

Responsiveness of Health State Utility Values in Knee Osteoarthritis

Erin M. Davis, Larry D. Lynd, Maja Grubisic, Jacek A. Kopec, Eric C. Sayre, Jolanda Cibere, John Esdaile, and Carlo A. Marra

ABSTRACT. Objective. Adaptive tests are increasingly being used to assess health-related quality of life in patients with a variety of medical conditions, including osteoarthritis (OA) of the knee. This approach has recently been used to assess health state utility valuations (HSUV) for use in quality-adjusted life-year calculations. To accurately assess incremental value for money, these tools must be responsive. Therefore, we examined the responsiveness of the Health Utilities Index mark 3 (HUI3) and Paper Adaptive Test-5D (PAT-5DQOL) in a group of patients with knee OA.

Methods. We used patient-level data from a randomized controlled trial evaluating a pharmacist-initiated multidisciplinary intervention in newly diagnosed patients with knee OA. The mean change for utility scores from baseline to 6 months was calculated, as well as effect size (ES) and standardized response mean (SRM) for the HUI3 and PAT-5DQOL, and generalized additive model plots, using the Western Ontario and McMaster Osteoarthritis index as a reference standard.

Results. When patients were assessed based on whether their condition had improved, remained unchanged, or worsened over time, the PAT-5DQOL showed greater responsiveness in patients whose condition had either improved or worsened. ES and SRM were generally small for both instruments.

Conclusion. The PAT-5DQOL is more responsive to change over time than the HUI3 in patients with knee OA. (First Release Nov 1 2013; J Rheumatol 2013;40:2075–82; doi:10.3899/jrheum.130176)

Key Indexing Terms:

KNEE OSTEOARTHRITIS HEALTH-RELATED QUALITY OF LIFE ADAPTIVE TESTS
QUALITY-ADJUSTED LIFE-YEARS HEALTH CARE EVALUATION MECHANISMS

The reduction of pain, maintenance or improvement of joint mobility, limitation of functional disability, and improvement of self-management are the primary goals in the

From the School of Pharmacy, Memorial University, St. John's, Newfoundland and Labrador; Collaboration for Outcomes Research and Evaluation (CORE), and School of Population and Public Health, University of British Columbia, Arthritis Research Centre, and Centre for Health Evaluation and Outcome Sciences, Providence Health Care Research Institute, Vancouver, British Columbia, Canada.

E.M. Davis, BSc(Pharm), PharmD, Assistant Professor, School of Pharmacy, Memorial University; L.D. Lynd, BSP, PhD, Associate Professor and Associate Director, CORE, Faculty of Pharmaceutical Sciences, University of British Columbia and Scientist, Centre for Health Evaluation and Outcome Sciences, Providence Health Care Research Institute; M. Grubisic, MSc, Statistician, CORE, Faculty of Pharmaceutical Sciences, University of British Columbia; J.A. Kopec, MD, PhD, Professor, School of Population and Public Health, Faculty of Medicine, University of British Columbia and Senior Scientist, Arthritis Research Centre; E.C. Sayre, PhD, Research Statistical Analyst, Arthritis Research Centre; J. Cibere, Associate Professor, Department of Rheumatology, Faculty of Medicine, and Scientist, Arthritis Research Centre; J.M. Esdaile, Professor of Medicine, University of British Columbia, and Scientific Director, Arthritis Research Centre; C.A. Marra, PharmD, PhD, Professor and Director, CORE, Faculty of Pharmaceutical Sciences, University of British Columbia and Scientist, Centre for Health Evaluation and Outcome Sciences, Providence Health Care Research Institute, and Scientist, Arthritis Research Centre.

*Address correspondence to Dr. C.A. Marra, 2146 East Mall, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada.
E-mail: Carlo.marra@ubc.ca*

Accepted for publication August 28, 2013.

treatment of osteoarthritis (OA), a condition for which there is no curative therapy¹. Thus health-related quality of life (HRQOL) and functional capacity is of paramount importance to assess the effect of therapies aimed at managing OA. As new therapies and strategies are developed, their relative cost-effectiveness needs to be estimated using standard techniques of economic evaluation, including the increment cost-effectiveness ratio (ICER) given by the formula:

$$\text{ICER} = \frac{\Delta C}{\Delta E}$$

In most economic evaluations, the numerator is the quality-adjusted life year (QALY). The QALY combines length with HRQOL such that each year of life is adjusted by a certain value according to decrements in HRQOL experienced. There are several different methods to assess HRQOL that can be used to calculate QALY; however, there are no specific recommendations for OA.

For optimal use of these methods, investigators must have sensitive instruments for quantifying changes in HRQOL as measured by health state utility values or HSUV (ΔE) for groups of patients over time, as a result of a given intervention. A tool insensitive to change over time may result in small or insignificant changes in ΔE compared to

changes in cost (ΔC) resulting in a larger ICER. This effect could result in decisions not to fund particular interventions owing to a perceived lack of cost effectiveness or cost utility. For HRQOL assessment tools, the ability of the tool to accurately measure changes in patient health states over time is referred to as responsiveness, or longitudinal validity^{2,3,4}.

Two such tools, the Health Utilities Index Mark 3 (HUI3) and the Paper Adaptive Test (PAT-5DQOL) were used in a recent randomized controlled trial (RCT) of patients with knee OA comparing standard care and pharmacist-initiated multidisciplinary care⁵. The HUI3 is a multiattribute, preference-based HSUV instrument that measures health status and HRQOL and produces HSUV⁶. The PAT-5DQOL is an adaptive measurement system using an item bank with domains developed to be specifically relevant to patients with arthritis, to produce HSUV and measurements of health status and HRQOL⁷. Adaptive measurement systems allow different respondents to answer a different set of questions from preexisting item banks, based on their current health state as measured by their previous responses, and final scores are determined using maximum likelihood estimation^{8,9}.

In general, disease-specific scales have been shown to be more responsive than generic instruments¹⁰. Responsiveness of multiple generic and disease-specific measures has also been compared in OA in a Thai population, finding superior responsiveness with disease-specific instruments¹¹. Other trials have compared the relative responsiveness of specific instruments in knee OA and knee replacement. The Western Ontario and McMaster Osteoarthritis index (WOMAC) has been shown to have superior responsiveness compared to the Lequesne's algofunctional index (AFI) and the Medical Outcomes Study Short Form-36 (SF-36) questionnaire in patients with lower extremity arthritis and knee replacement. Hart, *et al* studied the use of a computer adaptive test (CAT) model in a population of patients with a broad range of knee impairments, finding the CAT model to be valid, reliable, and responsive¹²; however, to our knowledge, no studies to date have compared the responsiveness of the HUI3 to the PAT-5DQOL in knee OA.

The objectives of our study were to (1) compare the PAT-5DQOL and the HUI3; and (2) assess the relative responsiveness of these measures in a cohort of patients with OA of the knee.

MATERIALS AND METHODS

Patient population. Patients who underwent randomization in the Pharmacist Initiated Intervention Trial in Osteoarthritis (PhIT-OA) and completed baseline HRQOL assessments as well as 6-month followup assessments for the HUI3, PAT-5DQOL, and WOMAC were included in this analysis. Baseline characteristics for patients included in this analysis are presented in Table 1. Detailed study methodology including patient inclusion and exclusion criteria for the PhIT-OA trial has been published elsewhere¹³. In brief, the results of the trial showed that the quality of care

Table 1. Baseline demographics for those with complete information (n = 94).

Age, yrs, mean (SD)	61.7 (8.2)
Sex, n (%)	
Male	39 (42)
Female	55 (59)
BMI*, n (%)	
< 18.5	0 (0)
18.5–24.9	34 (36)
25.0–29.9	37 (39)
≥ 30.0	23 (24)
Household income, CAN\$/yr	
< 20,000	7 (7)
20,000–50,000	20 (21)
> 50,000	67 (71)
Education, n (%)	
< High school	1 (1)
High school	16 (17)
> High school	77 (82)
Ethnicity, n (%)	
Aboriginal	1 (1)
White	75 (80)
Asian	10 (11)
South Asian	7 (7)
Other	1 (1)

* Body mass index (BMI) is based on self-reported weight.

was significantly improved in the intervention arm as compared to the control arm. In addition, significant improvements were observed for the intervention care group as compared to the usual care group in the WOMAC global, pain, and function scores at 3 and 6 months (all $p < 0.01$); the PAT-5D daily activity scores at 3 and 6 months (both $p < 0.05$); the PAT-5D pain scores at 6 months ($p = 0.05$); the HUI3 single-attribute pain scores at 3 and 6 months (all $p < 0.05$); and the Lower Extremity Function Scale scores at 6 months ($p < 0.05$).

Measures. Contact was made by blinded study staff at baseline, 3 months, and 6 months, which consisted of a predefined scripted questionnaire including the WOMAC, HUI3, and PAT-5DQOL. Changes in patient clinical status in the intervention arm were the result of a pharmacist-initiated multidisciplinary intervention including OA medication review, provision of an OA education booklet and education program, letter to their primary care physician, physiotherapist referral, and exercise program¹². Patients in the control arm received standard care, which consisted of provision of an OA education booklet.

HRQOL measures. The WOMAC is a 24-item OA-specific instrument originally designed for clinical trials of hip and knee OA, and has previously been validated in patients with OA^{13,14}. The WOMAC comprises 3 domains, including pain, stiffness, and physical function, rated from 0 (full health) to 4 (extreme disability). The score for individual domains was determined by summing the coded response scores for each domain, and the overall instrument utility was scored by summing the individual domain scores. Individual domain scores were then normalized to a scale of 0 to 10, resulting in a total score on a scale of 0 to 30. A ≥ 20% positive or negative change in the WOMAC has been validated as clinically significant¹⁵.

The HUI3 is a generic utility instrument comprising 8 domains (Vision, Hearing, Speech, Ambulation, Dexterity, Emotion, Cognition, and Pain), each with 4–6 descriptive levels, scored from 0.0 (death) to 1.0 (perfect health), with states worse than death measured as negative scores. The HUI3 describes a potential 972,000 unique health states. HSUV for the HUI3 were derived from a random sample of adults in Hamilton, Ontario, Canada, using both a visual analog technique and a standard gamble instrument⁶. A previously described multiplicative scoring model was used in this analysis¹⁷. A clinically significant change on the HUI3 has been

described as 0.03 by Grootendorst, *et al*¹⁸. Based on this analysis, and the fact that the smallest measurable change on the HUI3 is a change of 0.03, we used a cutoff of 0.03 on the HUI3 to represent a minimally important difference (MID)¹⁹.

PAT-5DQOL. The PAT-5DQOL is an adaptive measure of HRQOL designed for measuring HSUV in patients with musculoskeletal diseases, based on item response theory. It comprises 5 domains considered relevant to patients with arthritis and related conditions including daily activities, walking, handling objects, pain or discomfort, and feelings. The PAT-5DQOL utility scores are derived using an approach similar to that used for HUI, EQ-5D, or SF-6D, i.e., this is an indirect approach to measuring preferences. We had performed a preference study in which we used standard gamble to obtain preferences for different combination of health levels defined by the instrument. Scores were determined based on a previously described computer algorithm⁷.

Statistics. Baseline HRQOL scores were compared between those with complete (all HRQOL assessment instruments completed at both baseline and 6 months) and incomplete (those patients missing 1 or more HRQOL assessment score at either baseline, 6 months, or both) datasets using independent t tests. Patients were analyzed as 1 cohort, assuming that the multidisciplinary intervention offered to the intervention group would not affect validity, responsiveness, or discriminative ability of the instruments (i.e., independent of randomized intervention).

Analyses were performed using SAS statistical software version 9.2 (SAS Institute Inc.).

Assessment of responsiveness. All patients who had data on all 3 instruments at both baseline and 6 months were included in the analysis. Baseline and 6-month WOMAC scores were used as a standard with which to measure change in health state. The WOMAC was chosen as the reference standard because its validity, reliability, and responsiveness have already been demonstrated in this patient population^{14,15,16}. The following criteria were used to determine health status change based on the WOMAC: (1) Improvement: $\geq 20\%$ negative change; (2) No change: $< 20\%$ change; and (3) Worsening: $\geq 20\%$ positive change.

Responsiveness was measured using both distributive techniques and an anchor-based method. Distributive methods included effect size (ES) using the following formula:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{SD_{\text{Baseline}}}$$

Where X_1 is the mean score at 6 months, X_2 is the mean score at baseline for the entire group, and SD_{Baseline} is the SD of the mean baseline scores. An effect size of 1 indicates a mean change equal in magnitude to 1 SD.

Cohen's criteria were used in interpreting effect size with absolute changes: a small effect size was 0.2, a medium was 0.5, and a large effect size was 0.8^{20,21}.

The second distributive method was standardized response mean (SRM), using the following formula:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{SD_{\bar{X}_1 - \bar{X}_2}}$$

Where X_1 is the mean score at 6 months, X_2 is the mean score at baseline, and $SD_{\bar{X}_1 - \bar{X}_2}$ is the SD of the mean change between baseline and 6 months.

The absolute values of the SRM are interpreted in the same manner as ES (< 0.5 small, 0.5 – 0.8 medium, > 0.8 large).

To estimate the empirical 95% CI of each ES and SRM, bootstrap sampling with 1000 simulations was performed, because the standard errors of the distribution-based approaches are not defined.

To provide answers regarding the likelihood of a given magnitude of change in utility score representing an actual improvement or deterioration in patient status relative to the reference standard (i.e., How much of a change is actually clinically significant?), a generalized additive model

(GAM) was used as an anchor-based method. The GAM model used here was the additive logistic model; this means the nonparametric smoothing was used to estimate the log odds, that is, the log odds change smoothly with a predictor (as opposed to log odds being linear in a predictor).

This model was used to show how estimated probabilities of a specific response vary depending on the observed change in the scores of the instruments. GAM plots were produced in R version 2.11.1 (www.r-project.org).

RESULTS

The RCT in which these data were collected originally included 139 patients, the baseline characteristics of which are reported elsewhere⁵. Data from 45 patients (32%) were excluded from this analysis because they did not complete the HUI3, PAT-5DQOL, and WOMAC assessments at both baseline and 6-month followup. The following analyses were therefore conducted on the data from the remaining 94 patients (68%). There were some differences in baseline HRQOL between those patients with complete data and those with incomplete data. Those with incomplete data had a significantly lower global score on both the HUI3 and PAT-5DQOL compared with those who completed all assessments at both timepoints (HUI3 0.74 vs 0.58, $p < 0.001$, and PAT-5DQOL 0.87 vs 0.79, $p = 0.002$, respectively). Individual domain scores for pain were also significantly different for both the HUI3 and PAT-5DQOL between patients with complete and incomplete data (HUI3 pain 0.73 vs 0.62, $p = 0.029$ and PAT-5DQOL pain/discomfort 45.97 vs 42.16, $p = 0.022$, respectively). Of the 94 patients analyzed, 53% showed improvement, 28% were unchanged, and 19% had worsened, based on the previously described WOMAC cutpoints.

Responsiveness. Neither the HUI3 nor the PAT-5DQOL global scores showed a significant change from baseline to 6 months in the overall patient sample; however, when stratified into patients who improved, those who were unchanged, and those who worsened over time according to the WOMAC criteria, the HUI3 global score remained nonsignificant in all 3 groups, and the PAT-5DQOL detected significant changes in patients who improved (mean change 0.037, $p = 0.012$) and those who worsened (mean change 0.051, $p = 0.011$; Table 2). Interestingly, the HUI3 global score showed a numerically larger positive change from baseline for those whose clinical status remained unchanged compared to those who improved, although neither of these values was statistically significant.

The ES and SRM were small in general for both tools when considering global scores and are presented in Table 3. The HUI3 global score CI for ES and SRM included 0 for all patient groups. When the groups were broken down by the WOMAC response criteria, the PAT-5DQOL had larger ES and SRM for patients who either improved or worsened over time, and the CI did not cross 0. In contrast, for the group of patients who remained unchanged based on the WOMAC, the PAT-5DQOL CI for both ES and SRM included 0. This result is consistent with responsiveness in

Table 2. Mean HRQOL change from baseline.

	Mean Baseline (SD)	Mean 6 Mos (SD)	Mean Change (SD)	p
WOMAC Global, n = 94	7.41 (3.86)	6.07 (4.31)	-1.34 (3.78)	0.001
Improved, n = 50	7.36 (3.38)	3.51 (1.93)	-3.85 (2.69)	< 0.001
Same, n = 26	8.86 (4.55)	8.45 (4.46)	-0.41 (0.56)	0.001
Worse, n = 18	5.48 (3.32)	9.74 (4.40)	4.26 (1.98)	< 0.001
WOMAC Pain, n = 94	2.63 (1.51)	2.21 (1.55)	-0.42 (1.45)	0.006
Improved, n = 50	2.54 (1.45)	1.41 (1.12)	-1.13 (1.26)	< 0.001
Same, n = 26	3.27 (1.59)	3.00 (1.33)	-0.27 (0.78)	0.09
Worse, n = 18	1.94 (1.22)	3.28 (1.69)	1.33 (1.15)	< 0.01
WOMAC Stiffness, n = 94	2.93 (1.74)	2.35 (1.93)	-0.57 (2.10)	0.01
Improved, n = 50	3.08 (1.69)	1.30 (1.13)	-1.78 (1.26)	< 0.001
Same, n = 26	3.08 (1.98)	3.13 (2.01)	0.048 (0.97)	0.80
Worse, n = 18	2.29 (2.02)	4.17 (1.77)	1.88 (1.78)	< 0.001
WOMAC Function, n = 94	1.86 (1.64)	1.51 (1.48)	-0.35 (1.43)	0.02
Improved, n = 50	1.74 (1.63)	0.80 (0.75)	-0.94 (1.54)	< 0.001
Same, n = 26	2.56 (1.79)	2.33 (1.87)	-0.19 (0.55)	0.095
Worse, n = 18	1.24 (1.12)	2.29 (1.41)	1.05 (0.88)	< 0.001
HUI3 Global, n = 94	0.74 (0.18)	0.78 (0.16)	0.037 (0.20)	0.08
Improved, n = 50	0.76 (0.16)	0.81 (0.14)	0.046 (0.21)	0.13
Same, n = 26	0.69 (0.22)	0.75 (0.21)	0.055 (0.20)	0.17
Worse, n = 18	0.76 (0.15)	0.74 (0.15)	-0.015 (0.18)	0.74
HUI3 Pain, n = 94	0.73 (0.23)	0.79 (0.17)	0.057 (0.24)	0.021
Improved, n = 50	0.78 (0.18)	0.83 (0.10)	0.058 (0.21)	0.051
Same, n = 26	0.67 (0.30)	0.75 (0.25)	0.083 (0.28)	0.14
Worse, n = 18	0.72 (0.25)	0.74 (0.16)	0.018 (0.26)	0.78
HUI3 Ambulation, n = 94	0.98 (0.098)	0.97 (0.10)	0.0055 (0.063)	0.40
Improved, n = 50	1.00 (0.024)	1.00 (0.024)	0 (0.034)	1
Same, n = 26	0.94 (0.18)	0.94 (0.12)	0.0062 (0.088)	0.73
Worse, n = 18	1.00 (0)	0.96 (0.073)	-0.038 (0.073)	0.042
PAT-5DQOL Global, n = 94	0.87 (0.11)	0.88 (0.098)	0.011 (0.11)	0.32
Improved, n = 50	0.88 (0.10)	0.92 (0.068)	0.037 (0.10)	0.012
Same, n = 26	0.83 (0.12)	0.84 (0.12)	0.0033 (0.12)	0.89
Worse, n = 18	0.91 (0.066)	0.86 (0.099)	-0.051 (0.076)	0.011
PAT-5DQOL Pain/Discomfort, n = 94	45.97 (7.06)	47.80 (7.23)	1.82 (7.32)	0.018
Improved, n = 50	46.12 (6.29)	50.70 (5.70)	4.58 (6.10)	< 0.001
Same, n = 26	43.20 (7.37)	44.73 (7.81)	1.53 (7.21)	0.29
Worse, n = 18	49.53 (7.31)	44.11 (7.67)	-5.42 (5.60)	0.001
PAT-D Walk, n = 94	43.99 (7.50)	44.30 (7.53)	0.31 (5.88)	0.61
Improved, n = 50	45.38 (6.39)	47.03 (6.75)	1.66 (5.80)	0.049
Same, n = 26	40.56 (8.24)	41.75 (7.81)	1.19 (3.82)	0.13
Worse, n = 18	45.07 (8.07)	40.38 (6.40)	-4.70 (6.17)	0.005
PAT-5DQOL Daily Activities, n = 94	43.64 (8.37)	45.28 (8.38)	1.65 (9.17)	0.085
Improved, n = 50	43.73 (8.25)	48.21 (6.42)	4.48 (9.42)	0.002
Same, n = 26	40.90 (8.73)	42.27 (9.84)	1.36 (8.27)	0.41
Worse, n = 18	47.33 (7.02)	41.51 (8.29)	-5.82 (4.61)	< 0.001

HRQOL: health-related quality of life; WOMAC: Western Ontario and McMaster Osteoarthritis index; HUI3: Health Utilities Index mark 3; PAT-5DQOL: Paper Adaptive Test-5D.

those who either improve or worsen, while not showing inappropriate change in those who remain truly unchanged.

GAM. GAM plots were based on 2 groups: those who improved (n = 50) and those who either remained unchanged or worsened (n = 44). Using the GAM plot for the global HUI3 score, a change in HUI3 score of 0.4 (on a scale of 0.00 to 1.00) would result in a probability of improvement defined by WOMAC of only 0.6 (Figure 1). Deterioration on the HUI3 of about -0.4 would result in a probability of no improvement/worsening on the WOMAC

of about 0.5 (Figure 1). Conversely, in Figure 2, the GAM plots for the global PAT-5DQOL score showed a probability of improvement on the WOMAC of about 0.7, with a change of about 0.1 on the PAT-5DQOL global. A deterioration of PAT-5DQOL global of -0.2 was associated with a probability of not improving/worsening on the WOMAC of about 0.9. The individual domain scores for the PAT-5DQOL showed similar agreement with the WOMAC index (results not shown).

Single domain scores. Single domains reaching statistical

Table 3. Effect size (ES) and standardized responses mean (SRM).

	ES (95% CI)	SRM (95% CI)
HUI3 Global, n = 94	0.21 (−0.017–0.43)	0.18 (−0.014–0.42)
Improved, n = 50	0.29 (−0.12–0.58)	0.22 (−0.078–0.50)
Same, n = 26	0.24 (−0.11–0.60)	0.28 (−0.12–0.72)
Worse, n = 18	−0.10 (−0.88–0.44)	−0.080 (−0.58–0.41)
HUI3 Pain, n = 94	0.25 (0.069–0.41)	0.24 (0.062–0.42)
Improved, n = 50	0.33 (0.026–0.59)	0.28 (0.021–0.52)
Same, n = 26	0.28 (−0.10–0.59)	0.30 (−0.092–0.76)
Worse, n = 18	0.072 (−0.065–0.45)	0.068 (−0.54–0.46)
HUI3 Ambulation, n = 94	−0.056 (−0.48–0.086)	−0.088 (−0.31–0.11)
Improved, n = 50	0 (−0.28–0.25)	0 (−0.25–0.25)
Same, n = 26	0.035 (−0.28–0.33)	0.070 (−0.35–0.37)
Worse, n = 18	NA	−0.52 (−0.87– −0.24)
PAT-5DQOL Global, n = 94	0.104 (−0.11–0.26)	0.103 (−0.10–0.28)
Improved, n = 50	0.36 (0.18–0.71)	0.37 (0.16–0.80)
Same, n = 26	0.027 (−0.41–0.35)	0.028 (−0.43–0.38)
Worse, n = 18	−0.77 (−1.69– −0.32)	−0.67 (−1.27– −0.46)
PAT-5DQOL Pain/Discomfort, n = 94	0.26 (0.044–0.47)	0.25 (0.041–0.47)
Improved, n = 50	0.73 (0.44–1.22)	0.75 (0.47–1.11)
Same, n = 26	0.21 (−0.16–0.56)	0.21 (−0.15–0.70)
Worse, n = 18	−0.74 (−1.29– −0.40)	−0.97 (−1.66– −0.53)
PAT-5D Walk, n = 94	0.042 (0.11–0.21)	0.053 (−0.14–0.28)
Improved, n = 50	0.26 (0.00059–0.59)	0.29 (0.00080–0.63)
Same, n = 26	0.14 (−0.037–0.36)	0.31 (−0.088–0.66)
Worse, n = 18	−0.58 (−1.08– −0.27)	−0.76 (−1.38– −0.38)
PAT-5DQOL Daily Activities, n = 94	0.20 (−0.018–0.40)	0.18 (−0.015–0.36)
Improved, n = 50	0.54 (0.29–0.85)	0.48 (0.27–0.75)
Same, n = 26	0.16 (−0.19–0.59)	0.16 (−0.18–0.75)
Worse, n = 18	−0.83 (−1.29– −0.40)	−1.26 (−1.66– −0.53)

HUI3: Health Utilities Index mark 3; PAT-5DQOL: Paper Adaptive Test-5D.

significance in mean change from baseline for the overall patient population included the HUI3 pain domain and the PAT-5DQOL pain and discomfort domain. For those patients whose health status improved, the PAT-5DQOL pain and discomfort, walking, and daily activities domains all showed significant improvement from baseline to 6 months. For patients who worsened over time, the HUI3 Ambulation domain, and the PAT-5DQOL pain and discomfort, walking, and daily activities domains all showed significant change.

ES and SRM for individual domains are presented in Table 3. In general, for the pain domains, the PAT-5DQOL showed larger ES and SRM than did the HUI3 for those who either improved or worsened. The walking domain of the PAT-5DQOL showed larger ES and SRM for those who improved or worsened compared to the HUI3 Ambulation domain; however, CI included 0 in all patient groups. Finally, the daily activities domain of the PAT-5DQOL showed a moderate ES and SRM in patients who improved, and a large ES and SRM in those who worsened. As seen in Table 4, 58% of patients exceeded the MID for the PAT-5DQOL pain and discomfort domain, 56% exceeded the MID for the PAT-5DQOL usual daily activities domain, while 39% exceeded the MID for the HUI3 global score.

DISCUSSION

This analysis is the first to compare the responsiveness of the PAT-5DQOL, a disease-specific adaptive HSUV measurement tool to the HUI3, a generic, indirect HSUV measure, in patients with OA of the knee. In general, ES and SRM for both instruments were small. When patients were stratified into groups based on their change in WOMAC score (the reference standard), the PAT-5DQOL global score was more responsive for both patients who improved and those who worsened over time compared to the HUI3. The individual domains for pain and ambulation/walking were also more responsive in the PAT-5DQOL for those who improved or worsened, compared to the corresponding HUI3 domains. ES and SRM were considered large in the pain and discomfort and daily activities domains of the PAT-5DQOL for those who worsened over time. Pain and daily activity domains for both instruments were more responsive than the ambulation/walking domains.

GAM plots showed greater probability of correctly defining patients as improved, unchanged, or worsened, based on our reference standard, with the PAT-5DQOL compared to the HUI3. Neither the HUI3 global score nor the PAT-5DQOL global score showed significant changes from baseline in the overall population of patients with

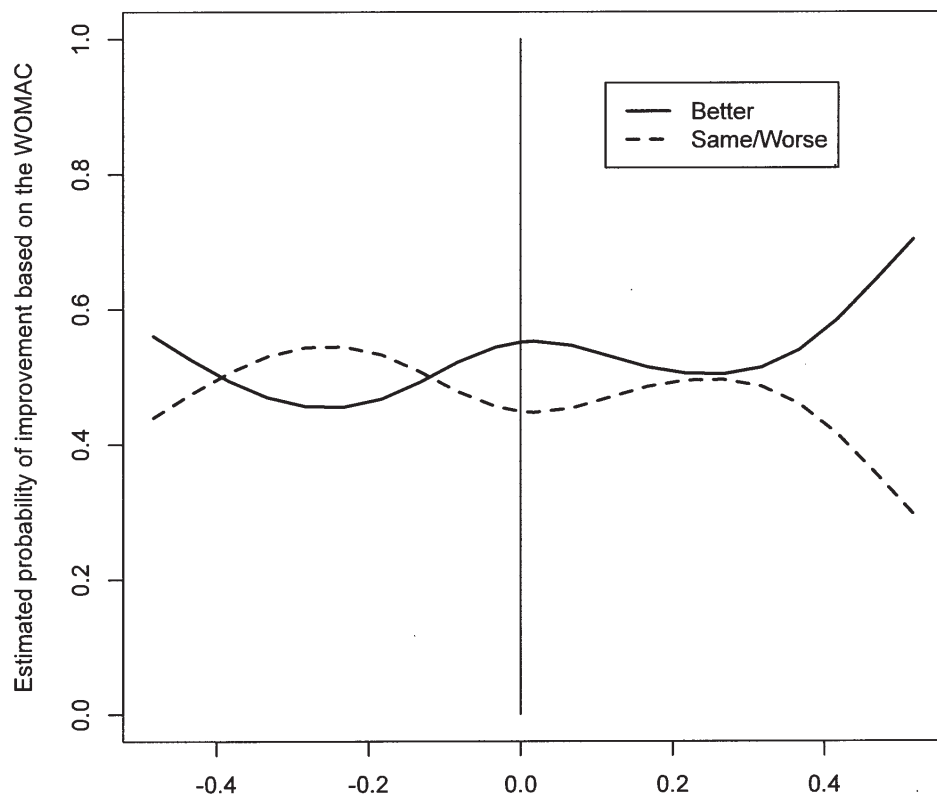


Figure 1. Change in Health Utilities Index mark 3 from baseline to 6 months. WOMAC: Western Ontario and McMaster Osteoarthritis index.

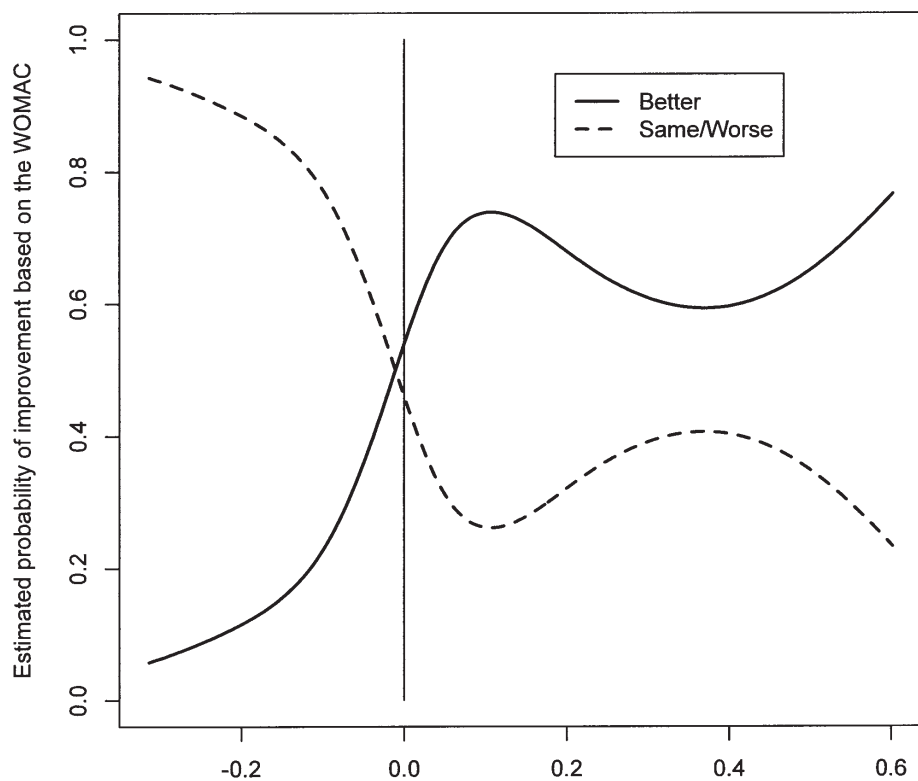


Figure 2. Change in Paper Adaptive Test-5D from baseline to 6 months. WOMAC: Western Ontario and McMaster Osteoarthritis index.

Table 4. Proportion of patients exceeding the minimally important difference (MID).

HRQOL	Improvement, n (%) / MID
PAT-5DQOL	
Pain and discomfort	55 (58) / 1E-7
Usual daily activities	53 (56) / 8.16E-7
HUI3 global score	39 (41) / 0.03

HRQOL: health-related quality of life; HUI3: Health Utilities Index mark 3; PAT-5DQOL: Paper Adaptive Test-5D.

complete data, and the HUI3 global score did not show a significant difference from baseline for any of the groups broken down by WOMAC cutpoint scores, although the PAT-5DQOL did measure significant change in both the improved and worsened patient groups. Our results also demonstrate that patients who failed to complete all HRQOL measurement tools at both baseline and 6 months had lower baseline HUI3 and PAT-5DQOL global scores, indicating a poorer HRQOL in this patient group.

Other investigators have examined the responsiveness of disease-specific and generic HRQOL instruments in knee OA and total knee arthroplasty (TKA). In 1999, Theiler, *et al* compared the responsiveness of the WOMAC and the AFI in patients with OA of the lower extremities and found that in general, pain scales were more responsive than function scales in both tests, and the WOMAC scales and global index were more responsive than the AFI over 12 months (SRM at 12 months 2 vs 1.5)²². Similarly, the WOMAC and SF-36 have been compared both in patients with OA of the legs undergoing a comprehensive rehabilitation intervention and knee OA patients undergoing TKA. In patients undergoing rehabilitation, pain scales were again found to be more responsive than function scales for both instruments, and the WOMAC was significantly more responsive than the SF-36 in both populations^{23,24,25,26}. Recently, Sweeney, *et al*²⁷ compared a WOMAC-derived HUI3 and the SF-6D in patients with knee OA undergoing arthroplasty. This group found that both instruments had good construct validity and correlated well with disease-specific measures. The HUI3 was found to be more responsive (ES 1.34, 95% CI 1.17, 1.52 at 3 months and 1.52, 95% CI 1.35, 1.68 at 6 months) than the SF-6D (ES 1.12, 95% CI 0.95, 1.30 at 3 months and 1.33, 95% CI 1.16, 1.51 at 6 months), although mean changes from baseline to 3 and 6 months did not differ significantly between instruments²⁷. This body of literature agrees with our finding that disease-specific tools such as the PAT-5DQOL tend to be more responsive than generic tools such as the HUI3, although both are able to detect change over time. In addition, the PAT-5DQOL differs from the HUI3 not only in being disease-specific, but also in being adaptive, which may make it more responsive to change.

Comparisons have been made between the respon-

siveness of CAT and nonadaptive instruments in other disease states. In a childhood physical activity intervention study, CAT was found to be less responsive than the Pediatric Evaluation of Disability Inventory (PEDI; ES 0.99 vs 0.87, SRM 1.64 vs 1.00, respectively), but this was offset by a large decrease in testing burden²⁷ (15 items vs 161 items). The authors speculate that a reason for the CAT being found less responsive in this patient population compared to the PEDI was that it was limited to 15 items per individual, which showed some reduced ability to detect changes in mobility in those with autism or cognitive deficits. This difference may be responsible for those findings being contrary to our own. The population of patients with knee OA examined in our study also differs greatly from the children studied by Haley, *et al*²⁸.

Limitations of our methodology must also be considered. First, our sample size was quite small, contributing to large CI around the ES and SRM, as well as the splines in the GAM plots. Second, we were unable to use a transition question as an external standard because it was not included in the questionnaires for the original RCT for which these data were gathered. Change in WOMAC score has previously been used to determine cutpoints for data analysis in similar patients²⁷. Third, the PAT-5DQOL has yet to undergo a validation study in this population, although several studies have used it in patients with knee OA for measuring HRQOL^{29,30}. Finally, in general, effect sizes were quite small for both the PAT-5DQOL and HUI3, potentially as a result of combining the patients from the intervention and control arms of an RCT, thereby resulting in a group of patients with more heterogeneous mean HRQOL changes from baseline. Strengths of our methods include the use of both ES and SRM to measure responsiveness, as well as GAM plots, which give a better indication of the magnitude of change required to classify a patient as either improved, unchanged, or worsened, based on the reference standard. We also analyzed the mean score changes and responsiveness from baseline in not only the entire patient population, but also in those who improved, remained unchanged, or worsened, based on their WOMAC scores. This allowed us to draw conclusions regarding the usefulness of these tools in populations of patients undergoing a known effective therapy, as well as those who remained untreated and followed the natural disease course of OA.

The PAT-5DQOL is more responsive than the HUI3 in patients who either improve or worsen from baseline. Pain domains and daily activity domains were more responsive than ambulation/walking domains. Single domain scores of the PAT-5DQOL are more responsive than corresponding domains of the HUI3. Economic evaluations in OA that use the HSUV from the PAT-5DQOL to determine QALY are likely to result in valid estimates in incremental cost-effectiveness ratios because the denominator is sensitive to relevant changes in quality of life.

REFERENCES

- Guidelines and Protocols Advisory Committee: Osteoarthritis in peripheral joints — diagnosis and treatment. 2008. [Internet. Accessed Sept 16, 2013.] Available from: www.bcguidelines.ca/guideline_osteoarthritis.html
- Wright JG, Young NL. A comparison of different indices of responsiveness. *J Clin Epidemiol* 1997;50:239–46.
- Guyatt GH, Deyo RA, Charlson M, Levine M, Mitchell A. Responsiveness and validity in health status measurement: a clarification. *J Clin Epidemiol* 1989;42:403–8.
- Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res* 2003;12:349–62.
- Marra CA, Cibere J, Grubisic M, Grindrod KA, Gastonguay L, Thomas JM, et al. Pharmacist-initiated intervention trial in osteoarthritis: a multidisciplinary intervention for knee osteoarthritis. *Arthritis Care Res* 2012;64:1837–45.
- Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties and applications. *Health Qual Life Outcomes* 2003;1:54.
- Kopec JA, Sayre EC, Davis AM, Badley EM, Abrahamowicz M, Sherlock L, et al. Assessment of health-related quality of life in arthritis: conceptualization and development of five item banks using item response theory. *Health Qual Life Outcomes* 2006;4:33.
- Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care* 2000;38 Suppl 9:II28.
- Fries JF, Bruce B, Cella D. The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. *Clin Exp Rheumatol* 2005;5 Suppl 39:S53–7.
- Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. *J Clin Epidemiol* 2003;56:52–60.
- Tangtrakulwanich B, Wiwatwongwana S, Chongsuvivatwong V, Geater AF. Comparison of validity, and responsiveness between general and disease-specific quality of life instruments (Thai version) in knee osteoarthritis. *J Med Assoc Thai* 2006;89:1454–9.
- Hart DL, Wang Y-C, Stratford PW, Mioduski JE. Computerized adaptive test for patients with knee impairments produced valid and responsive measures of function. *J Clin Epidemiol* 2008;61:1113–24.
- Marra CA, Tsuyuki RT, Soon JA, Gastonguay L, Oteng B, Cibere J, et al. Design of a randomized trial of a multidisciplinary intervention for knee osteoarthritis: Pharmacist Initiated Intervention Trial in Osteoarthritis (PhIT-OA). *Can Pharm J (Ott)* 2008;141:33–8.
- Bellamy N. WOMAC: a 20-year experiential review of a patient-centered self-reported health status questionnaire. *J Rheumatol* 2002;29:2473–6.
- Söderman P, Malchau H. Validity and reliability of Swedish WOMAC osteoarthritis index: a self-administered disease-specific questionnaire (WOMAC) versus generic instruments (SF-36 and NHP). *Acta Orthop Scand* 2000;71:39–46.
- Bellamy N, Bell MJ, Goldsmith CH, Pericak D, Walker V, Raynauld JP, et al. Evaluation of WOMAC 20, 50, 70 response criteria in patients treated with hylan G-F 20 for knee osteoarthritis. *Ann Rheum Dis* 2005;64:881–5.
- Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Med Care* 2002;40:113–28.
- Grootendorst P, Feeny D, Furlong W. Health Utilities Index Mark 3: evidence of construct validity for stroke and arthritis in a population health survey. *Med Care* 2000;38:290–9.
- Marra CA, Woolcott JC, Kopec JA, Shojania K, Offer R, Brazier JE, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Soc Sci Med* 2005;60:1571–82.
- Norman GR, Sloan JA, Wywich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003;41:582.
- Cohen J. A power primer. *Psychol Bull* 1992;112:155–9.
- Theiler R, Sangha O, Schaeren S, Michel BA, Tyndall A, Dick W, et al. Superior responsiveness of the pain and function sections of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) as compared to the Lequesne-Algofunctional Index in patients with osteoarthritis of the lower extremities. *Osteoarthritis Cartilage* 1999;7:515–9.
- Angst F, Aeschlimann A, Steiner W, Stucki G. Responsiveness of the WOMAC osteoarthritis index as compared with the SF-36 in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention. *Ann Rheum Dis* 2001;60:834–40.
- Escobar A, Quintana JM, Bilbao A, Aróstegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. *Osteoarthritis Cartilage* 2007;15:273–80.
- Brazier JE, Harper R, Munro J, Walters SJ, Snaith ML. Generic and condition-specific outcome measures for people with osteoarthritis of the knee. *Rheumatology* 1999;38:870–7.
- Jones CA, Voaklander DC, Johnston DW, Suarez-Almazor ME. Health related quality of life outcomes after total hip and knee arthroplasties in a community based population. *J Rheumatol* 2000;27:1745–52.
- Sweeney MK, Marra CA, Kendall R, Li L, Lynd L. An evaluation of the construct validity and responsiveness of generic utility measures in knee arthroplasty for the treatment of osteoarthritis. *J Arthroplasty* 2013 July 11 (E-pub ahead of print).
- Haley SM, Fragala-Pinkham M, Ni P. Sensitivity of a computer adaptive assessment for measuring functional mobility changes in children enrolled in a community fitness programme. *Clin Rehab* 2006;20:616–22.
- Garbuz DS, Tanzer M, Greidanus NV, Masri BA, Duncan CP. The John Charnley Award: metal-on-metal hip resurfacing versus large-diameter head metal-on-metal total hip arthroplasty: a randomized clinical trial. *Clin Orthop Relat Res* 2010;468:318–25.
- Grindrod KA, Marra CA, Colley L, Cibere J, Tsuyuki RT, Esdaile JM, et al. After patients are diagnosed with knee osteoarthritis, what do they do? *Arthritis Care Res* 2010;62:510–5.