

A simple visual analog scale for pain is as responsive as the WOMAC, the SF-36, and the EQ-5D in measuring outcomes of revision hip arthroplasty

A prospective cohort study of 45 patients followed for 2 years

Vasilis Zampelis^{1,3}, Ewald Ornstein¹, Herbert Franzén², and Isam Atroshi^{1,3}

¹Department of Orthopedics, Hässeleholm Hospital, Hässeleholm; ²Department of Orthopedics, Aleris, Ängelholm Hospital, Ängelholm; ³Department of Clinical Sciences, Lund University, Lund, Sweden.

Correspondence: vasilios.zampelis@skane.se

Submitted 13-03-18. Accepted 13-12-07

Background and purpose — Little is known about the comparative performance of patient-reported outcome measures in revision hip arthroplasty. We compared the performance of the WOMAC, the SF-36, the EQ-5D, and a pain-related visual analog scale (VAS) in revision hip arthroplasty.

Methods — 45 patients with aseptic prosthetic loosening following primary hip arthroplasty completed the WOMAC, the SF-36, the EQ-5D, and a VAS for pain—at baseline and 2 years after revision. Responsiveness of the measures was compared with the effect size (with ≥ 0.8 being considered large). Agreement between scales measuring the same type of outcome (pain or physical function) was assessed with the Bland-Altman method.

Results — The mean preoperative scores for the pain and physical function scales of WOMAC and SF-36, EQ-5D index, and VAS for pain improved statistically significantly 2 years after revision. The effect size for the WOMAC pain was 1.7, that for SF-36 pain was 1.4, that for WOMAC physical function was 1.6, that for SF-36 physical function was 0.8, and that for EQ-5D index was 1.2. The VAS for pain had an effect size of 2.1, which was larger than that for SF-36 pain and for the EQ-5D index ($p \leq 0.03$) but not for WOMAC pain ($p = 0.2$). The limits of agreement between WOMAC pain, SF-36 pain, and the VAS scale measuring pain—and between the WOMAC and SF-36 scales measuring physical function—were wide. Internal-consistency reliability was high for the WOMAC and SF-36 scales but low for the EQ-5D.

Interpretation — In patients with first-time revision hip arthroplasty done for aseptic loosening, the WOMAC, SF-36, and EQ-5D showed high responsiveness in measuring patient-reported outcomes and the simple VAS for pain performed equally well.

is also being increasingly used, for example in some national joint registries such as the Swedish Hip Arthroplasty Register (Rolfson et al. 2011). Several studies have shown good validity, reliability, and responsiveness of patient-reported outcome measures in primary hip arthroplasty (Nilsson et al. 2001). Although patient-relevant outcomes with regard to primary hip arthroplasty have been studied extensively, less is known about these outcomes following revision arthroplasty. In previous studies of revision arthroplasty, pain and physical function have been evaluated with clinician-based scores such as the Harris hip score and the Merle d'Aubigné score (Lubbeke et al. 2007). A few studies have used patient-based outcome measures such as the WOMAC and SF-36 (Davis et al. 2006, Lubbeke et al. 2007). Measures that have demonstrated good responsiveness in primary hip arthroplasty do not necessarily perform similarly in revision arthroplasty. Apart from responsiveness, the length of an outcome measure is an important factor with regard to the cost of administration and the response rate, 2 essential elements when using the measure in an arthroplasty registry. Head-to-head comparisons of patient-reported outcome measures in hip arthroplasty can provide important information, but there have been very few studies of that kind.

We compared the performance of the WOMAC, the SF-36, the EQ-5D, and a visual analog scale (VAS) for pain in patients undergoing revision hip arthroplasty. We hypothesized that these measures of patient-reported outcomes would vary in their responsiveness in measuring outcomes.

Patients and methods

Study design

This was a prospective cohort study carried out at one orthopedic department. The inclusion criteria were patients with

In clinical research involving primary hip arthroplasty, health and quality-of-life outcomes have commonly been measured with the WOMAC and the SF-36 questionnaires. The EQ-5D

hip osteoarthritis aged 60 years or older; aseptic prosthetic loosening following a primary total hip arthroplasty; first-time revision (replacement of the stem, cup, or of both components); surgery performed during a 2-year period (March 2006 through February 2008); and a minimum follow-up of 2 years. We included only patients who were revised with impaction bone grafting and with a cemented prosthesis because few patients were revised with other techniques during the study period.

Participants

Of the 57 consecutive patients who were eligible for inclusion, 1 patient died within a year of revision and 1 patient declined follow-up because of poor health. 10 patients (7 men), mean age 74 (62–86) years, could not be included because of missing preoperative and/or follow-up questionnaires. Thus, 45 patients (mean age 74 (60–89) years, 25 men) who completed the WOMAC, SF-36, EQ-5D, and the VAS for pain both before and 2 years after revision were included. Revision involved both components in 26 patients, the cup only in 17, and the stem only in 2. Most patients had 2 comorbidities (hypertension in 27 patients, coronary heart disease or heart failure in 19 patients, diabetes in 6, asthma in 6, and obesity in 1).

Surgery

The revision arthroplasty procedures were performed by 4 experienced orthopedic surgeons. The Exeter stem and/or cup components (Howmedica International, London, UK) were used in 42 patients and 3 patients were operated on with a hybrid technique in which a Revitan stem (Zimmer Inc., Warsaw, IN) was inserted without cement and the Exeter cup was inserted with impaction bone grafting and cement. Because of severe acetabular bone loss, a trabecular metal implant (Zimmer) was used in 3 patients and a Restoration graft augmentation prosthesis ring system was used in 1 patient in conjunction with impaction bone grafting and cement fixation of the components.

None of the 45 patients underwent re-revision within 2 years.

Outcome measures

The patients completed the questionnaires at the hospital prior to admission for surgery. At 2 years after surgery, the questionnaires were sent to all patients by post. During the follow-up visit, the patient handed the completed questionnaires to the examining surgeon (who was usually not the surgeon who performed the revision). The WOMAC, a disease-specific measure of symptoms and activity limitations associated with hip osteoarthritis (Bellamy et al. 1991), consists of 24 items grouped into 3 scales: pain (5 items), stiffness (2 items), and physical function (17 items). WOMAC version 3.1 was used (Bellamy 2005). The WOMAC scores were standardized to range from 0 (worst) to 100 (best). The SF-36 health-status and quality-of-life measure consists of 8 scales measuring

physical and mental health, including a bodily pain scale (2 items) and a physical functioning scale (10 items), each of which is scored from 0 (worst) to 100 (best) (Ware and Sherbourne 1992). Version 1.0 of the SF-36 was used. The EQ-5D health-status and quality-of-life measure consists of 5 items (mobility, self-care, usual activity, pain/discomfort, and anxiety/depression), with 3 possible response levels (no problems, some/moderate problems, extreme problems) (Dolan 1997). A single weighted score, the EQ-5D index, is calculated from the 5 dimensions, ranging from –0.594 (worst) to 1.0. In addition to the EQ-5D index, the EQ-5D includes a VAS for rating of current health status from 0 (worst) to 100 (best).

The baseline and 2-year questionnaires included a VAS for pain; the patients were asked to rate the severity of pain in the hip by marking on a 100-mm horizontal line that ranged from 0 (no pain) to 100 (worst possible pain). The 2-year questionnaire included a satisfaction VAS asking the patients to rate satisfaction with the results of the surgery from 0 to 100, divided into 5 evenly spaced anchors (very satisfied, satisfied, somewhat satisfied, uncertain, dissatisfied). For this study, we reversed the scoring so that a score of 0 would indicate lowest satisfaction and 100 would indicate highest satisfaction.

The WOMAC refers to hip pain or hip problems (with no side specified), the SF-36 and EQ-5D do not refer to a specific site, and the VAS for pain referred to the treated hip. For the WOMAC pain and physical function, the time frame referred to in the questions was “the past week”, for the SF-36 pain it was “the past 4 weeks”, for the SF-36 physical functioning it was “a normal day”, for the EQ-5D index it was “today”, and for the VAS pain it was “the past month”. These time frames are the standard time frames for the respective outcome measures. The EQ-5D, the VAS for pain, and the VAS for satisfaction used in the Swedish Hip Register (Rolfson 2010) were used in the present study.

Statistics

We calculated the preoperative and 2-year postoperative mean score and standard deviation (SD) for the WOMAC and SF-36 scales, the EQ-5D index and health status VAS score, and the VAS score for pain. We used paired t-test to compare the change in scores from baseline to 2 years, and Cohen's *d* as a measure of effect size. The effect size is computed as the mean change in preoperative-to-postoperative score divided by the SD of the preoperative score. We chose the effect size based on baseline SD because it has been shown to be superior to other indices of responsiveness (Norman et al. 2007). Effect sizes of 0.2, 0.5, and 0.8 generally indicate small, medium, and large changes in health, respectively. The 95% confidence intervals for the Hedges-adjusted effect sizes were calculated using Effect Size Generator Pro version 4.0 (Deville 2007). The effect sizes of different scales were compared with regard to whether the differences were statistically significant using a formula based on a z-test (Geoffrey Norman and David Streiner, personal communication). Because the effect size is

Scale scores before and 2 years after revision hip arthroplasty in 45 patients

Scale	Preoperatively ^a	Postoperatively ^a	Score change ^{a, b}	Effect size (95% CI) ^c
WOMAC				
Pain	43 (23)	81 (20)	38 (32)	1.7 (1.2–2.1)
Physical function	35 (18)	64 (22)	29 (26)	1.6 (1.1–2.1)
Stiffness	38 (21)	72 (21)	35 (34)	1.6 (1.2–2.1)
SF-36				
Bodily pain	31 (22)	61 (27)	29 (35)	1.4 (0.91–1.8)
Physical functioning	33 (22)	52 (25)	19 (29)	0.8 (0.40–1.3)
Physical Role	13 (27)	38 (44)	26 (52)	0.9 (0.49–1.4)
Vitality	43 (23)	57 (26)	14 (20)	0.6 (0.62–1.1)
EQ-5D				
Index	0.35 (0.31)	0.74 (0.17)	0.38 (0.32)	1.2 (0.78–1.7)
VAS	44 (25)	73 (19)	29 (29)	1.1 (0.67–1.6)
VAS pain	63 (20)	20 (20)	–43 (26) ^d	–2.1 (–2.7 to –1.6) ^d

^a Values are mean (SD) for preoperative and postoperative scores and score changes. Score ranges: WOMAC and SF-36 from 0 (worst) to 100 (best); EQ-5D index from –0.594 (worst) to 1.0 (best), and VAS pain from 0 (best) to 100 (worst).

^b $p < 0.001$ for all comparisons, except for Physical role ($p = 0.002$).

^c Effect size = mean score change divided by SD of baseline score.

^d Negative value indicates a decrease (improvement) in VAS pain score.

based solely on score distribution, we used patient satisfaction score as an external indicator. We calculated the Pearson correlation coefficient (r) between the patient satisfaction score and the 2-year score, and also the change score (preoperatively to 2 years) for the other measures. As the WOMAC is a specific outcome measure for patients with hip osteoarthritis, it was considered to be the standard scale for comparison purposes.

We examined the agreement between the scales measuring the same type of outcome (pain or physical function) using the Bland-Altman method. We assessed the internal-consistency reliability of the scales with Cronbach's alpha (α) coefficient; values between 0.70 and 0.95 have been proposed to indicate good internal consistency (Terwee et al. 2007). We also examined the scales with regard to the presence of floor and ceiling effects, which are considered to be present when 15% of the patients have the worst possible or best possible scores (Terwee et al. 2007). The distribution of the EQ-5D scores was also examined.

Informed consent was obtained from all patients.

Results

Responsiveness

The mean WOMAC scores for pain, stiffness, and physical function improved statistically significantly 2 years after revision arthroplasty, and the effect size was large for all scales (Table). The mean SF-36 pain, physical functioning, physical role, and vitality scores improved statistically significantly, with the first 3 showing large effect sizes. The improvements in mean scores for social functioning, emotional role, and mental health corresponded to small effect sizes and the mean score for general health perceptions showed little change

(data not shown). The EQ-5D index and EQ-5D VAS score improved significantly and the effect sizes of both were large. The VAS score for pain improved significantly, and had the largest effect size of 2.1.

Comparison of effect size

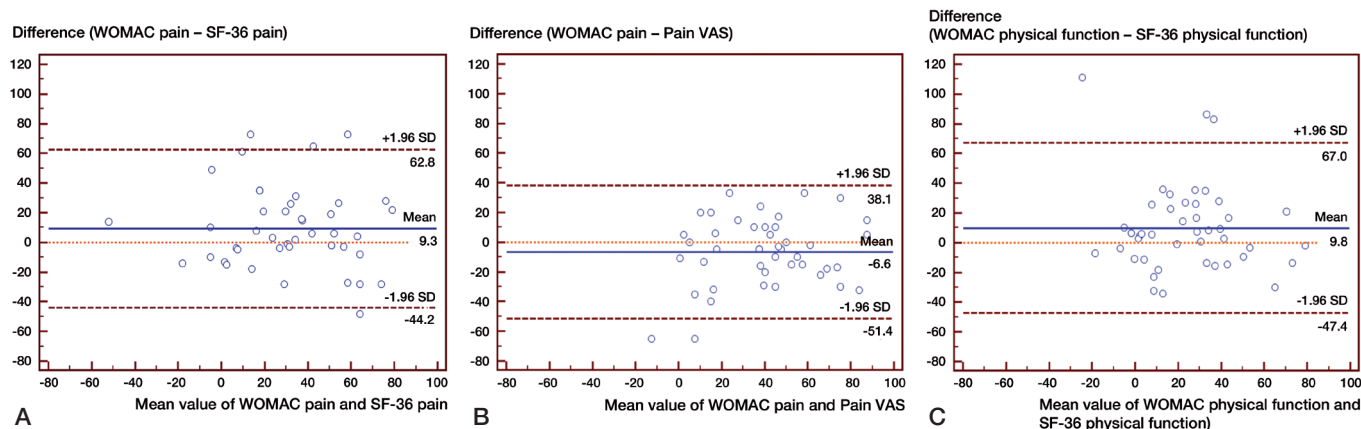
The VAS for pain had a larger effect size than that of the SF-36 pain scale and the EQ-5D index ($p = 0.03$ and $p = 0.01$, respectively) but its effect size was similar to that of the WOMAC pain scale ($p = 0.2$). The WOMAC physical function scale had a larger effect size than the SF-36 physical functioning scale ($p = 0.02$), but this effect size was similar to that of the EQ-5D index ($p = 0.2$).

Correlation with patient satisfaction

The mean VAS satisfaction score at 2 years (where 100 is best possible) was 81 (SD 19). The correlations between satisfaction and the 2-year pain scores were strong for the WOMAC ($r = 0.71$) and for the VAS pain ($r = -0.79$), but moderate for the SF-36 ($r = 0.47$). Correlations between satisfaction and score changes (preoperatively to 2 years) were moderate ($r = 0.58$, -0.49 , and 0.55 , respectively; $p < 0.001$ for all correlations). Correlations between satisfaction and both WOMAC physical function scores and score changes were moderate ($r = 0.49$ and $r = 0.62$; $p < 0.01$) and correlations between satisfaction and SF-36 physical function scores were weak ($r = 0.22$ and $r = 0.25$; $p > 0.1$). There was a moderate correlation between satisfaction and the 2-year EQ-5D index ($r = 0.49$; $p = 0.001$) but the correlation between satisfaction and change in EQ-5D index was weak ($r = 0.28$; $p = 0.07$).

Agreement between scales

The 95% limits of agreement between scales measuring the



Bland-Altman plots for pairs of scales measuring the same type of outcome. The differences between score changes (preoperatively to 2 years) for each pair of scales are plotted against the mean of the 2 score changes. The pairs of scales plotted are the WOMAC and SF-36 pain scores (panel A), the WOMAC and VAS pain scores (panel B), and the WOMAC and SF-36 physical function scores (panel C). The mean difference between score changes and the 95% limits of agreement are shown.

same type of outcome were wide, ranging from –44 to 63 for the WOMAC and SF-36 pain scales, from –47 to 67 for the WOMAC pain scale and the VAS for pain, and from –51 to 38 for the WOMAC and SF-36 physical function scales (Figure).

Internal consistency and ceiling/floor effects

Internal consistency was high for the WOMAC and SF-36 scales but low for the EQ-5D index. The preoperative and postoperative Cronbach α coefficients were both 0.90 for WOMAC pain, were 0.89 and 0.87 (respectively) for SF-36 pain, were both 0.96 for WOMAC physical function, were 0.88 and 0.87 for SF-36 physical functioning, and were 0.53 and 0.62 for EQ-5D. None of the scales showed ceiling effects preoperatively or floor effects postoperatively. A floor effect was found only for the preoperative SF-36 pain, with 16% ($n = 7$) having a worst possible score. For the postoperative scores, a ceiling effect was found in all scales measuring pain and in the EQ-5D; the percentage of best possible scores was 27% in the WOMAC ($n = 12$), it was 18% in the SF-36 ($n = 8$), it was 16% in the VAS ($n = 7$), and it was 18% in the EQ-5D ($n = 8$). No ceiling effects were found in the scales measuring physical function. The distribution of the EQ-5D was bimodal for the preoperative scores but not for the postoperative and change scores.

Discussion

We compared various measures of patient-reported outcomes in revision hip arthroplasty and showed that the WOMAC and SF-36 pain and physical function scales and the EQ-5D index had high responsiveness, and that a simple visual analog scale for was equally responsive or even more responsive. Also, the WOMAC appears to perform better than the SF-36 and EQ-5D in this patient group. These findings should be helpful when

designing clinical studies involving revision hip arthroplasty or when choosing outcome measures for use in registries.

The pain and physical function scales of the WOMAC and SF-36, which have previously been shown to have a large degree of responsiveness in primary hip arthroplasty (Nilsson et al. 2001), showed large effect sizes after revision arthroplasty. A previous study of revision arthroplasty in 126 patients with aseptic loosening of one or both prosthetic components (Davis et al. 2006) found improvement in the mean WOMAC pain score (converted here to a 0–100 scale) from 53 preoperatively to 81 postoperatively, and in the mean physical function score from 49 to 72, which is similar to our findings. In a study that assessed responsiveness of the SF-36 in 67 patients who were evaluated before and 6 months after revision hip arthroplasty (Shi et al. 2010), the effect size for pain was 0.41 and that for physical functioning was 1.2, as compared to our results of 1.4 and 0.8, respectively. In a study of revision hip arthroplasty (Dawson et al. 2001), the mean change in EQ-5D index from preoperatively to 1 year after first-time revision in 128 patients was 0.29, which is less than the improvement shown in our study. A previous study using the Nottingham Health Profile generic health-status measure also found large improvement in pain and moderate improvement in mobility following revision with impaction bone grafting and cement (Atroshi et al. 2004).

When comparing the WOMAC, SF-36, and EQ-5D results and interpreting the differences between scales that measure the same type of outcome (such as pain or physical function), the different time frames used in these measures should be considered. Similarly, the scores may be influenced by whether a scale refers to the treated hip, to the hip but without specifying which side, or to pain or function in general without specifying a location. These factors may at least partly explain the wide limits of agreement between the individual scores on the Bland-Altman plots.

Apart from good responsiveness, an outcome measure should be of proven validity and reliability, and should preferably be inexpensive and easy to administer. The disadvantages of the WOMAC and SF-36 include the length of the questionnaire with the risk of lower response rates and missing values, and possible licensing costs. Like the SF-36, the EQ-5D is applicable to a wide range of health conditions and provides a single index that can be used in health economic evaluation.

One limitation of the present study was that we did not assess the test-retest reliability of the scales. However, the patient-reported outcome measures that we used are well established and have been extensively tested previously regarding reliability and validity under various conditions. Moreover, the internal-consistency reliability was high for all scales except the EQ-5D.

Another limitation was the small sample size, which may restrict the extent to which the findings can be generalized. We wanted to compare the responsiveness of patient-reported outcome measures following revision hip arthroplasty in a well-defined patient group, and therefore included only patients with aseptic loosening who underwent first-time revision. In patients fulfilling these criteria, revision arthroplasty appears to result in a significant improvement in quality of life, which we believe is an important finding. Future studies are needed to examine outcomes in a larger population. A limitation shared by other, similar studies is that patient-reported outcomes measures may be analytically problematic when statistical assumptions are not fulfilled, such as departure from Gaussian distribution due to floor and ceiling effects, and bi- or multimodal distribution. However, floor and ceiling effects were uncommon and, in particular, the change scores did not show pronounced distribution problems, so this would not be a major issue.

The simple VAS for pain showed high responsiveness, which was equal to that shown by the WOMAC and SF-36 scales that measure pain. It might be argued that, in response to pain, patients may lower their activity level; therefore, measuring physical function may be equally important. In both the WOMAC and the SF-36, the pain scales had larger effect sizes than the physical function scales. Also, there was a stronger correlation between patient satisfaction and pain scores.

We have shown that in evaluating outcomes of revision hip arthroplasty, a VAS for pain is a highly responsive measure that is simple to use and that may enhance the practicality of outcome measurement. Responsiveness is, however, not the only factor to consider, and researchers may have other reasons for choosing longer disease-specific or general health-status measures or a combination of measures.

We thank Marie Davidsson, research coordinator at the Department of Orthopedics, Hässleholm Hospital, for administrative assistance.

This research was supported by Hässleholm Hospital Organization. The authors have no competing interests directly or indirectly related to this work.

Atroshi I, Ornstein E, Franzén H, Johnsson R, Stefánssdóttir A, Sundberg M. Quality of life after hip revision with impaction bone grafting on a par with that 4 years after primary cemented arthroplasty. *Acta Orthop Scand* 2004; 75 (6): 677-83.

Bellamy N. The WOMAC Knee and Hip Osteoarthritis Indices: development, validation, globalization and influence on the development of the AUSCAN Hand Osteoarthritis Indices. *Clin Exp Rheumatol (Suppl)* 2005; 23 (5): 148-53.

Bellamy N, Wells G, Campbell J. Relationship between severity and clinical importance of symptoms in osteoarthritis. *Clin Rheumatol* 1991; 10 (2): 138-43.

Davis A M, Agnidis Z, Badley E, Kiss A, Waddell J P, Gross A E. Predictors of functional outcome two years following revision hip arthroplasty. *J Bone Joint Surg (Am)* 2006; 88 (4): 685-91.

Dawson J, Fitzpatrick R, Frost S, Gundle R, McLardy-Smith P, Murray D. Evidence for the validity of a patient-based instrument for assessment of outcome after revision hip replacement. *J Bone Joint Surg (Br)* 2001; 83 (8): 1125-9.

Devilly G J. ClinTools Software for Windows: Version 4.1. 2007. Available from: www.clintools.com.

Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997; 35 (11): 1095-108.

Lubbeke A, Katz J N, Perneger T V, Hoffmeyer P. Primary and revision hip arthroplasty: 5-year outcomes and influence of age and comorbidity. *J Rheumatol* 2007; 34 (2): 394-400.

Nilsdotter A K, Roos E M, Westerlund J P, Roos H P, Lohmander L S. Comparative responsiveness of measures of pain and function after total hip replacement. *Arthritis Rheum* 2001; 45 (3): 258-62.

Norman G R, Wyrwich K W, Patrick D L. The mathematical relationship among different forms of responsiveness coefficients. *Qual Life Res* 2007; 16 (5): 815-22.

Rolfson O. Patient-reported outcome measures and health-economic aspects of total hip arthroplasty: a study of the Swedish Hip Arthroplasty Register. University of Gothenburg, Gothenburg 2010.

Rolfson O, Rothwell A, Sedrakyan A, Chenok K E, Bohm E, Bozic K J, Garellick G. Use of patient-reported outcomes in the context of different levels of data. *J Bone Joint Surg (Am) (Suppl)* 2011; 93: 66-71.

Shi H Y, Chang J K, Wong C Y, Wang J W, Tu Y K, Chiu H C, Lee K T. Responsiveness and minimal important differences after revision total hip arthroplasty. *BMC Musculoskelet Disord* 2010; 11: 261.

Terwee C B, Bot S D, de Boer M R, van der Windt D A, Knol D L, Dekker J, Bouter L M, de Vet H C. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60 (1): 34-42.

Ware J E Jr, Sherbourne C D. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992; 30 (6): 473-83.

VZ: design and conduction of the study, data analysis, and writing of the manuscript. EO: design and conduction of the study and critical revision of the manuscript. HF: conduction of study and critical revision of the manuscript. IA: design and conduction of the study, data analysis, and writing and revision of the manuscript.