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

ELENA M. ANDRESEN, GWENDELL W. GRAVITT, MARIE E. AYDELotte1, CAROL A. PODGORSKI2

Department of Community Health, St Louis University School of Public Health, 3663 Lindell Boulevard, St Louis, MO 63108, USA
Departments of 1Medicine and 2Psychiatry, 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.
and ADLs. 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 and SPSS statistical software. Intraclass correlations and lower 95 percent confidence limits (random effects, one-way model) were calculated for SF-36 scales. 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.

**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. 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>
<thead>
<tr>
<th>Table 1. Demographic characteristics of the 97 study participants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Value</strong></td>
</tr>
<tr>
<td>Mean age, years (SD)</td>
</tr>
<tr>
<td>Women</td>
</tr>
<tr>
<td>Caucasian</td>
</tr>
<tr>
<td>Education (high school or greater)</td>
</tr>
<tr>
<td>Skilled nursing level of care</td>
</tr>
<tr>
<td>Intermediate level of care</td>
</tr>
<tr>
<td>Mean no. of months in nursing home (SD)</td>
</tr>
<tr>
<td>Receiving Medicaid financial benefits</td>
</tr>
<tr>
<td>Mean no. of medications (SD)</td>
</tr>
</tbody>
</table>
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 Life Index are more appropriate for use in this setting.

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

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

Floor/ceiling effects

<table>
<thead>
<tr>
<th>Scale</th>
<th>No. scored</th>
<th>Mean score</th>
<th>% at 0</th>
<th>% at 100</th>
<th>ICC (95% lower CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical functioning</td>
<td>95</td>
<td>32.4</td>
<td>26.8</td>
<td>1.1</td>
<td>0.79 (0.74)</td>
</tr>
<tr>
<td>Social functioning</td>
<td>97</td>
<td>72.6</td>
<td>3.1</td>
<td>36.1</td>
<td>0.52 (0.43)</td>
</tr>
<tr>
<td>Role—physical</td>
<td>95</td>
<td>42.5</td>
<td>29.5</td>
<td>18.9</td>
<td>0.76 (0.70)</td>
</tr>
<tr>
<td>Role—emotional</td>
<td>95</td>
<td>69.5</td>
<td>11.6</td>
<td>49.5</td>
<td>0.55 (0.46)</td>
</tr>
<tr>
<td>Mental health</td>
<td>96</td>
<td>64.7</td>
<td>1.0</td>
<td>2.1</td>
<td>0.80 (0.76)</td>
</tr>
<tr>
<td>Vitality</td>
<td>96</td>
<td>49.9</td>
<td>1.0</td>
<td>2.1</td>
<td>0.74 (0.76)</td>
</tr>
<tr>
<td>Bodily pain</td>
<td>97</td>
<td>60.9</td>
<td>2.1</td>
<td>21.6</td>
<td>0.65 (0.58)</td>
</tr>
<tr>
<td>General health</td>
<td>97</td>
<td>52.6</td>
<td>0</td>
<td>0</td>
<td>0.65 (0.58)</td>
</tr>
<tr>
<td>Physical health—summary</td>
<td>90</td>
<td>33.3</td>
<td>0</td>
<td>0</td>
<td>0.82 (0.77)</td>
</tr>
<tr>
<td>Mental health—summary</td>
<td>90</td>
<td>50.5</td>
<td>0</td>
<td>0</td>
<td>0.79 (0.74)</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Scale</th>
<th>Mean (SD)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activities of daily living</td>
<td>11.5 (9.6)</td>
<td>0–32</td>
</tr>
<tr>
<td>Mini-Mental State Examination</td>
<td>23.1 (4.0)</td>
<td>17–30</td>
</tr>
<tr>
<td>Geriatric Depression Scale</td>
<td>10.4 (6.8)</td>
<td>0–28</td>
</tr>
</tbody>
</table>

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

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

SF-36, Short-Form 36; ADL, activities of daily living; MMSE, Mini-Mental State Examination; GDS, Geriatric Depression Scale.

\*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 Life Index are more appropriate for use in this setting.
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
This project was funded in part by the Rochester Area Pepper Center. The assistance of T. Franklin Williams, Paul R. Katz and George Abraham is acknowledged gratefully. We also would like to thank the Research Information Management Center at the University of Rochester for help with data entry and analysis.

Key points
- Only one in five nursing home residents met minimal participation criteria, suggesting that the SF-36 might be 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


Received 11 September 1998; accepted in revised form 28 January 1999