Is the SF-36 suitable for assessing health status of older stroke patients?

PAUL G. O'MAHONY1,2, HELEN RODGERS1,2, RICHARD G. THOMSON2, RUTH DOBSON1,2, OLIVER F. W. JAMES1

Departments of 1Medicine (Geriatric Medicine) and 2Epidemiology and Public Health, University of Newcastle, Newcastle upon Tyne, UK

Address correspondence to: P. O'Mahony, Elderly Annexe, Newcastle General Hospital, Westgate Road, Newcastle upon Tyne, NE4 6BE, UK. Fax: (+44) 191 219 5049

Abstract

Introduction: the Medical Outcomes Study short form 36 health survey (SF-36) is being increasingly used and recommended as a suitable measure of subjective health status. However, it is unlikely that any measure will be appropriate for all groups. We wished to determine the suitability of the SF-36 for assessing quality of life in older stroke patients.

Methods: a screening questionnaire was used to identify prevalent cases of stroke from a random sample of 2000 subjects aged 45 years and over. The SF-36 was included as part of a self-completion questionnaire posted to each stroke patient. Data quality indicators were analysed.

Results: we identified 104 cases of stroke and the response rate for the SF-36 questionnaires sent was 83%. Completion rates for individual items ranged from 66 to 96%. All items in the role physical and role emotional scales had completion rates <75%. The percentage of subjects for whom an individual scale score could be computed ranged from 67 to 96%, being lowest for the role physical and role emotional scales. Floor effects were high (>15%) for these two scales and for the social functioning and physical functioning scales. Ceiling effects were substantial (>15%) for the two role effect scales and for social functioning and bodily pain.

Conclusions: this study has shown high response rates from older stroke patients to a postal questionnaire incorporating the SF-36. The poor completion rates and consequent inability to compute scores for a large proportion of responders in certain scales raises concerns about the perceived relevance of these sections. Results for the response effects suggest that, on its own, the instrument is not suitable for assessing outcome. When data quality indicators were examined, it appears that postal administration of the SF-36 is not appropriate for assessing quality of life of older stroke patients.

Keywords: elderly patients, quality of life, SF-36, stroke

Introduction

Although quality of life measurement has been considered one of the most important ways to measure outcome after stroke, it has been the subject of little formal examination [1, 2]. The method of collecting data requires some trade-off between costs on the one hand and response rates, non-response bias and data quality on the other. In choosing a survey methodology, these often-neglected factors must be taken into account.

With the growing interest in health status measurement, a generic instrument, the short form 36 (SF-36), was developed for use in the Medical Outcomes Study in the USA [3]. The SF-36 is a short, 36-item questionnaire which measures eight dimensions of health status, covering three aspects of health of prime concern to individuals: functional status, well-being and overall evaluation of health. In total, 35 of the items contribute to these states and a further unscaled single item asks about change in health status over the previous year. For the eight dimensions, scores are coded, summed and transformed onto a scale from 0 (worst possible health status) to 100 (best possible health status). The survey has been anglicized and validated for use in a British population [4-7].

The SF-36 is increasingly being advocated as an appropriate subjective health status measure for use in a wide variety of settings. While some recent studies have assessed the performance of the SF-36 among elderly subjects, only a limited number have commented on features of data quality other than response rates.
Higher rates of missing data among older subjects suggest that the SF-36 may not be appropriate for postal surveys with this group [4, 8, 11]. Its use with different disease groups requires further research as it may be most useful for assessing health status in populations with less severe morbidity [4, 14]. Although the SF-36 has been used for assessing quality of life in young (15–45-year-old) stroke patients, there have been no detailed reports of its use in older subjects [15–17].

We wished, therefore, to answer the questions concerning the suitability of use of the SF-36, particularly in respect of response rate, data quality and relevance, before this instrument becomes adopted uncritically for assessing quality of life in this group of patients.

Methods

A random sample of 2000 subjects aged 45 years and over was chosen from a local Family Health Services Authority register. A valid screening questionnaire was used to identify cases of stroke in the community [18]. The SF-36 was included as part of a detailed self-completion questionnaire posted to each stroke patient. This questionnaire also incorporated other assessment scales and formed part of a wider study of the prevalence of stroke and associated disability and handicap and the consequent population health needs. Two reminders were sent at 2-week intervals. It was suggested that subjects could receive help to complete the questionnaire from a friend or relative if necessary—this has previously been reported [10].

The following data quality indicators were evaluated: (i) response rate for the questionnaire; (ii) completion rates for the 36 individual items (for each item, the number of subjects who completed it divided by the total number of respondents); (iii) proportion of respondents for whom a scale score could be computed in each of the nine parameters; and (iv) features of score distribution/response effects in the form of floor (the proportion of subjects getting the lowest possible score) and ceiling (the proportion of subjects getting the highest possible score) effects for the eight multi-item dimensions.

Results

The screening questionnaire identified 104 cases of stroke among the responders. Of these, 12 (12%) were not sent the more detailed questionnaire (incorporating the SF-36) because they were resident in nursing homes or hospitals at the time of the study or had died in the period between identification and posting of the questionnaire.

Of the remaining 92, five (5%) refused to complete the questionnaire, 10 (11%) subjects did not respond and a further four (4%) had died. The corrected response rate was therefore 83%, ranging from 67% among the 45–54-year-olds to 87% among those aged over 75 years.

Completion rates for individual items of the nine parameters ranged from 66 to 96% (Table 1). All items in the role physical (10 items) and role emotional (three items) scales had completion rates less than 75%. The table also shows the proportion of the 73 subjects for whom scale scores could be computed. It was not possible to obtain scores for approximately one-third of subjects in the role physical and role emotional scales. Floor effects were highest for the role physical and role emotional scales.

Table 1. Range of completion rates, percentage of subjects for whom a score could be computed and floor and ceiling effects for each scale

<table>
<thead>
<tr>
<th>Parameter</th>
<th>No. of items</th>
<th>Completion rate (%)</th>
<th>With computed scorea</th>
<th>At extremesb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical functioning</td>
<td>10</td>
<td>85–93</td>
<td>90 (66)</td>
<td>18 (12)</td>
</tr>
<tr>
<td>Role physical</td>
<td>4</td>
<td>66–73</td>
<td>68 (50)</td>
<td>54 (27)</td>
</tr>
<tr>
<td>Bodily pain</td>
<td>2</td>
<td>88–90</td>
<td>92 (67)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>General health</td>
<td>5</td>
<td>74–96</td>
<td>81 (59)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Vitality</td>
<td>4</td>
<td>86–92</td>
<td>93 (68)</td>
<td>10 (7)</td>
</tr>
<tr>
<td>Social functioning</td>
<td>2</td>
<td>85–95</td>
<td>97 (71)</td>
<td>17 (12)</td>
</tr>
<tr>
<td>Role emotional</td>
<td>3</td>
<td>69–73</td>
<td>67 (49)</td>
<td>35 (17)</td>
</tr>
<tr>
<td>Mental health</td>
<td>5</td>
<td>90–93</td>
<td>95 (69)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Health transition</td>
<td>1</td>
<td>96</td>
<td>96 (70)</td>
<td>-</td>
</tr>
</tbody>
</table>

a% of all questionnaires returned (n = 73).

bFloor effect, % of subjects (for whom a score could be computed) who achieved the lowest possible score (i.e. 0); ceiling effect, % of subjects (for whom a score could be computed) who achieved the highest possible score (i.e. 100).
physical and role emotional scales, while ceiling effects were most common for the role emotional, bodily pain and social functioning scales (Table 1).

Discussion

The SF-36 is rapidly becoming the generic health status measure of choice. Before applying such an instrument in disease-specific settings, it must be shown to be applicable for each of the diseases being studied. We have shown that the SF-36 can be incorporated in a self-completion questionnaire for postal administration to older stroke patients. The very high response rate achieved was higher than in other studies using a postal version with similar age groups [10, 19].

Few studies quote any information on data completeness [4, 8, 11]. Despite high response rates, data completeness may compromise the validity of measures because of non-response bias and because the resultant sample loss affects the generalizability of results. Missing data were common in our study and considerably higher than in the Medical Outcomes Study, where missing value rates for the 36 items in a general population were low, averaging just 3.9% (range 1.1-5.9%) [8]. A number of factors appear to influence the level of missing data, including age, disability and method of administration [4, 8, 9, 11].

Our completion rates might have been better if our postal questionnaire had not been so long. However, the Medical Outcomes Study itself embedded the SF-36 in a 245-item questionnaire and there is little evidence that questionnaire length affects response or completion rates [8, 14].

Although completion rates may not reach 100% in any study, the SF-36 provides a method for minimizing problems of missing data by having multi-item dimensions and methods for computing scores on the basis of available information. Despite this, our study demonstrated poor results for computation of scores in the two role functioning scales and the general health scale. For the other scales, the results are as good as those obtained in the Medical Outcomes Study, where it was possible to compute scores for over 90% of elderly subjects [8]. The percentage of respondents with computable scores has previously been shown to be significantly lower for postal respondents than telephone respondents [9].

To determine whether the range of health states defined by the SF-36 scales was appropriate for older stroke patients, we calculated the percentage of the sample achieving the lowest (floor effect) and highest (ceiling effect) possible scores. Floor effects in the role functioning scales have been found in studies of general and ill patient populations [8, 12]. We found more substantial floor effects for these two scales but have also demonstrated modest floor effects for the physical functioning, social functioning and vitality scales. Ceiling effects have previously been shown to be high for the two role functioning and the social functioning scales, and we have confirmed this among stroke patients and shown modest ceiling effects for the bodily pain and mental health scales [8].

Widely used health status measures must yield high-quality data across diverse populations, preferably with different methods of administration. Data completeness may be less problematic among elderly people with telephone or face-to-face administration but with a trade-off of substantially higher costs compared with postal surveys [20]. Data quality indicators, such as the extent of item non-response and missing score scales, are important because they reflect the respondents' understanding and acceptance of the survey. Features of score distributions such as floor and ceiling effects are infrequently evaluated but can vary across populations to the point of affecting the usefulness of the scale in specific applications.

The use of a standardized health survey measure across general and patient populations has many obvious advantages, including the comparison between groups differing in diagnosis, treatment or disease severity and the monitoring of change over time. We suggest that the SF-36 is not suitable for use as a postally administered quality of life instrument among older stroke patients. Although it may be suitable by interview administration, this would need further investigation. This study provides additional evidence that the SF-36 may not be appropriate for assessing the health status of certain groups of elderly patients [21].

Acknowledgements

We wish to acknowledge the financial support of the former Northern Regional Health Authority Research and Development Directorate. We are grateful to the Newcastle Family Health Services Authority for providing the population sample and especially to all the patients who participated in the study.

Key points

• The SF-36 is being widely used as a subjective health status measure.
• High response rates can be achieved for the postal version of the SF-36 when administered to older stroke patients.
• High ceiling and floor effects suggest that the SF-36 is inadequate on its own for assessing outcome.
• Results for response effects suggest that the SF-36 is not suitable for use among older stroke patients when administered as a postal questionnaire.

References


Received 29 January 1997