Osteoporosis
International

Original Article

Postal Administration Compared with Nurse-Supported Administration of the QUALEFFO-41 in a Population Sample: Comparison of Results and Assessment of Psychometric Properties

P. Murrell[1*], C. J. Todd[1†], A. Martin[1], J. Walton[1], P. Lips[2] and J. Reeve[3] (on behalf of the Working Party for Quality of Life of the International Osteoporosis Foundation)

[1]Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, UK; [2]Department of Endocrinology, Vrije Universiteit, Amsterdam, The Netherlands; [3]Department of Medicine and Institute of Public Health, University of Cambridge, UK

Abstract. The objective of this study was to test the psychometric properties of the QUALEFFO-41, an osteoporosis-specific health-related quality of life (HRQoL) measure, in a population sample. The study involved repeated administration of QUALEFFO-41 and another measure of HRQoL, the SF-36, to a stratified random population sample to test reliability and validity of instrument. The study was set in urban Harrow, Middlesex, UK and rural Cambridgeshire, UK. The participants were 83 males and 88 females aged 65–80+ years. Of the 230 individuals contacted 76.5% participated. The test–retest reliability of most QUALEFFO-41 items was good (Kappa 0.59–0.91) but two items had lower repeatability. Internal consistency was also generally good, but suggestive of some redundancy for three domains. The QUALEFFO-41 domains in general correlated well with SF-36 domains ($r = 0.57$ to $r = 0.87$), suggesting good validity. The QUALEFFO-41 has been shown to be a reliable and valid disease-specific HRQoL measure for osteoporosis. In population samples, it can be administered by post, or with a professional available to answer queries. Some specific items perform less well than may be desired and there is evidence of some redundancy, but further investigation is required and there are not yet grounds for textual change. We recommend continued use of this well-validated, reliable instrument.

Keywords: Disease-specific questionnaire; Osteoporosis; QUALEFFO-41; Quality of Life; SF-36

Introduction

Over recent years a number of generic instruments have been developed for the measurement of health-related quality of life (HRQoL). Such instruments as the SF-36 [1,2] the Nottingham Health Profile (NHP) [3–8] the Sickness Impact Profile (SIP) [9] and EuroQol [10] can give general indications of quality of life. Part of their value has been to allow comparisons across disease groups. Most researchers agree however that generic measures need to be supplemented by disease-specific measures. This is particularly true for trials comparing treatments within the same disease group since disease-specific measures should in principle be more sensitive and specific.

Lips et al. [11] described the initial development of the QUALEFFO-41 health-related quality of life questionnaire for patients with osteoporosis. They studied 159 patients aged 55–80 years and a similar number of controls. Patients had diagnosed vertebral osteoporosis and were attending clinics in seven centers. They reported good test–retest reliability and adequate

Current addresses: *Dr Paul Murrell, Department of Statistics, The University of Auckland, Auckland, New Zealand.
†Professor Chris Todd, Professor of Primary Care & Community Health, School of Nursing, Midwifery & Health Visiting, University of Manchester, Manchester, UK.

Correspondence and offprint requests to: Jonathan Reeve, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, UK.

internal consistency of the QUALEFFO-41 instrument. In their study, validation of the QUALEFFO-41 against the SF-36 was conducted and they reported significant correlations between comparable domains of QUALEFFO-41 and SF-36 especially as regards pain, physical function and mental function. All five QUALEFFO-41 domains discriminated well between osteoporotic fracture cases and controls. The odds ratios for QUALEFFO-41 pain and social function were greater than those for the SF-36, but the SF-36 general health-scale was more discriminating between groups than the corresponding QUALEFFO domain. All five QUALEFFO domains were predictive of vertebral osteoporotic deformity measured on radiographs; and the pain, physical performance and social function domains performed better than the SF-36 as predictors. The QUALEFFO-41 appears to be a repeatable, coherent instrument with reasonable criterion groups validity.

As part of the MORE study baseline measurement, Oleksik et al. [12] assessed 751 women across Europe with bone mineral densitometry (BMD) T-scores ≤ -2.5 with and without prevalent osteoporotic vertebral fractures. These osteoporotic women completed QUALEFFO-41 along with the NHP and EuroQol. Women with vertebral fractures had significantly worse quality of life as revealed by their higher QUALEFFO-41 scores. QUALEFFO-41 scores indicated worse quality of life amongst those with more vertebral fractures and this was progressive. Similar but weaker relationships were also found between both the NHP and EuroQol with the number of vertebral fractures. These authors conclude that this increase in relation to number of fractures indicates that QUALEFFO-41 is suitable for measuring quality of life amongst clinically identified postmenopausal osteoporotic patients.

The present study aimed to test whether the QUALEFFO-41 was a suitable instrument for self-administration and also to contribute data which could be used in a future process of questionnaire refinement, which might be undertaken with the aim of making QUALEFFO shorter and easier to administer. QUAL-EFFO-41 is reproduced in Lips et al. [11] but to aid the reader the questions are reproduced in Table 1.

Design and Method

Subjects

We aimed to recruit a population sample rather than a sample of diagnosed osteoporotic patients. However, we aimed for the sample to be directly comparable to known population samples with representative proportions of cases with vertebral deformities. We thus used two centers involved in the European Vertebral Osteoporosis Study (EVOS) [13–16], and recruited from general practice lists which had participated in EVOS, sampling from the same age-sex strata that had been used in the original study. In the event of the present study proving successful, the aim was to apply the QUALEFFO-41 in

Table 1. QUALEFFO-41 questions

Pain
1) How often have you had back pain in the last week?
2) If you have had back pain, for how long did you have back pain in the daytime?
3) How severe is your back pain at its worst?
4) How is your back pain at other times?
5) Has the back pain disturbed your sleep in the last week?

Physical function: activities of daily living
6) Do you have problems with dressing?
7) Do you have problems with taking a bath or shower?
8) Do you have problems with getting to or operating a toilet?
9) How well do you sleep?

Physical function: jobs around the house
10) Can you do the cleaning?
11) Can you prepare meals?
12) Can you wash the dishes?
13) Can you do your day-to-day shopping?
14) Can you lift a heavy object of 20 lbs/10 kg (e.g. a crate of 12 bottles of milk, or a 1-year-old child) and carry it for at least 10 yards/meters?

Physical function: mobility
15) Can you get up from a chair?
16) Can you bend down?
17) Can you kneel down?
18) Can you climb stairs to the next floor of a house?
19) Can you walk 100 yards/meters?
20) How often have you been outside in the last week?
21) Can you use public transport?
22) Have you been affected by the changes of your figure due to osteoporosis (for example loss of height, increase of waist measurement, shape of your back?

Social function
23) Do you play any sport now?
24) Can you do your gardening?
25) Do you perform any hobby now?
26) Can you visit a cinema, theatre, etc?
27) How often did you visit friends or relatives during the last 3 months?
28) How often did you participate in social activities (clubs, social gatherings, church activities, charity, etc.) during the last 3 months?
29) Does your back pain or disability interfere with intimacy (including sexual activity)?

General health perception
30) For your age, in general, would you say your health is (excellent, good, satisfactory, fair or poor).
31) How would you rate your overall quality of life during the last week?
32) How would you rate your overall quality of life compared with 10 years ago?

Mental function
33) Do you tend to feel tired?
34) Do you feel downhearted?
35) Do you feel lonely?
36) Do you feel full of energy?
37) Are you hopeful about your future?
38) Do you get upset over little things?
39) Do you find it easy to make contact with people?
40) Are you in good spirits most of the day?
41) Are you afraid of becoming totally dependent?

the assessment of the quality of life impact of osteoporotic fractures, such as those of the spine and hips, in community-based population samples. We aimed to recruit 180 subjects aged 50–80 years stratified by age in six strata and by sex. These subjects were

recruited from two general practices, one urban (Harrow, Middlesex, UK), the other rural (Upwell, Cambridgeshire, UK), 120 from Harrow, 60 from Cambridgeshire.

Subjects were selected at random from the relevant strata of the age-sex registers of the participating general practitioners. The only exclusion criteria were (1) previous participation in EVOS, (2) history of major psychiatric illness, (3) dementia, (4) institutionalized at time of survey (e.g., in acute hospital or long-term residential care), (5) non-fluent in English language. Once patients had been randomly selected from the list, their eligibility was checked with the general practitioner (GP) and the accuracy of the registered address checked as far as possible within the medical records system. We then wrote to patients using a standard letter, including an endorsement of the project from the GP, and inviting them to take part in a study of a research instrument. The QUALEFFO-41 and SF-36 were enclosed for completion at home and then returned by mail in an enclosed stamped addressed envelope. The covering letter also explained that they would receive a follow-up visit at home from a research nurse 14 days after completion of the questionnaires. An appointment date and time at which the nurse would visit was offered in the letter, and this could be confirmed when returning the questionnaires, or by telephone if another time would be more convenient. At the follow-up visit the nurse administered QUALEFFO-41 and SF-36 verbally and asked participants to comment in their own words on the QUALEFFO questions. Participants were specifically asked to comment on any ambiguities in the questionnaire items. The research nurses also asked some supplementary questions concerning socioeconomic status and confirming whether health status had changed over the period since completion of the first round of survey instruments.

Statistical Analysis

We assessed reliability and validity using standard statistical approaches [17–19]. The test–retest reliability of each item in the questionnaire was assessed using the weighted Kappa statistic. The internal consistency of the questionnaire domains was assessed using Cronbach's α and the external validity of the domains was assessed by calculating Pearson correlations with the corresponding domains of the SF-36. Internal consistency and external validity were assessed for both the postal administration and the interview administration of the questionnaire. Very high values of Cronbach's α indicate redundancy. For those scales with very high alphas we conducted corrected item-total correlations in order to identify which specific items contributed to redundancy.

The answers for each item were scored from 1 to 5, except for Questions 23, 24, 25, and 26 (scored 1 to 3), and Questions 27, 28, and 29 (scored 1 to 4). The scores for Questions 33, 34, 35, 37, 39, and 40 were reversed so that a low score always indicates better health and a high score always indicates worse health. Domain scores were calculated by summing the scores of the items in each domain and scaling to the range 0 to 100. For Questions 24, 26, and 29, the answers 'not applicable', 'no cinema or theatre within a reasonable distance', and 'not applicable' (respectively) were coded as missing values. This is consistent with the analysis of the QUALEFFO-41 questionnaire data reported by Lips et al. [11]. Individual item scores and total domain scores for the SF-36 questionnaire were calculated according to the instructions in the manual [2].

Cronbach's α values were calculated using SPSS for Windows, weighted Kappa values were calculated using StatXact [20], and summary statistics and plots were generated using S-Plus [21].

Results

A total of 230 individuals were contacted at least once to complete the postal administration of the questionnaire. Twenty-one individuals never responded and 33 refused to participate in the study – a response rate of 76.5%. Two individuals only completed the postal questionnaire and two only completed the interview administration. This left a total of 344 paired questionnaires for the main analysis.

Age and sex characteristics of the sample are presented in Table 2. Our sample is predominantly urban with an urban : rural ratio of 2.35:1. Based on a logistic regression of response rate on age-group, gender, and district, the non-response rate was significantly higher in rural Cambridgeshire (42%) compared with urban Harrow (8%; $p<0.001$) and there was a significant trend of higher non-response rate with increased age (chi-squared test for trend; $p<0.001$).

Table 2. Age sex characteristics of the sample

Cambridgeshire												Harrow											
55–59		60–64		65–69		70–74		75–79		80–84		55–59		60–64		65–69		70–74		75–79		80–84	
M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F
4	3	4	5	4	7	2	5	5	4	4	4	10	10	10	10	10	10	10	10	10	10	10	10
	7		9		11		7		9		8		20		20		20		20		20		20
						51												120					

This table totals to 171 rather than 172 because one person has a missing value for sex.

Based on the participants' comments to the nurses (as recorded on the interview report forms), approximately 10% of participants experienced some sort of adverse health-related event between the postal and the interview administration of the questionnaire. However, in only one case was this considered sufficiently severe for the patient to be excluded from the test–retest analysis because of major health status change.

The response rate for individual items was in excess of 95%, except for Questions 5 (88%), 22 (85%), 24 (93%), 26 (93%), and 29 (53%) in the postal administration, and Questions 24 (94%) and 29 (56%) in the interview administration.

Performance of the Questionnaire on a General Population

The test–retest reliability of most items was good, with values of the Kappa statistic ranging between 0.59 and 0.91 (Fig. 1). However, Question 29 had only a moderate Kappa value (0.47) and Question 22 had a poor value (0.19).

Internal consistency within domains was also generally good, with Cronbach's α values ranging between 0.70 and 0.91 (Table 3). The exceptions were the leisure and social activities domain in the postal administration (0.65) and the physical function (activities of daily
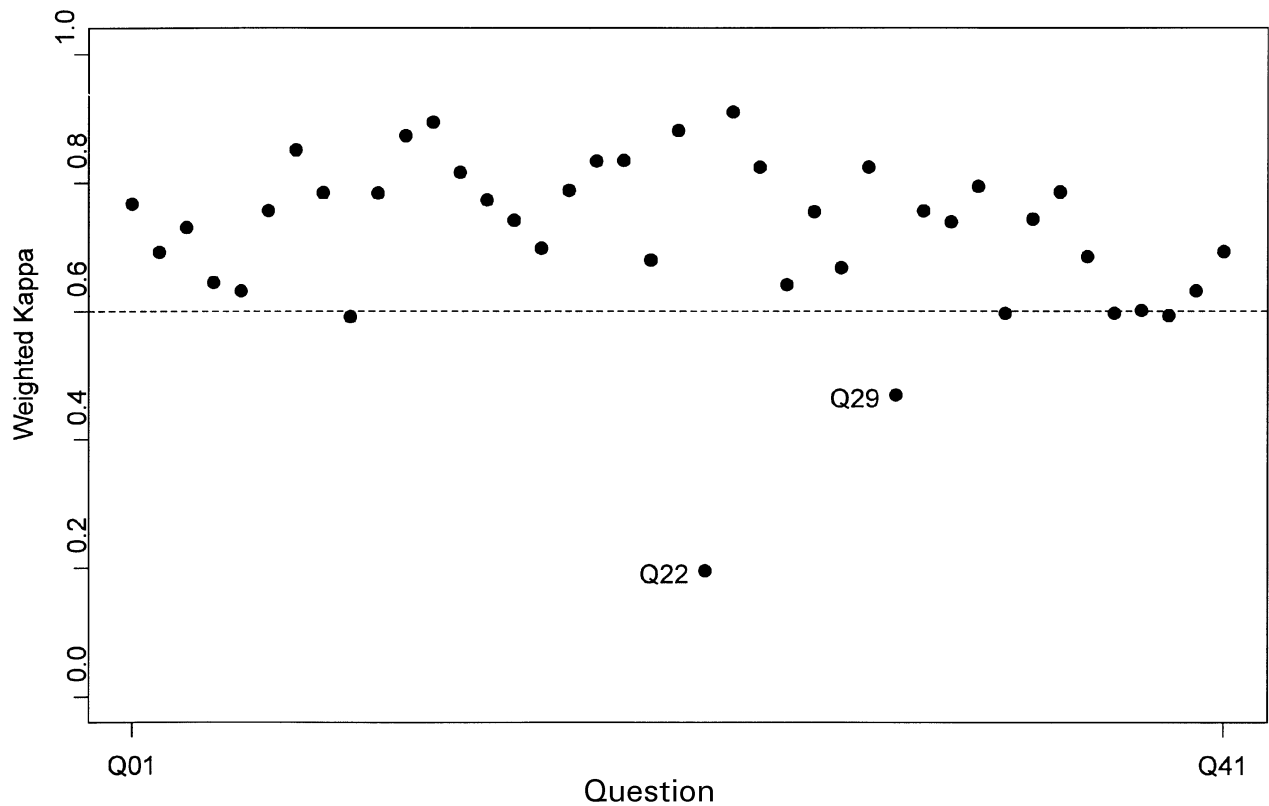

Fig. 1. Test-retest reliability (Kappa) for each item of the QUALEFFO-41.

Table 3: Internal consistency and external validity of QUALEFFO-41 domains

QUALEFFO-41 Domain	Internal consistency (Cronbach's α)		SF-36 domain	External validity (Pearson correlations)	
	Postal	Interview		Postal	Interview
Pain	0.89	0.91	Bodily pain	0.63	0.57
Activities of daily living	0.70	0.57	Physical functioning	0.70	0.72
Jobs around the house	0.90	0.88	Physical functioning	0.76	0.79
Mobility	0.88	0.87	Physical functioning	0.82	0.87
Leisure, social activities	0.65	0.74	Social functioning	0.40	0.66
General health perception	0.84	0.77	General health	0.76	0.78
Mental function	0.81	0.78	Mental health + vitality	0.84	0.83

living, Questions 6–9) domain in the interview admin-
istration (0.57). In these two cases, leave-one-out
diagnostics indicated that Question 29 and Question 9,
respectively, were most responsible for the lower
Cronbach's α values. The reproducibility (intraclass
correlation coefficient [19]) of the complete physical
function domain (Questions 6–22) was 0.88. The very
high Cronbach's α values observed for the domains
'pain' (Questions 1–5), 'jobs around the house'
(Questions 10–14) and 'mobility' (Questions 15-22)
are suggestive of redundancy within the domain. This
requires further analysis using corrected item-total
correlations to identify redundant questions.

The domain scores correlated well in general with
the corresponding domain scores from the administra-
tions of the SF-36, ranging from 0.57 to 0.87. The one
aberrant result was the leisure and social activities
domain in the postal administration, which only had a
correlation of 0.40 with the social functioning domain
of the SF-36, a finding which is similar to the
multicenter study [11].

Agreement of Questions with Their Own Domain Scores

Corrected item-total correlations were first used to
identify items, which did not agree well with other
items in the same domain. For these calculations, the
three physical function domains were combined into a
single domain. Item-total correlations should exceed 0.2
to be considered acceptable [18]. By this criterion,
Question 29 in the postal administration (correlation =
0.17) and Question 22 in the interview administration
(0.07) performed poorly.

Most items had the highest correlation with their own
domain score. Some items in the physical function
domains had their highest correlation with a physical
function domain other than their own (Question 8 in the
postal administration and Questions 6 and 7 in the
interview administration). More serious deviations
involved Question 8 in the interview administration
(which correlated better with the leisure and social
activities items than with activities of daily living items)
and Questions 24, 26 and 29 in both administrations
(which correlated better with physical function items
than with leisure and social activities items). In the
second administration, Question 22 did not correlate
well with any domain score, including its own.

In the single, combined physical function domain,
very high corrected item-total correlations were used to
detect items which may have been redundant, in that
they add little or no new information. In a future
version of QUALEFFO, items with excessive item-total
correlation coefficients might be candidates for
removal. An arbitrary criterion of 0.8 identified
Questions 7, 10, 18 and 19 in the postal administration
and Questions 10, 13, 18 and 19 in the interview
administration as potentially redundant (correlations
ranged from 0.80 to 0.89) (Table 4). However, it should

Table 4. Item total correlations for QUALEFFO-41 items 6-22 at each administration

Item number	Corrected item-total correlation	
	First administration	Second Administration
Q06	0.77	0.69
Q07	0.81	0.73
Q08	0.65	0.46
Q09	0.32	0.36
Q10	0.84	0.89
Q11	0.68	0.74
Q12	0.63	0.57
Q13	0.75	0.85
Q14	0.70	0.73
Q15	0.72	0.65
Q16	0.73	0.68
Q17	0.59	0.67
Q18	0.80	0.80
Q19	0.83	0.80
Q20	0.49	0.66
Q21	0.76	0.76
Q22	0.47	0.07

be emphasized that this was not a population of
osteoporotic patients, so no simplifications are justified
unless the same observation is made consistently in
such subjects in future studies.

Questions 9, 10, 13, 22, 26 and 29 had poor
psychometric properties. Question 9 was previously
associated with poor diagnostic performance in an
osteoporotic population (P. Lips, personal communica-
tion) and Questions 10, 13, 22, 26 and 29 were identified
by the nurses (JW and AM) administering the
questionnaire as ambiguous or difficult for participants
to understand. Despite their poorer performance,
Questions 18 and 19 should remain unchanged because
of their perceived importance for the diagnostic
performance of the questionnaire in an osteoporotic
population. In any future scoring algorithm, Question 24
should be considered for relocation to the physical
function domain. The effect of removing the activities of
daily living items was also investigated (Questions 6, 7,
8 and 9). In this data set, removing Questions 9, 10, 13
and 22 would not compromise the internal consistency of
the physical function domain (Cronbach's α = 0.93 and
0.92 for the two administrations). Removing Questions
6, 7 and 8 had little effect on internal consistency
(Cronbach's α = 0.92 and 0.91). Amongst the leisure and
social activities items, removal of Question 29 did not
damage internal consistency (although in postal admin-
istration Cronbach's α increased from 0.65 to 0.75).
However, for face validity it is important to retain this
item. Removal of Questions 24 and 26 as well resulted in
Cronbach's α values of less than 0.70. In terms of
external validity, the effect of removing Questions 9, 10,
13 and 22 from the physical function domain was minor.
Removing Question 29 from the leisure and social
activities domains had little effect in the interview
administration and improved the external validity in

postal administration (Pearson correlations increased from 0.40 to 0.52). However, also removing Questions 24 and 26 resulted in correlations of less than 0.5 for both administrations, which is not acceptable.

Performance of the SF-36

The SF-36 performed very well in terms of reliability and internal consistency in all of the domains that could be compared with the QUALEFFO questionnaire. Weighted Kappa values were all in the region of 0.6 or above and Cronbach's α values ranged from 0.77 to 0.93. In contrast, weighted Kappa values for the items in the two domains that were not comparable with QUALEFFO domains (Role–physical and Role–emotional) were all less than 0.6 (range: 0.35 to 0.58).

Residential District and Age Group Differences

A stratified analysis by district (urban Harrow versus rural Cambridgeshire) revealed only one major difference in terms of internal consistency or reliability; the reliability for Question 29 (Kappa = 0.47 overall) was much worse in Harrow (Kappa = 0.12) than in Cambridgeshire (Kappa = 0.79).

Adverse Health Events Between Administrations

An analysis excluding participants who reported an adverse health event between administrations did not reveal any major change in the reliability for Questions 22 and 29 (Kappa = 0.21 and 0.48).

Overview of Postal and Nurse-assisted Administrations

In overview, it appears that QUALEFFO-41 can be successfully administered both as a self-completed questionnaire as part of a clinical research interview and as a postal questionnaire (Table 5, Fig. 2).

Table 5. Means and standard deviations for the QUALEFFO-41 domains at each administration

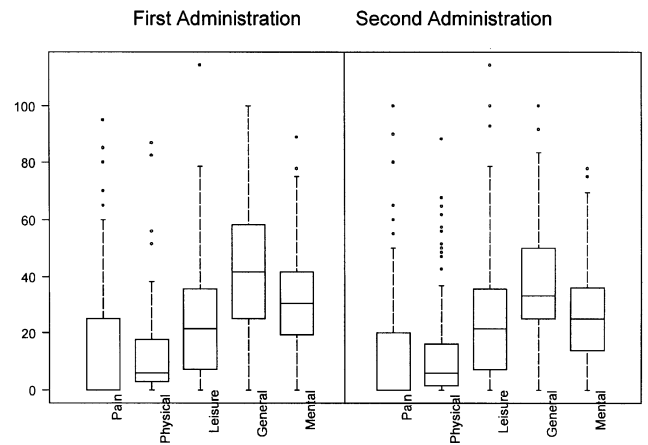| QUALEFFO-419 domain	Postal administration Mean (SD)	Interview administration Mean (SD)
Pain	16 (23)	12 (22)
Physical function	11 (14)	12 (15)
Social function	25 (21)	25 (24)
General health	42 (24)	40 (22)
Mental health	31 (17)	27 (17)
Total	19 (12)	16 (13)

Fig. 2. Box and whisker plots for the QUALEFFO-41 domains at each administration.

Discussion

The QUALEFFO-41 questionnaire has previously been shown to be a reliable, consistent and valid tool for assessing the quality of life of patients with osteoporosis [11,12]. The present study provides an opportunity to assess the performance of QUALEFFO-41, when self-administered, as well as to extend the assessment of its performance in a general population of men and women aged over 50. The two samples were drawn from registers which had been used previously to select two of the random samples of EVOS and the present sample had the same age structure. On average, these two EVOS centres had 12.7% males and 9.7% females, who showed evidence of at least one vertebral fracture according to the McCloskey–Kanis algorithm [16] and we have no reason to believe that our sample was substantially different in its prevalence of osteoporotic fractures. Our intention was that we should be able to use the QUALEFFO-41 questionnaire in EVOS and similar groups of subjects. However, for studies on groups with a higher prevalence of osteoporosis or who are substantially older, caution should be applied before assuming the direct applicability of our results.

The psychometric performance of the QUALEFFO-41 questionnaire, when self-administered as well as when supervised, was generally very good in terms of test–retest reliability, internal consistency, and external validity. It is particularly noteworthy that the test–retest reliability of 39 of the 41 QUALEFFO-41 items was good to excellent, despite the differences in method of administration of the questionnaires at the two measurement points. We take this to reflect the robust measurement properties of the QUALEFFO-41. Taken with previous results from testing the questionnaire on cases of osteoporosis and controls, this further establishes the QUALEFFO-41 questionnaire as an effective tool for measuring health-related quality of life in an elderly population. Our data suggest that the

QUALEFFO-41 is a suitable instrument for population-based studies of the impact of osteoporotic fracture both when self-administered as part of an interview and as a postal survey. However, further testing will be necessary in non-British populations to be sure that its good performance in the present study is not specific to UK populations.

In testing these results for internal consistency two domains performed less well, the activities of daily living part of the physical function domain, and social function. Social function also had the worst external validity measure in relation to SF-36. These domains also contained items that performed poorly in terms of convergent and discriminant validity. On the other hand, in constructing the questionnaire, the aim had been to provide domains that could represent a very broad range of activities, so these results for external validity, convergence and discrimination were to some extent inbuilt into the questionnaire in order to maintain face and content validity. Removing the activities of daily living items from the questionnaire did not adversely affect the performance of the physical function domain. It should be noted, however, that participants in this study represent a population sample, not a clinical sample with diagnosed osteoporosis. Activities of daily living are often compromised in osteoporotic patients and thus the psychometric properties of these domain items may be quite different in a clinical sample, as was demonstrated in the multicenter validation study of QUALEFFO-41 [11]. A further possible concern, is that the removal of these or other questions, which relate to relatively high-level or complex physical activities (compared to the more basic physical activities addressed in the mobility items), might compromise the content validity of the questionnaire as a whole, especially amongst osteoporotic patients. At the present time we do not recommend changing the QUALEFFO-41 in any respect until further evaluative work is available from additional studies in which QUALEFFO-41 has been tested. For example, the applicability and psychometric properties of QUALEFFO-41 amongst other osteoporotic populations, such as patients with osteoporotic peripheral fractures is still to be evaluated. We would caution that QUALEFFO-41 has been developed specifically with vertebral osteoporosis in mind, rather than aiming to measure the extreme effects that hip fracture has on quality of life [22,23] and thus is not necessarily expected to be of specific use amongst such patients. The present study however, has gone some way towards demonstrating the reliability and validity of QUALEFFO-41 amongst more 'healthy' persons. Future work will need to investigate further the psychometric properties of QUALEFFO-41 amongst population samples with specified fracture histories and identified prevalent, non-symptomatic osteoporotic vertebral deformity, as well as further clinic populations. Data on repeatability, internal consistency, and validity from several QUALEFFO studies should be compared and only consistent underperformance of items in more than one study used to justify making changes to the questionnaire.

In conclusion, this study provides further evidence for the good psychometric qualities of the QUALEFFO-41 in terms of excellent test–retest reliability, good internal consistency and very acceptable levels of validity when related to an established generic health status measure, the SF-36. In this respect, our findings were quite similar to those of the multicenter validation study of the QUALEFFO-41 [11]. We have also identified potential candidate items for modification or removal that are quite similar to those found in the multicenter validation study [11]. Along with evidence from other studies, these will be considered by the International Osteoporosis Foundation Working Party for Quality of Life for future modification of the QUALEFFO. We believe it would be useful to confirm that a postal version of the QUALEFFO-41 could be used successfully in other parts of Europe. If this proved to be the case, QUALEFFO-41 could be incorporated into cohort studies and randomized controlled trials of osteoporosis treatment at modest cost. The possibility of achieving further efficiencies by shortening the QUALEFFO-41 requires further evaluative work and it is premature at the present time to change the instrument. Nonetheless, it seems that QUALEFFO-41 is a frontrunner condition-specific health-related quality of life measure for osteoporosis, which can be used either as part of a postal survey or as a self-completion instrument during a research interview in a clinical or non-clinical setting.

References

1. Brazier JE, Harper R, Jones N M, O'Cathain A, Thomas K J, Usherwood T, Westlake L. Validating the SF-36 health survey questionnaire: new outcome measure for primary care BMJ, 1992;305:160–4.
2. Ware, J. E. SF-36 Health Survey Manual and Interpretation Guide. Boston, MA: Nimrod: 1993.
3. Hunt SM, McKenna SP, McEwen J, Backett EM, Williams J, Papp E. A quantitative approach to perceived health status: a validation study. J Epidemiol Community Health 1980;34:281–6.
4. Hunt SM, McKenna SP, Williams J. Reliability of a population

survey tool for measuring perceived health problems: a study of patients with osteoarthrosis. J Epidemiol Community Health 1981;35:297–300.

5. Hunt SM, McEwen J, McKenna SP, Backett EM, Pope C. Subjective health of patients with peripheral vascular disease. Practitioner 1982;226:133–6.

6. McKenna SP, McEwen J, Hunt SM, Papp E. Changes in the perceived health of patients recovering from fractures. Public Health 1984;98:97–102.

7. McKenna SP, Hunt SM, McEwen J. Weighting the seriousness of perceived health problems using Thurstone's method of paired comparisons. Int J Epidemiol, 1981;10:93–7.

8. Hunt SM, McKenna SP, McEwen J, Williams J, Papp E. The Nottingham Health Profile: subjective health status and medical consultations. Soc Sci Med 1981;15:221–9.

9. Bergner JE, Harper R, Jones NMB, et al. The Sickness Impact Profile: development and final revision of a health status measure. Med Carer; 1981;19;787–805.

10. EUROQoL Group. Euroqol: a new facility for the measurement of health related quality of life. Health Policy 1990;16:199–208.

11. Lips P, Cooper C, Agnusdei D, et al. Quality of life in patients with vertebral deformities: validation of the quality of life questionnaire of the European Foundation for Osteoporosis (QUALEFFO). Osteoporos Int 1999;10:150–60.

12. Oleksik A, Lips P, Dawson A, et al. Health related quality of life in postmenopausal women with low BMD with or without prevalent vertebral fractures. J Bone Miner Res 2000;15:1384–92.

13. O'Neill T W, Cooper C, Algra D, et al. Design and development of a questionnaire for use in a multicentre study of vertebral osteoporosis in Europe: The European vertebral osteoporosis study (EVOS). Rheumatol Eur 1995;24:75–81.

14. O'Neill TW, Marsden D, Matthis C, Raspe H, Silman AJ. Survey response rates: national and regional differences in a European multicentre study of vertebral osteoporosis. J Epidemiol Community Health 1995;49:87–93.

15. O'Neill TW, Cooper C, Cannata JB, et al. Reproducibility of a questionnaire on risk factors for osteoporosis in a multicentre prevalence survey: the European Vertebral Osteoporosis Study. Int J Epidemiol 1994;23:559–65.

16. O'Neill TW, Felsenberg D, Varlow J, et al. The prevalence of vertebral deformity in European men and women: the European Vertebral Osteoporosis Study. J Bone Miner Res 1996;11:1010–18.

17. Todd C, Bradley C. Evaluating the design and development of psychological scales. In Bradley C, (editor). Handbook of psychology and diabetes. Chur, Switzerland: Harwood Academic, 1994.

18. De Vellis RF. Scale development: theory and applications. Newbury Park: Sage, 1991.

19. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press, 1995.

20. StatXact-4 Version 4. Cambridge, MA: Cytel Software Corporation, 1998.

21. S-Plus 4.5, Seattle, WA: Mathsoft Inc, 1989–98.

22. Todd C. Outcomes assessment in context: Hip fracture. In Philp I, editor. Outcomes assessment in healthcare for elderly people. London: Ferrand Press, 1997.

23. Salkeld G, Cameron ID, Cumming RG, et al. Quality of life related to fear of falling and hip fracture in older women: a time trade-off study. BMJ 2000;320:341–6.