Limitations of the SF-36 in a sample of nursing home residents

Elena M. Andresen, Gwendell W. Gravitt, Marie E. Aydelotte[1], Carol A. Podgorski[2]

Department of Community Health, St Louis University School of Public Health, 3663 Lindell Boulevard, St Louis, MO 63108, USA
Departments of [1]Medicine and [2]Psychiatry, University of Rochester, Rochester, NY, USA

Address correspondence to E. M. Andresen. Fax: (+314) 977 8150. Email: Andresen@.slu.edu

Abstract

Objective: to assess test characteristics of the Medical Outcomes Study SF-36 (Short-Form 36) with residents of nursing homes.
Research design: nursing home residents with 17 or more points on the Mini-Mental State Examination (MMSE) and ≥ 3 months residence (128 of 552 screened) were selected randomly. Interviewers administered the SF-36 (repeated after 1 week), Geriatric Depression Scale and MMSE. We recorded activities of daily living and medication data from medical records. Data analysis included test–retest intraclass correlations, item completion, score distributions and SF-36 correlations with measures of physical and mental functioning.
Results: 97 nursing home residents (75.8%) consented. Test–retest intraclass correlation coefficients were good to excellent (range = 0.55 to 0.82). Convergent validity between SF-36 physical health scales and the activities of daily living index was modest (r range = −0.37 to −0.43). About 25% of residents scored zero (lowest score) on at least one SF-36 physical function measure. SF-36 mental health scales correlated strongly with the Geriatric Depression Scale (r range = −0.63 to −0.71) and modestly with bodily pain (r = −0.35). No SF-36 scales correlated strongly with the MMSE.
Conclusion: only one in five nursing home residents met minimal participation criteria, suggesting limited utility of the SF-36 in nursing homes. Reliability and validity characteristics were fairly good. Skewed scores were noted for some SF-36 scales. The utility of the SF-36 may be limited to assessments of subjects with higher cognitive and physical functioning than typical nursing home residents. The SF-36 might benefit from modification for this setting, or by tests of proxy ratings.

Keywords: *nursing home, quality of life, reliability, SF-36, validity*

Introduction

Self-reported health status is receiving increasing attention in epidemiological and outcomes research [1–4]. Maintenance of function and optimal quality of life is especially relevant to those with chronic illness, including elderly people [5–9]. Additionally, self-rated health status is recommended for geriatric assessment [10].

The Medical Outcomes Study SF-36 (Short-Form 36) was developed to assess self-perceived health in a variety of settings [11–13] and its performance generally is rated highly [14, 15]. We are aware of only one report of the SF-36 in nursing home residents, where correlation between residents' and health care providers' assessment of health status was relatively poor [16]. Measurement tools for older adults may need continued refinement, testing and further

development [17]. Since instruments like the SF-36 are being applied to assess older adults' health risks, for example in managed care [18–22], further research is needed to characterize SF-36 test attributes in these settings.

The 5–10% [23] of older Americans who live in nursing homes differ importantly from those living at home. Most have advanced chronic illnesses, and the dementia prevalence is about 40% [24]. Multiple functional dependencies are common, with one-third of residents dependent in all basic activities of daily living (ADLs) [25].

These unique attributes of nursing home residents were the basis for current study goals of evaluating the SF-36 for test–retest reliability, data completion, extreme score distributions and convergent validity in comparison with the Folstein Mini-Mental State Examination (MMSE) [26], the Geriatric Depression Scale

(GDS) [27] and ADLs [28]. The final objective was to examine practical considerations related to using the SF-36 in nursing homes.

Methods

Subjects

We obtained approval from the University of Rochester human subjects review committee and chose participants from two similarly sized nursing homes in Rochester, NY: a private, not-for-profit community home and a county-owned, university-affiliated facility. Both offered skilled nursing and intermediate care, catering for people with higher levels of self-care dependency and for those who were more independent. We selected randomly from the resident list using a simple random numbers table at each institution until 50 eligible subjects were enrolled. We eliminated those who were aged less than 58 or who had been in residence for less than 3 months. We also excluded residents with severe dementia or who were unable to communicate. Residents were then screened with the MMSE to exclude subjects with fewer than 17 points.

Protocol

Two interviewers administered informed consent and then the SF-36 and GDS to participants, repeating the SF-36 after 1 week. One investigator (M.E.A.) abstracted ADLs (transfer, locomotion, dressing, eating, toilet use, personal hygiene, bathing and bladder continence) and medications from the most recent MDS+ data within medical records. Each score ranged from 0 (independent) to 4 (completely dependent). Overall possible ADL summary scores ranged from 0 to 32.

Analysis

Analyses used SAS [29] and SPSS [30] statistical software. Intraclass correlations and lower 95 percent confidence limits (random effects, one-way model) were calculated for SF-36 scales [31–33]. Convergent validity was estimated using Pearson product-moment statistics (r) between the SF-36 scales and ADL, GDS and MMSE. We considered $r > 0.30$ as modest and $r > 0.49$ as strong correlations. We judged scales as having floor or ceiling effects if 20% of subjects had either the highest or lowest score [34].

Results

Seventy-five (18.8%) of the 400 residents in the group requiring skilled nursing and 53 of 152 intermediate level residents (34.9%) met study criteria. Ineligibility was most often because of dementia or communication problems. Thirty-one eligible residents declined to participate (75.8% response).

Demographic characteristics of the 97 participants are shown in Table 1. Most subjects were elderly (mean age 80), white (92%) and female (80.4%); 54.6% had completed high school. Mean duration of residency was about 3 years. Sixty-four percent received Medicaid (poverty) benefits: 48% of those receiving skilled nursing and 52% of those receiving intermediate care. Residents of the county-owned public facility were younger (mean age 73.6 *versus* 86.6, $P < 0.01$) and more likely to receive Medicaid (89% *versus* 41%, $P < 0.01$). Otherwise the groups were similar.

Although all participants completed the SF-36 at baseline and follow-up, up to 7% of SF-36 scales were missing because of skipped questions [11, 12]. Mean scores on the eight SF-36 scales and two summary scales are displayed in Table 2. The two physical health scales demonstrated floor effects (almost 25% received minimum, zero scores). Surprisingly, 14 respondents (18.9%) also scored in the top 1% of the role functioning–physical scale. The summary scale scores were relatively normally distributed (not shown). Test–retest reliability is summarized in Table 2. The intraclass correlation coefficients (ICC) generally were good to excellent (ICC range = 0.65 to 0.82), except for social functioning and role–emotional scales (ICC = 0.52 and 0.55, respectively). Despite screening for dementia and communication inability, substantial dependency was present: Table 3 displays the descriptive statistics for ADLs, cognition and the GDS.

SF-36 physical health domain scales showed the expected modest correlation with ADLs. Mental health scales and GDS scores generally correlated strongly; the role–emotional scale correlated modestly. Bodily pain and vitality scales correlated with GDS scores. Overall, appropriate SF-36 scales appeared to correlate moderately and sometimes strongly with ADL and GDS scores (convergent validity; Table 4).

Table 1. Demographic characteristics of the 97 study participants

	Value	Range
Mean age, years (SD)	80.1 (9.9)	58–101
Women	80.4%	
Caucasian	92%	
Education (high school or greater)	54.6%	
Skilled nursing level of care	48%	
Intermediate level of care	52%	
Mean no. of months in nursing home (SD)	40.7 (49.2)	3–256
Receiving Medicaid financial benefits	64%	
Mean no. of medications (SD)	9.1 (4.6)	1–26

Table 2. Descriptive statistics and test–retest estimates for Short-Form 36 scales in 97 nursing home residents

Scale	No. scored (n = 97)	Mean score (and SD)	Floor/ceiling effects		ICC (95% lower CI)
			% at 0	% at 100	
Physical functioning	95	32.4 (30.4)	26.8	1.1	0.79 (0.74)
Social functioning	97	72.6 (28.8)	3.1	36.1	0.52 (0.43)
Role–physical	95	42.5 (37.5)	29.5	18.9	0.76 (0.70)
Role–emotional	95	69.5 (35.5)	11.6	49.5	0.55 (0.46)
Mental health	96	64.7 (22.4)	1.0	2.1	0.80 (0.76)
Vitality	96	49.9 (19.9)	1.0	2.1	0.74 (0.76)
Bodily pain	97	60.9 (29.5)	2.1	21.6	0.65 (0.58)
General health	97	52.6 (18.6)	0	0	0.65 (0.58)
Physical health—summary	90	33.3 (10.9)	0	0	0.82 (0.77)
Mental health—summary	90	50.5 (11.6)	0	0	0.79 (0.74)

ICC, intraclass correlation coefficient; CI, confidence interval.

Table 3. Descriptive statistics for activities of daily living, cognition and the Geriatric Depression Scale in 97 nursing home residents

Scale	Mean (SD)	Range
Activities of daily living	11.5 (9.6)	0–32
Mini-Mental State Examination	23.1 (4.0)	17–30
Geriatric Depression Scale	10.4 (6.8)	0–28

Table 4. Pearson product moment correlation among measures in 97 nursing home residents

SF-36 scale	ADL	MMSE	GDS
Physical functioning	**−0.37**[a]	−0.16	−0.25
Social functioning	−0.22	0.06	**−0.50**
Role—physical	**−0.43**	0.08	−0.46
Role—emotional	−0.11	0.09	**−0.34**
Mental health	−0.05	0.11	**−0.71**
Vitality	−0.11	0.10	**−0.66**
Bodily pain	−0.06	−0.18	−0.35
General health	−0.17	−0.01	−0.41
Physical health—summary	**−0.38**	−0.14	−0.28
Mental health—summary	0.01	0.18	**−0.63**

SF-36, Short-Form 36; ADL, activities of daily living; MMSE, Mini-Mental State Examination; GDS, Geriatric Depression Scale.
[a]Hypothesized correlations for convergent validity (regardless of actual value obtained) are shown in bold.

Discussion

In this study of the characteristics of the SF-36 in nursing homes, we screened 552 randomly-chosen residents to obtain 128 eligible subjects. That only one in five residents was eligible is an important finding and suggests a very limited utility of the SF-36 in a setting where cognitive and communication ability is needed to obtain valid self-reported health status. In addition, test–retest reliability of the SF-36 scales was weaker than in a community-based study of older adults [35].

Extreme scale floor effects are more common among ill, disadvantaged and older patients [6, 35, 36]. This study confirms this problem, with five of the eight SF-36 scales showing floor effects. Additionally, several scales demonstrated ceiling effects. Summary scale scores, scored in reference to the fiftieth percentile, showed no floor or ceiling effects.

SF-36 convergent validity was satisfactory. Physical function scales demonstrated a modest correlation with the ADL index, whereas mental health scales were associated strongly with the GDS. The SF-36 scale and MMSE scores did not correlate well. This might imply there is not a strong association between cognitive and affective status (measured by the SF-36) in these subjects and correlations might also be attenuated due to the MMSE's restricted score range (the lower cut-off score of >17 points).

Face validity appears to be a major limitation for using the SF-36 in nursing homes. Nine items refer to activities not generally performed in this setting (e.g. carrying groceries) and six refer to 'work'. Lower face validity might affect the retest reliability because of uncertainty about the meaning and relevance of some questions. The Sickness Impact Profile and the Quality

of Well-Being are other measures of self-perceived health status modified for use with older adults in long-term care [37, 38]. The SF-36 also might benefit from modification.

In conclusion, using an MMSE cut-off of 17 and restricting to those who could self-report, only one in five nursing home residents could complete the SF-36. The SF-36 may be best used for assessing individuals who have less cognitive and physical impairment than typical nursing home residents. Furthermore, the SF-36 has important face validity problems. With modification, the SF-36 could potentially be used more broadly in nursing homes. The use of proxy responses for the SF-36 should receive formal evaluation.

Acknowledgements

Key points

- Only one in five nursing home residents met minimal participation criteria, suggesting that the SF-36 might of utility only in assessing residents who have higher cognitive and physical function than are typically found in this group.
- Five of eight scales had either floor or ceiling effects.
- Test–retest reliability and convergent (although not face) validity were fairly good.
- The SF-36 might benefit from modification for use in nursing homes or by tests of proxy ratings.

References

1. Ware, JE Jr. Measuring patient function and well-being: some lessons from the Medical Outcomes Study. In: Heitoff KA, Lohr KN, eds. Effectiveness and Outcomes in Health Care. Washington, DC: National Academy Press, 1990.

2. Andresen EM. Measures of general health status. In: Andresen EM, Rothenberg BM, Zimmer JG, eds. Assessing the Health of Older Adults. New York: Springer, 1997, 41–91.

3. Rothman ML, Diehr P, Hedrick SC et al. Effects of contract adult day health care on health outcomes and satisfaction with care. Med Care 1993; 31: SS75–83.

4. Patrick DL, Erickson P. Health Status and Health Policy: allocating resources to health care. New York: Oxford University Press, 1993.

5. Lyons RA, Perry HM, Littlepage BNC. Evidence for the validity of the Short-Form 36 Questionnaire (SF-36) in an elderly population. Age Ageing 1994; 23: 182–4.

6. McHorney CA. Measuring and monitoring general health status in elderly persons: practical and methodological issues in using the SF-36 health survey. Gerontologist 1996; 36: 571–83.

7. Sherbourne CD, Meredith LS. Quality of self-reported data: a comparison of older and younger chronically ill patients. J Gerontol 1992; 47: S204–11.

8. Stewart AL, Greenfield S, Hays RD et al. Functional status and well-being of patients with chronic conditions. Results from the Medical Outcomes Study. JAMA 1989; 262: 907–13.

9. Andresen EM, Rothenberg BM, Panzer R et al. Selecting a generic measure of health-related quality of life for use among older adults: a comparison of candidate instruments. Eval Health Profes 1998; 21: 244–64.

10. American College of Physicians, Health and Public Policy Committee. Comprehensive functional assessment for elderly patients. Ann Int Med 1988; 109: 70–2.

11. Medical Outcomes Trust. How to Score the SF-36 Health Survey. Boston: Medical Outcomes Trust, 1994.

12. Ware JE Jr, Kosinski M, Keller SD. SF-36 Physical and Mental Health Summary Scales: a user's manual. Boston: The Health Institute, 1994.

13. 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: 473–83.

14. McHorney CA, Ware JE Jr, Lu JFR et al. 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: 40–66.

15. McHorney CA, Ware JE Jr. Raczek AE. The MOS 36-item short-form survey (SF-36): II. Psychometric and clinical test of validity in measuring physical and mental health constructs. Med Care 1993; 31: 247–63.

16. Berlowitz DR, Du Wei, Kazis L et al. Health-related quality of life of nursing home residents: Difference in patient and provider perceptions. J Am Geriatr Soc 1995; 43: 799–802.

17. Feasely JC, ed. Health Outcomes for Older People: questions for the coming decade. Washington, DC: Institute of Medicine/National Academy Press, 1996.

18. Boult C, Dowd B, McCaffrey D et al. Screening elders for risk of hospital admission. J Am Geriatr Soc 1993; 41: 811–7.

19. Gouveia WA, Bungay KM, Massaro FJ et al. Paradigm for the management of patient outcomes. Am J Hosp Pharm 1991; 48: 1912–6.

20. HMO Workgroup on Care Management. Identifying high-risk Medicare HMO members: a report from the HMO Workgroup on Care Management, April 1996. Princeton, NJ: Robert Wood Johnson Foundation, 1996.

21. Pacala JT, Boult C, Boult L. Predictive validity of a questionnaire that identifies older persons at risk for hospital admission. J Am Geriatr Soc 1995; 43: 374–7.

22. Ware J Jr, Kosinski M, Keller SD. SF-12: How to score the SF-12 physical and mental health summary scales. Boston: The Health Institute, 1995.

23. Karuza J, Kane RA. Long-Term Care. In: Cassell CK, Cohen HJ, Larson EB eds. Geriatric Medicine, second edition. New York: Springer-Verlag, 1990.

24. Tariot PN, Podgorski CA, Blazina L et al. Leibovici A. Mental disorders in the nursing home: another perspective. Am J Pscyh 1993; 150: 1063–9.

25. Hing E, Sekscenski E, Strahan G. National Center for Health Statistics. The National Nursing Home Survey: 1985 summary for the United States. Vital and Health Statistics. Series 13, no. 97.

26. Folstein MF, Folstein SE, McHugh PR. Mini-Mental State: a practical method for grading cognitive-state of patients for the clinician. J Psychiatr Res 1975; 12: 189–98.

27. Yesavge JA, Brinak TL, Rose TL *et al.* Development and validation of a geriatric depression scale: a preliminary report. J Psychiatic Res 1988; 17: 37–49.

28. Morris JN, Hawes C, Fries BE *et al.* Designing the national resident assessment instrument for nursing home facilities. Gerontologist 1990; 30: 293–307.

29. SAS/STAT User's Guide, version 6, fourth edition. Cary, NC: SAS Institute Inc., 1989.

30. SPSS 7.0 TM for Windows. Chicago, IL: SPSS. Inc., 1996.

31. Armstrong BK, White E, Saracci R. Principles of Exposure Measurement in Epidemiology. Monographs in Epidemiology and Biostatistics, volume 21. New York: Oxford University Press, 1992.

32. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures: statistics and strategies for evaluation. Controlled Clin Trials 1991; 12: 142S–58S.

33. Fleiss JL. The Design and Analysis of Clinical Experiments. New York: John Wiley, 1986.

34. Andresen EM, Rothenberg BM, Kaplan RM. Performance of a self-administered mailed version of the Quality-of-Well-Being (QWB-SA) questionnaire among older adults. Med Care 1998; 36: 1349–60.

35. Andresen EM, Bowley N, Rothenberg BM *et al.* Test–retest performance of a mailed version of the Medical Outcomes Study 36-item short-form health survey among older adults. Med Care 1996; 34: 1165–70.

36. Weinberger M, Oddone EZ, Samsa GP *et al.* Are health-related quality of life measures affected by the mode of administration? J Clin Epidemiol 1996; 49: 135–40.

37. Gerety MB, Cornell JE, Mulrow CD *et al.* The Sickness Impact Profile for nursing homes (SIP-NH). J Gerontol 1994; 49: M2–8.

38. Hays RD, Siu AL, Keeler E *et al.* Long-term care residents' preferences for health states on the Quality of Well-Being Scale. Med Decis Making 1996; 16: 254–61.