Prevalence of Clinically Significant Improvement Following Total Knee Replacement

KHALID ALZAHRANI, RAJIV GANDHI, JUSTIN deBEER, DANIELLE PETRUCCHELLI, and NIZAR MAHOMED

ABSTRACT. Objective. Although total knee replacement (TKR) has a high reported success rate, the pain relief and functional improvement after surgery vary. The purpose of our retrospective cohort study was to determine the prevalence of patients showing no clinically important improvement 1 year after TKR, and patient factors that may predict this outcome.

Methods. We reviewed primary TKR registry data that were collected from 2 academic hospitals: the Toronto Western Hospital and the Hamilton Health Sciences Henderson Hospital in Ontario, Canada. Relevant covariates including demographic data, body mass index, and comorbidity were recorded. Knee joint pain and functional status were assessed at baseline and at 1-year followup with the Western Ontario McMaster University Osteoarthritis Index (WOMAC) and Oxford Knee Score (OKS) to measure the change using the minimal clinically important difference (MCID). Logistic regression modeling was used to identify the predictors of interest.

Results. Overall, 11.7% (373/3177) of patients reported no clinically important improvement 1 year after surgery. Logistic regression modeling showed that a greater patient age independently predicted no clinically important improvement on the WOMAC scale 1 year after surgery (p = 0.0003), while being male independently predicted no clinically important improvement on the OKS 1 year after surgery (p = 0.008).

Conclusion. Awareness of the prevalence of patients who may show no clinically important improvement and factors that predict this outcome will help patients and surgeons set realistic expectations of surgery. (First Release Jan 15 2011; J Rheumatol 2011;38:753–9; doi:10.3899/jrheum.100233)

Key Indexing Terms:
KNEE REPLACEMENT               OSTEOARTHRITIS               OUTCOME

Total knee replacement (TKR) is the “gold standard” treatment for endstage disabling degenerative arthritis, based on good reported outcomes for pain relief and functional improvement, and proven cost-effectiveness1,2,3,4,5,6,7. As the mean age of the population increases, the demand for TKR is expected to grow exponentially8,9.

The greatest improvement in pain and function following TKR occurs within the first 3 to 6 months10, with maximal recovery achieved by 1 year11,12. However, the prevalence and natural history of residual pain after TKR is not well established in the arthroplasty literature13. Many studies have reported up to a 30% dissatisfaction rate after TKR14,15,16,17. These reports identified many factors that may predict dissatisfaction after TKR, including residual pain after surgery, poor functional condition preoperatively, poor mental health status at the time of surgery, or unfulfilled patient expectation of surgery.

Despite unremarkable clinical or radiological findings, it has been reported that about 10% to 30% of patients who undergo TKR experience minimal or no improvement after surgery4,12,18,19,20. Although these studies used validated self-reported measures to record outcomes, some authors did not measure preoperative pain and function as a baseline, while others failed to provide a clear definition of a poor clinical outcome. Some authors categorize outcomes into excellent, good, fair, or poor; however, these categories are arbitrary and may not lead to accurate patient-level measurement21. Moreover, many studies compare mean outcome scores at the group level, a practice that likely has less clinical relevance at the patient level22.

The concept of the minimal clinically important difference (MCID) has been developed to obtain a more clinically oriented measurement of outcome at the individual level. The MCID is defined as the smallest difference in a score that a patient has perceived as beneficial22. It has been observed that the MCID in many self-reported outcome measures is about half the SD of the change23. Although the MCID has been used in different settings, this concept has only recently been introduced to the arthroplasty literature21,24,25.
The primary objective of our study was to report the prevalence of patients showing no clinically important improvement at 1 year post-TKR based on self-reported outcome scores. Our secondary objective was to identify the patient-level predictors for those reporting no clinical improvement following surgery. We hypothesized that despite the reported high success after TKR, there are a substantial number of patients who show no clinically important improvement 1 year after surgery.

MATERIALS AND METHODS
We conducted a retrospective review of TKR registry data collected from 2 academic hospitals: the Toronto Western Hospital (TWH) and the Hamilton Health Sciences Henderson Hospital (HHS) in Ontario, Canada. Both registries prospectively enroll patients on a waiting list to undergo knee replacement surgery to track longitudinal outcomes. All patients provide informed consent to participate. All data are collected by an independent assessor not involved in the medical care of the patients. The Human Subjects Review Committee approved the study protocol at each independent study site.

The inclusion criteria for our study were patients age 18 years and older with a diagnosis of primary or secondary knee osteoarthritis (OA) who had complete preoperative and 1-year self-reported outcome questionnaires. Patients undergoing TKR for inflammatory arthritis, revision surgery, or simultaneous bilateral TKR were excluded. Further, any patient who suffered a postoperative complication including deep infection, periprosthetic fracture, or medical complication necessitating prolonged hospital stay were excluded. All surgeries were performed between September 1998 and December 2007. Similar postoperative rehabilitation protocols were used for all patients at both study sites. Patients were encouraged to mobilize weight-bearing as tolerated on postoperative day 1, and began mobilization with a walker, progressing to independent ambulation as tolerated. All patients were encouraged to participate in formal physiotherapy for a minimum of 3 months.

The primary outcome of this study was to report the prevalence of patients who showed no clinically important improvement at 1 year following TKR using the MCID as measured by the Western Ontario McMaster University Osteoarthritis Index (WOMAC)\(^26\) or the Oxford Knee Score (OKS)\(^27\).

Data collection. Baseline demographic data including age, sex, and body mass index (BMI) were recorded. Baseline medical comorbidity encompassed the 9 categories of anemia, heart disease, lung disease, cancer, diabetes, ulcer disease, hypertension, kidney disease, and liver disease. Functional status and pain level were assessed preoperatively and at the 1-year followup using the WOMAC index\(^26\) for patients treated at the TWH, and the OKS\(^27\) for patients treated at HHS.

The WOMAC index covers the domains of pain (5 items), stiffness (2 items), and function (17 items) and each item is scored on a 5-point Likert scale representing a different degree of intensity (none, mild, moderate, severe, or extreme). The data were set to a range of values from 0 to 100, where a greater score represents greater pain and dysfunction. The psychometric properties of the WOMAC index with respect to reliability, validity, and responsiveness have been well established in the literature for patients with OA\(^28,29\). We defined the MCID for the WOMAC index as 7.5 points as suggested by others\(^30\).

The OKS is a self-administered disease-specific questionnaire developed to measure outcomes after TKR. It consists of 12 items that assess pain and functional impairment. Responses are ranked on a 5-point Likert scale (5 = worst, 1 = best) in which the final score is determined by adding the aggregate score for pain and function. The total scores range from 0 to 60; a higher OKS represents poorer function or greater pain. This instrument’s responsiveness, reliability, and content, and structural validity in this population are well established\(^31\). Further, the OKS may be compared with other similar questionnaires that have been successfully applied to the treatment of knee OA including the WOMAC\(^27,32\). We defined the MCID for the OKS as a 5-point change or less, as suggested by Murray, et al\(^33\).

Statistical analysis. Continuous data including age, BMI, number of comorbidities, WOMAC, and OKS were compared between groups using t-tests, after completing tests of normality. Means and SD are reported for all continuous variables. Binary data such as sex are reported with frequencies, and groups were compared with the chi-squared test.

Multivariable logistic regression modeling was used to determine the predictive factors for patients showing no clinically important improvement following TKR. The independent variables assessed were patient age, sex, BMI, and comorbidity. The dependent variable was defined as a binary measure; those with a clinically important improvement based on the respective change score and those without. The WOMAC and OKS change scores were calculated as the difference between the 1-year total WOMAC or OKS score and the preoperative total WOMAC or OKS score, respectively. Separate models were built for the 2 study sites. A sensitivity analysis to compare the mean total WOMAC scores and OKS before and after the midpoint of our study was conducted using the Mann-Whitney U test to determine any potential maturation effect. All statistical analyses were completed using SPSS version 13.0 (SPSS, Chicago, IL, USA). OR for regression modeling and their 95% CI are reported. All reported p values are 2-tailed, with an alpha of 0.05.

RESULTS
At TWH, 25 patients were excluded from our analysis due to incomplete outcomes data, as compared to 544 patients from the HHS. There were no significant differences in demographic data between included and excluded patients at either center (Table 1). The data of 3177 patients who had TKR and who had complete demographic and outcomes data were included. A total of 2720 patients were recruited from HHS, and 457 patients from TWH (Table 1). The mean age at time of surgery for the complete study population was 68.2 (SD 9.4) years, and 63% (2003/3177) of the patients were women. The mean BMI was 31 kg/m\(^2\) (SD 6.3), and the mean number of comorbidities was 2.1 (SD 1.3). Table 2 shows the mean baseline and 1-year WOMAC and OKS compared between improved and nonimproved patients. As expected, the improved group reported significantly better WOMAC and OKS 1 year after surgery as compared to patients defined as nonimproved (p < 0.05). Overall, 11.7% (373/3177) of patients showed no clinically important improvement in terms of pain and functional activities at 1 year postoperatively.

From the TWH, 11.2% (51/457) of patients showed no clinically important improvement 1 year after TKR based on the WOMAC index, and 12.2% (55/457) of patients showed no clinically important improvement 1 year after TKR based on the WOMAC pain subscale. Of those showing no clinical improvement, 12% (20/170) of patients were men, compared to 11% (31/287) women (p = 0.42). The mean age of patients with no improvement was 71.2 years (SD 7.2) as compared to 66.9 years (SD 9.7) for improved patients (p = 0.004; Table 3). Logistic regression modeling showed that a greater patient age was an independent predictor of no clinically important improvement on the WOMAC scale 1 year postoperatively.
after surgery (p = 0.0003, OR 1.06, 95% CI 1.02–1.10). For every 1-year increase in age at the time of surgery, the patient is 1.06 times more likely to show no improvement on the WOMAC scale at 1 year post-TKR.

In the HHS cohort, 11.8% (322/2720) of patients showed no clinically important improvement at 1 year following TKR (Table 3). The mean age at surgery for the nonimproved patients was 70.2 years (SD 8.7) as compared to 69.5 years (SD 9.3) for the improved patients (p = 0.03). Although this difference was statistically significant, it is not clinically relevant or clinically significant. Fourteen percent (142/1004) of men compared to 10% (180/1716) of women demonstrated no clinical improvement as measured by the OKS (p = 0.006). Logistic regression modeling showed that male sex was an independent predictor of no clinically important improvement on the OKS 1 year after surgery (p = 0.008, OR 0.72, 95% CI 0.57–0.92; Table 4). Men are 0.72 times more likely to show no clinically important improvement on the OKS compared to women, 1 year post-TKR.

Table 1. Demographic data and baseline functional scores for both study sites.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>TWH Included, n = 457</th>
<th>TWH Excluded, n = 25</th>
<th>p</th>
<th>HHS Included, n = 2720</th>
<th>HHS Excluded, n = 544</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, yrs (SD)</td>
<td>67.5 (9.6)</td>
<td>67.4 (9.3)</td>
<td>0.64</td>
<td>69 (9.1)</td>
<td>70 (9.3)</td>
<td>0.12</td>
</tr>
<tr>
<td>Women, n (%)</td>
<td>287 (62)</td>
<td>18 (72)</td>
<td>0.29</td>
<td>1719 (63)</td>
<td>323 (59)</td>
<td>0.10</td>
</tr>
<tr>
<td>BMI, kg/m² (SD)</td>
<td>30 (6.6)</td>
<td>28.6 (7.3)</td>
<td>0.28</td>
<td>32 (6.1)</td>
<td>31.5 (6.1)</td>
<td>0.06</td>
</tr>
<tr>
<td>Comorbidity, mean (SD)</td>
<td>2.8 (1.6)</td>
<td>2.7 (1.5)</td>
<td>0.53</td>
<td>1.5 (1.2)</td>
<td>1.6 (1.2)</td>
<td>0.17</td>
</tr>
<tr>
<td>Mean baseline OKS (SD)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>40.1 (7.6)</td>
<td>40.4 (7.1)</td>
<td>0.35</td>
</tr>
<tr>
<td>Mean baseline WOMAC, total score (SD)</td>
<td>51.6 (17.3)</td>
<td>48.1 (16.8)</td>
<td>0.43</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Mean baseline WOMAC pain (SD)</td>
<td>10.5 (3.6)</td>
<td>9.5 (3.5)</td>
<td>0.88</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 2. Preoperative and 1-year functional outcome scores compared between improved and nonimproved patients. Scores are mean (SD).

<table>
<thead>
<tr>
<th>Score Type</th>
<th>Improved, n = 2804</th>
<th>No Improvement, n = 373</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline WOMAC pain</td>
<td>10.7 (3.5)</td>
<td>7.4 (3.6)</td>
<td>p &lt; 0.0001</td>
</tr>
<tr>
<td>1-year WOMAC pain</td>
<td>3.9 (3.5)</td>
<td>8.8 (4.1)</td>
<td>p &lt; 0.0001</td>
</tr>
<tr>
<td>Baseline WOMAC total</td>
<td>52.1 (16.6)</td>
<td>42.6 (19.6)</td>
<td>p &lt; 0.0005</td>
</tr>
<tr>
<td>1-year WOMAC total</td>
<td>23.7 (16.8)</td>
<td>43.1 (23.3)</td>
<td>p &lt; 0.0001</td>
</tr>
<tr>
<td>Baseline OKS total</td>
<td>40.7 (7.1)</td>
<td>35.4 (9.1)</td>
<td>p &lt; 0.0001</td>
</tr>
<tr>
<td>1-year OKS total</td>
<td>21.6 (7.5)</td>
<td>35.2 (10.1)</td>
<td>p &lt; 0.0001</td>
</tr>
</tbody>
</table>

Table 3. Univariate analysis of demographic predictors comparing improved and nonimproved patients for both study sites.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>TWH (WOMAC)</th>
<th>HHS (OKS)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, yrs (SD)</td>
<td>66.9 (9.7)</td>
<td>71.2 (7.2)</td>
<td>0.004</td>
</tr>
<tr>
<td>Men, n (%)</td>
<td>150 (88)</td>
<td>20 (12)</td>
<td>0.42</td>
</tr>
<tr>
<td>Women, n (%)</td>
<td>256 (89)</td>
<td>31 (11)</td>
<td>0.42*</td>
</tr>
<tr>
<td>BMI, kg/m² (SD)</td>
<td>30.2 (6.7)</td>
<td>28.7 (5.1)</td>
<td>0.7</td>
</tr>
<tr>
<td>Comorbidity, mean (SD)</td>
<td>2.9 (1.7)</td>
<td>2.6 (1.6)</td>
<td>0.22</td>
</tr>
</tbody>
</table>

*p value comparing result between men and women with no improvement (chi-square test). TWH: Toronto Western Hospital; HHS: Hamilton Health Sciences Henderson Hospital; BMI: body mass index; OKS: Oxford Knee Score; WOMAC: Western Ontario McMaster University Osteoarthritis Index.
Additionally, the interaction between age and sex was considered and tested in both regression models; however, the interaction term did not reach significance in either model (p = 0.954 and 0.674, respectively).

We did not observe a statistically significant difference between the mean score of WOMAC at baseline or at 1 year postoperative, either before or after the midpoint of our study (Figure 1). Although the difference between OKS baseline before and after study midpoint was statistically significant, the difference was only 1 point, which is neither clinically significant nor clinically relevant.

DISCUSSION

Despite extensive literature that assesses outcomes following TKR, we would argue that the true rate of patients who show no clinically important improvement following TKR remains unknown. The MCID is considered an essential element to measure the clinical relevance of score changes for any intervention at the individual level. Therefore, we used the MCID to measure TKR outcomes to obtain more clinically relevant results. Our data show that 11.7% of patients showed no change in perceived pain and function at the 1-year followup compared to their preoperative condition. Older age and male sex predicted those with no clinically important improvement following surgery. As the increase in healthcare expenditures highlights the need for consistent assessment of the effectiveness of medical interventions, such findings may help healthcare providers and health service researchers to direct future research to improve outcomes after TKR. Further, we believe that patients should be counseled on a realistic expectation for the potential of persistent pain after surgery, as unfulfilled expectations have been shown to have a substantial effect on patient satisfaction after TKR.

Currently, there is no consensus in the arthroplasty literature as to which patient characteristics place patients at greater risk for a negative outcome. We found that older patient age at the time of surgery and male sex independently predicted no clinically important improvement at the 1-year followup as compared to preoperative status. Similar to our finding, others have shown that age greater than 65 years at the time of surgery is a predictor of a poor clinical outcome at the 2-year followup. This may be attributed to multiple comorbidity and the generalized character of OA in this age group. In a systematic review, identified male sex to be associated with higher risk of revision after TKR. In contrast to our findings, some authors have found that age and sex were not significant predictors of a poor clinical outcome or lack of improvement post-operatively based on self-reported outcome measures. However, many of these results should be interpreted with caution due to small sample sizes or the probability of residual confounding in these analyses that may affect their results.

Few studies have examined the effect of medical comor-

---

### Table 4. Logistic regression modeling predicting patient with no clinically important improvement after surgery, adjusted for age, sex, body mass index, and comorbidity.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>TWH, n = 457</th>
<th>HHS, n = 2720</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1.06 (1.02, 1.10)</td>
<td>0.99 (0.98, 1.01)</td>
</tr>
<tr>
<td>Male sex</td>
<td>1.31 (0.68, 2.56)</td>
<td>0.72 (0.57, 0.92)</td>
</tr>
<tr>
<td>BMI</td>
<td>0.99 (0.94, 1.05)</td>
<td>0.99 (0.98, 1.02)</td>
</tr>
<tr>
<td>Comorbidity</td>
<td>0.87 (0.87, 1.09)</td>
<td>0.98 (0.88, 1.09)</td>
</tr>
</tbody>
</table>

TWH: Toronto Western Hospital; HHS: Hamilton Health Sciences Henderson Hospital; BMI: body mass index.
bidity on functional outcomes following TKR. Similar to our finding, many authors have shown that comorbidity was not a significant predictor of outcome at 1-year follow-up.

In the short term, however, comorbidity was found to be a significant predictor of an increased complications rate and length of hospital stay after TKR. Further, coexisting ipsilateral hip OA could be a contributing factor to ongoing pain after TKR. Similar to others, our study has shown no significant association between BMI and clinical improvement 1 year post-TKR.

This observation of “it looks good but feels bad” remains a poorly understood phenomenon in knee replacement surgery. It is presently unclear if metal sensitivity is a contributing factor to ongoing pain and functional impairment after TKR. Metal sensitivity has been described with different materials used in TKR implants such as nickel, cobalt, chromium, and titanium. Histological findings of hypersensitivity reaction have been observed in the tissues around joint replacement implants in some revision cases. Recently, Granchi, et al. studied metal sensitization in 2 groups of patients who had TKR, some with and some without implant loosening, and compared the findings to a control group without TKR. Their results showed that the prevalence of positive skin sensitivity reaction increased significantly following TKR compared to the control group. Further, a higher prevalence of sensitization reaction was correlated with the presence of clinical symptoms such as moderate pain in patients with stable TKR. Other authors have also suggested that metal sensitivity may be a cause of poor clinical performance after TKR and this should be discussed with patients before surgery.

Although the prevalence of dermal reaction to metal is about 10% of the general population, and 20%–25% in the joint arthroplasty population, it is known that this does not predict a risk of deep tissue allergy to metal after surgery. Future improvement of diagnostic testing for metal sensitivity will likely help to identify patients vulnerable to metal sensitivity reaction and the potential for ongoing pain following knee replacement surgery.

There are potential limitations of our study. First, the reported data represent the experience of high-volume academic hospitals and therefore these findings can be generalized only to a similar setting. Second, we did not examine radiographs for component malalignment, which may be a possible unmeasured confounder of the outcome in our analysis. Third, the MCID for the WOMAC index has been suggested to range between 7.5 and 15 points, and we have chosen the more conservative difference for our study. This would then potentially underestimate the prevalence of poor reported outcomes following this surgery. Finally, there is no well-established value of MCID for the OKS within the literature; however, we used a 5-point difference as suggested by the OKS authors. This value was calculated using a distribution-based method that estimates the MCID at half the SD, which was consistent with our findings.

We acknowledge that use of 2 different outcome measures that represent slightly different forms and are structured in different ways are not comparable. However, it is not our intention to compare the outcome of surgery at 1 institution vs another, nor to compare the 2 outcome measures against each other. We simply reported the outcome from each center individually, which demonstrates a consistent finding across 2 sites. We believe the strength of presenting consistent findings from the 2 sites supports our conclusions.

The strength of this analysis is that we used the preoperative functional score as a baseline and compared it to 1-year followup to examine the effect TKR has on change in pain and function. Further, we reviewed large numbers of consecutive patients and adjusted for clinically relevant potential confounders of the outcome in our analysis. Finally, we used the MCID to test for significance to obtain more clinically relevant results.

Our study has shown that 11.7% of patients who had TKR reported no clinically important improvement 1 year after surgery, based on self-reported outcome measures using the MCID. We found that older patient age and male sex are independent predictors for this poor outcome. This knowledge should be considered when counseling patients preoperatively to help set realistic patient expectations of TKR.

REFERENCES

10. Naylor JM, Harmer AR, Heard RC, Harris IA. Patterns of recovery...


