Reliability, validity and responsiveness of the Western Ontario McMaster Osteoarthritis Index (WOMAC) in the elderly population with a femoral neck fracture

Paul T.P.W. Burgers¹, Rudolf W. Poolman², Theodorus M.J. Van Bakel¹, Wim E. Tuinebreijer¹, Stephanie M. Zielinski¹, Mohit Bhandari³, Peter Patka⁴, Esther M.M. Van Lieshout¹ on behalf of the HEALTH and FAITH trial investigators*

¹ Trauma Research Unit Dept. of Surgery, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, the Netherlands
² Department of Orthopaedic Surgery, Joint Research, Onze Lieve Vrouwe Gasthuis, Amsterdam, the Netherlands
³ Dept. of Clinical Epidemiology and Biostatistics, McMaster University, HSC 2C, 1200 Main Street West, Hamilton, ON, L8N 3Z5, Canada
⁴ Dept. of Emergency Medicine, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2010, 3000 CA Rotterdam, the Netherlands

Correspondence address:
E.M.M. van Lieshout, MSc PhD
Erasmus MC, University Medical Center Rotterdam
Trauma Research Unit, Department of Surgery
P.O. Box 2040, 3000 CA Rotterdam, The Netherlands
Phone: +31.10.7031050, Fax: +31.10.7032396
E-mail: e.vanlieshout@erasmusmc.nl
ABSTRACT

Background: Patient-reported outcome measures are gaining importance in clinical research. The WOMAC has been extensively evaluated in populations suffering from osteoarthritis, yet not in femoral neck fracture populations. This study aimed to determine the reliability, construct validity, and responsiveness of WOMAC, compared with SF-12 and EQ-5D, in elderly with a femoral neck fracture.

Methods: Reliability was tested by assessing Cronbach’s alpha. Construct validity was determined by the Pearson correlation coefficient. Change scores were calculated from the 10 week and 12 months follow-up. Standardized response means (SRM; responsiveness) and floor and ceiling effects were determined. Analyses were performed for patients <80 versus ≥80 years of age.

Results: The WOMAC mean total score was 89 points prefracture in younger patients, increasing from 70 (10 weeks) to 81 (24 months). In the oldest old (i.e., ≥ 80 years), these scores were 86, 75, and 78. The stiffness scores were 83, 67, and 76 in the younger group and 85, 80, and 80 in the oldest old. Pain scores were 92, 76, and 87 in the younger and 92, 84, and 93 in the older group. Function scores were 89, 68, and 79 in the younger and 84, 71, and 73 in the older group. Cronbach’s alpha for pain, stiffness, function, and the total scale ranged 0.83-0.98 for the younger and 0.79-0.97 for the older group. Construct validity was good with 82% and 79% of predefined hypotheses confirmed in the younger and older group, respectively. Responsiveness was moderate. No floor effects were found. Moderate to large ceiling effects were found for the Pain and Stiffness scales at 10 weeks and 12 months in younger patients (18-36%) and in the oldest old (38-53%).
Conclusions: The WOMAC showed a good reliability, construct validity, and responsiveness in both age groups with a femoral neck fracture who were physically and mentally fit prefracture. The instrument is suitable for use in future clinical studies in these populations.

Level of Evidence: Diagnostic studies, level I
INTRODUCTION

Most patients with a femoral neck fracture are treated surgically\(^1,2\). Objective determinants like mortality, complications, and revision surgery were traditionally used for evaluating quality of care. Nowadays, methods to assess the patient’s perspective of treatment results are gaining importance\(^3\). These subjective patients’ experiences are quantified using patient-reported outcome measures (PROMs). PROMs are available for assessing functional outcome and quality of life. For women at risk of hip fractures quality of life was profoundly threatened by falls and hip fractures\(^4\), regardless of the type of fracture\(^5\). The problem in the hip fracture population is a complex assortment of issues ranging from baseline health and frailty, social isolation and support, mental status and joint function and pain. Different constructs can be assessed in the evaluation of hip fracture treatment. A frequently used disease-specific PROM is the WOMAC (Western Ontario and McMaster Osteoarthritis Index). To also cover the complexity of issues, the SF-12 (Short Form-12) and the EQ-5D (EuroQuality of Life, EuroQol-5D) are used for evaluating general health and health-related quality of life (HRQoL).

The WOMAC is a disease-specific 24-item questionnaire (scored on a 5-point Likert scale), measuring three domains: pain, stiffness, and function\(^10\). The WOMAC was designed for patients with osteoarthritis of the hip and knee\(^11\). Translation and cross-cultural validation was performed for different countries, including the Netherlands\(^12-15\). The WOMAC has frequently been used in orthopaedic patients, including hip fracture patients\(^16-18\). It has been extensively validated in patients who underwent knee and hip arthroplasty following osteoarthritis, but measurement properties in hip fracture patients is undetermined.

The aim of this study was to determine the reliability, validity, and responsiveness of the WOMAC, compared with SF-12 and EQ-5D in elderly who sustained a femoral neck fracture.
MATERIALS AND METHODS

Population

Between March 3, 2008 and February 14, 2011, 400 patients with a femoral neck fracture were enrolled in two multicenter trials; 150 were enrolled in the Dutch branch of the HEALTH trial, and 250 in the FAITH trial. Adult patients (aged ≥50 years) with a low energy femoral neck fracture without other major trauma, who were ambulant prefracture were considered eligible. Patients with a suspected pathological fracture, associated major injuries of the lower extremities, retained hardware or infection around the hip, bone metabolism disorder other than osteoporosis, cognitive impairment, dementia, or Parkinson’s disease as well as patient who were not likely to be able to complete follow-up were excluded. Data on gender, age, American Society of Anesthesiologists (ASA) classification, walking status, and living situation were collected. The ethics committees approved this study and all patients signed informed consent.

Questionnaires

All 400 patients were asked to complete the three multidimensional questionnaires, the WOMAC, SF-12 and EQ-5D in clinic and at each visit to the outpatient department.

The SF-12 score ranges between 0 and 100 (best) and consists of a Physical Component Summary (PCS-12) and Mental Component Summary (MCS-12). The SF-12 represents a plausible alternative to the larger 36-item SF-36 for measuring health status, especially in large scale studies with a need to reduce questionnaire length. The reliability and validity were determined in different populations, and countries, including the Netherlands. The EQ-5D is a reliable and valid instrument for assessing HRQoL in elderly patients with a femoral neck fracture. The domains mobility, self-care, daily activities, pain, and anxiety are tested on a three-
point Likert scale, resulting in a utility score (EQ-US) that ranges from 0 to 1 (maximum). In addition, patients score their health status on a visual analog scale (EQ-VAS) that ranges from 0 (indicating worst possible health) to 100 (indicating best possible health).

Data about the prefracture situation were gathered during visits within one week after surgery, other follow-up moments were at 10 weeks, six, 12 and 24 months after surgery. Questionnaires were also completed during a telephone interview at nine and 18 months. Patients did not receive any assistance in their completion of the questionnaires and did not consider the burden of completion too high. If necessary (e.g., due to physical status early in recovery), interview-administered versions were used.

**Data analysis**

Data were analyzed with SPSS version 20.0 (IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Armonk, NY: IBM Corp). Data are reported using the COSMIN checklist.

Outcome scores per follow up moment for the all instruments and subdomains were calculated. Unless mentioned otherwise, all further analyses were performed for two age groups (i.e., <80 years and ≥ 80 years) and for 10 weeks and 12 months.

The reliability of the WOMAC instrument was tested by determining the Cronbach’s alpha as measure of internal consistency. Cronbach’s alpha was used as reliability parameter as there was no repetition measurement by different observers or at different time-points per follow-up. Questionnaires aimed at measuring the same construct (WOMAC Pain, SF-12 Pain and EQ-Pain) were used as repetition measurement. A Cronbach’s alpha within the range of 0.70 to 0.90 was considered acceptable.
In the absence of a gold standard for hip function after a hip fracture construct validity was determined by calculating the Pearson correlation coefficient. Construct validity refers to the extent to which scores on a particular measure relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts being measured\textsuperscript{21}. Only questionnaires of which the WOMAC total score was available were included. Hypotheses for construct validity were defined before data analysis (\textit{Appendix 2}). Correlation coefficients of 0.10-0.30 were considered weak, 0.3-0.6 moderate, and >0.6 strong\textsuperscript{22}.

Responsiveness is the ability of a questionnaire to detect clinically important changes over time\textsuperscript{21}. Change scores of the instruments were calculated from the 10 week and 12 months follow up. To assess responsiveness, standardized response means (SRM) were calculated by dividing the mean change by the SD of this change. These effect estimates were interpreted according to Cohen: a SRM of 0.2–0.4 was considered a small effect, 0.5–0.7 moderate, and 0.8 large\textsuperscript{23}.

A floor or ceiling effect was considered present if more than 15\% of the patients in a sample size of 50 patients achieved the lowest or highest possible score, respectively\textsuperscript{21}.

\textit{Missing data}

As the raw data for individual items were analyzed, missing data were not imputed.

\textit{Source of Funding}

Research grants were received from Stichting NutsOhra (The Netherlands) and The Netherlands Organization for Health Research and Development (The Netherlands). These organizations had no role in study design, data collection, data analysis, data interpretation, or writing of the report.
RESULTS

Population

Of all 400 included patients 275 were aged <80 years and 125 were aged ≥ 80 years (Table 1); 250 patients (62.5%) were female and 386 (96.5%) lived independently. This was similar in both age groups. Prefracture, 325 patients (81.2%) walked without aid; 248 (90.2%) and 77 (61.6%) in the younger and older age groups, respectively. Of the total group, 289 patients (72.2%) had ASA class I/II; 223 (81.8%) in the younger age group versus 66 (52.8%) in the oldest old. In the internal fixation group adverse events occurred in 41% of the patients, in the arthroplasty group this was the case in 49%, ranging from bladder infection to revision surgery.

Outcome scores

Patients aged <80 years reported a mean prefracture WOMAC Total score (Figure 1A and Table 2) of 89 points (SD 15), postfracture the score increased from 70 (SD 22) at 10 weeks to 81 (SD 20) after 24 months. In the oldest old, corresponding scores were 86 (SD 16), 75 (SD 21), and 78 (SD 18). Similar changes in scores were reported for WOMAC Stiffness, WOMAC Pain, and WOMAC Function (Figures 1B-D and Table 2).

SF-12 PCS scores decreased after fracture and increased again over time (Figure 2A and Table 2), yet SF-12 MCS remained more stable over time (Figure 2B and Table 2). Likewise, EQ-US scores decreased after fracture and increased again over time (Figure 2C and Table 2), whereas EQ-VAS remained more stable over time (Figure 2D and Table 2).
WOMAC reliability

Cronbach’s alpha for the domains pain, stiffness, function and the total scale were between 0.83 and 0.98 for the younger age group and between 0.79 and 0.97 for the oldest old (Table 3).

Construct validity

All Pearson correlations were statistically significant at a p<0.01 (2-tailed; Table 4). In six main hypotheses, 28 components were predicted, of which 23 (82%) were correctly hypothesized a priori in the younger age group and 22 (79%) in the oldest old. Unconfirmed predictions were mainly underestimations (hypothesis 3). WOMAC Stiffness correlated moderately (r>0.35-0.55) with all other scores, while weak correlations (r= 0.1-0.3) were expected. WOMAC Pain correlated strongly with EQ-US (r = 0.74 in the younger age group and 0.72 in the oldest old), whereas moderate correlation (r 0.3-0.6) was expected.

Responsiveness

Descriptive statistics and responsiveness is presented in Table 5. The SRM was moderate for WOMAC Function (SRM 0.64) and the WOMAC Total score (SRM 0.66). WOMAC subscales Pain and Function showed small SRMs, ranging from 0.21 to 0.30.

Floor and ceiling effects

None of the WOMAC domains nor the Total score showed a floor effect. No ceiling effects were found for the Function domain or the Total score in either of the age groups (Table 6). Pain and Stiffness showed a ceiling effect at 10 weeks and 12 months postsurgery for the younger age group (18-36%) as well as for the oldest old (38-53%).
DISCUSSION

The current study was the first to determine reliability, validity, and responsiveness of the WOMAC, compared with the SF-12 and EQ-5D, in elderly patients who sustained a femoral neck fracture and who were physically and mentally fit prefracture. Results indicated that WOMAC is a reliable and valid PROM for these patients in patients aged <80 years and ≥ 80 years. Responsiveness was also sufficient, indicating the instrument can be used for measuring changes in scores over time.

(Sub)scores of all PROMs showed similar patterns over time. Scores reduced (most likely immediately) post surgery; from 10 weeks after surgery onwards a gradual increase over time was noted. After two years the scores approximated the prefracture scores, indicating a small residual decrease in mental and physical functioning. Due to differences in symptom evolution, the clinimetric results from osteoarthritis populations are not readily applicable to hip fracture populations. The minimal clinically important change (MIC) for hip fracture patients has not been reported, yet changes in WOMAC Pain and Function were larger than the MIC reported for hip replacement after osteoarthritis (10 points for pain and 8-9 points for function)\(^24,25\). There may be two reasons why changes in WOMAC over time were smaller than observed after hip replacement for osteoarthritis (for pain: 53 points at baseline to 94 two years after surgery and from 50 to 92 for function)\(^26\). First, true baseline scores for hip fracture patients (i.e., between fracture and surgery) would have been much lower. Although these scores are unknown (as asking patients to complete questionnaires before surgery is not feasible), we expect these scores to be even lower than preoperative scores in osteoarthritis patients\(^26\). Calculating changes in WOMAC scores relative to baseline would have resulted in much larger changes over time. Second, at the first follow-up moment (10 weeks) rehabilitation likely has progressed
considerably already in the majority of patients, as participation required patients to be fit prefracture.

The construct validity of the WOMAC instrument was good, with 82% (<80 years) and 79% (≥ 80 years) of the predefined hypotheses being true. Especially the strong correlations of WOMAC Function and Total score with SF-12 PCS, SF-12 Total score and EQ-US support the preferred use of WOMAC for assessing functional recovery in this population.

Responsiveness of the WOMAC instrument as a whole was moderate in patients <80 years, as the SRM was 0.66. In the oldest old, the SRM was small (0.24). For hip fracture patients the SRM has not been shown before. However, for patients with hip osteoarthritis the SRM of the WOMAC exceeded 1.0\textsuperscript{27,28}, this is mainly due to much larger changes in score between the two measurements. Larger changes for those patients are expected, as the WOMAC score is at the lowest prior to the hip replacement procedure. These true baseline scores cannot be determined for hip fracture patients. The Dutch version of the non-hip-specific Short Musculoskeletal Function Assessment questionnaire (SMFA) appeared to be moderately responsive (SRMs for subscales 0.17-0.47) for patients with a variety of musculoskeletal disorders\textsuperscript{29}. In different Swedish hip fracture populations the SRMs were moderate to large (0.76-0.96) for the SMFA and small to large for the non-disease-specific instruments EQ-5D (SRM 0.01-1.14) and Nottingham Health Profile questionnaire (SRM 0.09-0.98)\textsuperscript{30,31}.

There was no evidence for any floor effect of the WOMAC instrument or for a ceiling effect of WOMAC Function and Total score. In contrast, moderate to large ceiling effects (18-53%) were found for Pain and Stiffness at 10 weeks and 12 months in both age groups. This is similar as reported\textsuperscript{32-34}. It may reflect the narrow discriminating capacity of these WOMAC domains in the studied population. Some selection bias might have played a role, as patients had
to be physically and mentally fit to participate. Participants were ambulant and lived independently prefracture.

**Strengths and limitations**

WOMAC has been used for assessing functional outcomes and quality of life in hip fracture populations before\(^{35,36}\). In the current study, it was validated in hip fracture populations for the first time. Rates of adverse events in these populations doesn’t seem to differ from literature, they may have affect the scores, but not the validity of the questionnaire. We consider the novelty of validating WOMAC for hip fracture patients a strength of our study. Moreover, we consider the studied populations representative for a daily practice hip fracture population and believe the results apply to a generally fit hip fracture population. Whether or not WOMAC is also valid for use in frail elderly remains unknown, as patients that were non-ambulatory prefracture and patients with e.g., Parkinson’s disease, pathological fractures, and dementia were excluded. Another strength was the built-in very short period, only several days postfracture, for the self-reported, pre-injury disability evaluation to be completed, minimizing the risk of recall bias\(^{37}\).

A limitation could be the use of arbitrary hypotheses for construct validation, although they were predefined in compliance with clinimetric evaluation guidelines\(^{19}\). Second, some selection bias might have led to overestimation of the outcomes, especially ceiling effects. This effect is also known from the Harris Hip Score which is frequently used in orthopedic research, but no gold standard exists for functional evaluation of hip fracture populations\(^{38,39}\).

Third, data completeness was not 100% at each time point. One can imagine that patients were not able to complete the forms especially if they were in bad condition. These 36% missing
items causing a missing WOMAC Total score might have influenced the outcomes resulting in the current more favorable mean outcome scores and large ceiling effects.

CONCLUSION
The WOMAC, a widely used disease-specific questionnaire, shows adequate reliability and construct validity in patients aged 50 years or older with a femoral neck fracture who were physically and mentally fit prefracture. Responsiveness was better for younger patients than for patients ≥80 years. It is therefore a suitable instrument for use in future clinical studies in this population.
REFERENCES


Figure Legends

Figure 1. Changes in functional outcome scores over time
Mean and SD are shown for the WOMAC Total score (A), WOMAC Stiffness score (B), WOMAC Pain score (C), WOMAC Function score (D), and per follow up moment. Higher scores represent better functional outcomes.
Figure 2. Changes in quality of life scores over time, separated by age group

(A) SF-12 Physical Component Summary (PCS) and (B) SF-12 Mental Component Summary (MCS) are from the Short Form-12 (SF-12) instrument; (C) EuroQol-5D Utility Score (EQUS) and (D) EuroQol-5D Visual Analog Scale (EQVAS) are from the EQ-5D score. Higher scores represent a better quality of life.

Mean and SD are shown per follow up moment.

For the SF-12, the dotted line shows the standardized mean (50 points) of the general population, the shaded area is the area within 1 SD deviation from this mean (10 points).
Table 1. Demographic description of the study population, separated by age group

<table>
<thead>
<tr>
<th></th>
<th>Total (N=400)</th>
<th>Age &lt;80 years (N=275)</th>
<th>Age ≥ 80 years (N=125)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>251 (62.8%)</td>
<td>173 (62.9%)</td>
<td>78 (62.4%)</td>
</tr>
<tr>
<td>Mean age (years)</td>
<td>74 (10)</td>
<td>69 (8)</td>
<td>85 (4)</td>
</tr>
<tr>
<td>ASA I/II</td>
<td>289 (72.3%)</td>
<td>223 (81.1%)</td>
<td>66 (52.8)</td>
</tr>
<tr>
<td>Walking without aids prefracture</td>
<td>325 (81.3%)</td>
<td>248 (90.2%)</td>
<td>77 (61.6%)</td>
</tr>
<tr>
<td>Living independently prefracture</td>
<td>386 (96.5%)</td>
<td>268 (97.5%)</td>
<td>118 (94.4%)</td>
</tr>
</tbody>
</table>

SD, standard deviation; ASA, American Society of Anesthesiologists.

Data are shown as number (%) or as mean (SD).
Table 2. Outcomes scores of the study population, separated by age group

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Age &lt;80 years</th>
<th>Age ≥ 80 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N=275)</td>
<td>(N=125)</td>
</tr>
<tr>
<td></td>
<td>Prefracture</td>
<td>10 weeks</td>
</tr>
<tr>
<td>WOMAC Stiffness</td>
<td>83 (21)</td>
<td>67 (25)</td>
</tr>
<tr>
<td>WOMAC Pain</td>
<td>92 (16)</td>
<td>76 (23)</td>
</tr>
<tr>
<td>WOMAC Function</td>
<td>89 (16)</td>
<td>79 (23)</td>
</tr>
<tr>
<td>WOMAC Total score</td>
<td>89 (15)</td>
<td>70 (22)</td>
</tr>
<tr>
<td>SF-12 PCS</td>
<td>47 (10)</td>
<td>37 (9)</td>
</tr>
<tr>
<td>SF-12 MCS</td>
<td>53 (9)</td>
<td>51 (11)</td>
</tr>
<tr>
<td>EQ US</td>
<td>0.86 (0.19)</td>
<td>0.66 (0.29)</td>
</tr>
<tr>
<td>EQ VAS</td>
<td>76 (16)</td>
<td>68 (17)</td>
</tr>
</tbody>
</table>

Data are shown as mean (SD).
Table 3. Reliability of the WOMAC instrument at 10 weeks and 12 months in patients who sustained a femoral neck fracture, separated by age group

<table>
<thead>
<tr>
<th>WOMAC domain of items</th>
<th>Number of items</th>
<th>Age &lt; 80 years (N=275)</th>
<th>Age ≥ 80 years (N=125)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>10 weeks</td>
<td>12 months</td>
</tr>
<tr>
<td>Stiffness</td>
<td>2</td>
<td>0.83 (229)</td>
<td>0.83 (204)</td>
</tr>
<tr>
<td>Pain</td>
<td>5</td>
<td>0.92 (191)</td>
<td>0.92 (186)</td>
</tr>
<tr>
<td>Function</td>
<td>17</td>
<td>0.97 (107)</td>
<td>0.97 (130)</td>
</tr>
<tr>
<td>Total scale</td>
<td>24</td>
<td>0.98 (105)</td>
<td>0.98 (128)</td>
</tr>
</tbody>
</table>

N; number of available questionnaires with all items completed per domain and with all items completed.

The Cronbach’s alpha is given with the number of patients included in the analysis between brackets.
Table 4. Construct validity of the WOMAC domains and WOMAC Total score and the domains of the SF-12 and EQ-5D instruments at 10 weeks, separated by age group

<table>
<thead>
<tr>
<th></th>
<th>WOMAC</th>
<th>SF-12 PCS</th>
<th>SF-12 MCS</th>
<th>SF-12 Pain</th>
<th>SF-12 Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stiffness</td>
<td>Age &lt;80 years (N=275)</td>
<td>Age ≥ 80 years (N=125)</td>
<td>Stiffness</td>
<td>Age &lt;80 years (N=275)</td>
</tr>
<tr>
<td>WOMAC</td>
<td></td>
<td>Pain</td>
<td>Function</td>
<td>Total score</td>
<td>Pain</td>
</tr>
<tr>
<td>SF-12 PCS r</td>
<td>0.35</td>
<td>0.56</td>
<td>0.71</td>
<td>0.70</td>
<td>0.41</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.23-0.46)</td>
<td>(0.46-0.65)</td>
<td>(0.62-0.77)</td>
<td>(0.61-0.77)</td>
<td>(0.24-0.55)</td>
</tr>
<tr>
<td>N</td>
<td>222</td>
<td>218</td>
<td>177</td>
<td>177</td>
<td>110</td>
</tr>
<tr>
<td>SF-12 MCS r</td>
<td>0.41</td>
<td>0.47</td>
<td>0.51</td>
<td>0.52</td>
<td>0.45</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.29-0.51)</td>
<td>(0.36-0.57)</td>
<td>(0.40-0.61)</td>
<td>(0.40-0.62)</td>
<td>(0.29-0.59)</td>
</tr>
<tr>
<td>N</td>
<td>222</td>
<td>218</td>
<td>177</td>
<td>177</td>
<td>110</td>
</tr>
<tr>
<td>SF-12 Pain r</td>
<td>0.55</td>
<td>0.74</td>
<td>0.77</td>
<td>0.79</td>
<td>0.51</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.45-0.63)</td>
<td>(0.68-0.80)</td>
<td>(0.70-0.82)</td>
<td>(0.72-0.84)</td>
<td>(0.36-0.63)</td>
</tr>
<tr>
<td>N</td>
<td>228</td>
<td>224</td>
<td>179</td>
<td>179</td>
<td>113</td>
</tr>
<tr>
<td>SF-12 Total r</td>
<td>0.50</td>
<td>0.67</td>
<td>0.78</td>
<td>0.78</td>
<td>0.55</td>
</tr>
<tr>
<td>95% CI</td>
<td>(0.39-0.59)</td>
<td>(0.59-0.74)</td>
<td>(0.71-0.83)</td>
<td>(0.71-0.83)</td>
<td>(0.40-0.67)</td>
</tr>
<tr>
<td>N</td>
<td>222</td>
<td>218</td>
<td>177</td>
<td>177</td>
<td>110</td>
</tr>
<tr>
<td>Domain</td>
<td>r</td>
<td>95% CI</td>
<td>95% CI</td>
<td>95% CI</td>
<td>95% CI</td>
</tr>
<tr>
<td>----------</td>
<td>-----</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>EQ-US</td>
<td>0.42</td>
<td>(0.30-0.52)</td>
<td>(0.59-0.79)</td>
<td>(0.71-0.82)</td>
<td>(0.37-0.64)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>228</td>
<td>223</td>
<td>179</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.74</td>
<td>(0.67-0.79)</td>
<td>(0.64-0.78)</td>
<td>(0.68-0.81)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>229</td>
<td>224</td>
<td>179</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.44</td>
<td>(0.33-0.54)</td>
<td>(0.74-0.84)</td>
<td>(0.64-0.78)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>229</td>
<td>224</td>
<td>179</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.39</td>
<td>(0.27-0.49)</td>
<td>(0.41-0.60)</td>
<td>(0.57-0.73)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>229</td>
<td>224</td>
<td>179</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>229</td>
<td>224</td>
<td>179</td>
</tr>
</tbody>
</table>

N: number of available questionnaires with all items completed per domain and with all items completed.

The Pearson correlation coefficient (r) is given with its 95% confidence interval in brackets.

Correlation is significant at the 0.01 level (2-tailed) for all comparisons.

Correlations that were predicted correctly are given in boldface.
Table 5. Responsiveness of the WOMAC domains and WOMAC Total score, separated by age group

<table>
<thead>
<tr>
<th></th>
<th>Age &lt;80 years (N=275)</th>
<th>Age ≥ 80 years (N=125)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10 weeks</td>
<td>12 months</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>WOMAC Stiffness</td>
<td>67 (24)</td>
<td>74 (23)</td>
</tr>
<tr>
<td>WOMAC Pain</td>
<td>76 (23)</td>
<td>85 (20)</td>
</tr>
<tr>
<td>WOMAC Function</td>
<td>68 (24)</td>
<td>80 (21)</td>
</tr>
<tr>
<td>WOMAC Total score</td>
<td>70 (22)</td>
<td>80 (20)</td>
</tr>
</tbody>
</table>

Scores at 10 weeks and at 12 months as well as the difference between these scores (change), are shown as median with SD.

The Standardized Response Mean (SRM) is given with the number of patients used in the analysis between brackets.
Table 6. Ceiling effect of the WOMAC domains and WOMAC Total score, separated by age group

<table>
<thead>
<tr>
<th></th>
<th>Age &lt;80 years (N=275)</th>
<th>Age ≥ 80 years (N=125)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N 10 weeks 12 months</td>
<td>N 10 weeks 12 months</td>
</tr>
<tr>
<td></td>
<td>N (%)</td>
<td>N (%)</td>
</tr>
<tr>
<td>WOMAC Stiffness</td>
<td>224 40 (17.9%) 72 (35.6%)</td>
<td>112 43 (38.4%) 50 (52.6%)</td>
</tr>
<tr>
<td>WOMAC Pain</td>
<td>229 46 (20.1%) 53 (26.1%)</td>
<td>113 48 (42.5%) 38 (40.0%)</td>
</tr>
<tr>
<td>WOMAC Function</td>
<td>179 7 (3.9%) 26 (13.7%)</td>
<td>73 3 (4.1%) 8 (10.3%)</td>
</tr>
<tr>
<td>WOMAC Total score</td>
<td>179 2 (0.1%) 13 (6.9%)</td>
<td>73 2 (2.7%) 7 (9.0%)</td>
</tr>
</tbody>
</table>

The number of patients reporting the maximum score of 100 points are given with the percentage given in brackets. N represents the total number of questionnaires used for the analysis. None of the instruments demonstrated a floor effect (i.e., 0 points).
Appendix 1. List of international collaborators

HEALTH and FAITH trial investigators*

HEALTH trial:
Research grants were received from the following: Canadian Institutes of Health Research (CIHR) (PI: M Bhandari, Co-PI: GH Guyatt), National Institutes of Health (NIH) (PI: TA Einhorn), The Netherlands Organisation for Health Research and Development (ZonMw) (PI: EMM van Lieshout), Sophies Minde Foundation for Orthopaedic Research (PI: L Nordsletten and F Frihagen), and McMaster Surgical Associates (PI: M Bhandari). Dr. Bhandari was also funded, in part, by a Canada Research Chair in Musculoskeletal Trauma which is unrelated to the present study (McMaster University, Hamilton, ON, Canada). The funding sources had no role in design or conduct of the study; the collection, management, analysis, or interpretation of the data; or the preparation, review, or approval of the manuscript. Dr. Mohit Bhandari had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Author Contributions:

Steering Committee: Mohit Bhandari (Chair), PJ Devereaux, Gordon H Guyatt, Thomas A. Einhorn, Lehana Thabane, Emil H. Schemitsch, Ken Korval, Frede Frihagen, Rudolph W. Poolman, Kevin Tetsworth.
Global Methods Centre: Mohit Bhandari (Principal Investigator); Sheila Sprague (Research Program Manager); Marilyn Swinton, Taryn Scott, Paula McKay, Kim Madden (Research Coordination); Diane Heels-Ansdell, (Statistical Analysis); Lisa Buckingham, Aravin Duraikannan (Data Management) (McMaster University)

US Methods Centre: Thomas A. Einhorn (Principal Investigator); Heather Silva (Research Coordination) (Boston University Medical Center).

Netherlands Methods Centre: Martin J. Heetveld (Principal Investigator); Rudolf W. Poolman (Co-Principal Investigator), Esther M.M. Van Lieshout (Research Coordination), Paul T.P.W. Burgers (Trial Coordination) (Erasmus Medical Centre, Rotterdam)

Central Adjudication Committee: Mohit Bhandari (Chair), Gregory J. Della Rocca, Susan Liew, Thomas A. Einhorn, Rudolph W. Poolman, Robert Haverlag, Martin Heetveld.

Data Safety Monitoring Board (CIHR): John Antoniou (Chair), Tim Ramsay, Earl R. Bogoch, Andrew Trenholm.

Data Safety Monitoring Board (NIH): Stephen Lyman (Chair), Madhu Mazumdar, Kevin J. Bozic, Mark Luborsky, Stuart Goodman, Susan Muray.
HEALTH Investigators:

The following persons participated in the HEALTH trial.

Canada


United States


**Netherlands**


International

The Alfred - Susan Liew, Harvinder Bedi, Ashley Carr, Andrew Chia, Steve Csongvay, Craig Donohue, Stephen Doig, Elton Edwards, Max Esser, Richard Freeman, Andrew Gong, Doug Li, Russell Miller, Lu Ton, Otis Wang, Ian Young, Adam Dowrick, Zoe Murdoch, Claire Sage.

FAITH trial:

Research grants were received from the following: Canadian Institutes of Health Research (CIHR) (PI: Mohit Bhandari); Stichting NutsOhra (PI: Martin J. Heetveld), The Netherlands Organisation for Health Research and Development (PI: Esther M.M. Van Lieshout); Physicians’ Services Incorporated (PI: Mohit Bhandari). Funding for the pilot phase of FAITH was supported, in part, by Stryker Inc. Dr. Bhandari was also funded, in part, by a Canada Research Chair in Musculoskeletal Trauma which is unrelated to the present study (McMaster University, Hamilton, ON, Canada). We would also like to acknowledge the support of The County Durham & Tees Valley Comprehensive Local Research Network which operates as part of the National Institute for Health Research Comprehensive Clinical Research Network in England. The funding sources had no role in design or conduct of the study; the collection, management, analysis, or interpretation of the data; or the preparation, review, or approval of the manuscript. Dr. Mohit Bhandari had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Steering Committee: Mohit Bhandari (Chair), Marc Swiontkowski, PJ Devereaux, Gordon Guyatt, Martin J. Heetveld, Kyle Jeray, Susan Liew, Emil H. Schemitsch, Lehana Thabane, Stephen Walter

Global Methods Centre: Mohit Bhandari (Principal Investigator); Sheila Sprague (Research Program Manager); Helena Viveiros, Paula McKay, Taryn Scott, Marilyn Swinton, (Research Coordination); Victoria Truong and Kaitlin Koo (Adjudication Coordination); Diane Heels-Ansdell, Qi Zhou (Statistical Analysis); Lisa Buckingham, Aravin Duraikannan (Data
Management); Deborah Maddock, Nicole Simunovic (Grant Management) (McMaster University)

US Methods Centre: Marc Swiontkowski (Principal Investigator); Julie Agel (Research Coordination) (University of Minnesota)

Netherlands Method Centre: Martin J. Heetveld (Principal Investigator); Esther M.M. Van Lieshout (Research Coordination); Stephanie M. Zielinski (Trial Coordination) (Erasmus MC, University Medical Center Rotterdam)

UK Method Centre: Amar Rangan (Principal Investigator), Birgit C. Hanusch, Lucksy Kottam, Rachel Clarkson (Research Coordination) (The James Cook University Hospital)

Central Adjudication Committee: Gregory J Della Rocca (Chair), Robert Haverlag, Susan Liew, Gerard Slobogean

Data Safety Monitoring Board: Jeffrey Katz (Chair), Brenda Gillespie, Gail A. Greendale, Pierre Guy, Curtis Hartman, Craig Rubin, James Waddell
FAITH Investigators

The following persons participated in the FAITH trial:

Canada:
Robert McCormack, Kelly Apostle, Dory Boyer, Farhad Moola, Bertrand Perey, Trevor Stone, Darius Viskontas and H. Michael Lemke, Mauri Zomar, Karyn Moon, Raely Moon, Amber Oatt (University of British Columbia/Fraser Health Authority); Richard E. Buckley, Paul Duffy, Robert Korley, Shannon Puloski, Kelly Johnston, James Powell, Kimberly Carcary (Foothills Medical Centre); David Sanders, Abdel Lawendy, Christina Tieszer (London Health Sciences Centre); David Stephen, Hans Kreder, Richard Jenkinson, Markku Nousiainen, Terry Axelrod, John Murnaghan, Diane Nam, Veronica Wadey, Albert Yee, Katrine Milner, Monica Kunz (Sunnybrook Health Sciences Centre); Emil H. Schemitsch, Henry Ahn, Jeremy A. Hall, Michael D. McKee, Daniel B. Whelan, Aaron Nauth, Milena R. Vicente, Lisa M. Wild, Ryan M. Khan, Jennifer T. Hidy (St. Michael’s Hospital); Chad Coles, Ross Leighton, Michael Biddulph, David Johnston, Mark Glazebrook, David Alexander, Cathy Coady, Michael Dunbar, David Amirault, Michael Gross, William Oxner, Gerald Reardon, Glen Richardson, Andrew Trenholm, Ivan Wong, Kelly Trask, Shelley MacDonald, Gwen Dobbin (Queen Elizabeth II Health Sciences Centre); Ryan Bicknell, Jeff Yach, Davide Bardana, Gavin Wood, Mark Harrison, David Yen, Sue Lambert, Fiona Howells, Angela Ward (Human Mobility Research Centre, Queen’s University and Kingston General Hospital); Paul Zalzal, Heather Brien, V. Naumetz, Brad Weening, Nicole Simunovic (Oakville Trafalgar Memorial Hospital); Eugene K. Wai, Steve Papp, Wade T. Gofton, Allen Liew, Stephen P. Kingwell, Garth Johnson, Joseph
O’Neil, Darren M. Roffey, Vivian Borsella (Ottawa Hospital); Victoria Avram (Juravinski Hospital and Cancer Centre)

United States:
Todd M. Oliver, Vicki Jones (Boone Hospital Center – Columbia Orthopaedic Group); Clifford B. Jones, James R. Ringler, Terrence J. Endres, Samuel G. Agnew, Debra L. Sietsema, Jane E. Walker (Orthopaedic Associates of Michigan); Kyle J. Jeray, J. Scott Broderick, David R. Goetz, Thomas B. Pace, Thomas M. Schaller, Scott E. Porter, Stephanie L. Tanner, Rebecca G. Snider, Lauren A. Nastoff, Shea A. Bielby (Greenville Hospital System); Julie A. Switzer, Peter A. Cole, Sarah A. Anderson, Paul M. Lafferty, Mengnai Li, Thuan V. Ly, Scott B. Marston, Amy L. Foley, Sandy Vang, David M. Wright (Regions Hospital-University of Minnesota); Andrew J. Marcantonio, Michael S.H. Kain, Richard Iorio, Lawrence M. Specht, John F. Tilzey, Margaret J. Lobo, John S. Garfi (Lahey Hospital & Medical Center); Heather A. Vallier, Andrea Dolenc, Chalitha Robinson (MetroHealth Medical Center); Michael J. Prayson, Richard Laughlin, L. Joseph Rubino, Jedediah May, Geoffrey Ryan Rieser, Liz Dulaney-Cripe, Chris Gayton (Miami Valley Hospital); James Shaer, Tyson Schrickel, Barbara Hileman (St. Elizabeth Health Center); John T. Gorczyca, Jonathan M. Gross, Catherine A. Humphrey, Stephen Kates, Krista Noble, Allison W. McIntyre, Kaili Pecorella (University of Rochester Medical Center); Craig A. Davis, Stewart Weinerman, Peter Weingarten, Philip Stull, Stephen Lindenbaum, Michael Hewitt, John Schwappach, Janell K. Baker, Tori Rutherford, Heike Newman, Shane Lieberman, Erin Finn, Kristin Robbins, Meghan Hurley, Lindsey Lyle, Khalis Mitchell, Kieran Browner, Erica Whatley, Krystal Payton, Christina Reeves (Colorado Orthopedic Consultants); Lisa K. Cannada, David Karges, Leslie Hill (St.
Louis University Hospital); Samir Mehta, John Esterhai, Jaimo Ahn, Annamarie D. Horan, Kelly McGinnis, Christine A. Kaminski, Brynn N. Kowalski (University of Pennsylvania); Jonathan P. Keeve, Christopher G. Anderson, Michael D. McDonald, Jodi M. Hoffman (Northwest Orthopaedic Specialists); Ivan Tarkin, Peter Siska, Gary Gruen, Andrew Evans, Dana J. Farrell, James Irrgang, Arlene Luther (University of Pittsburgh Medical Center); William W. Cross III, Joseph R. Cass, Stephen A. Sems, Michael E. Torchia, Tyson Scrabeck (Mayo Clinic); Mark Jenkins, Jules Dumais, Amanda W. Romero (Texas Tech University Health Sciences Center – Lubbock); Carlos A. Sagebien, Mark S. Butler, James T. Monica, Patricia Seuffert (University Orthopaedic Associates, LLC); Joseph R. Hsu, James Ficke, Michael Charlton, Matthew Napierala, Mary Fan (US Army Institute of Surgical Research); Paul Tornetta III, Chadi Tannoury, Hope Carlisle, Heather Silva (Boston University Medical Center); Michael Archdeacon, Ryan Finnan, Toan Le, John Wyrick, Shelley Hess (UC Health/University of Cincinnati Medical Center); Michael L. Brennan, Robert Probe, Evelyn Kile, Kelli Mills, Lydia Clipper, Michelle Yu, Katie Erwin (Scott and White Memorial Hospital); Daniel Horwitz, Kent Strohecker, Teresa K. Swenson (Geisinger Medical Center); Andrew H. Schmidt, Jerald R. Westberg (Hennepin County Medical Center); Kamran Aurang, Gary Zohman, Brett Peterson, Roger B. Huff (Kaiser Permanente); Joseph Baele, Timothy Weber, Matt Edison (OrthoIndy); Jessica McBeth (Santa Clara Valley Medical Center); Karl Shively, Janos P. Ertl, Brian Mullis, J. Andrew Parr, Ripley Worman, Valda Frizzell, Molly M. Moore, Erin Tobias, Emily Thomas (Indiana University – Wishard Health Services); Charles J. DePaolo, Rachel Alosky, Leslie E. Shell, Lynne Hampton, Stephanie Shepard, Tracy Nanney, Claudine Cuento (Mission Hospital Research Institute); Robert V. Cantu, Eric R. Henderson, Linda S. Eickhoff (Dartmouth-Hitchcock Medical Center); E. Mark Hammerberg, Philip
Stahel, David Hak, Cyril Mauffrey, Douglas Gibula, Hannah Gissel, Corey Henderson (Denver Health and Hospital Authority); David P. Zamorano, Martin C. Tynan, Deeba Pourmand, Deanna Lawson (University of California Irvine Medical Center); Gregory J. Della Rocca, Brett D. Crist, Yvonne M. Murtha, Linda K. Anderson (University of Missouri Health Care); Colleen Linehan, Lindsey Pilling (Covenant Healthcare of Saginaw); Courtland G. Lewis, Stephanie Caminiti, Raymond J. Sullivan, Elizabeth Roper (Hartford Hospital); William Obremskey, Philip Kregor, Justin E. Richards, Kenya Stringfellow (Vanderbilt University Medical Center); Michael P. Dohm, Abby Zellar (Western Slope Study Group)

The Netherlands:
Haverlag, Nico J.M. Out, Jan Bosma (Onze Lieve Vrouwe Gasthuis); Albert van Kampen, Jan Biert, Arie B. van Vugt, Michael J.R. Edwards, Taco J. Blokhuis, Jan Paul M. Frölke, Leo M.G. Geeraedts, Jean W.M. Gardeniers, Edward T.C.H. Tan, Lodewijk M.S.J. Poelhekke, Maarten C. de Waal Malefijt, Bart Schreurs (University Medical Center St. Radboud); Gert R Roukema, Hong A. Josaputra, Paul Keller, Peter D. de Rooij, Hans Kuiken, Han Boxma, Berry I. Cleffken, Ronald Liem (Maasstad Ziekenhuis); Steven J. Rhemrev, Coks H.R. Bosman, Alexander de Mol van Otterloo, Jochem Hoogendoorn, Alexander C. de Vries, Sven A.G. Meylaerts (Medisch Centrum Haaglanden); Michiel H.J. Verhofstad, Joost Meijer, Teun van Egmond, Frank H.W.M. van der Heijden, Igor van der Brand (St. Elisabeth Ziekenhuis); Peter Patka, Martin G. Eversdijk, Rolf Peters, Dennis Den Hartog, Oscar J.F. Van Waes, Pim Opred (Erasmus MC, University Medical Center Rotterdam); Harm M van der Vis, Martin Campo, Ronald Verhagen, G.H. Robert Albers, Arthur W. Zurcher (Tergooi Ziekenhuizen); Rogier K.J. Simmermacher, Jeroen van Mulken, Karlijn van Wessem, Taco J. Blokhuis, Steven M. van Gaalen, Luke P.H. Leenen (University Medical Center Utrecht); Maarten W.G.A. Bronkhorst, Onno R. Guicherit (Bronovo Ziekenhuis); J. Carel Goslings, Robert Haverlag, Kees Jan Ponsen (Academic Medical Center)

International:
Mahesh Bhatia, Vinod Arora, Vivek Tyagi (RLB Hospital and Research Centre, India); Susan Liew, Harvinder Bedi, Ashley Carr, Hamish Curry, Andrew Chia, Steve Csongvay, Craig Donohue, Stephen Doig, Elton Edwards, Greg Etherington, Max Esser, Andrew Gong, Arvind Jain, Doug Li, Russell Miller, Ash Moaveni, Matthias Russ, Lu Ton, Otis Wang, Adam Dowrick, Zoe Murdoch, Claire Sage (The Alfred, Australia); Frede Frihagen, John Clarke-Jenssen, Geir
Hjorthaug, Torben Ianssen, Asgeir Amundsen, Jan Egil Brattgjerde, Tor Borch, Berthe Bøe,
Bernhard Flatøy, Sondre Hasselund, Knut Jørgen Haug, Kim Hemlock, Tor Magne Hoseth, Geir
Jomaas, Thomas Kibsgård, Tarjei Lona, Gilbert Moatshe, Oliver Müller, Marius Molund, Tor
Nicolaisen, Fredrik Nilsen, Jonas Rydinge, Morten Smørsrud, Are Stødle, Axel Trommer, Stein
Ugland, Anders Karlsten, Guri Ekås, Elise Berg Vesterhus, Anne Christine Brekke (Oslo
University Hospital, Norway); Ajay Gupta, Neeraj Jain, Farah Khan (Nirmal Hospital, India);
Ateet Sharma, Amir Sanghavi, Mittal Trivedi (Satellite Orthopaedic Hospital and Research
Centre, India); Anil Rai, Subash, Kamal Rai (Highway Hospital, India); Vineet Yadav, Sanjay
Singh, Kamal Rai (Popular Hospital, India); Kevin Tetsworth, Geoff Donald, Patrick
Weinrauch, Paul Pincus, Steven Yang, Brett Halliday, Trevor Gervais, Michael Holt, Annette
Flynn (Royal Brisbane and Women’s Hospital, Australia); Amal Shankar Prasad, Vimlesh
Mishra (Madhuraj Nursing Home, India); D.C. Sundaresh, Angshuman Khanna (M.S.
Rammaiah Medical College & Hospital, India); Joe Joseph Cherian, Davy J Olakkengil,
Gaurav Sharma (St John’s Medical College Hospital, India); Marinis Pirpiris, David Love,
Andrew Bucknill, Richard J Farrugia (Royal Melbourne Hospital, Australia); Hans-Christoph
Pape, Matthias Knobe, Roman Pfeifer (University of Aachen Medical Center, Germany);
Peter Hull, Sophie Lewis, Simone Evans (Cambridge University Hospitals, England); Rajesh
Nanda, Rajanikanth Logishetty, Sanjeev Anand, Carol Bowler (University Hospital of North
Tees, England); Akhil Dadi, Naveen Palla, Utsav Ganguly (Sunshine Hospital, India); B.
Sachidananda Rai, Janakiraman Rajakumar (Unity Health Complex, India); Andrew Jennings,
Graham Chuter, Glynis Rose, Gillian Horner (University Hospital of North Durham
and Darlington Memorial Hospital, England); Callum Clark, Kate Eke (Wexham Park
Hospital, England); Mike Reed, Chris Herriott, Christine Dobb (Northumbria Healthcare NHS Foundation Trust, England)
Appendix 2. Hypotheses for evaluating the construct validity of the WOMAC Instrument

1. WOMAC Pain was expected to correlate strongly ($r > 0.6$) with SF-12 pain and EQ-Pain since they were expected to measure the same construct.

2. A moderate correlation ($r < 0.6$) was therefore predicted between WOMAC Pain and all other scores since they assess different or less specific constructs.

3. A weak ($r < 0.3$) correlation was expected for the specific WOMAC Stiffness and all other outcomes.

4. WOMAC Function was predicted to correlate strongly with SF-12 PCS, SF-12 pain, SF-12 Total score, EQ pain, EQ-US, and EQ-VAS.

5. A moderately to weak correlation was expected for WOMAC Function with SF-12 MCS.

6. WOMAC Total score was predicted to have a moderate correlation with SF-12 MCS and a strong correlation all other outcomes.