Osteoporos Int (1999) 10:150–160


**Osteoporosis International**

*Original Article*

# Quality of Life in Patients with Vertebral Fractures: Validation of the Quality of Life Questionnaire of the European Foundation for Osteoporosis (QUALEFFO)

P. Lips, C. Cooper, D. Agnusdei, F. Caulin, P. Egger, O. Johnell, J. A. Kanis, S. Kellingray, A. Leplege, U. A. Liberman, E. McCloskey, H. Minne, J. Reeve, J.-Y. Reginster, M. Scholz, C. Todd, M. C. de Vernejoul and I. Wiklund for the Working Party for Quality of Life of the European Foundation for Osteoporosis

**Abstract.** Vertebral fractures may be minor or lead to pain, decreased physical function, immobility, social isolation and depression, which together contribute to quality of life. A Working Party of the European Foundation for Osteoporosis has developed a specfic questionnaire for patients with vertebral fractures. This questionnaire, QUALEFFO, includes questions in the domains pain, physical function, social function, general health perception and mental function. QUALEFFO was validated in a multicenter study in seven countries. The study was done in 159 patients aged 55–80 years with clinical osteoporosis, i.e., back pain and other complaints with at least one vertebral fracture and lumbar bone mineral density *T*-score $< -1$. Patients with a recent vertebral fracture were excluded because of unstable disease. Controls were age- and sex-matched, and did not have chronic back pain or vertebral fractures. Subjects with conditions exerting a major influence on quality of life were excluded. The QUALEFFO was administered twice within 4 weeks and compared with a generic questionnaire, the Short Form 36 of the Medical Outcomes Study (SF-36). Standard spinal radiographs were made for assessment of vertebral height. Seven questions were removed from the analysis because of low response rate, linguistic ambiguities or redundancy. The 41 remaining questions were analyzed for repeatability, internal consistency and the capacity to discriminate between patients with vertebral fractures and controls. Comparison with the SF-36 was performed within similar domains by conditional logistic regression and by receiver operating characteristic (ROC) curves. The repeatability of QUALEFFO was good (kappa statistics 0.54–0.90) and 26 of 41 questions had a kappa score $\geqslant 0.70$. The internal consistency of the five domains was adequate, with Crohnbach $\alpha$ around 0.80. All except five questions discriminated significantly between patients and controls. The median scores of QUALEFFO were significantly higher in patients with vertebral fractures than in controls in all five domain ($p<0.001$), which is consistent with decreased quality of life in patients with osteoporosis. Spinal radiographs were assessed using the McCloskey–Kanis algorithm. According to this, 124 patients (78%) had vertebral fractures of $\geqslant 3$ SD severity, in contrast with 7 controls (4%). Significant correlations existed between scores of similar domains of QUALEFFO and the SF-36, especially for pain, physical function and mental function. All five domains within each questionnaire discriminated significantly between fracture cases and controls. The odds ratios for pain and social function were greater for QUALEFFO, while general health perception was more discriminating using the SF-36. The ROC curve analysis of QUALEFFO indicated that all five domains were significantly predictive of vertebral fractures. When comparing similar domains of the two questionnaires, QUALEFFO domains demonstrated significantly better performance for pain, physical

*Correspondence and offprint requests to:* P. Lips, Department of Endocrinology, Academic Hospital Vrije Universiteit, PO Box 7057, 1007 MB Amsterdam, The Netherlands. Tel: +31 20 444 0530. Fax: +31 20 444 0502. e-mail: p.lips@azvu.nl

function and social function. The QUALEFFO total score and SF-36 physical composite score showed similar performance. In conclusion, QUALEFFO is repeatable, coherent and discriminates well between patients with vertebral fractures and control subjects. The results of this study confirm the decreased quality of life in patients with vertebral fractures.

**Keywords:** Osteoporosis; Physical function; Quality of life; Vertebral fractures

## Introduction

The clinical impact of osteoporosis is determined by the fractures that occur and the subsequent morbidity [1,2]. Typical osteoporotic fractures include those of the hip, the distal radius and the vertebrae. Vertebral fractures may be minor and pass unnoticed, or they may lead to long-term immobility and disability. Recently it has been estimated that about 1 in 3 vertebral fractures comes to clinical attention [3]. On the other hand, vertebral fractures may cause local pain for 3 years or more [4], and lead to decreased physical functioning, immobility, social isolation and depression [5]. Since the sum of physical, social and mental functioning determines quality of life, its assessment plays an increasingly important role in clinical studies, and particularly as an outcome measure of clinical trials [5,6].

During the last 15 years, several generic instruments for measuring quality of life have been developed, such as the Nottingham Health Profile, the Sickness Impact Profile and the Short Form 36 [7–9]. These generic instruments give a general estimate of health and are not specific for any disease. They can be used to estimate the burden of disease in a population and to compare the consequences of different diseases. More recently, disease-specific instruments have been developed that may contain more relevant questions, are less time-consuming and may be more valid, in the sense that they measure quality of life more accurately in that particular disease than generic instruments [10,11]. In 1992 the European Foundation for Osteoporosis constituted a working party to develop a specific quality of life questionnaire for patients with vertebral fractures. Recently, we reported on the design and development of this questionnaire, called QUALEFFO [12].

The objective of the multicenter study presented in this paper is the validation of QUALEFFO in patients with vertebral fractures and matched control subjects. The main purpose of this validation study was to assess the repeatability, internal consistency and construct validity of the questionnaire. In addition, we compared the instrument with the Short Form 36 (SF-36) of the Medical Outcomes Study to assess the ability of the two questionnaires to discriminate between patients with vertebral fractures and controls. In addition, the SF-36 was used to assess the burden of disease in our study population.

## Subjects and Methods

The study was performed in seven centers (Cambridge, Malmö, Bad Pyrmont, Liège, Paris, Siena and Amsterdam). Radiologic morphometry was performed at Sheffield, statistical analysis was done in Southampton and the coordination center was in Amsterdam.

### *Patients and Control Subjects*

Patients eligible for the study were patients with clinical osteoporosis, i.e., they were recruited on clinical presentation of back pain and other complaints with vertebral fracture on radiographs. Inclusion criteria were an age between 55 and 80 years, lumbar bone mineral density (BMD) *T*-score $< -1$ and at least one vertebral fracture, i.e., reduction of vertebral height (anterior, middle or posterior) of more than 20% on clinical reading. The patients had to be ambulant and able to complete the questionnaire. Patients with a recent vertebral fracture (within 1 month) were excluded because of unstable disease. Patients with other recent fractures (interfering with pain or activity) were also excluded, as were patients with metabolic bone disease, disseminated malignancies, and recent treatment with calcitonin or fluoride.

Control subjects were healthy persons chosen either from the general population or (in most centers) from a hospital outpatient department (e.g., treated stable hypothyroidism, treated uncomplicated hypertension). They did not have chronic back pain or vertebral fractures according to the above-mentioned criteria. Control subjects with a recent nonspinal fracture were also excluded, as were those with other conditions exerting a major influence on quality of life, for example disseminated malignancy, similar to the fracture cases.

Each control subject was sex- and age-matched ($\pm 3$ years) to a patient in the same center. Informed consent was obtained from all patients and control subjects. The study protocol was approved by the local Ethical Review Boards.

### *Questionnaires*

The design of QUALEFFO has been published previously [12]. Briefly, a working party of clinicians and quality of life specialists from eight countries constructed a questionnaire with 48 questions and six visual analog scales in the following five domains: pain, physical function (activities of daily living, jobs around the house, mobility), social function, general health perception and mental function. Most questions were newly designed but two were taken from the MEDOS [13] and EVOS questionnaires [14]. The questionnaire was translated from English into French, German, Italian, Swedish and Dutch. Following standard proce-

dures, back-translations were made into English and the translations were corrected and adapted [15]. The layout of QUALEFFO was identical in all language versions.

The Short Form 36 of the Medical Outcome Study (SF-36) was used in the IQOLA format, available in the above-mentioned six languages (by kind permission of Dr J. Ware, International Quality of Life Project, IQOLA, New England Medical Center, Boston, MA).

### *Radiographs*

Standardized lateral radiographs of the thoracic and lumbar spine, centered at T8 and L3 respectively, were made at a film–focus distance of 105 cm. In the profile of each vertebra from T4 to L5, anterior, central and posterior height were measured. Vertebral deformity was defined when anterior, central, posterior or overall height loss was more than 3 standard deviations below the predicted value for that vertebra, according to the algorithm of McCloskey and Kanis [16].

### *Practical Conduct of the Study*

Patients and control subjects were included in the study when they met inclusion and exclusion criteria and after giving informed consent. The questionnaires were always administered in the same order: QUALEFFO followed by SF-36. The questionnaires were administered at the clinic before any other procedures. The subjects completed the questionnaires, alone in a quiet place, after instruction by the study nurse. Subsequently, the nurse checked whether all questions had been answered. The questionnaires were again administered at the clinic after 2–4 weeks. The patients were also questioned on intercurrent events.

### *Statistical Analysis*

The responses to all 48 questions and six visual analog scales were examined. There was a low response rate (less than 50%) to seven questions which were found during testing to have linguistic ambiguities (A6, A7, A8, A9, E27, E30, F37). These seven questions were removed from further analysis. The answers to each question were scored from 1 to 5, except for E28, 29, 31, 32 (score 1–3) and E33–35 (score 1–4); 'not applicable' was not scored. The response options for questions G40, 41, 42, 44, 46, 47 were reversed so that the order was always from 1 (healthy) to 5 (not healthy). Domain scores were calculated by summing the answer scores and submitting the sum to a linear transformation to a 100 scale.

A multitrait analysis was performed [17] to check the adequacy of score construction in Likert scales, in which the overall score is calculated by simple addition without weighting each question. This requires adequate convergent validity, i.e., the correlation between the score for each question and its own total domain score should be higher than 0.40 [18]. Discriminant validity also should be high, i.e., the correlation of the score for each question with its own domain score should be higher than with total scores of other domains. The floor and ceiling effects were calculated to check whether the answer scale is adequate for the investigated population. The floor effect indicates the percentage of subjects with the lowest possible domain score. The ceiling effect indicates the percentage of subjects with the highest possible domain score. The 41 questions grouped in five domains were further assessed for repeatability (test–retest) using the weighted kappa statistics [19]. The internal consistency was tested using the Crohnbach $\alpha$ reliability coefficient. The closer the proportion of variance due to error (in relation to real variance) is to zero, the closer Crohnbach $\alpha$ is to 1. The capacity of questions to discriminate between vertebral fracture cases and nonfracture controls was investigated, using conditional logistic regression to derive an odds ratio with 95% confidence intervals. For the conditional logistic regression, each domain score was treated as a continuous variable and divided into thirds of its distribution. These thirds were then entered in the logistic model assuming equivalence between lowest and middle third and between middle and upper third of each variable. Linear regression with the number of deformities treated as a continuous variable was used to assess trends in domain scores as the number of deformities increased. The 41 questions of QUALEFFO are listed in the Appendix. Comparison with the SF-36 was performed within similar domains: pain, physical function, social function, general health perception and mental function. Because the mental function domain of QUALEFFO also contains vitality questions, it was compared with the SF-36 mental health and vitality domains. Neither of the role functioning domains of the SF-36 were included in this between-domain comparison (except for the composite scores). The SF-36 was scored according to the manual [20]. For the comparison with the SF-36, we standardized the response options so that the trend toward greatest impact on quality of life was consistent (option 1, least impact; option 5, greatest impact). A total QUALEFFO score was computed by summing the standardized scores of all questions and submitting the sum to a linear transformation to a 100 scale. In a similar way, a QUALEFFO pain and physical composite score and a QUALEFFO mental and social composite score were calculated. For the SF-36, physical and mental composite scores were calculated as described [21].

Receiver operating characteristic (ROC) curves were constructed (STATA package) to compare the ability of QUALEFFO and SF-36 domains, total and composite scores, to discriminate between cases and controls over all possible cutoff values of the questionnaire scores [22]. The significance of differences between areas under curves was used to compare domains of QUALEFFO and SF-36, and also to compare total and

composite QUALEFFO scores with the composite scores of the SF-36 [22].

## Results

We studied 159 subjects with clinically diagnosed vertebral fractures and 159 age-matched controls. The mean age of the cases and control subjects was 67.4 years (SD 6.7 years) and 66.3 years (SD 7.3 years), respectively. The study included 18 male and 141 female case–control pairs. The origin of case–control pairs was as follows: Italy 32, Netherlands 30, United Kingdom 28, Belgium 26, France 25, Sweden 9, Germany 9. Thoracolumbar radiographs were evaluated for the presence of vertebral deformities using the method of McCloskey and Kanis. One hundred and twenty-four (78%) of the cases had vertebral deformities of $\geqslant 3$ SD severity using this algorithm, in contrast with 7 (4%) controls.

Table 1 shows the results of the multitrait analysis of the QUALEFFO questionnaire grouped according to domains: (a) pain; (b) physical function; (c) social function; (d) general health perception; (e) mental function.

Table 2 shows a summary of response rates, repeatability, internal consistency, prevalence of the lowest answer category and discriminative ability of the questions of QUALEFFO. The response rate for some questions in the social function domain was lower than for other question because a proportion of the responses was 'not applicable'. The Crohnbach $\alpha$ values derived are generally indicative of good levels of internal consistency (i.e., approximating 0.8). Almost all questions gave high degrees of reproducibility (test–retest repeatability) and 26 of 41 questions gave a kappa score $\geqslant 0.70$ indicating substantial agreement between the first and second administration. There was considerable variation in the frequency of reporting of the lowest answer category (compatible with absence of disease) in different questions. Most questions differentiated statistically significantly between vertebral fracture cases and nonfracture controls, except for five questions (B13, E31, G45–47). We included data for these questions in the statistical analyses presented here and subsequently, as there was no methodologic reason for their exclusion.

The questionnaire initially included six visual analog scales that inquired about: (a) severity of pain at its worst, (b) severity of pain at other times, (c) overall quality of life, (d) health during the previous week,

**Table 1.** Results of the multitrait analysis of QUALEFFO

| Domain | Floor effect | Ceiling effect | Convergent validity | Discriminant validity |
|---|---|---|---|---|
| Pain | 0 | 6.7% | 100% | 100% |
| Physical function | 0 | 0.6% | 94% | 72% |
| Social function | 1.9% | 5.7% | 100% | 96% |
| General health perception | 3.1% | 1.2% | 100% | 92% |
| Mental function | 0.6% | 0.6% | 78% | 92% |

The floor effect indicates the percentage of subjects with the lowest possible domain score. The ceiling effect indicates the percentage of subjects with the highest possible domain score. Convergent validity indicates the percentage of question scores correlating with the domain score better than 0.40. Discriminant validity is the percentage of question scores correlating better with their own domain score than with the scores of other domains.

**Table 2.** Evaluation of 41 questions in the five QUALEFFO domains

| Domain (no of question) | Response rate (%) | Repeatability | Internal consistency | Prevalence of lowest answer category | | Discrimination odds ratio |
|---|---|---|---|---|---|---|
| | | | | Control subjects (%) | Patients (%) | |
| | Median (range) | Median (range) | Crohnbach's $\alpha$ | Median (range) | Median (range) | Median (range) |
| Pain (5) | 99 (98–100) | 0.68 (0.64–0.70) | 0.82 | 47 (42–84) | 9 (8–59) | 2.2 (1.6–2.7) |
| Physical function (17) | 99 (98–100) | 0.74 (0.69–0.80) | 0.92 | 73 (23–94) | 45 (10–82) | 2.3 (1.1–3.2) |
| Social function (7) | 82 (69–100) | 0.79 (0.65–0.90) | 0.80 | 54 (38–76) | 39 (19–56) | 1.5 (1.4–2.6) |
| General health perception (3) | 100 (99–100) | 0.66 (0.64–0.74) | 0.72 | 13 (6–16) | 7 (5–8) | 1.9 (1.6–2.1) |
| Mental function (9) | 99 (98–100) | 0.65 (0.54–0.72) | 0.77 | 32 (10–69) | 22 (5–55) | 1.3 (1.1–1.8) |

The response rate indicates the number of valid responses for each question presented as median and range. The repeatability is the agreement (Cohen's kappa) between the first and second administration of QUALEFFO within 4 weeks. The internal consistency (Crohnbach's $\alpha$) was calculated for the entire domain. The prevalence of the lowest (healthiest) answer category is presented. The odds ratio represents the discrimination between patients and control subjects. All odds ratios were significant except for five questions (see text). All data except internal consistency are presented as the median and range for the different questions in the domain.

(e) social activities, and (f) financial situation. Five of these six scales (a–e above) discriminated significantly between patients with vertebral fractures and control subjects. However, the corresponding QUALEFFO questions (A3, A4, E34, F36 and F38) also discriminated well between cases and controls. We therefore eliminated the visual analog scales as they appeared to be redundant. The final adapted questionnaire contains 41 questions (see Appendix). The median domain scores ($\pm$ interquartile range) for patients with vertebral fractures and control subjects are shown in Table 3. The scores in the patients with osteoporosis clearly show significant impairment of quality of life.

As stated above, 35 of the 159 patients (22%) had vertebral deformities that were less severe than 3 SD by the McCloskey/Kanis algorithm. Of the 124 patients with McCloskey/Kanis-defined vertebral deformities, 43 (27%) had one deformity, 27 (17%) had two deformities and 54 (34%) had three or more deformities. One might expect an increasing QUALEFFO score with increasing number of deformities. When the patients were grouped according to the presence or absence of deformities (3 SD or more) or according to the number of deformities, the QUALEFFO domain scores did not show statistically significant differences or trends. However, when patients and control subjects were combined and grouped according to the number of deformities, significant trends for increasing scores with increasing number of deformities existed for all domains of QUALEFFO, but the significance was determined by the step from 0 to 1 fracture.

The response rate for the SF-36 was more than 95% for all questions. The five domains of QUALEFFO were directly comparable with corresponding sections of the SF-36. Table 4 reveals the Spearman correlation coefficient between domain scores on the QUALEFFO, and comparable scores on the SF-36 questionnaire. Table 5 compares the discriminatory capacity of the two instruments within these five domains. All five domains within each questionnaire discriminated significantly between fracture cases and controls. The odds ratios for pain and social function were greater for QUALEFFO, while general health perception was more discriminatory using questions from the SF-36. The total QUALEFFO score and SF-36 physical composite score discriminated well between patients and controls while the SF-36 mental composite score discriminated poorly.

Figure 1 shows the ROC curve analysis for the QUALEFFO total score and SF-36 physical and mental composite scores. Both instruments were significantly predictive of vertebral deformity. The area representing discrimination was similar ($p$=0.62) for QUALEFFO and for the SF-36 physical composite score, whereas the SF-36 mental composite score showed a poor discrimination. Figure 2 indicates the discriminatory capacities of the different QUALEFFO domains. The ROC curve analyses for QUALEFFO, SF-36 and the difference between the two questionnaires are summarized in Table 6. The three most discriminatory domains were pain, physical function and general health perception. However, all five domains were significantly predictive of vertebral fracture. When each domain of QUALEFFO

**Table 3.** Scores on five QUALEFFO domains in patients with vertebral fractures and control subjects

| Domain | Control subjects | Patients | $p$-value |
|---|---|---|---|
| Pain | 15.0 (0.0–35.0) | 45.0 (30.0–65.0) | <0.001 |
| Physical function | 10.3 (2.9–16.2) | 26.5 (14.7–38.2) | <0.001 |
| Social function | 28.4 (12.3–41.2) | 41.2 (24.7–64.7) | <0.001 |
| General health perception | 41.7 (33.3–58.3) | 58.3 (41.7–75.0) | <0.001 |
| Mental function | 30.6 (18.8–36.1) | 38.9 (25.0–50.0) | <0.001 |
| Total QUALEFFO score | 20.3 (12.3–29.5) | 35.3 (24.9–48.4) | <0.001 |

The scores were transformed to a 0–100 scale and are presented as median (25–75% range)

**Table 4.** Spearman rank correlation coefficients between scores of similar domains of QUALEFFO and SF-36 instruments

| QUALEFFO domain | SF-36 domain | Correlation coefficient | $p$ value |
|---|---|---|---|
| Pain | Bodily pain | 0.74 | <0.001 |
| Physical function | Physical functioning | 0.81 | <0.001 |
| Social function | Social functioning | 0.45 | <0.001 |
| General health perception | General health | 0.68 | <0.001 |
| Mental function | Mental health + vitality | 0.80 | <0.001 |
| Total QUALEFFO score | Physical composite score | 0.78 | <0.001 |
| Total QUALEFFO score | Mental composite score | 0.50 | <0.001 |

**Table 5.** Discriminatory capacity of QUALEFFO and SF-36 instruments as assessed by conditional logistic regression (see methods section)

| Domain[a] | QUALEFFO | | SF-36 | |
|---|---|---|---|---|
| | OR | (95% CI) | OR | (95% CI) |
| Pain | 4.9 | (3.0, 7.9) | 2.3 | (1.7, 3.3) |
| Physical function | 4.0 | (2.6, 6.0) | 4.1 | (2.6, 6.3) |
| Social function | 2.5 | (1.8, 3.7) | 1.4 | (1.1, 1.9) |
| General health perception | 2.6 | (1.8, 3.6) | 3.1 | (2.1, 4.7) |
| Mental function | 2.0 | (1.4, 2.7) | 2.1 | (1.5, 3.0) |
| Total QUALEFFO score | 3.7 | (2.4, 5.6) | | |
| SF-36 physical comp. score | | | 4.9 | (3.0, 8.0) |
| SF-36 mental comp. score | | | 1.3 | (1.0, 1.8) |

OR, odds ratio.
[a] See Table 4 for the SF-36 domains.

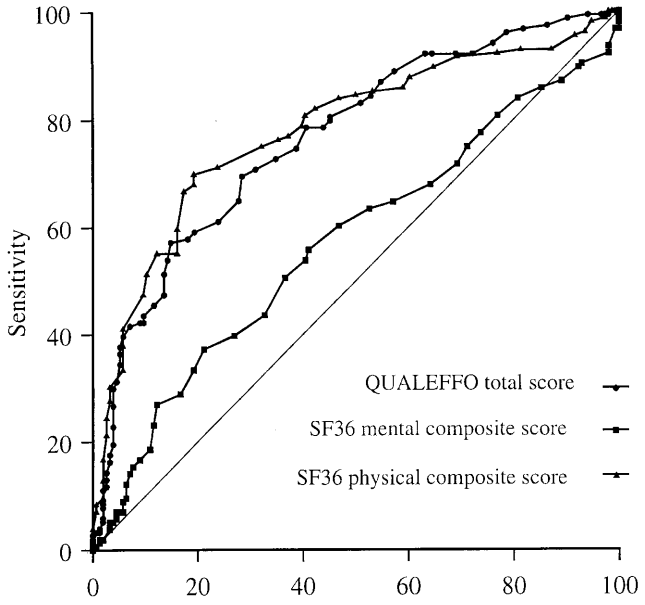


**Fig. 1.** Receiver operating characteristic curve for total QUALEFFO and SF-36 physical and mental composite scores distinguishing vertebral fracture cases from nonfractured controls.

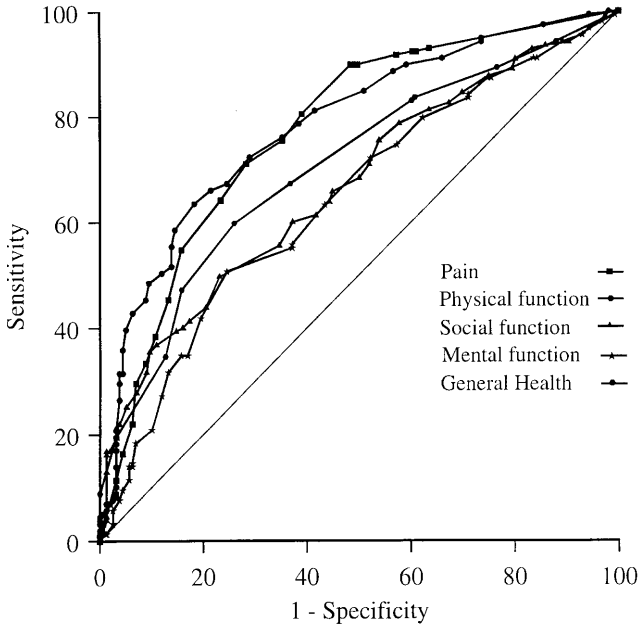


**Fig. 2.** Receiver operating characteristic curve for individual QUALEFFO domain performance in discriminating between vertebral fracture cases and nonfractured controls.

was compared with the corresponding domain of the SF-36, the QUALEFFO domains demonstrated significantly superior performance for pain, physical function ($p<0.01$) and social function ($p<0.05$). The results for mental function and general health perception did not differ significantly between the two questionnaires. The discriminant power was similar for QUALEFFO total score and the SF-36 physical composite score. A new QUALEFFO pain and physical composite score was created by summing the 5 pain and 17 physical function questions. This pain and physical function composite score performed slightly better than the SF-36 physical composite score, but the difference was not significant. The QUALEFFO mental and social composite score performed significantly better than the SF-36 mental composite score.

## Discussion

QUALEFFO was based on common problems affecting the life of patients with vertebral osteoporosis, i.e., patients with one or more vertebral fractures. The patients in the study did not have difficulty with

**Table 6.** Receiver operating characteristic (ROC) curve analysis for QUALEFFO and the SF-36

| Domain | QUALEFFO area* (SE) | SF-36 area* (SE) | Difference area (SE) | *p* value |
|---|---|---|---|---|
| Pain | 0.77 (0.03) | 0.70 (0.03) | 0.07 (0.02) | 0.001 |
| Physical function | 0.78 (0.03) | 0.73 (0.03) | 0.06 (0.02) | 0.006 |
| Social function | 0.67 (0.03) | 0.60 (0.03) | 0.07 (0.03) | 0.035 |
| General health perception | 0.71 (0.03) | 0.68 (0.03) | 0.02 (0.03) | 0.395 |
| Mental function | 0.64 (0.03) | 0.64 (0.03) | 0.00 (0.02) | 0.897 |
| QUALEFFO total score vs SF-36 physical composite score | 0.77 (0.03) | 0.78 (0.03) | −0.01 (0.02) | 0.617 |
| QUALEFFO total score vs SF-36 mental composite score | 0.77 (0.03) | 0.57 (0.03) | 0.20 (0.03) | 0.001 |
| QUALEFFO pain and physical composite score vs SF-36 physical composite score | 0.80 (0.02) | 0.78 (0.03) | 0.02 (0.02) | 0.254 |
| QUALEFFO mental and social composite score vs SF-36 mental composite score | 0.68 (0.03) | 0.57 (0.03) | 0.11 (0.03) | 0.001 |

The null hypothesis states that the questionnaire has no discriminant power (area under the curve = 0.5, line of unity). Higher values indicate better discrimination. The *p* value refers to the difference between the two questionnaires.
* All values are significantly different from 0.5 ($p<0.001$, except for SF-36 social function ($p<0.005$)).

completing the questionnaire after explanation by the study nurse. They were familiar with the topics of the questions and appreciated QUALEFFO, suggesting adequate face validity. Nevertheless, some questions could not be analyzed because of linguistic ambiguities (inviting multiple answers to one question). Other questions had 'not applicable' as an answer category (in the social function domain), leading to a low response rate. The wording and answer categories of these questions were adapted (see Appendix).

The results of the multitrait analysis indicate that the score construction of QUALEFFO is adequate. The scores of individual questions are better correlated with their own domain than with other domains, indicating good discriminant validity. The questions relating to the same concept have approximately the same variance, which leads to the conclusion that standardization or weighting of results is not required. Within each domain the questions contain a similar amount of information on the concept being measured. The results of this study indicate a good short-term repeatability after 3–4 weeks of most questions of QUALEFFO, with kappa statistics above 0.6 in 40 of 41 questions [19]. The internal coherence within the various domains is adequate without excessive redundancy, as indicated by the analyses of Crohnbach $\alpha$ [23]. These results indicate that the instrument is suitable for group comparisons. As shown in Table 5, QUALEFFO can discriminate between patients with vertebral fractures and age- and sex-matched control subjects, indicating worse quality of life in patients than in controls. In fact, in only five items was the odds ratio not significantly greater than 1. QUALEFFO discriminates particularly well in the domains of 'pain' and 'physical function'. The problems which are the subject of the questions in these domains are known to play an important role in the life of patients with vertebral osteoporosis [1,4].

The results of this study show significant impairment of quality of life in patients with vertebral osteoporosis compared with age- and sex-matched control subjects. The impaired quality of life was evident in all domains of both questionnaires. When comparing the discriminatory ability of QUALEFFO and the SF-36, QUALEFFO scored better on pain, physical and social function – items which are seriously affected in patients with osteoporosis. The SF-36 scored better on general health perception, which should be well measured by a generic questionnaire. The overall performance of the SF-36 in discriminating vertebral fracture cases from nonfracture controls is decreased by the poor results of the SF-36 mental composite score. On the other hand, the SF-36 physical composite score showed a similar discriminatory performance to QUALEFFO. The QUALEFFO pain and physical function composite score performed as well as total QUALEFFO and as the SF-36 physical composite score. This suggests the future development of a shorter version of QUALEFFO with similar discriminant power.

All patients in this study were patients with clinical osteoporosis who came to the hospital because of complaints or symptoms caused by vertebral fractures. As only around one-third of vertebral fractures comes to clinical attention [3], this questionnaire may not discriminate between patients with silent vertebral fractures and control subjects. The QUALEFFO domain scores did not show a significant trend according to the number of deformities. This may be explained by the fact that the patients in this study had stable disease and many fractures had probably occurred many years ago. This decreases the possibility of finding a significant correlation between QUALEFFO scores and number of deformities. Pain scores in patients with vertebral fractures decrease to stable levels after about 3 years [4].

Our controls were selected from hospital outpatient departments and from the general population. As the first source is more likely to be biased towards impaired quality of life, it is not surprising that the SF-36 domain scores in the control group were slightly worse than those reported in a previous population study from Sheffield [9]. However, any such bias would act in a conservative direction, and our results show that QUALEFFO is very well able to distinguish patients with vertebral fracture from hospital controls. One restriction has to be made for other causes of chronic back pain, as this was an exclusion criterion for the control subjects. The discriminative capacity of QUALEFFO between vertebral osteoporosis and osteoarthritis should be the subject of further study.

An important limitation of a disease-specific questionnaire such as QUALEFFO is that it cannot compare the impaired quality of life due to osteoporosis with that due to other diseases. A generic questionnaire is more appropriate for this purpose. Our results are of interest in this regard, and point to a substantial burden of illness attributable to osteoporosis. Thus, the mean SF-36 pain score of our cases was 49.9, while that of our controls was 68.5, and that of the previously published Sheffield population sample [9] was 79.0. Similarly, the mean SF-36 scores for physical function of our patients was 50.6 and that of our controls 71.7, while that of the Sheffield population was 86.2. There may be other advantages in combining generic and specific instruments in the study of osteoporosis. While a generic questionnaire can better assess osteoporosis as a general health problem in comparison with other diseases, a specific questionnaire is able to detect symptoms of osteoporosis.

In conclusion, QUALEFFO can be used in the patient group for which it is aimed, as it is repeatable, coherent and discriminates well between patients with vertebral fractures and control subjects, particularly in the domains of pain, physical function and social function. The sensitivity of QUALEFFO to change, e.g., improvement following treatment, is the subject of some longitudinal studies. The results of this study confirm the decreased quality of life in patients with vertebral fractures.

*Acknowledgements.* The authors gratefully acknowledge the assistance of: Anna Martin (Cambridge), Karin Önby (Malmö), Dr A. Paliwoda (Paris), Els Lommerse, Wilma Schaap, Ineke Theuwissen, Ania Oleksik (Amsterdam), Dr B. Zegels (Liège), W. Wandinger (Bad Pyrmont), Mrs L. Reany (Sheffield) and Emmanuel Ecosse (Le Kremlin Bicetre). This study was supported by an unconditional grant from Rhone-Poulence-Rorer to the European Foundation for Osteoporosis.

## References

1. Leidig G, Minne HW, Sauer P, et al. A study of complaints and their relation to vertebral destruction in patients with osteoporosis. Bone Miner 1990;8:217–29.
2. Chrischilles EA, Shireman T, Wallace R. Costs and health effects of osteoporotic fractures. Bone 1994;15:377–86.
3. O'Neill TW, Felsenberg D, Varlow J, Cooper C, Kanis JA, Silman AJ. The prevalence of vertebral deformity in European men and women: the European Vertebral Osteoporosis Study. J Bone Miner Res 1996;11:1010–8.
4. Ross PD, Davis JW, Epstein RS, Wasnich RD. Pain and disability associated with new vertebral fractures and other spinal conditions. J Clin Epidemiol 1996;47:231–9.
5. Kanis JA, Minne HW, Meunier PJ, Ziegler R, Allender E. Quality of life and vertebral osteoporosis. Osteoporos Int 1992;2:161–3.
6. Reginster JY, Compston JE on behalf of the Group for the Respect of Ethics and Excellence in Science. Recommendations for the registration of new chemical entities used in the prevention and treatment of osteoporosis. Calcif Tissue Int 1995;57:247–50.
7. Hunt S, McEwen J, McKenna SP. Measuring health status: a new tool for clinicians and epidemiologists. J R Coll Gen Pract 1985;35:185–8.
8. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The sickness impact profile: development and final revision of a health status measure. Med Care 1981;19:787–805.
9. Brazier JE, Harper R, Jones NMB, et al. Validating the SF-36 health survey questionnaire: new outcome measures for primary care. BMJ 1992;305:160–4.
10. Fitzpatrick R, Fletcher A, Gore S, Jones D, Spiegelhalter D, Cox D. Quality of life measures in health care I. Applications and issues in assessment. BMJ 1992;305:1074–7.
11. Fletcher A, Gore S, Jones D, Fitzpatrick R, Spiegelhalter D, Cox D. Quality of life measures in health care. II. Design, analysis and interpretation. BMJ 1992;305:1145–8.
12. Lips P, Cooper C, Agnusdei D, Caulin F, Egger P, Johnell O, et al. Quality of life as outcome in the treatment of osteoporosis; the development of a questionnaire for quality of life by the European Foundation for Osteoporosis. Osteoporos Int 1997;7:36–8.
13. Dequeker J, Ranstam J, Valsson J, Sigurgevisson B, Allander E. The Meditteranean Osteoporosis (MEDOS) Study questionnaire. Clin Rheumatol 1991;10:54–72.
14. O'Neill TW and the EVOS group. Reproducibility of the questionnaire on risk factors for osteoporosis in a multicentre prevalence survey: the European Vertebral Osteoporosis Study. Int J Epidemiol 1994;23:559–65.
15. Bradley C. Translation of questionnaires for use in different languages and cultures. In: Bradley C, editors. Handbook of psychology and diabetes: a guide to psychological measurement in diabetes research and practice. Chur, Switzerland: Harwood Academic, 1994.
16. McCloskey EV, Spector TD, Eyres KS, Fern ED, O'Rourke N, Vasikaran S, et al. The assessment of vertebral deformity: a method for use in population studies and clinical trials. Osteoporos Int 1993;3:138-47.
17. Ware JE, Harris WJ, Gandek B, Rogers BW, Reese PR. MAP-R for Windows: Multitrait/multi-item Analysis Program-Revised. Boston Health Assessment Lab, 1997.
18. Stewart AL, Ware JE. Measuring functioning and well-being: the Medical Outcomes Study approach. Durham, NC: Duke University Press, 1992.
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1997;33:159–174.
20. Medical Outcomes Trust. How to score the SF-36 Health Survey. Boston; MA: The Medical Outcomes Trust, 1993:1–25.
21. Ware JE, Kosinski M, Keller SD. SF-36 physical and mental health summary scales: a user's manual. Boston, MA: The Health Institute, 1994.
22. Hanley JA, McNeil BJ. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. Radiology 1982;143:29–36.
23. Todd C, Bradley C. The design and development of psychological scales. In: Bradley C, editor. Handbook of psychology and diabetes: a guide to psychological measurement in diabetes research and practice. Chur, Switzerland: Harwood Academic, 1994.

## Appendix. European Foundation for Osteoporosis: quality of life questionnaire: QUALEFFO-41 (10 December 1997)



Copies of this questionnaire and authorized translations can be obtained after signing a user's agreement.

### A Pain

The five questions in this section regard the situation in the last week.

1(A1) How often have you had back pain in the last week?
- ☐ never
- ☐ 1 day per week or less
- ☐ 2–3 days per week
- ☐ 4–6 days per week
- ☐ every day

2(A2) If you have had back pain, for how long did you have back pain in the daytime?
- ☐ never
- ☐ 1–2 hours
- ☐ 3–5 hours
- ☐ 6–10 hours
- ☐ all day

3(A3) How severe is your back pain at its worst?
- ☐ no back pain
- ☐ mild
- ☐ moderate
- ☐ severe
- ☐ unbearable

4(A4) How is your back pain at other times?
- ☐ no back pain
- ☐ mild
- ☐ moderate
- ☐ severe
- ☐ unbearable

5(A5) Has the back pain disturbed your sleep in the last week?
- ☐ less than once per week
- ☐ once a week
- ☐ twice a week
- ☐ every other night
- ☐ every night

### B Physical function: Activities of daily living

The next 4 questions regard the situation at present.

6(B10) Do you have problems with dressing?
- ☐ no difficulty
- ☐ a little difficulty
- ☐ moderate difficulty
- ☐ may need some help
- ☐ impossible without help

7(B11) Do you have problems with taking a bath or shower?
- ☐ no difficulty
- ☐ a little difficulty
- ☐ moderate difficulty
- ☐ may need some help
- ☐ impossible without help

8(B12) Do you have problems with getting to or operating a toilet?
- ☐ no difficulty
- ☐ a little difficulty
- ☐ moderate difficulty
- ☐ may need some help
- ☐ impossible without help

9(B13) How well do you sleep?
- ☐ sleep undisturbed
- ☐ wake up sometimes
- ☐ wake up often
- ☐ sometimes I lie awake for hours
- ☐ sometimes I have a sleepless night

### C Physical function: Jobs around the house

The next 5 questions are concerned with the *present* situation. If someone else does these things in your house, please answer as though you were responsible for them.

10(C14) Can you do the cleaning?
- ☐ without difficulty
- ☐ with a little difficulty
- ☐ with moderate difficulty
- ☐ with great difficulty
- ☐ impossible

11(C15) Can you prepare meals?
- ☐ without difficulty
- ☐ with a little difficulty
- ☐ with moderate difficulty
- ☐ with great difficulty
- ☐ impossible

12(C16) Can you wash the dishes?
- ☐ without difficulty
- ☐ with a little difficulty
- ☐ with moderate difficulty
- ☐ with great difficulty
- ☐ impossible

13(C17) Can you do your day-to-day shopping?
- ☐ without difficulty
- ☐ with a little difficulty
- ☐ with moderate difficulty
- ☐ with great difficulty
- ☐ impossible

14(C18) Can you lift a heavy object of 20 lb (e.g. a crate of 12 bottles of milk, or a 1-year-old child) and carry it for at least 10 yards?
- ☐ without difficulty
- ☐ with a little difficulty

☐ with moderate difficulty
☐ with great difficulty
☐ impossible

**D Physical function: Mobility**

The next 8 questions also regard the present situation.

15(D19) Can you get up from a chair?
☐ without difficulty
☐ with a little difficulty
☐ with moderate difficulty
☐ with great difficulty
☐ only with help

16(D20) Can you bend down?
☐ easily
☐ fairly easily
☐ moderately
☐ very little
☐ impossible

17(D21) Can you kneel down?
☐ easily
☐ fairly easily
☐ moderately
☐ very little
☐ impossible

18(D22) Can you climb stairs to the next floor of a house?
☐ without difficulty
☐ with a little difficulty
☐ with at least one rest
☐ with help only
☐ impossible

19(D23) Can you walk 100 yards?
☐ fast without stopping
☐ slowly without stopping
☐ slowly with at least one stop
☐ only with help
☐ impossible

20(D24) How often have you been outside in the last week?
☐ every day
☐ 5–6 days/week
☐ 3–4 days/week
☐ 1–2 days/week
☐ less than once/week

21(D25) Can you use public transport?
☐ without difficulty
☐ with a little difficulty
☐ with moderate difficulty
☐ with great difficulty
☐ only with help

22(D26) 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)?
☐ not at all
☐ a little
☐ moderately
☐ quite a bit
☐ very much

**E Social function**

23(E28) Do you play any sport now?*
☐ yes
☐ yes with restrictions
☐ not at all

24(E29) Can you do your gardening?
☐ yes
☐ yes with restrictions
☐ not at all
☐ not applicable

25(E31) Do you perform any hobby now?*
☐ yes
☐ yes with restrictions
☐ not at all

26(E32) Can you visit a cinema, theatre, etc.?
☐ yes
☐ yes with restrictions
☐ not at all
☐ no cinema or theatre within a reasonable distance

27(E33) How often did you visit friends or relatives during the last 3 months?
☐ once a week or more
☐ once or twice a month
☐ less than once a month
☐ never

28(E34) How often did you participate in social activities (clubs, social gatherings, church activities, charity, etc.) during the last 3 months?
☐ once a week or more
☐ once or twice a month
☐ less than once a month
☐ never

29(E35) Does your back pain or disability interfere with intimacy (including sexual activity)?
☐ not at all
☐ a little
☐ moderately
☐ severely
☐ not applicable

**F General health perception**

30(F36) For your age, in general, would you say your health is:
☐ excellent
☐ good
☐ satisfactory
☐ fair
☐ poor

31(F38) How would you rate your overall quality of life during the last week?
☐ excellent
☐ good

- ☐ satisfactory
- ☐ fair
- ☐ poor

32(F39) How would you rate your overall quality of life compared with 10 years ago?
- ☐ much better now
- ☐ slightly better now
- ☐ unchanged
- ☐ slightly worse now
- ☐ much worse now

## G Mental function

The next 9 questions regard the situation in the last week.

33(G40) Do you tend to feel tired?
- ☐ in the morning
- ☐ in the afternoon
- ☐ only in the evening
- ☐ after strenuous activity
- ☐ almost never

34(G41) Do you feel downhearted?
- ☐ almost every day
- ☐ 3–5 days a week
- ☐ 1 or 2 days a week
- ☐ once in a while
- ☐ almost never

35(G42) Do you feel lonely?
- ☐ almost every day
- ☐ 3–5 days a week
- ☐ 1 or 2 days a week
- ☐ once in a while
- ☐ almost never

36(G43) Do you feel full of energy?
- ☐ almost every day
- ☐ 3–5 days a week
- ☐ 1 or 2 days a week
- ☐ once in a while
- ☐ almost never

37(G44) Are you hopeful about your future?
- ☐ never
- ☐ rarely
- ☐ sometimes
- ☐ quite often
- ☐ always

38(G45) Do you get upset over little things?
- ☐ never
- ☐ rarely
- ☐ sometimes
- ☐ quite often
- ☐ always

39(G46) Do you find it easy to make contact with people?
- ☐ never
- ☐ rarely
- ☐ sometimes
- ☐ quite often
- ☐ always

40(G47) Are you in good spirits most of the day?
- ☐ never
- ☐ rarely
- ☐ sometimes
- ☐ quite often
- ☐ always

41(G48) Are you afraid of becoming totally dependent?
- ☐ never
- ☐ rarely
- ☐ sometimes
- ☐ quite often
- ☐ always

* In the original version questions E28 and E31 were: Do you play your sport now? Do you perform your hobby now?

*Received for publication 4 August 1998*
*Accepted in revised form 28 December 1998*