**Clinical Orthopaedics
and Related Research®**
A Publication of The Association of Bone and Joint Surgeons®

CrossMark

SYMPOSIUM: 2014 INTERNATIONAL HIP SOCIETY PROCEEDINGS

Women Demonstrate More Pain and Worse Function Before THA but Comparable Results 12 Months After Surgery

Anne F. Mannion PhD, Franco M. Impellizzeri PhD, Florian D. Naal MD,
Michael Leunig MD, PhD

Abstract

Background   Many studies report differences in patient-reported outcome measures (PROMs) for men and women undergoing total hip arthroplasty (THA). Few studies have evaluated whether these are explained by corresponding differences in important preoperative factors.

Questions/purposes   (1) Are there differences between men and women in PROM scores preoperatively and 12 months after THA? (2) Do baseline differences in comorbidity, age, body mass index (BMI), and mental health status explain these differences in PROM scores?

Methods   Preoperatively, 300 patients completed the Oxford Hip Score (OHS), WOMAC, and SF-12; 261 (86%) of them (129 women, 64 ± 11 years; 132 men, 66 ± 10 years) completed the same questionnaires 12 months postoperatively and also rated the acceptability of their current symptoms and change in general health.

Results   Preoperatively, women showed worse scores than men in the OHS (−1.9; 95% confidence interval, −3.6 to −0.3) and WOMAC (−6.3; −10.9 to −1.7). At 12 months postoperatively, the absolute scores for all PROMs were not significantly different. After controlling for BMI, age, comorbidity, SF-12 mental health scores, and sociodemographic characteristics, the baseline differences remained.

Conclusions   Surgeons may be more reluctant to operate on women than men because they perceive that, because of their worse baseline status, women are likely to have worse outcomes; however, given that we found no evidence for differences in patient-reported outcomes at 12 months, these suspicions would appear to be unfounded. Women and men can be expected to benefit to a similar extent from THA.

Level of Evidence   Level III, therapeutic study.

A. F. Mannion, F. M. Impellizzeri (✉)
Department of Research and Development, Schulthess Clinic,
Lengghalde 2, Zurich 8008, Switzerland
e-mail: franco.impellizzeri@gmail.com

F. D. Naal
Department of Orthopaedic Surgery, Schulthess Clinic, Zurich,
Switzerland

M. Leunig
Department of Orthopaedic Surgery, University of Berne, Berne,
Switzerland

Introduction

THA is one of the most common orthopaedic surgical procedures and generally is associated with high success rates. A good patient outcome after orthopaedic surgery is usually dependent, among other things, on an appropriate indication for surgery [34, 37]. However, there is currently

🖄 Springer

no consensus on objective indications for THA; most experts consider the severity of pain and disability to be important [33], but there is little agreement regarding the actual severity of symptoms that indicates the need for surgery [11].

From the patient's perspective, pain—and the hope of its resolution—is one of the main driving forces for seeking care [19]. Pain is always subjective, and it has long been known that differences exist between the genders in their response to painful stimuli. Women typically have a lower pain threshold (that is, a greater sensitivity to experimentally induced pain) than men [14]. This may be mediated by sociocultural factors (including age, family history, and gender roles), psychological factors (such as anxiety, depression, cognition, behavior), or biological factors (like genetics, sex hormones, endogenous pain inhibition), all of which can also interact with each another in complex ways [27].

Gender may be an important factor to consider when examining indications for surgery, because differences in pain sensitivity may influence accepted "thresholds for surgery" in relation to pain and disability. Similarly, gender may have an influence on the interpretation of patient-rated outcomes when judging the success of surgery. Many [18, 30, 32] but not all [31] studies have shown that women have higher pain levels and worse function when they present for surgery for the treatment of degenerative musculoskeletal conditions. However, it is unclear whether this necessarily translates into a worse overall outcome after surgery or a lower level of satisfaction with treatment. Moreover, the role of baseline factors such as age, mental health status, comorbidity, body mass index (BMI), and patient expectations in explaining gender differences in patient-reported outcome measure (PROM) scores is not well known. Because these factors are themselves typically correlated with PROM scores [12, 26, 28, 39] and may differ between men and women preoperatively, they represent potential confounding effects in any analysis of gender differences in PROM scores.

We therefore asked the following questions: (1) Are there significant differences in PROM scores between men and women preoperatively and 12 months after THA? (2) Do baseline differences in age, mental health status, comorbidity, BMI, and sociodemographic characteristics explain any significant differences in PROM scores?

Patients and Methods

This was a retrospective analysis of prospectively collected data from patients undergoing THA in a tertiary care center (Specialized Orthopaedic Hospital) in Switzerland between July 2008 and December 2009. Inclusion criteria were: German-speaking, end-stage osteoarthritis requiring primary THA, and surgery planned to be carried out by one of the three most experienced hip surgeons of the clinic. The only exclusion criterion was the presence of neurological disorders. A total of 300 patients (150 men, 150 women), equivalent to 84% of all who were eligible, completed the questionnaires preoperatively and were thus included in the study. Two hundred sixty-one (86%) of those completing the baseline questionnaires (129 women, 64 ± 11 years; 132 men, 66 ± 10 years; Table 1) also completed the followup questionnaire 12 months postoperatively. The main reasons for loss to followup reported by the patients contacted by phone were a lack of time and the questionnaire battery being too long. Fewer than 3% were no longer interested or had pain for other reasons and hence did not consider it appropriate to complete the questionnaires. Two men and two women underwent revision during the study period (within 12 months). Four men underwent revision from 14 months to 5 years after the first operation.

The patients had clinical and radiographic evidence of end-stage hip osteoarthritis and underwent THA using either a posterolateral or direct anterior approach. Postoperative precautions to prevent dislocation were given for the posterolateral approach, but otherwise there was no difference in the postoperative care. Typically, the patients had undergone prior conservative treatment comprising physiotherapy, oral nonsteroidal antiinflammatory drugs, and therapeutic hip injections. One to 3 weeks before admission, the patients were asked to complete a preoperative questionnaire booklet that included the instruments used in the present study (see subsequently). These were completed by the patient at home and brought into the hospital on the day of admission. Twelve months postoperatively, the questionnaire booklet was sent out by mail to those who had returned a preoperative questionnaire and had not undergone any surgery on the spine or lower extremities in the preceding 4 months with the request to complete it and return it using the stamped, addressed envelope enclosed. At its inception, the study did not require ethics committee approval in our country, because it concerned the reuse of routinely collected data with the written consent of the patient.

At baseline, patients completed crossculturally adapted and validated German versions of the following questionnaires: (1) Oxford Hip Score (OHS) [7, 29], which consists of 12 questions on hip pain and function in the past 4 weeks. Each item uses a 5-point response scale with values from 0 to 4. An overall score is created by summing the responses to each of the 12 questions. The total score can range from 0 to 48, where 0 is the worst possible score indicating severe hip symptoms and 48 is the best score suggesting excellent joint function. (2) WOMAC, a 24-item self-administered, disease-specific instrument for

Table 1. Baseline data of the patients

Variable	Women	Men	p value
Number (%)	129 (50%)	132 (50%)	
Age (years)	63.9 (11.0)	66.0 (9.7)	0.408
BMI (kg/m^{-2})	25.5 (4.4)	27.4 (3.8)	< 0.001
Number of comorbidities	1.2 (1.1)	1.2 (1.0)	1.000
Mental health (SF-12 MCS) score	51.7 (10.8)	54.8 (9.6)	0.015
ASA			0.343*
I category	9.8%	13.7%	
II category	63.4%	64.7%	
III category	26.8%	21.6%	
Education			0.005*
At most, high school	23%	11%	
Professional school or university	77%	89%	
Civil status			< 0.001*
Married	53%	78%	
Single, separated, divorced, widowed	47%	22%	
Employment status			< 0.001*
Full- or part-time	32%	39%	
Homemaker	25%	2%	
Retired, disabled, other	43%	59%	
Contralateral hip			0.789*
Involved but not operated	23%	23%	
Involved and operated	19%	21%	
Not involved	58%	56%	

Response options: much better (4), better (3), somewhat better (2), unchanged (1), worse (0); * p values from chi square indicating significance of the gender difference in distribution of response categories for the given variable; BMI = body mass index; MCS = mental component summary; ASA = American Society of Anesthesiologists.

assessing pain, stiffness, and physical function in patients with osteoarthritis (OA) [3, 40]. We used the version with the 0 to 10 numeric scale. Five items address pain with a total score for this subscale ranging from 0 to 50. Two items address stiffness (subscale score range, 0–20 points) and 17 items assess physical function (subscale score range, 0–170). The scores for each subscale were converted to a 0 to 100 scale with the highest scores indicating the best health status. (3) SF-12, a 12-item self-administered measure of general quality of life [15, 42]. Scores are transformed into two weighted summary scores for physical function (physical component summary [PCS]) and mental health (mental component summary [MCS]). The standardized scores for the SF-12 range from 0 to 100 with higher scores indicating a better health state. We used the SF-12 MCS for comparing the mental health between genders.

At 12 months' followup, the patients again completed the OHS, WOMAC, and SF-12 and also rated the acceptability of their current symptoms and change in general health by answering the following questions: (1) "If you had to spend the rest of your life with the symptoms you have now, how would you feel about it?" Response options: very satisfied, somewhat satisfied, neither satisfied nor dissatisfied, somewhat dissatisfied, very dissatisfied. (2) "Compared with 1 year ago, how would you rate your health in general now? Would you say it is: (a) much better now than 1 year ago, (b) somewhat better now than 1 year ago, (c) about the same as 1 year ago, (d) somewhat worse now than 1 year ago, (e) much worse than 1 year ago."

Question 1 is the Symptom-Specific Well-Being (SSWB) item and has been used by the Patient Outcome Research Teams in studies on coronary heart disease, prostate disease, and cataract disease. The SSWB was also used in the Lumbar Cluster of the American Academy of Orthopaedic Surgeons and is included in the Core Outcome Measures Index validated for use in patients with femoroacetabular impingement and in patients undergoing THA [9, 20, 21]. The health change question is question 2 of the SF-36 [43].

Descriptive data included proportions (%) for categorical data and means and SDs for the instrument scores. The significance of gender differences was analyzed using independent t-tests. Differences were also reported as mean

Table 2. Comparison of the unadjusted PROM scores between women and men preoperatively

Instrument	Women		Men		p value	Mean difference	95% CI	
	Mean	SD	Mean	SD			Lower	Upper
OHS (0–48)	24.2	7.3	26.1	7.3	0.023	−1.9	−3.6	−0.3
WOMAC pain (0–100)	47.6	20.9	53.9	21.4	0.011	−6.3	−11.2	−1.5
WOMAC stiffness (0–100)	45.7	21.8	53.4	22.4	0.003	−7.6	−12.7	−2.6
WOMAC function (0–100)	48.9	21.2	54.4	21.0	0.026	−5.5	−10.3	−0.7
WOMAC total score (0–100)	48.4	19.5	54.6	20.2	0.007	−6.3	−10.9	−1.7

Higher values indicate better state/condition; PROM = patient-reported outcome measure; CI = confidence interval; OHS = Oxford Hip Score.

Table 3. Comparison of the unadjusted PROM scores between women and men 12 months postoperatively

Instruments	Women		Men		Difference	p value	95% CI	
	Mean	SD	Mean	SD			Lower	Upper
OHS	42.6	7.1	43.2	7.4	−0.6	0.517	−2.4	1.2
WOMAC pain	94.8	10.2	93.9	13.9	0.9	0.577	−2.2	3.9
WOMAC stiffness	92.3	11.1	92.2	12.9	1.2	0.550	−2.6	4.9
WOMAC function	91.8	11.8	91.5	13.7	0.3	0.835	−2.8	3.5
WOM total score	92.3	10.9	91.7	13.1	0.5	0.732	−2.5	3.5

Higher values indicate better state/condition; PROM = patient-reported outcome measure; CI = confidence interval; OHS = Oxford Hip Score.

differences and the corresponding 95% confidence intervals. Multiple regression analyses were carried out with the chosen PROMs (OHS and WOMAC) as the dependent variables and with all the variables identified as possible "confounders" (age, BMI, comorbidity, and mental health [SF-12 MCS scores] and education, civil status, and employment status [transformed as dummy variables]) being force-entered as independent variables together with gender. The assumptions of all statistical tests were verified before analysis. The aforementioned analyses were carried out using SPSS (Version 17; SPSS Inc, Chicago, IL, USA) and Medcalc (MedCalc Statistical Software, Mariakerke, Belgium). Probability values < 0.05 were considered to be statistically significant.

Results

Differences Between Men and Women in PROM Scores Before and After Surgery

Preoperatively, women showed worse scores than men for all of the PROMs (Table 2). Twelve months postoperatively, the absolute scores for all PROMs were not

significantly different between the men and women (Table 3). The PROM improvements (preoperatively to 12 months postoperatively) were not different between men and women for the OHS, whereas women showed greater improvement in WOMAC pain (6.5; 95% confidence interval [CI], 1.1–12.0), stiffness (7.9; 1.8–14.1), and total score (5.9; 0.7–11.0) (Fig. 1). However, when adjusting for baseline scores of the corresponding instrument, there were no significant differences between men and women for any of their change scores (Table 4).

No significant differences were found in the proportion of patients in each response category for the SSWB (p = 0.806) and health changes (p = 0.395) questions (Figs. 2A and 2B, respectively).

Differences in Baseline Factors in Relation to Differences in PROMs Between Men and Women

At baseline, the women had lower BMI (−1.9; 95% CI, −1.0 to −2.8) and lower SF-12 MCS scores (−3.1; −0.8 to 5.4) than the men (Table 1).

Significant differences in preoperative PROM scores between men and women were present for all PROMS

even when controlling for age, BMI, comorbidity, SF-12 MCS scores, and sociodemographic characteristics (Table 5).

There was no significant difference between men and women for the condition-specific PROM scores (OHS and WOMAC) 12 months postoperatively when controlling for baseline values of the corresponding PROM, age, BMI, comorbidity, SF-12 MCS, and sociodemographic characteristics (Table 6).

Discussion

Severe pain and disability are generally considered to be the primary indications for THA [11]. Some previous studies have reported differences in PROMs in men and women undergoing THA, and hence, sex and gender may
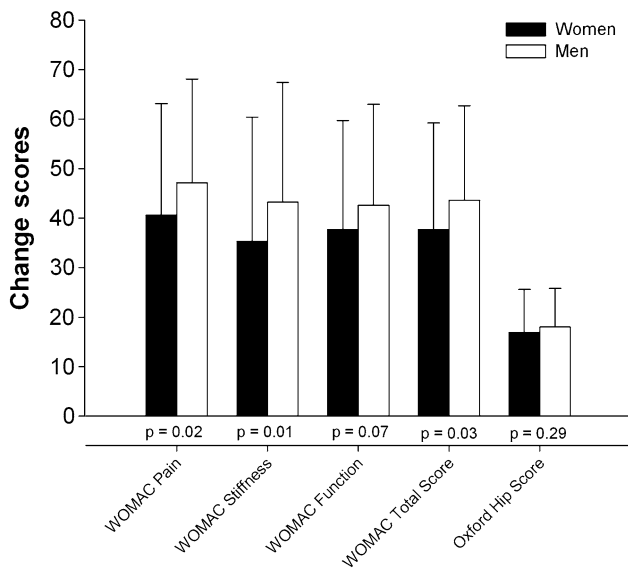

Fig. 1 The figure shows the change scores for the PROMs from preoperatively to 12 months postoperatively for women and men.
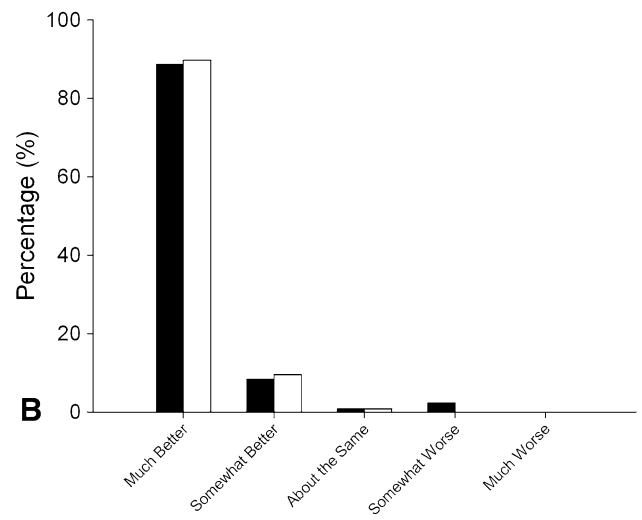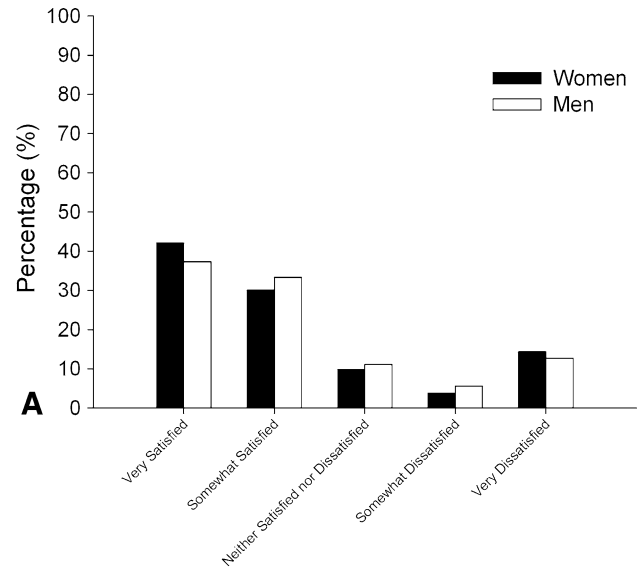

Fig. 2A−B The figure shows the global treatment outcome ratings of the women and men 12 months postoperatively. (A) If you had to spend the rest of your life with the symptoms you have now, how would you feel about it? (B) Compared to 1 year ago, how would you rate your health in general now?

Table 4. PROM change scores from preoperatively to 12 months postoperatively, adjusted by baseline values

Instruments	Women		Men		p value	Difference	95% CI	
	Mean	SD	Mean	SD			Lower	Upper
OHS	17.4	0.6	17.5	0.6	0.958	0.0	−1.7	1.7
WOMAC pain	44.7	1.1	43.0	1.1	0.255	1.7	−1.2	4.7
WOMAC stiffness	40.0	1.4	38.6	1.4	0.443	1.5	−2.3	5.3
WOMAC function	40.6	1.1	39.8	1.1	0.615	0.8	−2.3	3.9
WOMAC total score	41.2	1.0	40.2	1.1	0.501	1.0	−2.0	4.0

PROM = patient-reported outcome measure; CI = confidence interval; OHS = Oxford Hip Score.

Table 5. Results of the multiple regression analysis showing the influence of gender on each of the PROMs at baseline while controlling for the covariates comorbidity, age, BMI, SF-12 MCS, and sociodemographic differences

Predictors	OHS		WOMAC total	
	B (95% CI)	p value	B	p value
Gender	−2.5 (−4.3 to −0.6)	0.008	−7.1 (−12.1 to −2.0)	0.006
Age	−0.1 (−0.2 to 0.1)	0.274	−0.1 (−0.3 to 0.2)	0.743
BMI	−0.3 (−0.5 to −0.1)	0.011	−0.9 (−1.5 to −0.3)	0.002
Number of comorbidities	−1.0 (−1.8 to −0.2)	0.011	−3.1 (−5.2 to −1.0)	0.004
SF-12 MCS	0.2 (0.1–0.3)	< 0.001	0.5 (0.3–0.8)	< .001
High education versus lower education	0.1 (−2.2 to 2.3)	0.958	−1.8 (−7.9 to 4.3)	0.567
Living with partner versus living alone	−1.1 (−2.8 to 0.7)	0.235	−1.0 (−5.9 to 3.9)	0.681
Full-time employed versus retired	0.5 (−1.7 to 2.7)	0.667	1.0 (−5.1 to 7.0)	0.759
Homeworker versus retired	1.4 (−1.3 to 4.2)	0.984	0.6 (−7-1 to 8.2)	0.887

PROMs = patient-reported outcome measures; BMI = body mass index; MCS = mental component summary; OHS = Oxford Hip Score.

Table 6 Results of the multiple regression analysis showing the influence of gender on the 12-month postoperative scores for each of the PROMs while controlling for baseline scores of the given PROM and the covariates comorbidity, age, BMI, SF-12 MCS, and sociodemographic differences

Predictors	OHS		WOMAC	
	B	p value	B	p value
Gender	−0.2 (−2.2 to 1.9)	0.873	−0.5 (−3.9 to 3.0)	0.779
Baseline PROM	0.3 (0.2−0.4)	< 0.001	0.1 (0.0−0.2)	0.010
Age	−0.0 (−0.1 to 0.1)	0.959	−0.1 (−0.3 to 0.1)	0.513
BMI	−0.2 (−0.4 to 0.1)	0.166	−0.5 (−0.9 to −0.2)	0.018
Number of comorbidities	−0.4 (−1.3 to 0.5)	0.397	−1.2 (−2.7 to 0.4)	0.141
SF-12 MCS	0.1 (0.0–0.2)	0.159	0.1 (−0.1 to 0.2)	0.534
High education versus lower education	−0.7 (−3.2 to 1.8)	0.576	−0.5 (4.6–3.6)	0.812
Living with partner versus living alone	0.5 (−1.5 to 2.4)	0.643	1.8 (−1.5 to 5.1)	0.299
Full-time employed versus retired	0.8 (−1.7 to 3.3)	0.529	0.8 (−3.8 to 5.0)	0.725
Homeworker versus retired	−0.7 (−3.9 to 2.4)	0.640	2.3 (−2.9 to 7.6)	0.382

PROMs = patient-reported outcome measures; BMI = body mass index; MCS = mental component summary; OHS = Oxford Hip Score.

be important factors to consider when examining indications for surgery and interpreting outcomes after surgery [18, 30, 32]. We found that women undergoing THA had worse preoperative PROM scores than did men, and these were not explained by differences in BMI, age, comorbidity, or mental health scores. Women showed a greater absolute improvement in PROM scores for many instruments, but this was predominantly the consequence of their worse preoperative scores. Controlling for all baseline variables, there were no gender-related differences in outcomes 1 year after THA.

The study had some limitations. First, we did not obtain any objective measures of the severity of hip arthritis to see whether lower baseline PROM scores were matched by more severe radiographic findings in women. Some previous studies suggest that the grade of radiographic OA is unrelated to self-rated pain and function (WOMAC scores),

at least for moderate to severe OA [31], whereas others show a relationship between radiographic and clinical variables [5, 10]. Second, this study was conducted in Switzerland; differences between countries in their healthcare systems or social structure might govern whether gender differences in baseline PROMs are observed. Furthermore, other psychosociocultural factors that we did not measure such as anxiety, catastrophizing, coping strategies, and the like may have influenced the significant differences seen in baseline PROM scores. A final limitation—and one that besets other studies of this type—is the inability to interpret the gender differences in relation to "minimal clinically important differences" (MCID). To our knowledge, no cross-sectional studies have evaluated what constitutes a clinically important score difference between groups. The MCID values reported in the literature refer to changes over time (mostly for improvement) within a

group, usually in relation to some external criterion of change, ie, they reflect "change" rather than "differences." In addition, the MCID depends on the characteristics of the population and the treatment used and should be used only for general guidance in interpreting individual, not group change [24]. Depending on the method used for calculation, the MCID values reported in the literature range from 6 to 29 for WOMAC and from 1 to 11 for OHS [2, 6–8, 13, 16, 36, 41, 44, 45]; if these values are used to interpret the clinical relevance of our findings (despite the previously mentioned caveats), the CIs of our baseline gender differences would include clinically important differences, whereas the differences at 12 months are highly unlikely to be clinically relevant.

Our findings confirmed previous reports of greater pain and disability in women than men presenting for THA [18, 30, 32]. This finding extends to other medical and orthopaedic conditions requiring surgery and is often considered to indicate that women present for surgery at a later stage in their disease or (for various reasons) choose to wait longer than men before deciding to undergo surgery. Reasons may include a more risk-averse approach to treatment in women (persevering with conservative measures rather than undergoing the risks of surgery), barriers to access (being deterred from undergoing surgery by family, clinicians, or self), personal preferences, or familial pressures and caregiving responsibilities [22, 23]. Indeed, previous reports suggest a threefold greater underuse of arthroplasty for severe arthritis in women than men [17]. However, we cannot rule out the alternative explanations that women are simply more sensitive to the pain associated with a given severity of disease or that there are sociocultural gender differences in the reporting of pain. Social learning has a strong influence on the pain response, and it is generally perceived that men are less willing than women to report pain [38]. Some authors contend that, because PROM scores do not differ between men and women after surgery, overreporting and/or enhanced pain sensitivity in women is an unlikely explanation for the differences seen [23]; however, THA is such a successful intervention, leaving most patients with negligible pain 12 months postoperatively, that any residual differences between men and women would likely not be perceptible.

In the present study we evaluated whether the differences seen between men and women were explained by differences in various baseline factors at presentation. It is known that distressed patients have worse scores than nondistressed subjects for most PROMs [26] and that women tend to display worse preoperative mental health scores than men [25]. Hence, mental health might theoretically mediate the observed gender difference. However, this could not be substantiated in the present study; although mental health was a unique predictor of baseline PROM scores, and women presented with worse mental health than men, the differences in PROM scores remained, even after controlling for mental health scores. BMI and the number of comorbidities had a significant influence on baseline PROMs, but sex or gender remained a significant independent (unique) predictor in the model alongside these variables.

Women undergoing THA had worse preoperative PROM scores than did men, and these were not explained by differences in BMI, age, comorbidity, or mental health scores. Whether this is a reflection of sex or gender differences in pain reporting, of sensitivity to the same physical insult (such as the severity of OA), or of women simply waiting longer and developing more extreme symptoms before presenting for surgery cannot be ascertained from the present study.

However, these differences need to be considered when evaluating indications and thresholds for surgery, because many guidelines are based on the degree of pain and function on presentation [1, 35]. Further studies to elucidate the nature of the differences therefore are required; if they reflect barriers to access, these should be addressed and eliminated. Previous studies have suggested that physicians identify women's complaints as psychosomatic more frequently than they do men's [4]. Surgeons may be more reluctant to operate on women than men because they perceive that women are likely to have worse outcomes; however, given that we found no significant differences in patient-reported outcomes at 12 months postoperatively, these suspicions would appear to be unfounded. Women and men can be expected to benefit to a similar extent from THA.

References

1. Allepuz A, Quintana JM, Espallargues M, Escobar A, Moharra M, Arostegui I. Relationship between total hip replacement appropriateness and surgical priority instruments. *J Eval Clin Pract.* 2011;17:18–25.
2. Beard DJ, Harris K, Dawson J, Doll H, Murray DW, Carr AJ, Price AJ. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. *J Clin Epidemiol.* 2015;68:73–79.
3. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. 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.* 1988;15:1833–1840.
4. Bernstein B, Kane R. Physicians' attitudes toward female patients. *Med Care.* 1981;19:600–608.
5. Croft P, Cooper C, Wickham C, Coggon D. Defining osteoarthritis of the hip for epidemiologic studies. *Am J Epidemiol.* 1990;132:514–522.

6. Davidson D, Yen YM, Jette AM, Olarsch S, Kim YJ. Responsiveness of the SF-36 and WOMAC following periacetabular osteotomy for acetabular dysplasia. *J Bone Joint Surg Am.* 2011;93:2214–2218.

7. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br.* 1996;78:185–190.

8. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol.* 2006;59:1033–1039.

9. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G. Outcome measures for low back pain research. A proposal for standardized use. *Spine.* 1998;23:2003–2013.

10. Dowsey MM, Nikpour M, Dieppe P, Choong PF. Associations between pre-operative radiographic osteoarthritis severity and pain and function after total hip replacement: radiographic OA severity predicts function after THR. *Clin Rheumatol.* 2014. doi:10.1007/s10067-014-2808-7.

11. Dreinhofer KE, Dieppe P, Sturmer T, Grober-Gratz D, Floren M, Gunther KP, Puhl W, Brenner H. Indications for total hip replacement: comparison of assessments of orthopaedic surgeons and referring physicians. *Ann Rheum Dis.* 2006;65:1346–1350.

12. Dunbar MJ, Robertsson O, Ryd L. What's all that noise? The effect of co-morbidity on health outcome questionnaire results after knee arthroplasty. *Acta Orthop Scand.* 2004;75:119–126.

13. Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities osteoarthritis index questionnaire and global assessments in patients with osteoarthritis. *J Rheumatol.* 2000;27:2635–2641.

14. Fillingim RB, King CD, Ribeiro-Dasilva MC, Rahim-Williams B, Riley JL. Sex, gender, and pain: a review of recent clinical and experimental findings. *J Pain.* 2009;10:447–485.

15. Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE, Bullinger M, Kaasa S, Leplege A, Prieto L, Sullivan M. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol.* 1998;51:1171–1178.

16. Gosens T, Hoefnagels NH, de Vet RC, Dhert WJ, van Langelaan EJ, Bulstra SK, Geesink RG. The "Oxford Heup Score": the translation and validation of a questionnaire into Dutch to evaluate the results of total hip arthroplasty. *Acta Orthop.* 2005;76:204–211.

17. Hawker GA, Wright JG, Coyte PC, Williams JI, Harvey B, Glazier R, Badley EM. Differences between men and women in the rate of use of hip and knee arthroplasty. *N Engl J Med.* 2000;342:1016–1022.

18. Holtzman J, Saleh K, Kane R. Gender differences in functional status and pain in a Medicare population undergoing elective total hip arthroplasty. *Med Care.* 2002;40:461–470.

19. Impellizzeri FM, Mannion AF, Naal FD, Leunig M. Acceptable symptom state after surgery for femoroacetabular impingement compared with total hip arthroplasty. *Hip Int.* 2013;23(Suppl 9):S54–60.

20. Impellizzeri FM, Mannion AF, Naal FD, Leunig M. A Core Outcome Measures Index (COMI) for patients undergoing hip arthroplasty. *J Arthroplasty.* 2013;28:1681–1686.

21. Impellizzeri FM, Mannion AF, Naal FD, Leunig M. Validation of the Core Outcome Measures Index in Patients With Femoroacetabular Impingement. *Arthroscopy.* 2015;31:1238–1246.

22. Karlson EW, Daltroy LH, Liang MH, Eaton HE, Katz JN. Gender differences in patient preferences may underlie differential utilization of elective surgery. *Am J Med.* 1997;102:524–530.

23. Katz JN, Wright EA, Guadagnoli E, Liang MH, Karlson EW, Cleary PD. Differences between men and women undergoing major orthopedic surgery for degenerative arthritis. *Arthritis Rheum.* 1994;37:687–694.

24. Katz NP, Paillard FC, Ekman E. Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. *J Orthop Surg Res.* 2015;10:24.

25. Kostamo T, Bourne RB, Whittaker JP, McCalden RW, MacDonald SJ. No difference in gender-specific hip replacement outcomes. *Clin Orthop Relat Res.* 2009;467:135–140.

26. Lavernia CJ, Alcerro JC, Brooks LG, Rossi MD. Mental health and outcomes in primary total joint arthroplasty. *J Arthroplasty.* 2012;27:1276–1282.

27. Lund I, Lundeberg T. Is it all about sex? Acupuncture for the treatment of pain from a biological and gender perspective. *Acupunct Med.* 2008;26:33–45.

28. Mancuso CA, Sculco TP, Salvati EA. Patients with poor preoperative functional status have high expectations of total hip arthroplasty. *J Arthroplasty.* 2003;18:872–878.

29. Naal FD, Sieverding M, Impellizzeri FM, von Knoch F, Mannion AF, Leunig M. Reliability and validity of the cross-culturally adapted German Oxford hip score. *Clin Orthop Relat Res.* 2009;467:952–957.

30. Ng CY, Ballantyne JA, Brenkel IJ. Quality of life and functional outcome after primary total hip replacement. A five-year follow-up. *J Bone Joint Surg Br.* 2007;89:868–873.

31. Nilsdotter AK, Aurell Y, Siosteen AK, Lohmander LS, Roos HP. Radiographic stage of osteoarthritis or sex of the patient does not predict one year outcome after total hip arthroplasty. *Ann Rheum Dis.* 2001;60:228–232.

32. Nilsdotter AK, Lohmander LS. Age and waiting time as predictors of outcome after total hip replacement for osteoarthritis. *Rheumatology (Oxford).* 2002;41:1261–1267.

33. Quintana JM, Arostegui I, Azkarate J, Goenaga JI, Elexpe X, Letona J, Arcelay A. Evaluation of explicit criteria for total hip joint replacement. *J Clin Epidemiol.* 2000;53:1200–1208.

34. Quintana JM, Escobar A, Arostegui I, Bilbao A, Azkarate J, Goenaga JI, Arenaza JC. Health-related quality of life and appropriateness of knee or hip joint replacement. *Arch Intern Med.* 2006;166:220–226.

35. Quintana JM, Escobar A, Azkarate J, Goenaga JI, Bilbao A. Appropriateness of total hip joint replacement. *Int J Qual Health Care.* 2005;17:315–321.

36. Quintana JM, Escobar A, Bilbao A, Arostegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after hip joint replacement. *Osteoarthritis Cartilage.* 2005;13:1076–1083.

37. Riddle DL, Perera RA, Jiranek WA, Dumenci L. Using surgical appropriateness criteria to examine outcomes of total knee arthroplasty in a United States sample. *Arthritis Care Res (Hoboken).* 2015;67:349–357.

38. Robinson ME, Riley JL 3rd, Myers CD, Papas RK, Wise EA, Waxenberg LB, Fillingim RB. Gender role expectations of pain: relationship to sex differences in pain. *J Pain.* 2001;2:251–257.

39. Stickles B, Phillips L, Brox WT, Owens B, Lanzer WL. Defining the relationship between obesity and total joint arthroplasty. *Obes Res.* 2001;9:219–223.

40. Stucki G, Meier D, Stucki S, Michel BA, Tyndall AG, Dick W, Theiler R. [Evaluation of a German version of WOMAC (Western Ontario and McMaster Universities) Arthrosis Index] [in German]. *Z Rheumatol.* 1996;55:40–49.

41. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, Bombardier C, Felson D, Hochberg M, van der Heijde D, Dougados M. Evaluation of clinically relevant states in patient

reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. *Ann Rheum Dis.* 2005;64:34–37.

42. Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care.* 1996;34:220–233.

43. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care.* 1992;30:473–483.

44. White DK, Keysor JJ, Lavalley MP, Lewis CE, Torner JC, Nevitt MC, Felson DT. Clinically important improvement in function is common in people with or at high risk of knee OA: the MOST study. *J Rheumatol.* 2010;37:1244–1251.

45. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol.* 1999;52:861–873.