of in-depth interviews of patients (irrespective of their disease) and their families attending government medical outpatient clinics. Here, we explored their perception, understanding and interpretation of translated items of the Malay version, the various terminology used, and the formatting of the self-administered questionnaire. Results of the cognitive debriefing were then used to further refine the translated version both in terms of words used, and the format/layout of the questionnaire.

Phase II of the study implemented in 2000 saw the administration of the further refined version on the population in a nationwide household community survey. A total of 5,238 living quarters[c] (LQ) had been selected with the help of the Department of Statistics Malaysia and each LQ was given two questionnaires together with self-addressed envelopes. Details of the methodology, including the typological errors detected[d], are reported elsewhere[14].

In both phases of the study, the eight scales of SF-36 were calculated using the formula specified by Ware *et al*[4]. Higher scores on each scale signify a better health state.

Distributions and floor-ceiling effects were explored using frequency distributions, and Cronbach's α for internal consistency. Principal component analysis with orthogonal rotation (varimax method) was done for both disease and population data, to look at the pattern of correlations between the eight scales and the physical and mental health domains put forward by Ware *et al*[4,8,9].

Construct validity was looked at based on a *priori* hypotheses[15]. Based on other findings elsewhere[16-18], with rising age, lower scores are expected for scales which measure physical health, namely physical functioning (PF), Role physical (REP) and Bodily pain (BP). Lower scores are also expected with presence of medical conditions[16,18]. The Mann-Whitney U or Kruskal-Wallis test of significance for non-normally distributed continuous variables was used to assess significance of observed differences as all the scales did not have a normal distribution (using the Kolmogorov-Smirnov Test of Normality).

Footnote:

a: A research team under the aegis of International Quality of Life Assessment (IQOLA) Project had developed a translated version.

b: Malay, Chinese Indian and other Bumiputera (from Sarawak).

c: A living quarters (LQ) is a living unit, structurally separate (surrounded by walls, fences etc. and is covered by a roof) and independent (has direct access via a public staircase, communal passages or landing) and are meant for living.

d: Typological errors were noted in questions 6 and 8 which could affect the accuracy of responses for social functioning (SF) and bodily pain (BP) scales respectively. Question 6-one response category left out (moderately"). Question 8-had six response categories instead of the specified five. "Mild" and "very mild" were combined.

Results

In Phase I, 1634 asthmatic patients were eligible and 1612 (98.7%) responded. The respondents differed from that of the population of asthmatics in Malaysia[19], with males, the young, Chinese and other ethnicity, unemployed and those with lower education under-represented[13].

The cognitive debriefing and in-depth interviews revealed some phrases that non-Malays and those with lower educational status had difficulty with, such as responses for questions on PF, which has a positive *("yes")* and negative expression *("limited")* in a particular response. In the final translated version, we omitted the *"yes"*, as we noted they had misinterpreted the *"yes"* to mean, *"yes"*, they can do that activity[e]. Other phrases found to be confusing had also been modified. Those with lower educational level also encountered problems with the tabular form used for some questions (such as for PF and MH items), resulting in wrong application, with columns seen as the "main questions", and rows as the possible responses, hence the puzzle of multiple responses for some.

Subsequently, in Phase II, of 10,041 questionnaires sent out, 3072 were returned in a usable form. Response rate was 30.6%. Compared to the Malaysian population in 2000[20], urban dwellers, young males, Chinese and Other ethnicity were under-represented[14].

Distribution and Data Quality

Table I shows distribution indicators for the scales. Role limitations for physical (REP) and emotional (REE) scales had considerable floor and ceiling effects. The ceiling effects for these scales were substantially greater for the general population data from the phase II study. Physical functioning (PF) and social functioning (SF) also had substantial ceiling effects. However, this picture of substantial ceiling effects for PF was not seen amongst the older population 60 years and above. Whilst PF showed greater variability with increasing age, minimal floor and ceiling effects were reported across all ages for bodily pain (BP), general health (GH), vitality (VT) and mental health (MH); neither did they show much difference in distribution of scores across age groups. Additional details of distribution for the scales have been described elsewhere[14].

In terms of distribution, in one hand the asthma data reported both REP and REE to have a bimodal distribution, with the peaks at the minimum and maximum values. On the other hand, the population data set did not show a similar distribution. Instead, PF, REP, SF and REE in the population data showed a J-shaped distribution, whilst the remaining scales (including those of the asthma data), were more normally distributed but negatively skewed. Similarly shaped distributions were found within different sexes, age groups and ethnicity for both data, and asthma severity grades for asthma data.

At item level, the missing value rates were noted to be consistently low; of the highest being only 0.46% for question 5c, and the lowest percent of 0.07% for questions 3b and 9g (Table II).

Tests of Scaling Assumptions and Reliability

The results of tests of scaling assumptions for general population data are summarized in Table II. Correlations between each item and its hypothesized scale were all above 0.50, except for item 3j in PF (0.38) and both items in SF (0.45). The highest score was 0.81 (5b) for the REE scale. Item-scale correlations were more or less similar within each scale, the exception being item 3j (bathing and dressing – PF), and to a lesser extent, item 1 (health in general – GH). The means and standard deviations were more or less similar within every scale.

For item discriminant validity, items in VT, SF and MH had high correlations to scales other than that hypothesized for the particular item. Almost all items in VT and SF overlapped with MH scale. Both items in SF had higher correlations with MH scale (0.49 and 0.56) than with its hypothesized scale (0.45). In addition, item 6 and item 10 in SF had higher correlations with REE (0.48) and VT (0.49) respectively than with its own scale. In MH, item 9d and 9h had similar correlation to VT as to its hypothesized scale.

The internal consistency reliability statistics for the 8 scales are presented in Table III. All scales exceeded the recommended 0.70 level for group comparison[21], except for SF scale which had lower Cronbach's αs for both data sets (0.55 and 0.62). Most of the inter-scale correlation coefficients were medium to low, though the correlation between MH and VT (0.69 and 0.70 for the two data) was almost as high as the two scales' reliability statistics.

Footnote:
e: They registered the meaning of "yes" yet ignored the subsequent phase of "limited..."

Latent dimensions and validity

Table IV presents the results of the principal component analysis. The eight scales were highly correlated with a single component, with only 52-53% of variance accounted for both datasets. If two components as hypothesized are forced[f], the inclusion of the second component increased the variance to 63-64%. As expected, PF and REP had high correlations with the physical component for the asthma data; while in the general population data, it was the PF, BP and GH scales. Scales highly correlated with the mental component were MH, VT in the asthma data, and SF, REE and MH in the general population data.

There were some contrasting findings, however. In the asthma data, BP and GH loaded higher in the mental component, whilst REE loaded higher on the physical component and SF had relatively equal loading on both components. In the population data, VT and REP scales have moderate to high correlations with both components. There is no difference noted for this between the sexes; across ethnic groups, REP had higher factor loadings in mental component amongst the Chinese, Indians and Other Bumiputera ethnic groups (0.54 to 0.71), while Malays had almost equal loadings for this scale (0.51 and 0.45 for physical and mental component respectively). As expected, GH scale loaded higher in the physical component in the general population data, in contrast to the asthma data in which it loaded higher in the mental component.

Results of item level factor analysis is shown in Table V. Items in PF, REP, GH scales are grouped together, as expected. Positively worded items in VT ("full of life", "a lot of energy") and MH ("calm and peaceful", "happy") are loaded together in the same factor as BP items. Alternately, negatively worded items in VT ("worn out", "tired") and MH ("nervous", "in the dumps", "downhearted") are loaded in another factor, together with SF.

Table VI shows means of the scales across several variables. Amongst asthmatics, age significantly affects PF, GH and MH scores, though increasing age shows a continual decline only in the PF scale. PF, REP and VT scales demonstrate deteriorating scores with increasing disease severity in the asthma data. In the general population data, though all domains are affected by age, a trend in reduction of PF, REP, BP and GH was seen with increasing age (also in Figure 1). Conversely, MH scale demonstrates no obvious pattern with age, also the case for VT, REE and SF. Contrasting those 50 years and above with the other two age groups, besides PF, REP, BP and GH scales were also significantly lower in those 50 years and above. Presence of self-reported disease and handicap also appreciably affect means of all the scales. Considerable inter-ethnic differences were seen for both data, with the Others category, followed by the Chinese, having higher scores for most scales. Malays had the highest score for SF in the general population.

Footnote:
f: The eigenvalue cut-off value was lowered to 0.8.

Table I: Distribution indicators for the SF-36 scales

Indicators	Physical functioning (PF)	Role limitation-physical (REP)	Bodily Pain (BP)	General Health (GH)	Vitality (VT)	Social Functioning (SF)	Role limitation-Emotional (REE)	Mental Health (MH)
Asthma Data								
Number of cases	1611	1601	1612	1611	1610	1612	1609	1612
Floor (%)	0.4	28.6	0.4	0.2	0.1	0.2	27.7	0.1
Ceiling (%)	4.8	25.0	0	0.3	2.5	16.7	40.0	5.7
Skewness	-0.4	0.1	-0.1	0.0	-0.1	-0.2	-0.2	-0.5
Mean	64.1	47.7	58.9	52.2	62.0	69.5	56.1	69.8
SD	22.4	39.3	20.2	20.4	18.5	20.8	41.8	18.9
General Population Data								
Number of cases	3072	3064	3070	3070	3071	3070	3060	3071
Floor (%)	0.2	8.2	0.3	0.3	0.1	0.2	12.7	0
Ceiling (%)	32.3	70.4	0	2.4	4.0	43.2	71.3	9.2
Skewness	-1.87	-1.6	-0.6	-0.4	-0.3	-1.3	-1.4	-0.5
Mean	86.0	82.0	70.0	66.7	66.8	83.7	79.2	74.7
SD	17.9	32.1	17.6	20.0	17.7	19.3	35.9	17.2

Table II: Item Descriptive Statistics and Pearson Item-Scale Correlations for General Population Data

Scale	Item **	Valid n***	Missing	Mean	SD	PF	REP	BP	GH	VT	SF	REE	MH
Physical Functioning	3a	3067	0.16%	2.31	0.67	0.53*	0.33	0.34	0.42	0.36	0.26	0.19	0.25
	3b	3070	0.07%	2.82	0.45	0.68*	0.32	0.29	0.35	0.33	0.31	0.21	0.26
	3c	3067	0.16%	2.87	0.41	0.62*	0.27	0.26	0.30	0.27	0.31	0.17	0.24
	3d	3066	0.20%	2.65	0.58	0.74*	0.39	0.35	0.43	0.40	0.30	0.23	0.29
	3e	3069	0.10%	2.81	0.45	0.72*	0.33	0.30	0.36	0.34	0.30	0.21	0.28
	3f	3065	0.23%	2.71	0.51	0.61*	0.35	0.34	0.36	0.34	0.28	0.21	0.25
	3g	3067	0.16%	2.52	0.65	0.69*	0.43	0.37	0.43	0.41	0.31	0.26	0.30
	3h	3066	0.20%	2.74	0.53	0.71*	0.39	0.34	0.38	0.37	0.29	0.24	0.26
	3i	3066	0.20%	2.86	0.40	0.57*	0.30	0.25	0.26	0.26	0.25	0.20	0.22
	3j	3066	0.20%	2.95	0.27	0.38*	0.17	0.16	0.13	0.17	0.17	0.14	0.14
Role Physical	4a	3064	0.26%	1.83	0.38	0.38	0.70*	0.33	0.32	0.34	0.36	0.37	0.32
	4b	3065	0.23%	1.82	0.38	0.40	0.73*	0.36	0.37	0.38	0.37	0.39	0.34
	4c	3060	0.39%	1.82	0.38	0.41	0.73*	0.38	0.36	0.39	0.34	0.36	0.31
	4d	3062	0.33%	1.81	0.39	0.39	0.65*	0.39	0.37	0.39	0.35	0.40	0.33
Bodily Pain	7	3061	0.36%	4.79	1.11	0.39	0.39	0.68*	0.50	0.57	0.39	0.30	0.46
	8	3067	0.16%	4.22	0.78	0.43	0.42	0.68*	0.48	0.55	0.44	0.32	0.45
General Health	1	3069	0.10%	3.36	0.90	0.44	0.31	0.43	0.53*	0.46	0.29	0.22	0.34
	11a	3068	0.13%	3.78	1.07	0.35	0.32	0.38	0.60*	0.44	0.36	0.26	0.40
	11b	3063	0.29%	3.70	1.09	0.40	0.32	0.40	0.68*	0.47	0.33	0.25	0.42
	11c	3060	0.39%	3.91	1.05	0.33	0.32	0.37	0.57*	0.42	0.34	0.26	0.39
	11d	3065	0.23%	3.63	1.08	0.43	0.35	0.44	0.71*	0.51	0.36	0.27	0.45
Vitality	9a	3070	0.07%	4.27	1.32	0.40	0.38	0.48	0.49	0.58*	0.36	0.32	0.52
	9e	3067	0.16%	4.39	1.22	0.40	0.37	0.49	0.53	0.62*	0.41	0.32	0.60
	9g	3070	0.07%	4.53	1.01	0.35	0.34	0.46	0.41	0.57*	0.43	0.34	0.55
	9i	3068	0.13%	4.18	1.01	0.30	0.29	0.42	0.39	0.52*	0.40	0.30	0.47

cont'd... next page

cont'd... Table II

Scale	Item **	Valid n***	Missing	Mean	SD	PF	REP	BP	GH	VT	SF	REE	MH
Social Functioning	6	3060	0.39%	4.49	0.84	0.31	0.37	0.36	0.33	0.39	0.45*	0.48	0.49
	10	3065	0.23%	4.23	0.95	0.37	0.37	0.40	0.42	0.49	0.45*	0.38	0.56
Role Emotional	5a	3065	0.23%	1.78	0.41	0.25	0.41	0.27	0.27	0.37	0.44	0.74*	0.46
	5b	3061	0.36%	1.78	0.41	0.25	0.42	0.30	0.29	0.37	0.45	0.81*	0.47
	5c	3058	0.46%	1.82	0.38	0.26	0.38	0.29	0.29	0.37	0.43	0.68*	0.46
Mental Health	9b	3066	0.20%	4.81	1.03	0.26	0.29	0.36	0.35	0.48	0.47	0.41	0.61*
	9c	3064	0.26%	5.19	0.97	0.26	0.29	0.35	0.35	0.46	0.50	0.44	0.64*
	9d	3062	0.33%	4.39	1.30	0.29	0.33	0.41	0.46	0.62	0.44	0.39	0.63*
	9f	3065	0.23%	5.01	1.02	0.24	0.26	0.34	0.35	0.46	0.50	0.40	0.61*
	9h	3064	0.26%	4.32	1.22	0.29	0.32	0.42	0.48	0.62	0.46	0.38	0.62*

* denotes correlation between an item and its hypothesized scale, corrected for overlap.

** Numbers correspond to the Standard form of SF-36[4]

*** Values of n for the particular item, not list wise.

Table III: Reliability Statistics and Correlations between SF-36 Scales

Scale	PF	REP	BP	GH	VT	SF	REE	MH
Asthma Data								
Physical Functioning (PF)	(0.85)							
Role-Physical (REP)	0.42	(0.80)						
Bodily Pain (BP)	0.34	0.44	(0.73)					
General Health (GH)	0.29	0.42	0.47	(0.72)				
Vitality (VT)	0.34	0.46	0.55	0.55	(0.74)			
Social Functioning (SF)	0.34	0.44	0.52	0.38	0.45	**(0.55)**		
Role-Emotional (REE)	0.33	0.65	0.47	0.40	0.49	0.48	(0.80)	
Mental Health (MH)	0.30	0.39	0.51	0.50	0.69	0.49	0.49	(0.81)
General Population Data								
Physical Functioning (PF)	(0.88)							
Role-Physical (REP)	0.47	(0.86)						
Bodily Pain (BP)	0.43	0.43	(0.78)					
General Health (GH)	0.50	0.42	0.52	(0.82)				
Vitality (VT)	0.47	0.45	0.60	0.60	(0.77)			
Social Functioning (SF)	0.39	0.43	0.45	0.44	0.51	**(0.62)**		
Role-Emotional (REE)	0.29	0.45	0.33	0.32	0.42	0.50	(0.87)	
Mental Health (MH)	0.35	0.39	0.49	0.52	0.70	0.61	0.52	(0.82)

Note: Cronbach's Alpha denoted in parenthesis, on the diagonal. Internal consistency reliability lower than the recommended level of 0.70[21] is shown in bold.

Table IV: Principal Component Analysis showing the associations between the SF-36 scales

SF-36 Scales	Single Component		Hypothesized two components				
			Physical		Mental		
		h^2	Expected*	Observed**	Expected*	Observed**	h^2
Asthma Data							
Physical Functioning (PF)	0.55	0.31	●	0.77	○	0.10	0.60
Role-Physical (REP)	0.73	0.53	●	0.79	○	0.31	0.72
Bodily Pain (BP)	0.75	0.56	●	0.37	○	0.67	0.58
General Health (GH)	0.69	0.48	◐	0.20	◐	0.73	0.57
Vitality (VT)	0.80	0.64	◐	0.25	◐	0.82	0.73
Social Functioning (SF)	0.71	0.50	◐	0.47	●	0.53	0.51
Role-Emotional (REE)	0.75	0.57	○	0.66	●	0.44	0.63
Mental Health (MH)	0.77	0.59	○	0.19	●	0.83	0.73
Variance	0.52			0.11		0.52	
General Population Data							
Physical Functioning (PF)	0.66	0.43	●	0.80	○	0.10	0.64
Role-Physical (REP)	0.68	0.47	●	0.55	○	0.41	0.47
Bodily Pain (BP)	0.74	0.54	●	0.71	○	0.31	0.60
General Health (GH)	0.74	0.55	◐	0.77	◐	0.26	0.66
Vitality (VT)	0.82	0.68	◐	0.66	◐	0.51	0.68
Social Functioning (SF)	0.75	0.56	◐	0.35	●	0.73	0.65
Role-Emotional (REE)	0.64	0.41	○	0.10	●	0.85	0.73
Mental Health (MH)	0.80	0.63	○	0.40	●	0.75	0.72
Variance	0.53			0.53		0.11	

h^2 is the proportion of total variance of each scale explained by the extracted component(s).

* Expected picture of validity. ● =Strong association ($r \geq 0.70$) ◐ = moderate to substantial association ($0.3<r<0.7$), and ○ = weak association ($r \leq 0.3$)[4].

** Factor loadings observed.

Table V: Item level Factor Analysis of the Malay version of SF-36 in the general population data

	Factor Loadings of individual SF-36 items					
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
PF01 (3a)	0.464					
PF02 (3b)	0.739					
PF03 (3c)	0.743					
PF04 (3d)	0.706					
PF05 (3e)	0.779					
PF06 (3f)	0.615					
PF07 (3g)	0.616					
PF08 (3h)	0.685					
PF09 (3i)	0.673					
PF10 (3)j	0.549					
REP1 (4a)				0.763		
REP2 (4b)				0.766		
REP3 (4c)				0.784		
REP4 (4d)				0.700		
BP1 (07)			0.513			
BP2 (08)			0.461			
GH1 (01)			0.424		0.485	
GH2 (11a)					0.698	
GH3 (11b)					0.732	
GH4 (11c)					0.676	
GH5 (11d)					0.726	
VT1 (9a)			0.699			
VT2 (9e)			0.683			
VT3 (9g)		0.600				
VT4 (9i)		0.563				
SF1(06)		0.446				0.415
SF2 (10)		0.583				
REE1 (5a)						0.803
REE2 (5b)						0.829
REE3 (5c)						0.757
MH1 (9b)		0.662				
MH2 (9c)		0.692				
MH3 (9d)		0.418	0.575			
MH4 (9f)		0.715				
MH5 (9h)		0.419	0.564			

Principal Component Analysis with varimax rotation was used. Factor loadings $\geq$0.40 were considered significant.

Table VI: Construct Validity of the Malay Version of SF-36

	Mean SF-36 scale scores, with standard deviation (in parenthesis below the mean)								
	n	PF	REP	BP	GH	VT	SF	REE	MH
ASTHMA DATA									
Gender									
Male	589	67.36	49.19	60.90	53.10	64.47	70.03	57.61	71.69
		(22.52)	(39.47)	(20.53)	(20.87)	(18.86)	(21.56)	(41.34)	(19.26)
Female	1023	62.17	46.89	57.75	51.63	60.51	69.16	55.29	68.78
		(22.15)	(39.12)	(19.94)	(20.15)	(18.14)	(20.33)	(41.99)	(18.69)
p-value*		<0.001	0.247	0.002	0.070	<0.001	0.233	0.322	0.001
Age Group(years)									
18 - 29	301	68.06	48.67	58.29	48.74	61.00	69.23	54.37	67.80
		(22.60)	(37.73)	(20.42)	(19.55)	(18.01)	(20.83)	(40.37)	(17.87)
30 - 39	413	65.13	48.42	58.24	51.10	60.80	68.22	55.42	67.26
		(21.76)	(39.47)	(20.02)	(20.23)	(17.48)	(21.76)	(41.47)	(18.28)
40 - 49	537	63.45	47.50	58.86	52.83	62.63	69.39	58.40	70.88
		(22.94)	(39.24)	(20.43)	(20.00)	(19.02)	(19.89)	(42.21)	(18.85)
50 - 59	250	60.70	48.19	59.13	54.64	63.13	71.30	56.27	72.91
		(21.64)	(40.64)	(20.22)	(21.23)	(19.37)	(20.90)	(42.12)	(19.34)
≥60	110	60.14	42.20	62.89	56.56	63.12	71.48	51.99	72.87
		(21.78)	(39.48)	(19.15)	(22.22)	(18.88)	(20.82)	(43.61)	(22.07)
p-value**		<0.001	0.623	0.242	<0.001	0.262	0.410	0.470	<0.001
Ethnic Group									
Malay	1160	64.10	45.40	58.45	51.61	62.13	68.63	54.81	69.85
		(22.08)	(39.45)	(20.13)	(20.14)	(18.78)	(20.75)	(41.89)	(19.25)
Chinese	127	70.16	61.02	69.10	56.28	65.00	76.67	66.67	73.73
		(24.31)	(38.92)	(18.98)	(21.42)	(19.36)	(21.33)	(40.93)	(17.18)
Indian	224	58.75	51.24	55.54	54.57	58.71	69.81	58.89	68.09
		(22.77)	(37.24)	(20.16)	(20.98)	(16.97)	(20.34)	(40.63)	(18.65)
Other Bumiputera	93	67.47	49.73	57.69	47.01	62.80	68.28	50.54	67.74
		(21.09)	(38.22)	(18.52)	(19.10)	(16.61)	(19.90)	(42.17)	(17.58)
Others	8	72.50	53.13	71.00	60.00	69.38	82.81	70.83	80.00
		(13.09)	(41.05)	(22.78)	(28.23)	(14.00)	(17.60)	(37.53)	(16.84)
p-value**		<0.001	<0.001	<0.001	0.003	0.018	<0.001	0.009	0.032

cont'd... next page

cont'd... Table VI

	Mean SF-36 scale scores, with standard deviation (in parenthesis below the mean)								
	n	PF	REP	BP	GH	VT	SF	REE	MH
Severity of Disease									
Mild	478	67.03	50.69	60.15	53.11	63.76	70.37	56.04	69.91
		(21.88)	(38.14)	(19.61)	(19.96)	(18.25)	(20.44)	(40.98)	(18.23)
Moderate	754	64.12	47.93	59.17	52.37	62.08	69.86	57.76	70.15
		(21.84)	(39.18)	(19.81)	(20.45)	(17.99)	(20.73)	(41.70)	(18.74)
Severe	380	60.24	43.67	56.79	50.57	59.45	67.60	53.07	69.15
		(23.68)	(40.49)	(21.60)	(20.89)	(19.53)	(21.27)	(42.76)	(20.24)
p-value**		<0.001	0.028	0.100	0.237	0.012	0.115	0.229	0.940
GENERAL POPULATION									
Gender									
Male	1563	87.38	82.48	70.91	67.39	68.46	84.48	81.37	75.99
		(17.23)	(31.78)	(17.56)	(19.82)	(17.68)	(19.01)	(34.51)	(16.64)
Female	1498	84.52	81.47	68.96	66.03	65.10	82.94	76.92	73.30
		(18.52)	(32.55)	(17.56)	(20.15)	(17.54)	(19.55)	(37.25)	(17.63)
p-value*		<0.001	0.323	0.001	0.071	<0.001	0.014	<0.001	<0.001
Age group(years)									
18 - 29	709	91.16	84.96	71.93	70.18	66.45	83.29	74.29	72.37
		(13.66)	(28.51)	(16.61)	(17.40)	(16.40)	(18.71)	(38.25)	(17.16)
30 - 39	877	89.14	83.60	70.94	67.81	67.57	84.08	80.25	75.17
		(14.87)	(31.54)	(17.07)	(19.09)	(17.30)	(18.54)	(35.74)	(16.48)
40 - 49	781	86.27	83.97	69.62	68.12	68.38	85.40	81.45	76.39
		(16.51)	(29.54)	(17.42)	(19.37)	(17.86)	(17.88)	(34.33)	(16.79)
50 - 59	411	80.10	79.21	69.11	62.56	66.31	83.70	81.83	74.85
		(20.73)	(34.51)	(17.87)	(22.16)	(17.80)	(20.30)	(33.80)	(18.17)
≥60	253	69.78	67.59	63.23	55.81	60.59	78.41	77.08	73.25
		(23.55)	(41.69)	(20.20)	(22.90)	(19.90)	(23.96)	(38.05)	(18.38)
p-value**		<0.001	<0.001	<0.001	<0.001	<0.001	0.002	<0.001	<0.001

cont'd... next page

cont'd... Table VI

	Mean SF-36 scale scores, with standard deviation (in parenthesis below the mean)								
	n	PF	REP	BP	GH	VT	SF	REE	MH
Ethnic Group									
Malay	2019	87.00 (16.98)	81.81 (32.33)	69.58 (17.32)	67.09 (19.96)	67.81 (17.59)	84.44 (18.98)	79.58 (35.86)	75.83 (17.03)
Chinese	505	85.52 (17.50)	85.69 (29.87)	74.52 (17.27)	67.32 (19.28)	65.37 (17.36)	82.43 (19.53)	82.67 (33.54)	73.26 (16.89)
Indian	165	78.76 (20.99)	79.24 (31.07)	66.00 (19.29)	65.02 (19.72)	61.55 (17.75)	79.73 (20.83)	74.80 (36.20)	67.85 (17.96)
Other Bumiputera	354	84.04 (20.72)	78.35 (34.61)	66.98 (17.63)	64.57 (21.20)	65.20 (18.18)	83.53 (19.25)	74.83 (38.58)	73.01 (17.58)
Others	29	87.24 (21.11)	93.97 (19.66)	75.69 (14.68)	68.38 (20.02)	70.52 (14.10)	81.90 (23.76)	73.56 (39.22)	76.28 (12.65)
p-value**		<0.001	<0.001	<0.001	0.184	<0.001	0.008	0.003	<0.001
Self-reported disease									
Absent	2282	89.78 (14.25)	87.05 (27.45)	73.09 (16.40)	71.80 (17.09)	70.03 (16.49)	86.04 (17.51)	82.94 (33.12)	76.72 (16.37)
Present	790	75.00 (22.34)	67.49 (39.43)	60.89 (17.79)	52.12 (20.58)	57.46 (17.68)	77.04 (22.35)	68.49 (41.19)	68.69 (18.10)
p-value*		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Self-reported handicap									
Absent	2996	86.31 (17.49)	82.51 (31.66)	70.19 (17.42)	67.11 (19.73)	67.04 (17.60)	84.01 (18.95)	79.48 (35.79)	74.86 (17.13)
Present	76	73.03 (27.25)	63.49 (43.10)	60.99 (21.58)	51.88 (24.42)	57.24 (18.25)	72.53 (27.16)	69.30 (39.91)	66.53 (17.83)
p-value*		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.010	<0.001

*Mann-Whitney U test, **Kruskal-Wallis test.

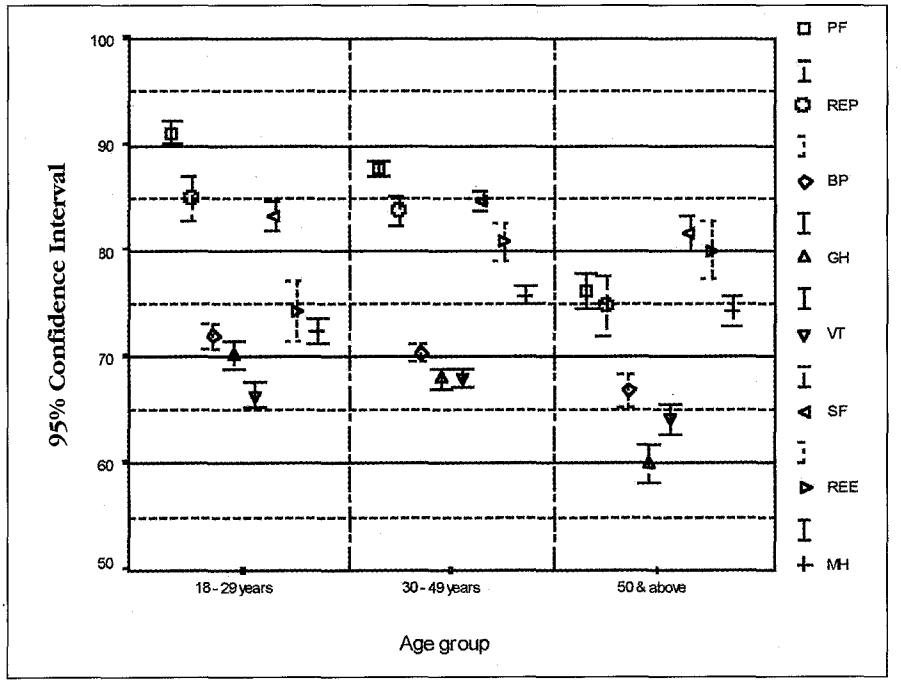

Fig. 1: Comparison of the SF-36 scale profiles (of the Malay version) across age groups in the general population data.

Discussion

There is a need for a standard list of measures that can be used internationally for purposes of cross-country comparison. It is for this reason that this study was undertaken. Health outcomes, from the patient's perspective, have become increasing more essential in health care in Malaysia. Both generic and specific measures of self-reported health-related quality of life questionnaires have been gradually used more. Hence, the applicability of SF-36 instrument in the Malaysian context is an important issue given its generic nature and use across multiple health conditions internationally.

In general terms, the study substantiates the applicability of the SF-36 concept in Malaysia, with a few caveats. The SF-36 Malaysia version has a high rate of data completion, with good quality data obtained not only in institutional settings, but also in the context of self-administration in the general population. Percent missing was less than 0.5%, lower than that reported elsewhere[22,25]. The floor-ceiling effects and the skewed distributions of the scales seen here is similar to findings in Taiwan[22], Hong Kong[29] and Australia[23]. As expected, domains of physical functioning, role functioning and social functioning have substantial ceiling effects for the general population, similar to Taiwan[22] and Hong Kong[29]. Ceiling effects were most considerable for both scales of role functioning, not surprising given they are comparatively coarse role disability scales[24]. Thus, the scales of role functioning, and to some extent physical and social functioning, might not be sensitive enough to detect changes at the upper end of the health state in the general population.

Item-scale correlations were generally good, except for item 3j in physical functioning which looks at bathing and dressing oneself, and both items in social functioning scale. Item 3j addresses physical functioning at its lowest performance level in the scale, thus suggesting that people might perceive it to be discordant in this domain, perhaps interpreting it as too basic; an activity not addressing their perception of physical performance. The item is not correlated with other scales, implying it still fits best in the physical functioning scale, than in the other seven. However, most SF-36 scales, physical functioning included, had high Cronbach's αs, suggestive of internally consistent scales.

An exception is the social functioning scale, with internal consistency reliability lower than that recommended[21], a finding encountered in Taiwan[22] and Hong Kong[25]. The items in social functioning have higher correlations with items in mental health and vitality scales than with its own. With these high inter-correlations with other scales, items in social functioning, mental health and vitality have some problems with item discriminant validity. Except for these 3 scales, inter-scale correlation analysis shows that other scales were generally distinct for the Malay version of SF-36.

This implies the possibility of cultural differences at work in item interpretation, as noted by Tseng *et al.*[22]. Tseng *et al.*[22] had noted that social functioning as a concept may be more westernised. Similar to the Chinese[25] and Taiwanese[22], Malaysians would not generally consider using health problems to evade social, let alone family gatherings. Culturally, family and kin play a fundamental role in people's lives, one separate from that of friends and other social interactions. Thus, as recommended elsewhere in Asia[22,25], perhaps a specific "family functioning scale" might need to be added to recognise the influence of health on family life in Malaysia.

Additionally, vitality and mental health scales were highly inter-correlated. Both vitality scale items ("have a lot of energy" and "feel worn out") had high correlations with mental health scale, while two mental health items ("felt calm and peaceful" and "been a happy person") had high correlations with vitality scale. Furthermore, item level factor analysis had shown that loadings of the items in VT and MH are split into two factors, interestingly according to the way the items were worded, either "positively" or "negatively". This implies that Malaysians have difficulty distinguishing between the two domains. Such a finding had been reported in Singapore[18] as well as studies amongst Asians[22,25] elsewhere. Moreover, vitality scale shows a high correlation in the mental component and a low correlation with the physical component for the asthma data, as seen also in Taiwan[22] and Japan[26]. In contrast, amongst the general population, vitality scale had roughly similar correlation (moderate to substantial association[4]) with both components, as that reported in United States by Ware *et al.*[4]. Ren *et al.*[25] had stated that for the Chinese, "'happiness', a healthy mental state, is a sign of possessing vitality", hence "vitality is central to the concept of a healthy mental state". Non-Western cultures also "express emotions in ways that merge mind and body", with mind and body integrated with one another as well as with social milieu, and psychological, physical and social factors perceived as indivisible[27]. This could also possibly explain the difficulty Malaysians have in discriminating between items in the vitality and mental health scales, unable to conceptually distinguish between having energy, liveliness with happiness or a positive affective state.

The eight scales were found to form two distinct clusters of physical and mental health in America[4,9,24], Western Europe [28] and Japan[26]. However, in Malaysia, only one component (higher order cluster) was identified for both disease and general population data. The higher order cluster of one component had also been noted in Australia[23]. Forcing two components yielded scales that were clustered differently from the expected in the asthma data. This disparity could perhaps be due to the translation process, as difficulties were encountered during administration in the study, hence the subsequent cognitive debriefing to further refine the language and terminology used. In the second phase of the study, (i.e. the general population data), the pattern of the hypothesized two components is more apparent as that reported by Ware *et al.*[4].

In the Malaysian general population data, general health scale is primarily a physical scale and does not have as strong an association with mental health as that seen in Singapore[18], Taiwan[22] and Japan[26]. Instead, the Malaysian picture approximates that of Hong Kong and America[4]. Role-physical scale had a moderate correlation with the mental component, with this picture more obvious amongst non-Malays. The differences seen in factor structure could be due to linguistic and cultural differences. Perhaps people are more willing to attribute limitations in physical role functioning to affective states. Another point is the amount of variance accounted for by the two components, which is lower than that found in Japan[26, 30], (58%-66%) and America[4] (more than 80%). Possibly, there may be other areas of importance which was not included, such as spirituality, sexual health and family relationships, the last as had been mentioned earlier.

Lower scores seen for physical functioning, role physical and bodily pain with increasing age supports the construct validity of the Malay version of SF-36, as does the lower scores associated with increased severity of disease, and presence of self-reported disease and handicap. The differences seen in the SF-36 scale profiles across age groups, with evidence of physical

functioning declining with age whilst mental health show no particular trend, is indicative of discriminant validity of the Malay version of SF-36.

The asthma data had further demonstrated that the Malay version of SF-36 was sensitive enough to detect changes amongst mild asthmatics, specifically in the expected domains of role functioning, general health, physical functioning and bodily pain, even after adjusting for the effects of gender, age, ethnicity, education, marital state, education, employment, duration of having suffered from asthma and co-morbidity, in an article published earlier[13]. Given the nature of the disease, as expected, mental health and vitality were not affected in mild disease. Furthermore, the instrument was also responsive to different grades of severity of disease, as shown by increasingly severe drop in quality of life with worsening disease severity; all eight domains affected among moderate and severe asthmatics; and with the poorest quality of life scores amongst the severe group[13], further evidence of its construct validity, albeit only in asthma disease.

Perhaps Malaysians view the concept of HRQoL differently, in that the two higher order clusters of Physical and Mental Health might not be fully applicable or entirely appreciated in the Malaysian setting. This implies that one may not use these summary measures to represent HRQoL, but should instead use the eight scales. Arguably, perhaps further refinement of the questionnaire may yield a picture similar to that seen in America, though one may dispute that culturally, the mores, customs, traditions and way of life of both Asians and Americans are distinct and separate. On the other hand, with the current wave of globalisation, the gap may narrow over time. Given the current scenario, the domains affecting health related quality of life of people in Malaysia need to be explored further, with efforts to identify domains or scales partially, or not addressed, in the SF-36. In essence, the Malay version of SF-36 could be used in Malaysia, with its generally acceptable internal consistency and validity. The caveat is in the call for additional domains of importance to Malaysians that is not covered by the instrument, and in the caution to be employed when using and construing the meaning of some scales.

This study is limited to Malaysians aged 18 years and above, and who are able to understand either Malay or English language. There is the possibility of response bias with the poor response rate for population data; though reported rates for postal surveys have been reported to range from a low of 24% to a high of 92%[31]. Other possible bias include under-representation of males, the young, Chinese and Other ethnicity for both data and the minor errors in the questionnaire involving 1 item each in social functioning and bodily pain scale, all or one of which might potentially affect the results. Furthermore, items in physical functioning and role functioning scales contain three or less responses, and hence, by stringent standards, is inappropriate for factor analysis. Lastly, stability/reproducibility and content validity of the instrument was not assessed in this study.

Acknowledgements

This study received a grant from the Intensified Research Priority Areas Grant, Government of Malaysia (IRPA Code: 06-05-05-0177). We thank all respondents for their participation in these studies. The authors would like to express their gratitude to all the respiratory consultants, physicians and the team of nurses who had made the data collection possible in Phase I, and to the Department of Statistics Malaysia including Hjh. Normah Mohd Aris, Head, Communications And Operations Division, Kate Van Kit Yee, Rosilawati, Rafliza and Kho Sian Eng and all the state statisticians for their invaluable assistance in planning and implementation of the data collection phase of this survey. Their assistance and support is very much appreciated and valued. The researchers are also grateful to Prof Wan Manan and team, Universiti Sains Malaysia (USM) for their assistance and for providing their developmental version of the translated Malay version of SF-36. All their support and cooperation given is very much appreciated.

References

1. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. JAMA 1995; 1995: 59-65.
2. Guyatt GH and Cook DJ. Health status, quality of life and the individual. Journal of American Medical Association. 1994; 272: 630-31.
3. Bowling A. Measuring Health. A Review of Quality Of Life Measurement Scales. Buckingham: Open University Press, 1991.
4. Ware JE Jr, Kristin KS, Kosinski M, et al. SF-36 Health Survey Manual and Interpretation Guide. Boston, Massachusetts: Nimrod Press, 1993.
5. Shiely JC, Boyliss MS, Keller SD, et al. SF-36 Health Survey Annotated Bibliography: First Edition (1988 - 1995). Boston, Massachusetts: The Health Institute, New England Medical Centre, 1996.
6. Hemingway H, Stafford M, Stansfeld S, Shipley M & Marmot M. Is the SF-36 a valid measure of change in population health? Results from the Whitehall II study. BMJ 1997; 315: 1273-279.
7. Watson EK, Firman DW, Baade PD, Ring I. Telephone administration of the SF-36 Health Survey: validation studies and population norms for adults in Queensland. Aust NZ J Public Health 1996; 306: 1437-40. [MEDLINE]
8. Ware JE Jr, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): I. Conceptual framework and item selection. Med Care 1992; 30(6): 473-83.
9. McHorney CA, Ware JE Jr, 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(2): 247-63.
10. McHorney CA, Ware JE Jr, Lu JFR & Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 1994; 32(1): 40-66.
11. Ware JE Jr, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary results from the Medical Outcomes Study. Med Care 1995; 33(4 Suppl): AS264-AS279.
12. Wagner AK, Gander B, Aaronson NK, et. al. Cross-cultural comparisons of the content of SF-36 translations across 10 countries: result from the IQOLA Project. J Clin Epidemiology 1998; 51 (11): 925-32.
13. Sararaks S, Rugayah B, Azman AB, Karuthan C & Low LL. Quality of life: How do Malaysian asthmatics fare? Med J Malaysia 2001; 56(3): 350-58.
14. AB Azman, S Sararaks, B Rugayah, LL Low, AA Azian, S Geeta, CT Tiew. Quality of life of the Malaysian general population: Results from a postal survey using the SF-36. Med J Malaysia 2003: 58(5); 694-711.
15. Ware JE, Gandek B. Methods for testing data quality, scaling assumptions and reliability: The IQOLA project approach. J Clin Epidemiol 1998; 51: 945-59.
16. Brazier J, Harper R, Jones N et al. Validating the SF-36 health survey questionnaire: New outcome measure for primary care. BMJ 1992; 305: 160-64.
17. Sullivan M, Karisson J. The Swedish SF-36 Health Survey III. Evaluation of criterion-based validity: Results from normative population. J Clin Epidemiol 1998; 51: 1105-113.
18. Thumboo J, Fong KY, Machin SP et al. A community-based study of scaling assumptions and construct validity of the English (UK) and Chinese (HK) SF-36 in Singapore. Quality of Life Research 2001; 10: 175-88.
19. Rugayah B, Nora'i MS, Mohan J, et al. National Health and Morbidity Survey 1996: Volume 11. Asthma. Kuala Lumpur: Institute of Public Health, Ministry of Health Malaysia, 1999.
20. Population and Housing Census of Malaysia 2000: Population Distribution and Basic Demographic Characteristics. Malaysia: Department of Statistics, July 2001.
21. Nunally J, Bernstein I. Psychometric Theory. New York: McGraw-Hill; 1994.
22. Tseng HM, Lu JFR, Gandek B. Cultural issues in Using SF-36 Health Survey in Asia: Results from Taiwan. Health & Quality of Life Outcomes 2003; 1: 72. (http://www.hqlo.com/content/1/1/72).
23. Shadbolt B, McCallum J & Singh M. Health outcomes by self-report: validity of the SF-36 among Australian hospital patients. Quality of Life Research 1997; 6: 343-52.
24. Ware JE Jr & Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. J Clin Epidemiol 1998; 51(11): 903-12.
25. Ren XS, Amick B 3rd, Zhou L & Gandek B. Translation and psychometric evaluation of a Chinese version of the SF-36

Health Survey in the United States. J Clin Epidemiol 1998; 51(11): 1129-138.

26. Fukuhara S, Ware JE Jr., Kosinski M, Wada S & Gandek B. Psychometric and clinical tests of validity of the Japanese SF-36 Health Survey. J Clin Epidemiol 1998; 51(11): 1045-053.

27. Ryder AG, Yang J & Heini S. Somatization vs. psychologization of emotional distress: A paradigmatic example for cultural psychopathology. In Lonner WJ, Dinnel DL, Hayes SA & Sattler DN (Eds.), Online Readings in Psychology and Culture, Centre for Cross-Cultural Research, Western Washington University, Bellingham, Washington, USA. (http://www.wwu.edu/~culture). 8 Jun 2004.

28. Ware JE, Kosinski M, Gandek B, Aaronson NK, Apolone G, Bech P, et al. The Factor structure of the SF-36 Health Survey in ten countries:Results from the IQOLA Project. J Clin Epidemiol 1998; 51 (11): 1159-165.

29. Lam CK, Gandek B, Ren XS & Chan MS. Tests of scaling assumptions and construct validity of the Chinese (HK) version of the SF-36 Health Survey. J Clin Epidemiol 1998; 51 (11): 1139-147.

30. Fukuhara S, Bito S, Green J, Hsio A & Kurokawa K. Translation, adaptation, and validation of the SF-36 Health Survey for use in Japan. J Clin Epidemiol 1998; 51(11): 1037-44.

31. Basic Survey Design & Implementation. Module 4 – Health surveys. Workshop on Evidence for Health Policy: Burden of Disease, Cost-effectiveness, Health Systems Performance and Health Surveys, Dikhololo, South Africa. World Health Organization & Burden of Disease Unit, Harvard Center for Population and Development Studies, Oct 2001.