Age and Ageing 2001; **30**: 337–343

Using the SF-36 with older adults: a cross-sectional community-based survey

STEPHEN J. WALTERS, JAMES F. MUNRO[1], JOHN E. BRAZIER

Sheffield Health Economics Group and [1]Medical Care Research Unit, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield S1 4DA, UK

Address correspondence to: S. J. Walters. Fax: (+44) 114 272 4095. Email: s.j.walters@sheffield.ac.uk

Abstract

Objectives: to assess the practicality and validity of using the 36-item short-form health survey (SF-36) in a community-dwelling population over 65, and to obtain population scores in this age group.
Design: postal survey, using a questionnaire booklet containing the SF-36 and other health-related items, of all those aged 65 or over registered with 12 general practices in Sheffield. Non-respondents received up to two reminders at 3-weekly intervals.
Sample: 9897 subjects, aged 65–104 years.
Main outcome measures: scores for the eight dimensions of the SF-36 and a modified version of the physical functioning dimension.
Results: the SF-36 achieved a response rate of 82% ($n = 8117$) and dimension completion rates of 86.4–97.7%. Internal consistency measured by Cronbach's α exceeded 0.80 for all dimensions except social functioning. These results compare favourably with postal surveys of younger adults. We calculated scores for older adults by age and sex. Comparison with data from younger people showed how physical health declines steeply with age, in marked contrast to mental health.
Conclusions: the SF-36 is a practical and valid instrument for use in postal surveys of older people living at home. The population scores provided here may facilitate its use in future surveys of older adults.

Keywords: *postal survey, health-related quality of life, community study, SF-36*

Introduction

Measures of 'health-related quality of life' are commonly used in determining the effects of medical and community-based health interventions, comparing the health of different populations and assessing health needs. However, although people over 65 makeup a growing proportion of the population and are major users of health care, they have frequently been neglected in both the development and use of such measures. This may be because of uncertainty over the practicality and validity of using quality of life measures with older adults, and particular concerns over low response rates, inability to complete postal surveys independently and the inclusion of questions which may be irrelevant to retired or severely impaired respondents [1–4].

Given the increasing importance of older people as service users and potential beneficiaries of public health interventions, quality of life measures should be adequately assessed and, if necessary, adapted for use in this population. The aims of this study were therefore to (i) establish whether the 36-item short-form health survey (SF-36), a popular self-completed generic health status instrument, could be successfully used in a large postal survey of older adults and (ii) provide population scores for a representative sample of community-dwelling adults aged over 65 years.

Methods

Instruments used

Although the SF-36 is American in origin [5], there is a validated version adapted to British use [1]. It asks 36 questions about health over the past 4 weeks, measuring eight dimensions of health status: physical functioning,

role limitation because of physical health, social functioning, vitality or energy, bodily pain, mental health, role limitation because of emotional problems and general health. The responses to questions within each dimension are summed and transformed to generate dimension scores ranging from 0 ('poor health') to 100 ('good health').

Because our study population was retired and potentially in poor health, we made three minor changes to the standard SF-36 layout. First, the usual order of items in question 3 (typical daily activities) was reversed, so that instead of moving from most to least vigorous activity, the items began with the least vigorous. The aim was to overcome the problem of older people seeing a very vigorous activity (running) at the top of the list and deciding that the question, or the entire survey, was irrelevant to them. Secondly, three additional items were added to this question because some respondents might not be capable of even the most limited activity in the standard list ('bathing and dressing yourself'). The three items added were 'feeding yourself', 'getting up from a chair' and 'walking in your home'. To allow our results to be compared with those using the usual questions, we report scores for both the standard and 'extended' physical dimensions. Thirdly, we altered the wording of question 4 to remove the phrase 'work or other', which would not be widely applicable in a retired population.

Data collection

We randomly selected 12 general practices in Sheffield from those with two to five partners, in the context of a study of physical activity in older adults. Of 13 practices approached, 12 agreed to participate. All individuals aged 65 or over registered with these practices were sent a 10-page postal questionnaire containing the SF-36, along with a physical activity questionnaire and a small number of socio-demographic questions. Up to two reminders were sent, at 3-weekly intervals.

Analysis and statistical methods

The usability of the SF-36 in older adults was assessed by examining response and completion rates by age group. Dimension scores were calculated only if all items contributing to the dimension had been completed. The internal consistency—the extent to which items correlate with items within their own dimension—was examined in terms of Cronbach's α and item-to-own-dimension correlation. Cronbach's α is a summary measure, generating a value between 0 and 1 for each dimension [6]: values >0.8 are usually regarded as acceptable [7].

We examined evaluative properties in terms of potential sensitivity to change or 'responsiveness'. To some degree, the ability to respond to change can be assessed in terms of the proportion of respondents at the floor (worst score) or the ceiling (best score) of each scale. A large proportion at either extreme will clearly limit the ability to register deterioration or improvement, respectively.

The ability to discriminate between different levels of ill health is an essential requirement for any health-related quality of life instrument and we examined this in terms of construct validity, where score distributions are compared between groups with known or expected health differences. For example, one might expect physical functioning scores to be lower—reflecting poorer perceived health—for women, for older people and for those who have recently consulted a doctor. We tested differences between these groups with a *t*-test or one-way analysis of variance, as appropriate.

The value of the three additional items in the physical dimension was examined in terms of completion, correlation with the physical score and effect on the floor and ceiling of the physical dimension.

Finally, we calculated data in the form of the mean, median and standard deviation for each age–sex group for all eight dimensions of the SF-36, and plotted age-specific mean scores for each dimension, alongside scores from younger respondents to a large postal survey conducted in the same city [1].

Results

Response and completion rates

Response

We sent surveys to 9887 individuals, of whom 8117 [82%; 95% confidence interval (CI) 81.3–82.8] returned a completed survey. The characteristics of respondents were broadly similar to those of people over 65 in the UK as a whole [8] (Table 1). The response rate was $>80\%$ for all age groups except those aged 85 and over,

Table 1. Characteristics of sample

	% of sample	
	Sheffield[a]	UK[b]
Age group, years		
65–69	26.3	30
70–74	28.4	27
75–79	22.5	22
80–84	14.1	13
85+	8.7	8
Female	57.6	56
Living alone	38.7	39
Current smoker	17.1	18[b]
Saw doctor during past fortnight		
At surgery	22.9	21
At home	6.4	5

[a]Mean age 74.6 years (SD 6.1); sample size, 7961–8117.
[b]Aged 60 or over; taken from Office for National Statistics, 1998 [8].

for whom the rate was 69% (95% CI 66.2–71.9). Non-responders tended to be slightly older than responders (76.3 *versus* 74.5 years, difference 1.8; 95% CI 1.4–2.2; $P = 0.001$). Response rates did not vary by sex.

Item and dimension completion

Item completion was high, reaching >93% for all questions. Dimension completion (i.e. completion of all items in a dimension, allowing it to be calculated without interpolation) ranged from 86.4 to 97.7%. Dimension completion was lowest among those aged 85 or over, ranging from 83.3 to 94.4%.

For all dimensions, non-completion was associated with increasing age ($P < 0.001$). Men were more likely ($P < 0.001$) than women to complete all dimensions except pain ($P = 0.96$). Completion was also lower ($P < 0.003$) for respondents living alone. All eight dimensions were calculable for 5841 respondents (72%; 95% CI 71.0–72.9).

Internal consistency

Cronbach's α exceeded 0.8 for all dimensions except social functioning (0.79), which is similar to results reported in studies of younger populations [1, 2]. The result for social functioning may partly reflect the low number of items (two) in that dimension.

Extra physical function items

Completion rates for the three additional items were comparable to those for the other items in the physical dimension and had little impact on the overall dimension completion rate (87.7% for the original dimension, 86.4% for the extended dimension). The added items correlated significantly with the original dimension: 'feeding yourself' showed the weakest correlation ($r = 0.36$), but 'getting up from a chair' and 'walking in your home' had correlations of 0.68 and 0.67 respectively. The item-to-own-dimension correlation exceeded the correlation with all other dimensions, supporting the logic of adding these items to the original scale. Cronbach's α for the extended physical dimension was 0.95.

Population scores

Tables 2 and 3 show the mean and standard deviation of each SF-36 dimension by age and sex. Overall, women reported poorer health than men on all dimensions ($P < 0.001$), even after adjusting for age. People living alone reported worse health than those not living alone ($P < 0.001$, except general health: $P = 0.02$). Respondents consulting their doctor during the past 2 weeks had poorer perceived health on all dimensions of the SF-36 ($P < 0.001$) than those who had not. For each dimension, scores decreased with increasing age. These data support the construct validity of the SF-36 in older adults.

For each dimension and age group, Table 4 gives median scores and the proportion of respondents at the floor and ceiling. While 6.5% of respondents (95% CI 6.0–7.1) were at the floor of the original physical dimension, extending the dimension with three additional items left only 1.2% (95% CI 1.0–1.5) of respondents at the floor.

Table 2. Mean scores and sample sizes for SF-36 dimensions, by age for men

	Mean (and SD) by age, years				
SF-36 dimension	65–69	70–74	75–79	80–84	85+
Physical function					
Standard	67.2 (29.8) ($n = 925$)	61.9 (30.3) ($n = 932$)	58.4 (30.4) ($n = 694$)	48.4 (30.5) ($n = 377$)	39.2 (28.7) ($n = 158$)
Extended	72.5 (26.2) ($n = 917$)	68.4 (26.8) ($n = 916$)	64.5 (27.5) ($n = 685$)	55.6 (28.1) ($n = 372$)	47.6 (26.4) ($n = 155$)
Role physical	56.6 (42.6) ($n = 975$)	47.6 (43.3) ($n = 980$)	43.5 (42.0) ($n = 739$)	30.2 (38.4) ($n = 386$)	29.8 (37.1) ($n = 168$)
Bodily pain	68.2 (27.7) ($n = 1001$)	65.6 (28.0) ($n = 1012$)	64.0 (29.1) ($n = 767$)	59.6 (28.8) ($n = 414$)	63.9 (26.4) ($n = 172$)
General health	57.4 (24.3) ($n = 941$)	56.0 (24.2) ($n = 947$)	56.2 (23.1) ($n = 706$)	52.1 (24.0) ($n = 374$)	51.5 (22.1) ($n = 160$)
Vitality	58.6 (23.3) ($n = 995$)	55.8 (22.4) ($n = 994$)	53.8 (22.3) ($n = 747$)	48.5 (24.1) ($n = 412$)	46.5 (21.1) ($n = 170$)
Social function	78.6 (28.9) ($n = 997$)	75.1 (29.9) ($n = 1004$)	70.4 (31.5) ($n = 759$)	63.0 (34.3) ($n = 410$)	59.7 (32.5) ($n = 173$)
Role emotional	67.3 (42.0) ($n = 961$)	61.8 (43.7) ($n = 972$)	55.4 (44.0) ($n = 723$)	44.5 (44.6) ($n = 377$)	43.8 (44.5) ($n = 163$)
Mental health	74.9 (19.4) ($n = 989$)	74.3 (19.1) ($n = 995$)	73.8 (18.5) ($n = 746$)	71.8 (19.4) ($n = 404$)	71.2 (19.0) ($n = 165$)

The dimensions of the SF-36 are scored on a scale from 0 (worst possible health) to 100 (best possible health).

Table 3. Mean scores and sample sizes for SF-36 dimensions by age for women

	Mean (and SD) by age, years				
SF-36 dimension	65–69	70–74	75–79	80–84	85+
Physical function					
Standard	61.9 (29.1) ($n = 1000$)	55.9 (29.4) ($n = 1112$)	45.9 (28.9) ($n = 873$)	36.6 (29.3) ($n = 601$)	23.8 (25.0) ($n = 444$)
Extended	68.1 (25.7) ($n = 991$)	62.9 (26.1) ($n = 1092$)	53.8 (26.3) ($n = 860$)	44.6 (27.0) ($n = 588$)	32.2 (23.7) ($n = 435$)
Role physical	50.7 (43.4) ($n = 1041$)	42.3 (42.6) ($n = 1179$)	34.3 (40.5) ($n = 940$)	28.5 (38.5) ($n = 630$)	21.0 (34.0) ($n = 465$)
Bodily pain	63.4 (28.1) ($n = 1096$)	60.1 (28.3) ($n = 1249$)	57.7 (28.2) ($n = 1028$)	52.5 (29.5) ($n = 693$)	48.8 (29.7) ($n = 502$)
General health	56.9 (23.8) ($n = 997$)	55.8 (23.3) ($n = 1127$)	52.6 (22.4) ($n = 885$)	47.7 (22.4) ($n = 576$)	44.4 (21.4) ($n = 433$)
Vitality	53.2 (23.2) ($n = 1068$)	51.9 (22.6) ($n = 1209$)	47.6 (21.5) ($n = 977$)	42.1 (21.7) ($n = 663$)	37.0 (21.7) ($n = 472$)
Social function	75.5 (29.0) ($n = 1083$)	72.7 (30.2) ($n = 1228$)	66.5 (31.9) ($n = 998$)	56.3 (33.2) ($n = 669$)	44.5 (33.1) ($n = 495$)
Role emotional	61.0 (43.4) ($n = 1033$)	55.6 (44.2) ($n = 1171$)	46.8 (45.0) ($n = 916$)	41.4 (43.6) ($n = 611$)	39.9 (45.4) ($n = 444$)
Mental health	67.8 (21.1) ($n = 1059$)	68.1 (20.2) ($n = 1224$)	66.8 (19.3) ($n = 965$)	64.7 (19.9) ($n = 647$)	63.9 (21.6) ($n = 478$)

The dimensions of the SF-36 are scored on a 0 (worst possible health) to 100 (best possible health) scale.

Table 4. Medians and floor and ceiling effects for SF-36 dimensions, by age[a]

	Median (and % on floor, % at ceiling) by age, years				
SF-36 dimension	65–69	70–74	75–79	80–84	85+
Physical function					
Standard	75.0 (2.9, 5.9) ($n = 1925$)	65.0 (4.3, 3.5) ($n = 2044$)	55.0 (5.9, 2.0) ($n = 1567$)	35.0 (11.1, 0.8) ($n = 978$)	20.0 (19.6, 0.5) ($n = 602$)
Extended	80.8 (0.5, 5.8) ($n = 1908$)	73.1 (0.7, 3.4) ($n = 2008$)	61.5 (0.9, 1.9) ($n = 1545$)	50.0 (2.1, 0.5) ($n = 960$)	30.8 (4.1, 0.3) ($n = 590$)
Role physical	50.0 (30.9, 38.6) ($n = 2016$)	25.0 (38.7, 30.8) ($n = 2159$)	25.0 (44.6, 23.9) ($n = 1679$)	0.0 (54.9, 16.4) ($n = 1016$)	0.0 (61.0, 11.7) ($n = 633$)
Bodily pain	66.7 (2.2, 21.1) ($n = 2097$)	66.7 (2.7, 18.5) ($n = 2261$)	55.6 (2.3, 17.4) ($n = 1795$)	55.6 (4.4, 14.1) ($n = 1107$)	44.4 (3.6, 14.8) ($n = 674$)
General health	60.0 (0.8, 1.5) ($n = 1938$)	57.0 (0.9, 1.7) ($n = 2074$)	55.0 (0.5, 0.7) ($n = 1591$)	50.0 (1.7, 0.7) ($n = 950$)	45.0 (1.2, 10.8) ($n = 593$)
Vitality	55.0 (1.9, 1.6) ($n = 2063$)	55.0 (2.6, 1.4) ($n = 2203$)	50.0 (2.1, 0.5) ($n = 1724$)	45.0 (4.5, 0.2) ($n = 1075$)	40.0 (3.9, 0.2) ($n = 642$)
Social function	88.9 (2.5, 46.7) ($n = 2080$)	88.9 (3.5, 40.8) ($n = 2232$)	77.8 (4.0, 32.8) ($n = 1757$)	66.7 (6.6, 24.6) ($n = 1079$)	44.4 (11.7, 15.0) ($n = 668$)
Role emotional	100.0 (25.6, 53.2) ($n = 1994$)	66.7 (30.5, 47.5) ($n = 2143$)	33.3 (37.5, 40.2) ($n = 1639$)	33.3 (44.7, 32.2) ($n = 988$)	0.0 (50.4, 32.5) ($n = 607$)
Mental health	76.0 (0.1, 5.0) ($n = 2048$)	72.0 (0.2, 5.2) ($n = 2219$)	72.0 (0.2, 3.9) ($n = 1711$)	68.0 (0.1, 3.9) ($n = 1051$)	68.0 (0.5, 3.7) ($n = 643$)

[a]Respondents at the 'floor' are those who scored 0, and at the 'ceiling' those who scored 100.

Quality of life at different ages

The mean score for each 5-year age band is shown for each dimension in Figure 1, alongside scores for adults aged 16–64 years, taken from a previous postal survey of younger adults in Sheffield [1]. Health-related quality of life declines steadily with age, but the decline is much sharper for the physical and physical role limitation dimensions than for others. The mental health score remains almost constant, even in very old age.

One notable feature, which is apparent in all dimensions, is the higher than expected score in the 60–64-year age group. When sex-specific scores are plotted, this 'blip' is apparent for women but not for men.

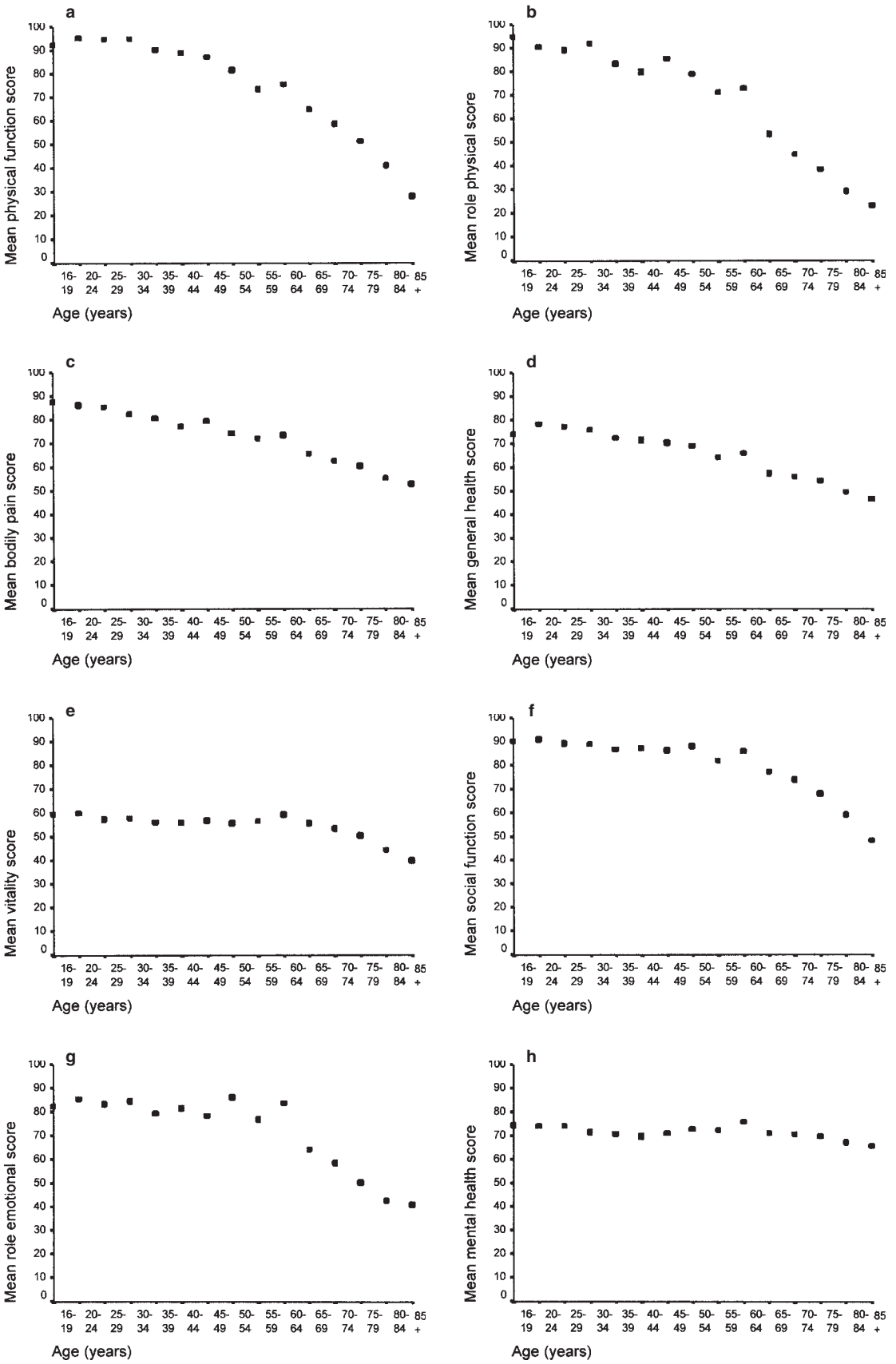

Figure 1. Mean age profiles for SF-36 dimensions: **a** physical function, **b** role physical, **c** bodily pain, **d** general health, **e** vitality, **f** social function, **g** role emotional and **h** mental health. Data for ages 16–64 are from Brazier *et al.* 1992 [1].

Discussion

This postal survey, using the SF-36 in a large representative population of community-dwelling older adults, achieved a response rate of 82%, which is as good as or better than previous experience in younger age groups (83% in Sheffield [1], 72% in Oxford [9] and 75% in Aberdeen [10]). Even among adults aged 85 or over, the response rate reached 69%. Dimension completion (88–98%) also compared favourably with that of other surveys [1].

In this study we avoided using substitution procedures for missing data, but if this were done the number of respondents for whom all eight dimensions are calculable would increase from 5841 (72%) to 7076 (87%). This has little impact on the normative scores, but could be useful if it is important to maximize the number of competed questionnaires.

These results demonstrate the practicality of using the SF-36 as a self-completed instrument in community-based surveys of older people and contradict the findings of a previous smaller postal survey [4]. The lower response rates reported from other studies have tended to relate to older people in outpatient or hospital settings with serious illness, rather than to an unselected community-dwelling population [2, 11, 12]. For some patient groups, particularly those with significant visual or cognitive disabilities, self-completing the SF-36 would be impractical [12], and interview administration should be considered [13, 14]. Administering the SF-36 by interview would substantially increase the expense of using it and limit comparability with data from younger populations.

Our findings and those of others have confirmed the internal consistency of items with their own dimensions in older populations [13]. The score distributions by age, gender and recent use of general practitioner services support the construct validity of this instrument for the over-65s. However, because of multiple testing, some caution should be applied in the interpretation of the *P* values we have reported.

Previous studies have found that items in the role functioning dimensions are problematic for older people in terms of lower completion and high proportions on the floor and ceiling of the dimension scores [1, 2], and our findings confirm this. The reference to 'work' in the wording of these questions may be the cause of the lower completion rates [2], and the floor and ceiling effects result from the use of dichotomous (yes or no) response choices. Version 2 of the SF-36 partly overcomes this problem by increasing the number of response choices to five [15].

We would advise potential users of the SF-36 among older adults to use the ordering of physical functioning items we used in this study and to amend the wording of questions 4 and 5 to remove the phrase 'work or other'. Indeed, given the diversity of all populations, we believe there is a strong case for these changes to be made irrespective of the age of respondents. The addition of three extra questions to the physical functioning dimension was successful in reducing the 'floor' of this dimension and did not appear to affect the validity or completion of the rest of the questionnaire. Adding these items still allows the standard physical function score to be calculated and comparisons with younger age groups to be made, while improving the relevance of this dimension to older people.

Combining our data with those from a previous large survey of younger adults living in the same city has allowed us to plot mean scores for each dimension across the entire adult age range. Although these data are cross-sectional rather than longitudinal, and may therefore be subject to cohort as well as age effects, they suggest that health-related quality of life may fall much more sharply with age for physical than for mental functioning (which seems hardly to fall at all). It is possible that this difference between dimensions is the result of a differential response bias, if those with mental health problems are less likely to respond than those with physical health problems, although the satisfactory response rates in all age groups suggests that the impact of response bias is unlikely to be great. Other dimensions decline at rates between these.

For each dimension, 60- to 64-year-olds show a slightly higher score than would be expected from the overall decline, which has not previously been described. Although possibly artefactual, this 'blip' is also observable in the published Oxford norms [9] and in a national survey of health-related quality of life using a different instrument, the EQ-5D [16]. It occurs before the datasets join and seems to occur in women rather than men. Taken together, these observations suggest that health-related quality of life of women in this age group may indeed be higher than that of those 5 years older or younger, although whether this is an age or cohort effect could only be determined through a longitudinal study.

Overall, the findings from this and other studies of the SF-36 in older adults suggest that age *per se* should not be a bar to its use. Good response rates can be achieved in community populations with only minor modifications, without compromising validity or losing comparability with scores from younger age groups. Those undertaking population surveys who do not wish to exclude people over 65 should feel confident in using the SF-36.

Key points

- The SF-36 is a practical and valid instrument to use in postal surveys of older people living in the community.
- Good response and completion rates can be achieved with minor modifications to the layout of the SF-36 questionnaire.

- The provided SF-36 population scores by age and sex for older adults should facilitate the use of this instrument in future surveys of older adults.
- Comparison with data from younger people shows how physical health declines steeply with age, in marked contrast to mental health.

Acknowledgements

We are grateful to the 12 practices who assisted us in this study, and to Gwyneth Askham for her clerical support.

References

1. Brazier JE, Harper R, Jones NM *et al*. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. Br Med J 1992; 305: 160–4.

2. Hayes V, Morris J, Wolfe C, Morgan M. The SF-36 health survey questionnaire: is it suitable for use with older adults? Age Ageing 1995; 24: 120–5.

3. Brazier JE, Walters SJ, Nicholl JP, Kohler B. Using the SF-36 and Euroqol on an elderly population. Qual Life Res 1996; 5: 195–204.

4. Mallinson S. The Short-Form 36 and older people: some problems encountered when using postal administration. J Epidemiol Community Health 1998; 52: 324–8.

5. Ware JEJ, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992; 30: 473–83.

6. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951; 16: 297–334.

7. Streiner DL, Norman GR. Health Measurement Scales: a practical guide to their development and use. Oxford: Oxford University, Press, 1989.

8. Office for National Statistics. Living in Britain: results from the 1996 General Household Survey. London: The Stationery Office, 1998.

9. Jenkinson C, Coulter A, Wright L. Short form 36 (SF-36) health survey questionnaire: normative data for adults of working age. Br Med J 1993; 306: 1437–40.

10. Garratt AM, Ruta DA, Abdalla MI *et al*. The SF-36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? Br Med J 1993; 306: 1440–4.

11. Parker SG, Peet SM, Jagger C *et al*. Measuring health status in older patients. The SF-36 in practice. Age Ageing 1998; 27: 13–18.

12. Hill S, Harries U, Popay J. Is the short form 36 (SF-36) suitable for routine health outcomes assessment in health care for older people? Evidence from preliminary work in community based health services in England. J Epidemiol Community Health 1996; 50: 94–8.

13. Lyons RA, Perry HM, Littlepage BN. Evidence for the validity of the short-form 36 questionnaire (SF-36) in an elderly population. Age Ageing 1994; 23: 182–4.

14. Lyons RA, Crome P, Monaghan S *et al*. Health status and disability among elderly people in three UK districts. Age Ageing 1997; 26: 203–9.

15. Jenkinson C, Stewart-Brown S, Petersen S, Paice C. Assessment of the SF-36 version 2 in the United Kingdom. J Epidemiol Community Health 1999; 53: 46–50.

16. Kind P, Hardman G, Macran S. UK Population Norms for EQ-5D. Centre for Health Economics Discussion Paper. York: University of York, 1998.

Received 16 May 2000; accepted in revised form 23 October 2000