# Validity and Reliability of MOS Short Form Health Survey (SF-36) for Use in India

Richa Sinha, Wim J A van den Heuvel, Perianayagam Arokiasamy[1]

Department of Community and Occupational Health, Research Institute SHARE, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands, [1]Department of Development Studies, International Institute for Population Sciences (IIPS), Mumbai, India

## ABSTRACT

**Background:** Health is defined as the state of complete physical, mental and social well-being than just the absence of disease or infirmity. In order to measure health in the community, a reliable and validated instrument is required. **Objectives:** To adapt and translate the Medical Outcomes Study Short-Form Health Survey (SF-36) for use in India, to study its validity and reliability and to explore its higher order factor structure. **Materials and Methods:** Face-to-face interviews were conducted in 184 adult subjects by two trained interviewers. Statistical analyses for establishing item-level validity, scale-level validity and reliability and tests of known group comparison were performed. The higher order factor structure was investigated using principal component analysis with varimax rotation. **Results:** The questionnaire was well understood by the respondents. Item-level validity was established using tests of item internal consistency, equality of item-scale correlations and item-discriminant validity. Tests of scale-level validity and reliability performed well as all the scales met the required internal consistency criteria. Tests of known group comparison discriminated well across groups differing in socio-demographic and clinical variables. The higher order factor structure was found to comprise of two factors, with factor loadings being similar to those observed in other Asian countries. **Conclusion:** The item- and scale-level statistical analyses supported the validity and reliability of SF-36 for use in India.

**Keywords:** Health surveys, reliability, questionnaires, quality of life, translations, validity

## Introduction

Health is defined as the state of complete physical, mental and social well-being than just the absence of disease or infirmity.[1] Therefore, patient's perspective about his/her state of health is recognized as an important parameter for assessing health outcomes and efficacy of an intervention. Validated and reliable self-reported questionnaires[2,3] are needed to enable this assessment.

The Medical Outcomes Study Short-Form Health Survey (SF-36)[4] is a widely used generic health-related quality of life (QoL) instrument consisting of 36 questions and measuring health in eight dimensions: physical functioning (PF), role limitations due to physical health problems (RP), bodily pain (BP), social functioning (SF), general mental health covering psychological distress and well-being (MH), role limitations due to emotional problems (RE), vitality, energy and fatigue (VT) and general health perceptions (GH). SF-36 has been adapted and translated into several languages,[5,6] and its validity and reliability established in several countries.[7]

SF-36 has been used in India to assess health outcomes in several diseased populations.[8-15] However, no studies regarding the validity and reliability of SF-36 in the general Indian population have been cited in electronic scientific databases. The primary objectives of this study were to adapt and translate SF-36 for use in India and to study its validity and reliability. Additionally, the study aimed to explore the higher order factor structure of the eight SF-36 scales.

**Address for correspondence:**
Ms. Richa Sinha, Melkzwamsingel 16, 3451HN Vleuten, The Netherlands. E-mail: sinha.richa@hotmail.com

## Materials and Methods

### Adaptation and translation

Conceptually equivalent cultural adaptations[5] were made to the questionnaire to make it suitable in the Indian context. In PF01, vigorous activities such as lifting heavy objects was exemplified as lifting a bucket of water to improve the understandability of the question. In PF02, moderate activities such as moving a table, pushing a vacuum cleaner, bowling or playing golf were replaced by brooming, shopping for groceries and cooking, as these activities can be considered to be equivalently moderate.

Climbing several flights of stairs was explained as climbing five to six floors (PF04), and one flight of stairs was expressed as one floor (PF05). One mile was expressed as 1 km (PF07) in accordance with the metric standards in India. A block was represented by a distance of 100 m (PF08 and PF09). Words like pep, dumps and blue in VT01, MH02 and MH04, respectively, were translated appropriately to capture their essence.

The questionnaire items were translated from English into Hindi by a professional experienced in translating health survey questionnaires. The translated version was back-translated into English by an independent person followed by a review to control for possible discrepancies. The final translated version evolved following an iterative process,[6] which included pilot-testing and incorporating appropriate changes.

### Data collection

The study sample was selected from Mumbai, its suburbs and a village in its vicinity following purposive sampling. Face-to-face interviews were conducted by two trained interviewers in residential areas, industrial estates, business centers, offices and market place by selecting the units systematically and interviewing the people present at the time of visit.

The Institutional Review Board of one of the co-authors' institute in India approved the study. Subjects 18 years and above who could communicate in Hindi were included. The study purpose was explained to the subjects and interviews were performed after obtaining written informed consent. Two people did not participate in the study, one because of lack of time and the other due to being psychologically disturbed due to the death of a family member.

### Statistical analyses

The statistical analyses were performed using Statistical Package for Social Sciences (SPSS), version 15.[16]

Tests of item internal consistency, equality of item-scale correlations and item-discriminant validity[17] were carried out to establish item-level validity. Item internal consistency is considered satisfactory if an item correlates 0.40 or more with its hypothesized scale after correction for item-scale overlap. Equality of item-scale correlations is considered satisfied when items have similar item-scale correlations. Item-discriminant validity is supported if the correlation between an item and its hypothesized scale is significantly higher than the correlations between that item and the other scales.

Scale-level validity and reliability were assessed by testing whether the SF-36 scale scores showed substantial variability and if their reliability estimated using the internal consistency method (Cronbach alpha coefficient) was acceptable, namely 0.70 or higher for group comparisons,[18] and whether the internal consistency of each scale was higher than the correlation between that scale and the other scales, which tests if each scale measured a distinct health concept.[17]

Tests of known group comparison were performed to assess the ability of SF-36 to distinguish between groups differing in factors known to affect QoL. Gender, age, residential area and comorbidities were used as grouping variables. Non-parametric tests (Mann-Whitney U or Kruskal-Wallis one-way ANOVA) were performed to test the statistical significance of observed differences between the groups.

Exploratory factor analyses were performed on the eight SF-36 scales using principal component analysis with varimax rotation[19] to examine the existence of higher order factor structure in the Indian general population.

## Results

### Sample characteristics

Table 1 summarizes the socio-demographic and clinical attributes of the study population. The population ranged in age from 18 to 76 years, with a mean age of 37.44 years.

### Statistical analyses

Table 2 summarizes the descriptive statistics for the eight SF-36 scales. Within each scale, the correlation between items and their hypothesized scale were of similar magnitude. Item internal consistency and item-discriminant validity criteria were met for most of the scales. The floor effect was found to be less than 10% for all the scales except RP and RE. A ceiling effect of more than 50% was observed for all the scales except GH, VT and MH. For all the scales, the internal consistency (Cronbach alfa) equaled or exceeded 0.70 [Table 3]. The internal consistency of each scale exceeded the correlations between that scale and the other scales.

Table 4 summarizes the tests for known group comparisons. Females scored lower on all the eight scales, with the differences being statistically significant in PF, RP, VT, MH and GH. With age, the range of scores related to physical health scales (PF, RP, BP and GH) were more than those related to mental health scales (RE, VT, MH and SF). Scales associated with mental health showed significantly different scores in groups classified as per residential area, and only one physical health scale (RP) was found to be significantly different. All the scales were found to be lower in the sub-group having comorbidities, and these differences were significant for all the scales except for RE.

Exploratory factor analysis of the eight SF-36 scales yielded a two-factor solution [Table 5]. PF, RP and BP correlated more with factor 1, whereas SF, RE and MH correlated more with factor 2. VT and GH scales correlated moderately with both the factors. The two principal components explained 63.42% of the total variance, which was within the prescribed range.[20,21] The obtained two-factor structure was similar to that obtained in other studies (factor 1 as "physical" and factor 2 as "mental" domains of health underlying the SF-36).[22]

## Discussion

The study sample can be considered adequate because a sample of 100 or more has been recommended for reliability and validity[23,24] studies, and to perform second-order analysis on SF-36 scales.[6,25] SF-36 translation and validation studies in several languages[26-29] have been carried out using similar sample sizes. The overrepresentation of males (72%) in the study population is not an issue as the study did not aim at generating normative QoL data.

SF-36 was adapted to make it suitable for use in India. Similar adaptations have also been done while translating SF-36 for use in other countries in order to preserve the conceptual meaning of the original question to make it relevant within each country and language. In Iran,[30] in items regarding activities, bowling and playing golf were changed to light sport activities, mile was changed to kilometer and walking one block or walking several blocks were changed to walking one alley or several alleys to refer to a similar distance. In the United Kingdom,[31] walking one block was changed

### Table 1: Socio-demographic and clinical characteristics of the study population

| Variable | *n* | % |
|---|---|---|
| Sex | | |
| Male | 132 | 71.7 |
| Female | 52 | 28.3 |
| Age (years) | | |
| 18-28 | 56 | 30.4 |
| 29-38 | 50 | 27.2 |
| 39-48 | 44 | 23.9 |
| 49+ | 34 | 18.5 |
| Education | | |
| No schooling | 23 | 12.5 |
| Primary education | 76 | 41.3 |
| Secondary education | 23 | 12.5 |
| Tertiary education | 61 | 33.2 |
| Missing values | 1 | 0.5 |
| Residential area | | |
| Village | 36 | 19.6 |
| Suburbs | 55 | 29.9 |
| Metropolitan | 93 | 50.5 |
| Marital status | | |
| Never married | 47 | 25.5 |
| Married | 131 | 71.2 |
| Separated/widowed | 6 | 3.2 |
| Comorbidity | | |
| Present | 52 | 28.3 |
| Not present | 132 | 71.7 |
| Comorbidities* | | |
| Musculo-skeletal | 29 | 55.8 |
| Diabetes | 9 | 17.3 |
| Hypertension | 14 | 26.9 |
| Others | 10 | 19.23 |

*Includes single as well as multiple comorbidities

### Table 2: SF-36 scale descriptives, tests of item internal consistency and discriminant validity

| Scale | Mean (SD) | Ceiling/floor (%) | Range of item-scale correlations | | Internal consistency tests[c] | Discriminant validity tests[d] |
|---|---|---|---|---|---|---|
| | | | Item-internal consistency[a] | Item-discriminant validity[b] | Success rate (%) | Success rate (%) |
| PF | 93.59 (14.34) | 62.0/1.1 | 0.36-0.76 | 0.00-0.49 | 90 | 100 |
| RP | 78.53 (35.37) | 67.4/10.3 | 0.68-0.83 | 0.15-0.56 | 100 | 100 |
| BP | 83.80 (26.98) | 63.6/1.6 | 0.70 | 0.12-0.57 | 100 | 100 |
| GH | 79.41 (20.42) | 16.3/1.6 | 0.36-0.61 | 0.11-0.50 | 80 | 89 |
| VT | 80.82 (19.22) | 27.2/0.5 | 0.47-0.57 | 0.19-0.56 | 100 | 96 |
| SF | 90.42 (17.70) | 69.0/0.5 | 0.63 | 0.10-0.48 | 100 | 100 |
| RE | 79.89 (35.71) | 72.3/12.5 | 0.69-0.80 | 0.00-0.44 | 100 | 100 |
| MH | 86.16 (14.92) | 29.9/0.5 | 0.36-0.61 | 0.09-0.63 | 80 | 89 |

[a]Correlations between items and hypothesized scale corrected for overlap; [b]Correlation between items and other scales; [c]Number of items having correlation more than 0.40; [d]Number of correlations significantly greater/total number of correlations

to walking 100 yards and downhearted and blue was changed to downhearted and low. In Sweden,[31] walking in the forest or gardening was used as an example for

moderately strenuous activity instead of bowling or playing golf.

**Table 3: Interscale correlations and reliability coefficients for the eight SF-36 scales**

|    | PF | RP | RE | VT | MH | SF | BP | GH |
|----|----|----|----|----|----|----|----|----|
| PF | (0.88) | | | | | | | |
| RP | 0.52** | (0.88) | | | | | | |
| RE | 0.05 | 0.41** | (0.86) | | | | | |
| VT | 0.43** | 0.45** | 0.42** | (0.73) | | | | |
| MH | 0.20** | 0.38** | 0.45** | 0.67** | (0.71) | | | |
| SF | 0.17* | 0.20** | 0.46** | 0.43** | 0.49** | (0.76) | | |
| BP | 0.43** | 0.56** | 0.24** | 0.39** | 0.41** | 0.17* | (0.82) | |
| GH | 0.36** | 0.49** | 0.34** | 0.61** | 0.48** | 0.28** | 0.44** | (0.73) |

Reliability coefficients (Cronbach alfa) provided in round braces; **Correlation is significant at the 0.01 level (2-tailed); *Correlation is significant at the 0.05 level (2-tailed)

**Table 4: Summary of principal component analysis for the eight SF-36 scales**

| Scale | Correlation with components[a] | | Relative validity[b] | | Variance explained | |
|-------|---------|---------|---------|---------|---------|---------|
|       | Factor 1 | Factor 2 | Factor 1 | Factor 2 | Total[c] | Reliable[d] |
| PF | 0.81 | -0.03 | 1.00 | -0.04 | 0.66 | 0.75 |
| RP | 0.78 | 0.26 | 0.96 | 0.33 | 0.67 | 0.76 |
| BP | 0.75 | 0.18 | 0.92 | 0.23 | 0.60 | 0.68 |
| MH | 0.32 | 0.76 | 0.40 | 0.98 | 0.69 | 0.97 |
| RE | 0.10 | 0.77 | 0.13 | 0.98 | 0.60 | 0.69 |
| SF | 0.02 | 0.78 | 0.03 | 1.00 | 0.61 | 0.80 |
| VT | 0.52 | 0.65 | 0.63 | 0.83 | 0.68 | 0.93 |
| GH | 0.65 | 0.45 | 0.80 | 0.57 | 0.57 | 0.78 |

[a]Correlation between each SF-36 scale and varimax rotated principal components [b]Ratio of the common factor variance of each scale relative to the scale with the greatest common factor variance [c]$h^2$, proportion of the total variance of each scale explained by the two components [d]$h^2/r_{tt}$, proportion of the reliable variance of each scale explained by the two components

The adapted and translated version was well understood by the respondents as there was a low (1%) missing data at both item and scale levels, and all the response choices were used. The suggested Cronbach alfa values of 0.70 for group-level comparison[7] were met for all the scales, thus, supporting the item homogeneity and internal consistency across scales. Tests of known group comparison discriminated well among the groups, differing in socio-demographic and clinical variables. These findings were similar to those reported in other studies.[30,32-35] In Greece,[32] women reported worse health than men. Moreover, age was an important health status factor affecting physical health relatively more than mental health. Also, in Iran,[30] SF-36 discriminated well between sub-groups of people differing in gender and age. The findings demonstrated that women and old people had poorer health as compared with men and younger people.

Exploratory factor analyses were performed as the higher order factor structure of SF-36 has not been established in the general Indian population.[25] The higher order factor structure was found to be similar to that in other countries.[22] The two factors explained 63.42% of the total variance of the SF-36 scale scores and 68-97% of the reliable variance of each scale. SF and MH showed, respectively, lower and higher association with physical domain in congruence with the findings from other Asian countries.[26,27,30,36]

To conclude, the item and scale level analyses supported the validity and reliability of the translated and

**Table 5: Mean SF-36 scale scores in relation to socio-demographic and clinical characteristics**

| Variable | PF | RP | RE | VT | MH | SF | BP | GH |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|
| Sex | | | | | | | | |
| Male | 95.19 | 84.28 | 81.06 | 85.19 | 89.00 | 91.19 | 85.06 | 71.69 |
| Female | 89.41 | 63.24 | 76.47 | 69.83 | 78.94 | 88.24 | 80.25 | 82.35 |
| *P* value* | 0.035 | <0.001 | 0.171 | <0.001 | <0.001 | 0.061 | 0.217 | 0.007 |
| Age (years) | | | | | | | | |
| 18-28 | 96.25 | 83.48 | 83.93 | 79.23 | 85.23 | 91.52 | 88.62 | 82.86 |
| 29-38 | 96.50 | 74.00 | 75.33 | 81.60 | 84.72 | 89.00 | 79.40 | 78.10 |
| 39-48 | 93.87 | 83.72 | 79.85 | 85.23 | 88.86 | 91.86 | 85.06 | 83.95 |
| 49+ | 84.56 | 69.85 | 79.41 | 77.21 | 86.59 | 88.60 | 80.29 | 69.74 |
| *P* value** | 0.038 | 0.174 | 0.692 | 0.461 | 0.686 | 0.784 | 0.227 | 0.104 |
| Residential area | | | | | | | | |
| Village | 91.81 | 61.11 | 68.52 | 70.05 | 81.61 | 86.46 | 81.81 | 78.92 |
| Suburbs | 94.81 | 85.00 | 84.24 | 85.64 | 88.61 | 91.36 | 84.18 | 83.27 |
| Metropolitan | 93.53 | 81.25 | 81.52 | 82.34 | 86.61 | 91.30 | 84.18 | 77.23 |
| *P* value** | 0.851 | <0.001 | 0.008 | <0.001 | 0.015 | 0.010 | 0.895 | 0.238 |
| Comorbidity | | | | | | | | |
| Present | 82.60 | 61.06 | 73.08 | 70.22 | 79.15 | 85.09 | 65.38 | 67.21 |
| Not present | 97.94 | 85.31 | 82.44 | 85.15 | 88.99 | 92.46 | 90.99 | 84.21 |
| *P* value* | <0.001 | <0.001 | 0.058 | <0.001 | 0.002 | 0.040 | <0.001 | <0.001 |

*Mann-Whitney U test; **Kruskal-Wallis test

adapted version of SF-36 for use in India. Studies with larger sample size with representative general Indian population are encouraged to generate normative QoL data.

## References

1. World Health Organization. Handbook of basic documents. 46th ed. World Health Organization: Geneva.

2. Saw SM, Ng TP. The design and assessment of questionnaires in clinical research. Singapore Med J 2001;42:131-5.

3. Hays RD, Anderson R, Revicki D. Psychometric considerations in evaluating health-related quality of life measures. Qual Life Res 1993;2:441-9.

4. McDowell I. General health status and quality of life. Measuring health: A guide to rating scales and questionnaires. 3rd ed. Oxford University Press, Inc.: New York, NY; 2006. p. 520-703.

5. Wagner AK, Gandek B, Aaronson NK, Acquadro C, Alonso J, Apolone G, *et al*. Cross-cultural comparisons of the content of SF-36 translations across 10 countries: Results from the IQOLA project. International Quality of Life Assessment. J Clin Epidemiol 1998;51:925-32.

6. Bullinger M, Alonso J, Apolone G, Leplège A, Sullivan M, Wood-Dauphinee S, *et al*. Translating health status questionnaires and evaluating their quality: The IQOLA project approach. J Clin Epidemiol 1998;51:913-23.

7. Gandek B, Ware JE Jr, 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;51:1149-58.

8. Muniyandi M, Rajeswari R, Balasubramanian R, Nirupa C, Gopi PG, Jaggarajamma K, *et al*. Evaluation of post-treatment health-related quality of life (HRQoL) among tuberculosis patients. Int J Tuberc Lung Dis 2007;11:887-92.

9. Rajeswari R, Muniyandi M, Balasubramanian R, Narayanan PR. Perceptions of tuberculosis patients about their physical, mental and social well-being: A field report from south India. Soc Sci Med 2005;60:1845-53.

10. Wig N, Lekshmi R, Pal H, Ahuja V, Mittal CM, Agarwal SK. The impact of HIV/AIDS on the quality of life: A cross sectional study in north India. Indian J Med Sci 2006;60:3-12.

11. Kohli RM, Sane S, Kumar K, Paranjape RS, Mehendale SM. Assessment of quality of life among HIV-infected persons in Pune, India. Qual Life Res 2005;14:1641-7.

12. Kohli RM, Sane S, Kumar K, Paranjape RS, Mehendale SM. Modification of medical outcome study (MOS) instrument for quality of life assessment and its validation in HIV infected individuals in India. Indian J Med Res 2005;122:297-304.

13. Bedi GS, Gupta N, Handa R, Pal H, Pandey RM. Quality of life in Indian patients with rheumatoid arthritis. Qual Life Res 2005;14:1953-8.

14. Haroon N, Aggarwal A, Lawrence A, Agarwal V, Misra R. Impact of rheumatoid arthritis on quality of life. Mod Rheumatol 2007;17:290-5.

15. Ray J, Das SK, Gangopadhya PK, Roy T. Quality of life in Parkinson's disease-Indian scenario. J Assoc Physicians India 2006;54:17-21.

16. SPSS for Windows, Rel. 15.0.1. SPSS Inc: Chicago, IL; 2006.

17. Ware JE Jr, Gandek B. Methods for testing data quality, scaling assumptions, and reliability: The IQOLA project approach. International Quality of Life Assessment. J Clin Epidemiol 1998;51:945-52.

18. Nunnally JC, Bernstein IR. Psychometric theory. 3rd ed. McGraw-Hill: New York, NY; 1994.

19. Ware Je Jr, Kosinski M, Keller SD. SF-36 physical and mental health summary scales: A user's manual. Boston, MA: The Health Institute, New England Medical Center; 1994.

20. Ware JE Jr, Gandek B, Kosinski M, Aaronson NK, Apolone G, Brazier J, *et al*. The equivalence of SF-36 summary health scores estimated using standard and country-specific algorithms in 10 countries: Results from the IQOLA project. International Quality of Life Assessment. J Clin Epidemiol 1998;51:1167-70.

21. Norman GR, Streiner DL. Exploratory factor analysis. PDQ Statistics. 3rd ed. Hamilton, ON: BC Decker Inc; 2003. P. 144-55.

22. Ware JE Jr, Kosinski M, Gandek B, Aaronson NK, Apolone G, Bech P, *et al*. The factor structure of the SF-36 health survey in 10 countries: Results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol 1998;51:1159-65.

23. Pagano M, Gauvreau K. Principles of biostatistics. 2nd ed. Pacific Grove, CA: Duxbury Press; 2000.

24. Peat J, Mellis C, Williams K, Xuan W. Health science research: A handbook of quantitative methods. 1st ed. London: Sage Publications Ltd.; 2002.

25. de Vet HC, Adèr HJ, Terwee CB, Pouwer F. Are factor analytical techniques used appropriately in the validation of health status questionnaires? A systematic review on the quality of factor analysis of the SF-36. Qual Life Res 2005;14:1203-18.

26. Lam CL, 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;51:1139-47.

27. Fukuhara S, Bito S, Green J, Hsiao A, Kurokawa K. Translation, adaptation, and validation of the SF-36 health survey for use in Japan. J Clin Epidemiol 1998;51:1037-44.

28. Leplège A, Ecosse E, Verdier A, Perneger TV. The French SF-36 health survey: Translation, cultural adaptation and preliminary psychometric evaluation. J Clin Epidemiol 1998;51:1013-23.

29. Ren XS, Amick B 3rd, Zhou L, Gandek B. Translation and psychometric evaluation of a Chinese version of SF-36 health survey in the United States. J Clin Epidemiol 1998;51:1129-38.

30. Montazeri A, Goshtasebi A, Vahdaninia M, Gandek B. The short form health survey (SF-36): Translation and validation study of the Iranian version. Qual Life Res 2005;14:875-82.

31. Keller SD, Ware JE Jr, Bentler PM, Aaronson NK, Alonso J, Apolone G, *et al*. Use of structural equation modeling to test the construct validity of the SF-36 health survey in ten countries: Results from the iqola project. J Clin Epidemiol 1998;51:1179-88.

32. Pappa E, Kontodimopoulos N, Niakas D. Validating and norming of the Greek SF-36 health survey. Qual Life Res 2005;14:1433-8.

33. Taft C, Karlsson J, Sullivan M. Performance of the swedish SF-36 version 2.0. Qual Life Res 2004;13:251-6.

34. Apolone G, Mosconi P. The italian SF-36 health survey: Translation, validation and norming. J Clin Epidemiol 1998;51:1025-36.

35. Sabbah I, Drouby N, Sabbah S, Retel-Rude N, Mercier M. Quality of life in rural and urban populations in lebanon using SF-36 health survey. Health Qual Life Outcomes 2003;1:30.

36. Lam CL, Tse EY, Gandek B, Fong DY. The SF-36 summary scales were valid, reliable, and equivalent in a Chinese population. J Clin Epidemiol 2005;58:815-22.