

Osteoporosis Assessment Questionnaire-Physical Function (OPAQ-PF): a psychometrically validated osteoporosis-targeted patient reported outcome measure of daily activities of physical function

A. Nixon · C. Kerr · H. Doll · A. N. Naegeli ·
S. L. Shingler · K. Breheny · R. Burge ·
D. T. Gold · S. Silverman

Received: 14 October 2013 / Accepted: 14 March 2014 / Published online: 16 April 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract

Summary The purpose of this study was to evaluate the measurement properties of the Osteoporosis Assessment Questionnaire-Physical Functioning (OPAQ-PF). Based on this study, the OPAQ-PF has confirmed unidimensionality and acceptable reliability, construct validity, and sensitivity to change in a recent fracture/no recent fracture osteoporosis sample.

Methods Dimensionality was established through exploratory and confirmatory factor analysis. Patients completed three patient reported outcome (PRO) measures and four performance-based measures (PBMs) at baseline to enable an evaluation of construct validity. Patients without a recent fracture completed the OPAQ-PF 2 weeks after baseline to enable an evaluation of test–retest reliability. Ability to detect change and interpretation of change were investigated following completion of the OPAQ-PF 12 and 24 weeks postbaseline by patients with a recent fracture.

Results A prospective psychometric validation study in 144 postmenopausal women, with moderate to severe osteoporosis, 37 of whom had experienced a recent fragility fracture (<6 weeks). Unidimensionality was established for the OPAQ-PF by factor analysis. The OPAQ-PF had good internal consistency ($\alpha=0.974$) and test–retest reliability (mean intraclass correlation coefficient (ICC) 0.993). The OPAQ-PF differentiated between patients with/without recent fracture, and by severity of osteoarthritis; it correlated strongly with hypothesized-related scales and PBMs ($r>0.3$, $p<0.001$). Ability to detect change was established with high correlations between changes in OPAQ-PF score and changes in global concept scores in recent fracture patients ($r\geq 0.6$, 24-week change). Effect size of change on OPAQ-PF score increased by level of global change ($p<0.001$). Anchor-based methods identified an OPAQ-PF change of 10 at an individual patient level and 20 at a group level as meaningful to patients.

Conclusions The OPAQ-PF has confirmed unidimensionality and acceptable reliability, construct validity, and sensitivity to change in a recent fracture/no recent fracture osteoporosis sample.

A. Nixon · C. Kerr · H. Doll · S. L. Shingler · K. Breheny
Oxford Outcomes, an ICON Plc. Company, Oxford, UK

A. N. Naegeli (✉) · R. Burge
Eli Lilly and Company, Lilly Corporate Center Drop Code 1730,
Indianapolis, IN 46285, USA
e-mail: naegelian@lilly.com

D. T. Gold
Departments of Psychiatry and Behavioral Sciences, Sociology, and
Psychology and Neuroscience, Duke University Medical Center,
Durham, NC, USA

S. Silverman
Cedars-Sinai Medical Center, Los Angeles, CA, USA

Keywords OPAQ-PF · Osteoporosis · Outcomes
assessment · Patient reported outcomes · Physical function ·
Psychometrics

Introduction

Commonly identified consequences associated with osteoporotic-related fractures are impaired physical functioning, disability, depression, social isolation, pain, and loss of independence. These impacts of osteoporosis on a patient's

life are often measured in clinical trials using patient-reported outcome (PRO) instruments assessing overall health-related quality of life (HRQL) [1–3]. However, specific aspects of HRQL need to be well measured in the development of new health technologies as a means to demonstrate and meet patient expectations to maintain and live an active, healthy life with the ability to independently perform activities of daily living, even after experiencing a fracture event [4].

New treatments in development for osteoporosis are subject to regulatory review of demonstrated effectiveness and safety, evaluated using measures and methods that conform to published regulatory guidance. The US Food and Drug Administration (FDA) requires evidence of a specific treatment benefit, defined as a favorable effect on a meaningful aspect of how a patient feels, functions, or survives [5]. Meaningful capture of how a patient feels (a patient's physical sensation or perceived mental state related to health within typical "daily" life, e.g., pain, low mood) or functions (a patient's ability to perform an activity that is a meaningful part of a typical "daily" life) relies on direct report from the patient, known as a PRO [5].

A number of measures are available to assess health-related outcomes in osteoporosis that are reported from the patient perspective. Generic measures most commonly used to evaluate patient-reported aspects of HRQL include the Short Form 36 (SF-36 and the EQ-5D [6–12]. Disease-targeted measures have also been developed and include, but not limited to, the Osteoporosis Assessment Questionnaire (OPAQ) [13], the Osteoporosis Quality of Life Questionnaire (OQLQ) [14, 15], the Osteoporosis Functional Disability Questionnaire (OFDQ) [16], the Osteoporosis-Targeted Quality of Life Questionnaire (OPTQOL), the Quality of Life Questionnaire of the European Foundation for Osteoporosis (QUALEFFO) [17, 18], and the Questionnaire Quality of Life in Osteoporosis (QUALIOST) [19].

The FDA's definition of what constitutes PRO evidence of specific treatment benefit requires instruments with a conceptual focus on feelings or function suggests measurement of proximal concepts (e.g., symptoms) rather than distal concepts (e.g., HRQL) to be most appropriate. The Osteoporosis Assessment Questionnaire-Physical Functioning (OPAQ-PF) has been developed specifically to meet the requirements of evaluating osteoporosis treatment effectiveness [20]. This new 15-item measure captures impact of osteoporosis on patients ability to perform daily activities of physical function and was previously adapted from the OPAQ v 2.0 [13] through a process of item reduction based on item response theory, input from key thought leaders on quality of life issues and measurement in osteoporosis, concept elicitation, and cognitive debriefing interviews conducted with osteoporotic patients with and without experience of fracture [20]. The aim of this current study was to evaluate the measurement properties of the OPAQ-PF following published psychometric standards [21–23].

Methods

Design

This prospective, clinical site-based psychometric validation study was designed to confirm the conceptual framework of the OPAQ-PF (using factor analysis) and evaluate the following measurement properties of the instrument: reliability (internal consistency, item homogeneity, test–retest); construct validity (known groups and convergent validity); ability to detect change; and interpretation of change.

Patients completed the OPAQ-PF and other measures at baseline and then either 2 weeks later (no recent fracture group) or at 12 and 24 weeks (recent fracture group). All psychometric properties were evaluated on baseline data apart from test–retest reliability which was evaluated using data collected at 2 weeks, ability to detect change, and interpretation of change which used data collected at 12 and 24 weeks in the recent fracture group.

The study protocol and materials were developed with input from two expert clinicians (DTG and SS). Institutional review board (IRB) approval was obtained for the study (Protocol OXO2550; original approval from Independent Investigational Review Board, 21 October 2011).

Sample

The study aimed to recruit approximately 150 participants in total: 100 osteoporosis patients without recent fracture, and 50 recent osteoporotic fracture patients. This would enable 80 % power to be achieved to detect differences (at two-sided $p < 0.05$) to a moderate effect of around 0.5. Participants were recruited through ten clinical sites distributed across eight US states; most were specialist orthopedic sites, generally based within physician-led practices which varied in size.

Eligibility criteria were developed in consultation with the two key thought leaders. All participants were:

- (1) Postmenopausal women (aged ≥ 50 years of age, reporting no menstruation in previous 12 months).
- (2) Diagnosed with moderate to severe osteoporosis (T-score of ≤ -2.5) or a history of a nontraumatic or fragility clinical fracture (symptomatic and diagnosed by X-ray).

Participants were classified into one of two subgroups: those with or without recent osteoporotic fracture. Participants without fracture were expected to experience stability in their ability to perform daily activities of physical functioning for evaluation of test–retest reliability of the OPAQ-PF. Participants with recent osteoporotic fracture (within 6 weeks prior to baseline, fracture date taken as the date of diagnostic X-ray) were expected to experience change in their ability to perform daily activities of physical

functioning over 12–24 weeks for evaluation of the OPAQ-PF's ability to detect change.

Participants without a recent fracture had to be able to ambulate independently (with or without a walking aid). Participants with a recent fracture had to have been able to ambulate independently prior to the fracture and to be able to ambulate with or without a walking aid at their baseline visit. Patients with Parkinson's disease and/or other neuromuscular disorders were excluded. Patients who had received procedures potentially impacting physical functioning were also excluded (vertebroplasty and/or kyphoplasty).

Recruitment of the recent fracture group aimed for a balance of upper body (e.g., wrist, shoulder, and upper arm), lower body (e.g., hip, tibia, ankle, and foot), and trunk fracture (e.g., rib, and vertebral). Types of fracture considered difficult to classify as fragility, versus traumatic, not to result in sufficient physical impairment for OPAQ-PF validation purposes, or to have unpredictable healing rates were excluded. Examples of excluded fracture types were toe, finger, clavicle, pelvis, or face fracture.

Measures

The OPAQ-PF was previously adapted from the OPAQ v2.0 and was designed to evaluate participant's ability to perform their daily activities of physical function during the past 7 days covering mobility (five items), physical positions (six items) and transfers (four items) [20]. Items are rated on a six-point Likert response scale ranging from no difficulty to completely avoided doing this. All 15 items are summed and normalized to a 0–100 scale to provide a total score, where 0 indicates the worst health status and 100 no difficulties.

Three additional PROs were included to evaluate the construct validity of the OPAQ-PF: The Western Ontario and McMaster Universities Arthritis Index (WOMAC), the SF-36, and QUALEFFO-41. The WOMAC [24] was designed to assess pain, disability, and joint stiffness in osteoarthritis during the last 48 hours. The SF-36 [25] was designed to measure health status in a wide range of conditions as well as in the general population. Eight SF-36 domain scores provide a health profile (bodily pain, general health perceptions, mental health, physical functioning, role limitations due to emotional problems, role limitations due to physical problems, social functioning, and vitality), as well as two component summary scores for physical (PCS) and mental health (MCS). The QUALEFFO-41 [26, 27] was designed to measure HRQL in patients with vertebral fractures and produces five domain scores (pain, physical function, social function, general health perception, mental function) and a total score.

Three global concept items were developed to evaluate the ability of the OPAQ-PF to detect change and to evaluate interpretation of change [28]. Global concept items were self-completed by participants to reflect overall difficulty in

the last 7 days with mobility, physical positions, and transfer activities due to osteoporosis. Participants rated difficulty on a five-point scale ranging from “no difficulty” (0) to “severe difficulty” (5).

Four performance-based measures (PBMs) were included in the study to help evaluate the construct validity of the OPAQ-PF: ten-meter walk test (10MWT); timed up and go (TUG); functional reach (FR), and timed unsupported steady stand (TUSS). 10MWT is a measure of gait speed; the individual walks for 10 m without assistance (other than with their usual walking aid) at their usual pace. TUG is a test of balance and mobility; the participant is asked to stand up from a chair and walk a distance of 3 m at their normal pace, turn around, walk back to the chair and sit down again, using their usual walking aid. For 10MWT and TUG, longer times to complete tests reflect worse health status. FR is a test of balance impairment and change in balance performance: the maximal forward reach by the participant is measured, using a fixed base of support. Distance is recorded in centimeters, with a shorter reach reflecting a worse health status. TUSS is test of balance impairment associated with an increased risk for falling [29]. The participant is asked to stand holding onto a support (e.g., a table or chair), then put their hands by their sides and stand as long as they feel safe and steady (a maximum of 60 s), putting their hands back on the support if they feel unsteady, at which point timing stops. Shorter times reflect poorer balance and risk of falling. The PBMs were completed by participants under the supervision of site staff. Each test was repeated three times with the score being the average of the three attempts.

Procedures

Site staff received in-person training to ensure familiarity with protocol requirements, study procedures, and materials. Clinical site staff screened for eligibility using chart review and discussion with individual patients.

Baseline visit

All participants (with and without recent fracture) attended the clinical site for a baseline visit. Participants completed all PROs and PBMs and provided sociodemographic and other background information. Site staff completed a medical history form at the baseline visit recording details of bone mass density, previous fractures, current medication, and comorbidities.

Week 2 visit

Participants without a recent fracture completed the OPAQ-PF and global concept items 2 weeks (median 14 days, IQR 14 to 17 days) after baseline.

Weeks 12 and 24 visits

Participants with a recent fracture attended visits at 12-week (median 12.0 weeks, IQR 12.0 to 12.8 weeks) and 24-week (median 24.1 weeks, IQR 24.0 to 24.9 weeks) postbaseline and completed the OPAQ-PF and global concept items.

Statistical analysis

Scale structure and conceptual framework of the OPAQ-PF

To investigate whether the three aspects of daily activities of physical functioning measured by the OPAQ-PF are related, the scale structure was evaluated by both Exploratory and Confirmatory Factor Analysis (EFA and CFA, respectively), as CFA commonly requires a minimum sample size of 200 [30]. For the EFA, SAS PROC FACTOR was used with maximum likelihood extraction. The CFA was performed by fitting structural equation models using SAS PROC TCALIS, utilizing a range of measures of fit to compare two competing models: either a one general factor or a three-factor structure.

Distribution of scores

OPAQ-PF item and score variability (frequency and percentage of endorsement) was assessed to evaluate score distributions, floor, and ceiling effects.

Reliability

The internal consistency of the OPAQ-PF was evaluated with baseline data using Cronbach's alpha [31] for the total sample and each subgroup. Individual item scores were correlated with the OPAQ-PF total score further to assess the homogeneity of the items (item-total correlation). Test-retest reliability (intraclass correlation coefficient, ICC) of the OPAQ-PF score was assessed using baseline and week 2 follow-up data from patients with no recent fracture and reporting no change on each of three global concept items between baseline and week 2.

Construct validity (known groups)

Known-groups analysis was conducted on baseline data. It was hypothesized that patients who have had a recent fracture, had osteoarthritis, or had high WOMAC scores (0 indicates better WOMAC health status) would have poorer OPAQ-PF scores than comparator groups. In addition, patients in the no-recent-fracture group but who had experienced fractures at some point previously were expected to have poorer OPAQ-PF scores if fracture was experienced more recently than patients experiencing fractures longer ago (split at the median). Comparisons were also made by fracture location.

Significant group differences were tested using Mann-Whitney *U* tests or independent sample *t* tests and effect sizes.

Construct validity (convergent validity)

Analysis of convergent validity was conducted on baseline data with calculation of Spearman's correlation coefficients (r_s). The OPAQ-PF was expected to correlate positively with the SF-36 physical functioning, role functioning, bodily pain, vitality, and PCS domain scores. No relationships were expected between the OPAQ-PF and the remaining SF-36 domains. Negative correlations were hypothesized between the OPAQ-PF and (i) the WOMAC total and domain scores and (ii) the QUALEFFO-41 pain, physical function, and total scores, in line with their reversed scoring schemes. No relationships were expected between the OPAQ-PF and the social function, general health perception, and mental function QUALEFFO-41 domains. The 10MWT and the TUG were expected to correlate negatively with the OPAQ-PF, in line with the reversed scoring schemes. The FR and TUSS were both expected to correlate positively with the OPAQ-PF. The existence of observed relationships with a correlation coefficient ≥ 0.3 and greater than correlations, where no relationship was hypothesized, were taken to support the convergent validity of the OPAQ-PF. Observed relationships < 0.3 and smaller than correlations, where relationships were expected, were taken to support divergent validity.

Ability to detect change

The ability of the OPAQ-PF to detect change (at weeks 12 and 24) was assessed by relating the changes in OPAQ-PF to changes in the perceived global concept scores (mobility, physical positions, and transfers) in the recent fracture subgroup. First, the changes in OPAQ-PF scores and global concept scores at each follow-up point were calculated and Spearman's correlation coefficients used to assess their degree of association. Second, patients were then classified into change categories based on changes in the global concept items. Finally, the patterns of mean change, and the associated effect sizes, in OPAQ-PF score by global change category were then evaluated. Differences between patient change categories were tested using parametric and nonparametric analysis of variances with tests for linear trend.

Interpretation of change

The level of change in OPAQ-PF scores likely to represent meaningful change to patients at the group level was investigated using both distribution (one standard error of measurement (1SEM) [32]; MDC_{90}) and anchor-based approaches (mean change score in OPAQ-PF from participants reporting one unit improvement on a patient-reported global concept

item). Individual level change, receiver operator characteristic (ROC) curves were used to identify the OPAQ-PF change score (the best cut point), which best distinguishes patients who change to a minimal extent from those who do not. The area under the curve (AUC) indicates the overall usefulness of the prediction (the observed AUC is compared with the AUC of 0.5 expected from a “useless” test), and the best cut point is identified as the value which maximizes the values of sensitivity (proportion of true “positives” detected) and specificity (proportion of true “negatives” detected) [33, 34].

Results

In total, 144 osteoporosis participants were recruited into this study: 107 without recent fracture and 37 with recent fracture. Recruitment numbers varied between the sites; mean 14.4 (range 2–37).

Patient demographics and clinical characteristics

Homogeneity between the two study groups was confirmed for most demographic/medical history characteristics but with expected differences in terms of time since most recent fracture (more recent for recent fracture group), pain medication (more use for recent fracture group), and use of bisphosphonates (more use for recent fracture group; Table 1). The majority of the sample reported taking some sort of pharmaceutical agent for their OP ($n=117$, 81.3 %). For the combined sample, mean age at diagnosis was 62.8 years (SD 7.08, range 43.4 to 84.0 years). Fracture location for the recent fracture group was well distributed: upper body $n=15$ (40.5 %), lower body $n=10$ (27.0 %), trunk $n=12$ (32.4 %), and the most commonly reported comorbid conditions were hypertension ($n=68$, 47.2 %), high cholesterol ($n=55$, 38.2 %), osteoarthritis ($n=53$, 36.8 %), depression/anxiety ($n=39$, 27.1 %), and eye conditions ($n=36$, 25.0 %).

Scale structure and conceptual framework of the OPAQ-PF

EFA maximum likelihood extraction identified one factor with an eigenvalue >1 , with 75 % of the variability being explained by this one factor on which all items loaded strongly (>0.74). CFA found a significant deterioration in model fit when fitting three (uncorrelated) factors compared with the one factor model (overall χ^2 fit statistics = 644.74 vs. 764.00, $p < 0.0001$; χ^2/df ratios 7.16 vs. 8.49).

Distribution of scores

Mean OPAQ-PF scores were 75.7 (SD 24.9, range 5.33–100), with 22 % achieving the maximum score (Table 2). Participants with recent fracture reported greater impairment

on the OPAQ-PF than those with no recent fracture (mean score 57.0 vs. 82.2, $p < 0.001$), with 2.7 % achieving the maximum score compared with 29.0 % of the no recent fracture group. Greater variability in scores was found in the recent fracture group. A similar pattern was identified at the item level, with greater evidence of ceiling effects in the no recent fracture group.

Reliability

Internal consistency reliability

Good internal consistency was demonstrated with a Cronbach's alpha value of 0.974 for the entire sample; similar values were reported for the no recent fracture group ($\alpha=0.974$) and the recent fracture group ($\alpha=0.961$) (Table 3).

Item homogeneity

Item-total correlations ranged from 0.733 to 0.923 (Table 3).

Test-retest reliability

Good test-retest reliability was demonstrated for the OPAQ-PF score with a mean ICC of 0.933 (95 % CI 0.87 to 0.97) among the 37 patients reporting no change on each of the global concept scales.

Construct validity

Known groups validity

OPAQ-PF was able to discriminate well between the no recent fracture and recent fracture groups (mean 82.19 SD 21.08 vs. 57.05 SD 26.09, $p < 0.001$), and between patients defined by severity of osteoarthritis (WOMAC ≥ 40 vs. < 40 : mean 90.2 SD 12.6 vs. 53.3 SD 21.3, respectively, $p < 0.001$). The OPAQ-PF was notable to discriminate between participants in terms of whether or not they had osteoarthritis, time since last fracture for fractures experienced >6 weeks from baseline, and for the recent fracture participants, between fracture location (upper body, lower body, and trunk).

Convergent validity

As hypothesized, the OPAQ-PF was positively correlated with the SF-36 physical functioning, role functioning, bodily pain, vitality, and PCS domain scores. While the OPAQ-PF also correlated with the remaining SF-36 domains where no relationship had been expected, correlations were consistently higher with the physical domains and the PCS ($r_s=0.69$) than with the mental domains and the MCS ($r_s=0.38$). The OPAQ-PF was negatively correlated with the WOMAC total score

Table 1 Participant characteristics

<i>n</i> (percent)	No recent fracture (<i>n</i> =107)	Recent fracture ^a (<i>n</i> =37)	Total (<i>n</i> =144)
Age, years, mean (SD) [min–max, years]	69.7 (8.70) [52.0–88.8]	69.4 (9.70) [51.3–87.9]	69.6 (8.93) [51.3–88.8]
<i>Education</i>			
High school or less	36 (33.7 %)	15 (40.5 %)	51 (35.5 %)
Some college	33 (30.8 %)	12 (32.4 %)	45 (31.3 %)
College degree	25 (23.4 %)	4 (10.8 %)	29 (20.1 %)
Graduate degree	13 (12.1 %)	6 (16.2 %)	19 (13.2 %)
<i>Ethnicity</i>			
Asian	2 (1.9 %)	0 (0 %)	2 (1.4 %)
Black	1 (0.9 %)	3 (8.1 %)	4 (2.8 %)
Hispanic	4 (3.7 %)	2 (5.4 %)	6 (4.2 %)
White or Caucasian	100 (93.5 %)	32 (86.5 %)	132 (91.7 %)
<i>Employment</i>			
Employed (part or full time)	22 (20.6 %)	6 (16.2 %)	28 (19.5 %)
Retired	64 (59.8 %)	21 (56.8 %)	85 (59.0 %)
Looking after home/family	11 (10.3 %)	2 (5.4 %)	13 (9.0 %)
Not working/not retired	10 (9.3 %)	8 (21.6 %)	18 (12.5 %)
Disease duration [min–max, years]	6.83 (4.80), <i>n</i> =94 [0.45–19.8]	8.02 (6.75), <i>n</i> =25 [0–30.6]	7.08 (5.26), <i>n</i> =119 [0–30.6]
Time (months) since most recent fracture [min–max, months]	123.9 (193.5), <i>n</i> =53 [1.97–703.8]	0.75 (0.41), <i>n</i> =37 [0–1.35]	73.3 (160.0), <i>n</i> =90 [0–703.8]
<i>Medication</i>			
Calcium	69 (64.5 %)	24 (64.9 %)	93 (64.6 %)
Vitamin D	73 (68.2 %)	24 (64.9 %)	97 (67.4 %)
Bisphosphonates	68 (63.6 %)	12 (32.4 %)	80 (55.6 %)
Other	29 (27.1 %)	9 (24.3 %)	38 (26.4 %)
Pain medications ^b	64 (59.8 %)	28 (75.7 %)	92 (63.9 %)
<i>Fracture location</i>			
Shoulder/humerus	–	4 (10.8 %)	
Radius/wrist	–	9 (24.4 %)	
Ulna and radius	–	2 (5.4 %)	
Ribs	–	6 (16.2 %)	
Vertebra	–	6 (16.2 %)	
Hip/femoral neck	–	3 (8.1 %)	
Ankle/malleolus	–	4 (10.8 %)	
Foot	–	3 (8.1 %)	

^a Recent fracture was defined as a nontraumatic or fragility clinical fracture (symptomatic and diagnosed by X-ray) in the 6 weeks prior to baseline. Fracture date taken as the date of diagnostic X-ray

^b Includes eight patients who reported taking pain medications their clinicians were unaware of and five patients who did not report taking medications prescribed by their clinicians

and all WOMAC domain scores. While large and negative correlations were observed as hypothesized between the OPAQ-PF and the QUALEFFO-41 pain, physical function, and total scores, large correlations were also observed with the social function, general health perception, and mental function QUALEFFO-41 domains. As hypothesized, the OPAQ-PF correlated negatively with the 10MWT and the TUG. All reported relationships were significant at $p < 0.001$ with correlations ≥ 0.3 .

Ability to detect change

The global concept scores indicated that a large proportion of recent fracture patients were either unchanged or improved at weeks 12 and 24. At 24 weeks, 34 % reported no change in mobility, 28 % in physical positions, and 52 % in transfers, while 59 % reported improvement in mobility, 59 % in physical positions, and 47 % in transfers.

Table 2 OPAQ-PF summary statistics

	Total (0–100) ^a
Overall (<i>n</i> =144)	
Mean (SD)	75.7 (24.9)
Median (IQR)	83.3 (58.0–98.7)
Min, max	5.33, 100
<i>n</i> (percent) max	32 (22.2 %)
No recent fracture (<i>n</i> =107)	
Mean (SD)	82.2 (21.1)
Median (IQR)	92.0 (69.3–100)
Min, max	28, 100
<i>n</i> (percent) max	31 (29.0 %)
Recent fracture (<i>n</i> =37)	
Mean (SD)	57.0 (26.1)
Median (IQR)	58.7 (38.7–78.7)
Min, max	5.33, 100
<i>n</i> (percent) max	1 (2.7 %)

^a Lower scores indicate greater impairment

Large correlations were observed between changes in OPAQ-PF score and changes in global concept scores. Strong, statistically significant associations were observed at 12 weeks, and even greater associations at 24 weeks: transfers $r_s=0.50$ at 12 weeks and $r_s=0.65$ at 24 weeks, mobility $r_s=0.46$ at 12 weeks and $r_s=0.60$ at 24 weeks, and physical positions $r_s=0.75$ at 12 weeks and $r_s=0.73$ at 24 weeks. The amount of change in OPAQ-PF scores and the associated effect sizes generally increased with perceived degree of change (improvement), observed associations were strong, and the tests for linear trend were statistically significant (all $p<0.01$).

Interpretation of change

Distribution-based approaches identified a value of 4.0 for interpreting a minimum change on the OPAQ-PF based on the SEM approach, and an MDC₉₀ of 9.6 (Table 3). This is the smallest value that could be used to indicate minimal change on an individual patient basis. The anchor-based approaches

Table 3 OPAQ-PF reliability: item-total correlations and internal consistency

OPAQ-PF item, item-total correlation ^a	Overall (<i>n</i> =144)	No fracture (<i>n</i> =107)	Fracture (<i>n</i> =37)
1. Walking to do your daily chores or errands (e.g., grocery shopping, taking out garbage, house work, going to post office, walking the dog)	0.923	0.927	0.863
2. Walking unaided so you can do your day to day activities	0.797	0.808	0.718
3. Carrying objects in order to perform your day to day activities (e.g., a bag of groceries, a bag of garbage)	0.870	0.843	0.829
4. Walking one block	0.860	0.867	0.747
5. Climbing one flight of stairs or steps	0.863	0.867	0.785
6. Bending or stooping to do your daily chores or errands (e.g., grocery shopping, taking out the garbage, housework, going to post office, walking the dog)	0.906	0.887	0.876
7. Lifting objects in order to perform your day to day activities (e.g., a bag of groceries, a bag of garbage)	0.877	0.868	0.815
8. Reaching overhead in order to perform your day to day activities	0.827	0.871	0.689
9. Picking things up from the floor	0.842	0.846	0.782
10. Standing as much as you needed in order to perform your day to day activities	0.863	0.877	0.807
11. Sitting as much as you needed in order to perform your day to day activities	0.733	0.753	0.716
12. Getting in or out of bed	0.761	0.830	0.580
13. Getting in or out of a chair	0.852	0.830	0.841
14. Getting on or off the toilet	0.857	0.839	0.817
15. Getting in or out of cars on your own	0.844	0.800	0.832
Cronbach's alpha ^b	0.974	0.974	0.961
SD	24.95	21.08	26.09
SEM ^c	4.02	3.40	4.21
MDC ₉₀ ^d	9.55	8.08	10.00

^a Items should have a significant correlation ≥ 0.30 [36]

^b A minimum value of 0.80 is a guideline for demonstrating internal consistency [35], and alpha scores in excess of 0.90 may indicate an overly homogenous scale, where items may be redundant due to excessive similarity [37]

^c Standard error of measurement, SEM=SD \times sqrt(1-Cronbach's alpha)

^d Minimal detectable change 90, MDC₉₀=1.68 \times SEM \times sqrt(2)

described above indicated that, for interpretation at a group level, the mean OPAQ-PF scores for patients reporting at least one point of change on the mobility, physical positions, and transfers global concept items were 19.0, 14.3, and 13.3, respectively (overall mean 15.5) at 12 weeks and 23.3, 19.2, and 23.6 (overall mean 22.0) at 24 weeks: overall mean over 12 and 24 weeks=18.8 (~20). The ROC analyses demonstrated that the OPAQ-PF change values that can be used to identify minimal change at an individual level on the mobility, physical positions, and transfers global concept items are 4.0, 10.0, and 12.7, respectively (overall mean 8.89) at 12 weeks and 9.3, 9.3, and 16.0, respectively (overall mean 11.6) at 24 weeks: mean of overall means at 12 and 24 weeks=10.25 (~10).

Discussion

This study sought to evaluate the psychometric measurement properties of the OPAQ-PF, a PRO instrument developed with patients with osteoporosis and designed to assess a patients' ability to perform daily activities of physical function for evaluating osteoporosis treatment effectiveness. The study design allowed for a comprehensive assessment of the psychometric properties of the OPAQ-PF in line with current FDA regulatory guidelines [5].

The conceptual focus of the OPAQ-PF is a unidimensional measure of daily activity of physical function; the EFA strongly indicated that the variability in the data was best explained by one underlying construct and that this single factor alone was sufficient to explain the majority of the variance in the data (75 % on principal component extraction). This was supported by the CFA, i.e., the OPAQ-PF is most appropriately scored in terms of a total score, and the OPAQ-PF conceptual framework is confirmed.

The OPAQ-PF had a good distribution of scores for patients with a recent fracture (within 6 months) with scores ranging across almost the entire spectrum of the 0–100 scale, and a median score of 58.7. The OPAQ-PF therefore has good capacity to measure change (improvement or decline) in daily activities of physical function of osteoporosis patients who have experienced a recent fracture. However, the OPAQ-PF did demonstrate a possible ceiling effect for those with no recent fracture, with 29.0 % scoring the maximum, with a narrower scoring range (28–100). The OPAQ-PF may be most suitable for measuring the maintenance of the ability to perform daily activities of physical function.

The OPAQ-PF demonstrates excellent internal consistency ($\alpha=0.974$). A minimum value of 0.80 is a guideline for demonstrating internal consistency [35]. Alpha scores in excess of 0.90 may indicate an overly homogenous scale, where items may be redundant due to excessive similarity [36]. Items were not deleted as it was felt that they represent clinical

relevant concepts that should be measured as part of the overall construct. The OPAQ-PF has good stability over a 2-week period among no recent fracture participants who reported experiencing no change in the previous 7 days, with an ICC of 0.93. Guidelines suggest an ICC > 0.70 [35] or 0.90 [22] as minimum requirements, both of which are exceeded by the OPAQ-PF.

In terms of construct validity, the OPAQ-PF was able to differentiate between patients with no recent fracture (>6 weeks) and patients with a recent fracture (<6 weeks) and severity of osteoarthritis. The inability of the OPAQ-PF to discriminate between other identified known groups does not necessarily reflect poor construct validity. It could be that the location of fracture and time since fracture > 6 weeks (the no fracture group on average having their most recent fracture 4 years ago) has no bearing on physical activities of daily living. The small sample sizes for each identified known group may also explain why no significant differences were observed.

Convergent validity was evaluated by exploring the relationship of the OPAQ-PF with three PRO instruments: the WOMAC, SF-36, QUALEFFO-41, and PBMs. As hypothesized, scores on the OPAQ-PF were most strongly correlated with physical dimensions of the WOMAC, the physical dimensions of the SF-36, and the physical function domain of the QUALEFFO-41, with lower correlations being observed with the mental dimensions of the SF-36. The size of the correlations observed in the relationships between the OPAQ-PF and the PBMs (specifically, the 10MWT and TUG) indicate a good level of convergent validity for the OPAQ-PF.

Results demonstrated that where an osteoporosis patient has experienced change in their health status related to ability to perform daily activities of physical function (specifically mobility, transfers, and physical positions), the OPAQ-PF captures that change. In patients with a recent fracture, high correlations were observed between changes in OPAQ-PF score and changes in patient-reported global concept scores. Strong statistically significant associations were observed at 12 weeks postbaseline with even greater associations at the 24-week follow-up. The reported effect size statistics at 12 and 24 weeks clearly demonstrate that the greater the degree of improvement reflected by the change categories, the greater and more positive the effect sizes for change in the OPAQ-PF score.

Analysis was undertaken to enable OPAQ-PF change scores to be interpreted using two established approaches: distribution and anchor-based. The 1SEM distribution-based approach identified an OPAQ-PF change score of 4.0 (1SEM) as being the minimum needed for the change to be meaningful, with a 90 % confidence interval (MDC_{90}) of 9.6. Distribution-based approaches are useful supporting information for interpretation of PRO change scores, as the change

scores need to be at least as large as a distribution-based value to rule out the possibility of change in score by chance. However, values generated by distribution-based approaches do not necessarily indicate clinically relevant change or change of a magnitude that is meaningful to patients. Greater credibility is placed in the anchor-based approaches [28], which in this analysis compared a one-point change on global concept items to corresponding score changes on the OPAQ-PF. The overall mean of the mean changes in OPAQ-PF score across the three global concept items suggests that, at a group level, a change in OPAQ-PF score of ~20 points needs to be achieved before the change can be considered meaningful to patients. At an individual patient level, a change in OPAQ-PF score of ten points appears meaningful.

This study had several limitations. Recruitment of the recent fracture subgroup fell short of the target sample size of 50. The power of the study was therefore less than planned. While this may partially account for the inability of the OPAQ-PF to differentiate between certain groups, e.g., with/without osteoarthritis and fracture location, the observed differences between these groups were small and other statistically significant group differences were identified. Thus while the demonstration of the ability of the OPAQ-PF to detect change was limited to the relatively small “recent fracture” group, strong and highly statistically significant associations were observed between changes in the OPAQ-PF and changes in perceived degree of improvement, indicating meaningful results despite the small sample. Another limitation is that study design for the change analysis focused on improvements. Further work is needed to verify the ability of the OPAQ-PF to capture decline in osteoporosis patients, as well as to gather further evidence for the ability to capture improvements on a larger sample. In addition, for interpretation of change, the change in global concept scores was framed in terms of hypothesized domains which were not verified in the factor analysis. Nevertheless, effect sizes of change within categories of perceived change (with approximately one third of patients reporting a minimal degree of improvement) showed clear highly statistically significant group heterogeneity and linear trends; further work is planned to confirm interpretation of change scores based on a global concept item for daily activities of physical function which will be more closely aligned to the unidimensional structure confirmed for the OPAQ-PF. The study may not be generalizable to all osteoporosis patients, and so the validity of the OPAQ-PF beyond the patient sample in this study is uncertain. Missing from the sample were representative males, premenopausal female, and glucocorticoid-induced osteoporosis patients. Further validation work would be required in order to confirm that the psychometric measurement properties reported in this study are maintained in these other osteoporosis patient populations. Finally, the impact of comorbidities on OPAQ-PF scores is not fully understood and warrants further

investigation. Many participants reported comorbid conditions, which is a common scenario in osteoporosis. It is known that for some comorbid conditions it can be very difficult for the patient to attribute the consequences of these in their OPAQ-PF responses to their osteoporosis [20]. To overcome this, when using the OPAQ-PF to evaluate treatment benefit, analyses could be adjusted for presence of musculoskeletal or other comorbidities (based on clinical examination or self-report).

Conclusions

The OPAQ-PF is a new PRO instrument uniquely tailored to the assessment of the daily activities of physical function in osteoporosis patients. This study demonstrated the strong psychometric measurement properties of the OPAQ-PF by providing evidence to support the conceptual framework, reliability, validity, and sensitivity to change in a combined recent fracture/no recent fracture osteoporosis sample through a purpose-designed psychometric validation study which gathered data from patients recruited through clinical sites in the USA. A minimum change score of 10 at an individual patient level, and 20 at a group level, was identified as potentially representing a meaningful change from a patient perspective. The OPAQ-PF has been developed to meet FDA regulatory requirements for PRO instruments intended to be used in a phase 3 registration trial to support a label claim. It is a suitable PRO instrument for capturing change in daily activities of physical function of osteoporosis patients who have experienced a recent fracture and maintenance in daily activities of physical function of osteoporosis patients who have not experienced a recent fracture.

Conflicts of interest Drs. Naegeli and Burge are fulltime employees and hold stocks and/or stock options with Eli Lilly and Company. Eli Lilly and Company is investigating new pharmaceutical therapies for the treatment of osteoporosis. Drs. Nixon, Kerr, Doll, Shingler, Breheny, Gold, and Silverman have no disclosures.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any non-commercial use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Johnell O (1996) Advances in osteoporosis: better identification of risk factors can reduce morbidity and mortality. *J Intern Med* 239:299–304
2. Silverman S (2005) Quality-of-life issues in osteoporosis. *Curr Rheumatol Rep* 7:39–45
3. Gold DT, Solimeo S (2006) Osteoporosis and depression: a historical perspective. *Curr Osteoporos Rep* 4:134–139
4. Wylie CD (2010) Setting a standard for a “silent” disease: defining osteoporosis in the 1980s and 1990s. *Stud Hist Phil Biomed Sci* 41: 376–385

5. Food and Drug Administration (FDA) (2009) Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims
6. Tanaka K, Fujii A, Kuwabara A (2012) SF-36 and EQ-5D in the evaluation of QOL in the osteoporotic patients as the generic questionnaires. *Clin Calcium* 22:259–262
7. Dallanezi G, Nahas EA, Freire BF, Nahas-Neto J, Corrente JE, Mazeto GM (2011) Quality of life of women with low bone mass in postmenopause. *Rev Bras Ginecol Obstet* 33:133–138
8. Wildner M, Sangha O, Clark DE, Clark DE, Doring A, Manstetten A (2002) Independent living after fractures in the elderly. *Osteoporos Int* 13:579–585
9. Hallberg I, Rosenqvist AM, Kartous L, Lofman O, Wahlstrom O, Toss G (2004) Health-related quality of life after osteoporotic fractures. *Osteoporos Int* 15:834–841
10. Voight K, Tache S, Hofer M (2012) Health related quality of life in male patients with osteoporosis: results of a cross sectional study. *Aging Male* 15:220–226
11. Roux C, Wyman A, Hooven FH, Gehlbach SH, Adachi JD, Chapurlat RD, Compston JE, Cooper C, Diez-Perez A, Greenspan SL, Lacroix AZ, Netelenbos JC, Pfeilschifter J, Rossini M, Saag KG, Sambrook PN, Silverman S, Siris ES, Watts NB, Boonen S (2012) Burden of non-hip, non-vertebral fractures on quality of life in postmenopausal women: the global longitudinal study of osteoporosis in women (GLOW). *Osteoporos Int* 23:2863–2871
12. Yoh K, Hamaya E, Urushihara H, Iikuni N, Yamamoto T, Taketsuna M, Miyauchi A, Sowa H, Tanaka K (2012) Quality of life in raloxifene-treated Japanese women with postmenopausal osteoporosis: a prospective, postmarketing observational study. *Curr Med Res Opin* 28:1757–1766
13. Silverman S (2000) The osteoporosis assessment questionnaire (OPAQ): A reliable and valid disease-targeted measure of health-related quality of life (HRQOL) in osteoporosis. *Qual Life Res* 9:767–774
14. Cook DJ, Guyatt GH, Adachi JD, Clifton J, Griffith LE, Epstein RS, Juniper EF (1993) Quality of life issues in women with vertebral fractures due to osteoporosis. *Arthritis Rheum* 36:750–756
15. Osteoporosis quality of life study group (1997) Measuring quality of life in women with osteoporosis. *Osteoporos Int* 7:478–487
16. Helmes E, Hodsman A, Lazowski D, Bhardwaj A, Crilly R, Nichol P, Drost D, Vanderburgh L, Pederson L (1995) A questionnaire to evaluate disability in osteoporotic patients with vertebral compression fractures. *J Gerontol A Biol Sci Med Sci* 50:M91–M98
17. Lips P, Cooper C, Agnusdei D, Caulin F, Egger P, Johnell O, Kanis JA, Liberman U, Minne H, Reeve J, Reginster JY, de Verjoul MC, Wiklund I (1997) 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* 7:36–38
18. Lips P, Cooper C, Agnusdei D, Caulin F, Egger P, Johnell O, Kanis JA, Kellingray S, Leplege A, Liberman UA, McCloskey E, Minne H, Reeve J, Reginster JY, Scholz M, Todd C, de Verjoul MC, Wiklund I (1999) Quality of life in patients with vertebral fractures: validation of the quality of life questionnaire of the European Foundation for Osteoporosis (QUALEFFO). Working party for quality of life of the European Foundation for Osteoporosis. *Osteoporos Int* 10:150–160
19. Marquis P, Cialdella P, De la Loge C (2001) Development and validation of a specific quality of life module in post-menopausal women with osteoporosis: the QUALIOST. *Qual Life Res* 10:555–566
20. Naegeli A, Nixon A, Burge R, Gold DT, Silverman S (2014) Development of the osteoporosis assessment questionnaire-physical function (OPAQ-PF): an osteoporosis-specific, patient-reported outcomes (PRO) measure of physical function. *Osteoporos Int* 25:579–588, Epub 2013 Aug 9
21. Patient reported health instruments group (2012) A structured review of patient-reported measures in relation to selected chronic conditions, perceptions of quality of care and carer impact. 2006. Available at http://phi.uhce.ox.ac.uk/pdf/ChronicConditions/DH_REVIEWS_NOVEMBER_2007.pdf Accessed 17 December
22. Scientific Advisory Committee of the Medical Outcomes Trust (2002) Assessing health status and quality-of-life instruments: Attributes and review criteria. *Qual Life Res* 11:193–205
23. Streiner DL, Norman GR (1995) Health measurement scales: a practical guide on their development and use, 2nd edn. Oxford University Press, Oxford
24. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW (1988) Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 15:1833–1840
25. Ware JE Jr, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care* 30:473–483
26. International Osteoporosis Foundation (2012) Quality of life questionnaires QUALEFFO-41
27. Lips P, Leplege A (2000) Development and validation of a quality of life questionnaire for patients with vertebral fractures: QUALEFFO-41. *Qual Life Res* 9:763–766
28. Wyrwich KW, Norquist JM, Lenderking WR, Acaster S (2013) Methods for interpreting change over time in patient-reported outcome measures. *Qual Life Res* 22:475–483
29. Studenski S, Duncan PW, Chandler J, Samsa G, Prescott B, Hogue C, Bearon LB (1994) Predicting falls: the role of mobility and nonphysical factors. *J Am Geriatr Soc* 42:297–302
30. Kline RB (2010) Principles and practice of structural equation modeling, 3rd edn. Guilford Publications, New York
31. Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* 17:297–334
32. Norman GR, Sloan JA, Wyrwich KW (2003) Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 41:582–592
33. Dawson J, Doll H, Coffey J, Jenkinson C (2007) Responsiveness and minimally important change for the Manchester-Oxford foot questionnaire (MOXFQ) compared with AOFAS and SF-36 assessments following surgery for hallux valgus. *Osteoarthritis Cart* 15:918–931
34. Dawson J, Doll H, Boller I, Fitzpatrick R, Little C, Rees J, Carr A (2008) Comparative responsiveness and minimal change for the Oxford elbow score following surgery. *Qual Life Res* 17:1257–1267
35. Nunnally J, Bernstein I (1994) Psychometric theory, 3rd edn. McGraw-Hill Humanities/Social Science/Languages, London
36. Kline P (1993) The handbook of psychological testing, 1st edn. Routledge, London