Indication for total knee arthroplasty based on preoperative functional score: Are we operating earlier?

Peter Dust, MD, MSc  
Jan Kruijt, MD  
Nikolaos A. Stavropoulos, MD, PhD  
Olga Huk, MD, MSc  
David Zukor, MD  
John Antoniou, MD, PhD  
Stephane G. Bergeron, MD, MPH

Background: The number of total knee arthroplasty (TKA) procedures performed annually is increasing for reasons not fully explained by population growth and increasing rates of obesity. The purpose of this study was to determine the role of patient functional status as an indication for surgery and to determine if patients are undergoing surgery with a higher level of preoperative function than in the past.

Methods: A systematic review and meta-analysis of the MEDLINE, Embase and Cochrane databases was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. Functional status was assessed using the 36-Item Short Form Health Survey’s physical component summary (PCS) score. Only primary procedures were included. Articles were screened by 2 independent reviewers, with conflicts resolved with a third reviewer. Meta-regression analysis was performed to determine the effect of time, age and sex on preoperative PCS score. Subgroup analysis was performed to compare results for the United States with those for the rest of the world.

Results: A total of 1502 articles were identified, of which 149 were included in the study. Data from 257 independent groups including 57,844 patients recruited from 1991 to 2015 were analyzed. The mean preoperative PCS score was 31.1 (95% confidence interval 30.6–31.7) with a 95% prediction interval of 22.8–39.5. The variance across studies was found to be significant (p < 0.001) with 99.01% true variance. Year of enrolment, age, the percentage of female patients and geographic region did not have any significant effect on preoperative PCS score.

Conclusion: Patients are undergoing TKA with a level of preoperative function similar to their level of function in the past. Patient age, sex and location did not influence the functional status at which patients were considered to be candidates for surgery.

Contexte : La hausse du nombre annuel d’interventions pour prothèse totale du genou (PTG) ne peut s’expliquer entièrement par la croissance de la population et l’augmentation du taux d’obésité. Cette étude avait pour but de déterminer dans quelle mesure le statut fonctionnel constitue une indication de la chirurgie et si à présent la chirurgie est effectuée chez des malades plus fonctionnels que par le passé.

Méthodes : Une revue systématique et méta-analyse des articles recensés à partir des bases de données MEDLINE, Embase et Cochrane a été effectuée conformément à la Norme de réalisation de revue systématique de la littérature PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). Le statut fonctionnel a été établi selon le score du sommaire de la dimension physique (SDP) du questionnaire SF-36. Seules les interventions primaires ont été incluses. Les articles ont été passés en revue par 2 réviseurs indépendants, et tout désaccord était résolu par une troisième personne. Une méta-analyse de régression a été effectuée pour évaluer l’effet du temps, de l’âge et du sexe sur le score préopératoire du SDP. Une analyse de sous-groupes a permis de comparer les résultats entre les États-Unis et les autres pays.

Résultats : En tout, 1502 articles ont été recensés, et 149 ont été inclus dans l’étude. Les données de 257 groupes indépendants regroupant 57,844 participants recrutés entre 1991 et 2015 ont pu être analysées. Le score préopératoire moyen du SDP a été de 31,1 (intervalle de confiance de 95 % 30,6–31,7), avec un intervalle de prévision de 95 % à 22,8–39,5. La variance s’est révélée significative entre les études (p < 0,001) avec une variance vraie de 99,01 %. L’année de l’inclusion, le pourcentage de participants féminin et le lieu géographique n’ont eu aucun effet significatif sur le score préopératoire du SDP.

Conclusion : Les PTG sont effectuées chez des participants qui ont un score de fonctionnement similaire aux scores d’autrefois. L’âge, le sexe et le lieu géographique n’ont pas exercé d’influence sur le score de statut fonctionnel à partir duquel les personnes étaient considérées candidates à la chirurgie.
Total knee arthroplasty (TKA) is one of the most common orthopedic procedures. It is cost effective in relieving pain and improving function of patients with advanced joint disease, usually due to osteoarthritis.\(^1,2\) The number of procedures performed annually is rapidly increasing, with a 162\% increase in volume among United States Medicare enrollees alone between 1991 and 2010.\(^3\) Population growth and the obesity epidemic cannot fully explain this increase.\(^4\) The purpose of this study was to examine whether changes in the indications for primary TKA account for the increases in the number of procedures performed each year. Specifically, we sought to determine if patients were undergoing primary TKA at a better level of function than in the past. Differences in patient functional status could indicate changes in patient expectations and a desire to undergo surgery to remain active or a change in surgeon practice. Subgroup analysis was performed to compare the findings in the United States with those in the rest of the world.

### Methods

To study how the preoperative functional health status of patients undergoing primary TKA has changed over time, we performed a systematic review and meta-analysis. Preoperative functional status was assessed using the 36-Item Short Form Health Survey (SF-36). This general health survey was developed in 1992 as part of the Medical Outcomes Study, a 4-year study designed to assess the influence of specific factors on the outcomes of care and develop practical tools for monitoring patient outcomes.\(^5-7\) It was designed to produce data relevant for both research and clinical practice and generates 8 individual scale scores: physical functioning, role limitations due to physical problems, social functioning, bodily pain, general mental health, role limitations due to emotional problems, vitality and general health perceptions.\(^8\) Each score ranges from 0 to 100, with higher scores indicating better health status. The 8 individual scores can be combined into 2 aggregate scores: the physical component summary (PCS) and mental component summary (MCS) scores. These aggregate scores offer the advantages of reducing the number of analyses while producing smaller confidence intervals and smaller floor and ceiling effects.\(^6,8\)

The SF-36 scoring system was chosen for several reasons. First, it is the only scoring system that permits comparisons across multiple decades. Although multiple derivations of the SF-36 have been developed, the traditional SF-36 scoring system has remained largely unchanged since 1993.\(^6\) The only substantial change was the development of the SF-36 version 2 in 1996, but this version has been shown to produce results comparable to those produced by the original survey.\(^5,9\) Second, the SF-36 survey has been used extensively in the orthopedics literature to assess patients who undergo TKA.\(^6,10\) Finally, the SF-36 scoring system has been translated into more than 170 languages,\(^6,9\) thereby supporting international assessments. Although other scoring systems, such as the EQ-5D, exist, none have been as ubiquitously used, have been validated in as many languages and have remained as consistent in their scoring system as the SF-36.\(^11,12\)

The systematic review and meta-analysis was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.\(^13\) A literature search of the MEDLINE, Embase and Cochrane databases was performed on Dec. 12, 2016. Search terms included “arthroplasty, replacement, knee,” “knee prosthesis,” “knee replacement,” “knee arthroplasty,” “knee endoprosthesis,” “TKR,” “TKA,” “SF-36” and “Short Form-36.” No restrictions were placed on the search, including no language restrictions. Titles and abstracts were screened by 2 independent reviewers (J.K., N.A.S. or P.D.). The main inclusion criterion was primary TKA. Articles were excluded if they did not include patients with TKA; if they did not report SF-36 scores; if they consisted of conference proceedings only, a published abstract only or a protocol only; or if the only form of TKA reported was revision TKA. Conflicts between the reviewers were resolved by consensus with the assistance of a third independent reviewer (S.G.B.).

After abstract review, eligible articles underwent full manuscript review by the same reviewers (J.K., N.A.S. or P.D.). Additional exclusion criteria were applied as follows: articles with no preoperative SF-36 scores; articles with incomplete scale scores to calculate the PCS, if this score was not explicitly reported in the article; articles that reported SF-36 scores for patients who underwent TKA combined with scores for patients who underwent other procedures (e.g., total hip arthroplasty); articles that reported SF-36 scores for patients whose data had been reported in other publications that were already included in this meta-analysis; review articles; meta-analyses; and letters to the editor or editorials.

Conflicts were again resolved by consensus with the third reviewer (S.G.B.). Articles in languages not spoken fluently by the main reviewers were discussed with individuals fluent in the those languages. We had the articles fully translated, and then Google Translate was employed to confirm parts of the translations. The reference lists of all relevant articles, including review articles and meta-analyses, were manually cross-referenced to identify potential papers missed in the literature search. Articles that satisfied the inclusion criteria and did not meet any of the exclusion criteria were included in the analysis.

Data from all included articles were extracted by the 2 independent reviewers (J.K., N.A.S. or P.D.) and then the work of the 2 reviewers was compared for accuracy. Extracted data included earliest year of patient enrolment, sample size, region from which the patients were sampled (categorized as United States or other), patient age,
percentage of female patients, percentage of patients diagnosed with primary osteoarthritis, and PCS and MCS scores. If the year of enrolment was not provided, the date of publication or the date the manuscript was received by the publisher was used. If the PCS and MCS scores were not directly reported, they were calculated from the 8 scale scores using the algorithm provided by Laucis and colleagues. We employed the orthogonal method of calculation as this is the most commonly used method, to maximize the comparability of our results with the PCS scores explicitly reported in the literature.

The level of evidence of each study was assessed using the guidelines provided by Marx and colleagues. As the purpose of this study was to examine the trend in preoperative SF-36 scores over time rather than to examine the effects on the SF-36 score of any individual study, both experimental and observational studies were included. For the same reason, the risk of bias in each study was not assessed as the bias in any individual study design would not have an impact on the simple reporting of the preoperative SF-36 scores.

**Statistical analysis**

Statistical analysis was performed using RStudio (Posit) and Comprehensive Meta-Analysis version 3 (Biostat). Inter-rater reliability was assessed using the Cohen $\kappa$. The overall PCS score across all studies was calculated using the mean and standard deviation or 95% confidence interval (CI), depending on which metric was reported. Missing standard deviation values were imputed using a weighted mean of the reported standard deviations. This mean was calculated using the sample size of the studies with reported standard deviations as weights. The validity of this imputation technique was tested by rerunning the analysis using the true mean of the reported standard deviations as the imputed value. As there is no definitive technique for imputing missing standard deviations in meta-analyses, we chose this technique supported by the literature. The PCS scores were combined using a random-effects model. The variance in the PCS scores was assessed using the Q value and $I^2$ statistic. The 95% prediction interval was calculated using the formula provided by the Comprehensive Meta-Analysis program.

The associations of year of patient enrolment, mean patient age and percentage of female patients in each group with the PCS score were assessed through separate meta-regression analyses using Comprehensive Meta-Analysis. The effect of region of patient enrolment was assessed by performing a subgroup analysis on the overall PCS scores and comparing the groups using the Q value and $F$ statistic. A mixed-effects model was used. A $p$ value less than 0.05 was considered significant.

**RESULTS**

The results of the literature search are summarized in Figure 1. A total of 149 studies, of which 16 were
randomized controlled trials (RCTs), were included in the analysis. From the included studies, a minimum of year of enrolment, sample size and PCS score were extracted for 257 independent patient groups. Some studies contributed more than 1 group as separate arms of the same study (e.g., in a study comparing different implants); these were extracted separately, provided sufficient data were reported. Figure 2 shows the distribution of the levels of evidence of the articles from which the groups’ data were extracted. Data for 32 groups were extracted from the RCTs. These groups include data from 57,844 patients first enrolled between 1991 and 2015. The PCS score was explicitly reported for 184 groups and calculated for the remaining 73 groups. Sixty-one groups were from the US; the remaining 196 groups were from the rest of the world. The sex distribution was reported for 217 groups. The overall population contained data from at least 36,038 women and 15,361 men. The mean age was reported for 239 groups. It ranged from 35.1 to 89.0 years, and the weighted mean for the groups in which patient ages were known was 67.7 years. The percentage of patients diagnosed with primary osteoarthritis was reported for 186 groups; the data of at least 17,526 patients with primary osteoarthritis were included in the analysis.

Combining all of the PCS scores produced a mean score of 31.1 (95% CI 30.6–31.7). The 95% prediction interval was 22.8–39.5. The Q value was 25,938.105 (256 degrees of freedom, \( p < 0.001 \)). The \( I^2 \) statistic was 99.01. These results indicate that the mean PCS score varied across the included studies and 99.01% of this variance was true variance and not due to chance.

The meta-regression analyses of year of enrolment, mean age and percentage of female patients per group are shown in Figure 3, Figure 4 and Figure 5, respectively.
The preoperative PCS score was found to not change significantly with any of these variables. The correlation coefficient for year of enrolment was 0.09 ($p = 0.11$). The coefficient for mean age was 0.03 ($p = 0.54$) and the coefficient for the percentage of female patients per group was –0.005 ($p = 0.73$).

Subgroup analysis comparing data from the US with data from the rest of the world did not reveal any significant differences in preoperative PCS score. The American groups had a mean PCS score of 30.7 (95% CI 29.8–31.7). The groups from the rest of the world had a mean PCS score of 31.3 (95% CI 30.3–32.0). When these 2 groups were compared, the $Q$ value was 0.815 (1 degree of freedom, $p = 0.37$).

**Discussion**

The analysis of preoperative PCS scores revealed several interesting results. While the overall mean score of 31.1 had a fairly tight 95% confidence interval, the analysis of variance and wide 95% prediction interval suggest that there were substantial differences among the preoperative scores reported in the included studies. Furthermore, very little of this variance was due to chance as 99.01% of the variance was true variance. The differences among the groups, however, were not due to differences in year of enrolment, mean age or the percentage of female patients in each group. Regional differences also failed to account for this variance as no significant difference was found when we compared the preoperative scores from the US with those from the rest of the world. Therefore, 1 or more additional factors not identified by our study clearly had an influence on the preoperative PCS scores. One potential factor could be body mass index or obesity. Obesity rates have been steadily increasing, and obesity is a well-recognized contributing factor to the development of disabling joint disease.

Of course, this variability could also be the result of the currently subjective nature of determining when a patient is a candidate for surgery. Given the substantial amount of true variance among the preoperative PCS scores, we caution against over-interpreting the mean PCS score produced by this meta-analysis. While we had hoped our analysis would reveal a target PCS score at which patients could be considered to be candidates for surgery, given the 95% prediction interval of 22.8 to 39.5, we do not believe such a recommendation can be made. The 95% prediction interval estimates the range in which the results of similar studies will fall in the future. We consider a range of 16.7 points to be too large to make any recommendations, particularly as the minimal clinically important difference in the PCS score is approximately 10 points.

![Fig. 4. Regression of preoperative PCS score on mean age. Each circle corresponds to a group included in the analysis. The size of the circle is proportional to the group’s contribution to the analysis. PCS = physical component summary.](image_url)
That our meta-regression analysis revealed no significant changes in preoperative PCS score over time is important as it indicates that patients are undergoing TKA at a level of function similar to that of their counterparts in the past. Changes in patient functional status therefore cannot account for the steadily increasing volume of TKA procedures performed annually. This result is surprising as we anticipated that technological improvements in implant longevity and an unwillingness to accept substantial declines in function would entice patients to undergo surgery earlier in their disease process. Additional studies will be necessary to determine the as-yet-unidentified variables contributing to this rise. One potential factor could be the number of surgeons performing primary TKA. Both the total number and the number of surgeons per capita have been generally increasing.\textsuperscript{21–26} Reassuringly, however, our results suggest that despite the increasing number of surgeons, the indications for surgery are not being eroded by operating on healthier patients to fill operating room time.

That age, sex and regional distribution were also found to have no significant relationship with preoperative PCS score is interesting because it implies a certain global uniformity in terms of the functional status at which patients are considered to be candidates for surgery. Although significant true variance did exist in the PCS scores, no specific patterns or biases were revealed when these variables were considered. With respect to sex, this result slightly contradicts previous studies that have found that arthroplasty tends to be significantly more underused by women than men.\textsuperscript{27,28} The difference, however, is likely due to the point in patients’ treatment course at which our data were collected. These studies\textsuperscript{27,28} surveyed populations to determine the need for and use rates of arthroplasty in the community. Our study assessed the functional status of patients of both sexes once they were considered to be candidates for surgery. Our results suggest that once a patient is evaluated by an orthopedic surgeon, the functional status at which the decision to proceed with surgery is made is the same for men and women; however, barriers to accessing this evaluation may still exist.

The lack of regional differences is surprising, particularly with regard to the US. Given that the US has some of the highest rates of obesity and chronic disease burden in the world, the highest health care expenditures as a percentage of gross domestic product and a population generally intolerant to long wait times,\textsuperscript{29} we expected the preoperative PCS score of American patients to be substantially higher than that of patients in other regions. Our results, however, suggest the US is on par with the rest of the world in this regard.

\textbf{Fig. 5.} Regression of preoperative PCS score on percentage of female patients per group. Each circle corresponds to a group included in the analysis. The size of the circle is proportional to the group’s contribution to the analysis. PCS = physical component summary.
Limitations

Although we do not believe that it had a substantial impact, a potential limitation of our study is the fact that the floor effects of the PCS score in the evaluation of patients who undergo TKA are not well defined. Several studies have found floor effects for some of the scale scores that contribute to the PCS.\textsuperscript{20,30} While 1 study found the PCS score to be unaffected by floor effects, this result was based on only 24 patients.\textsuperscript{10} If floor effects had substantially affected our results, we would have anticipated a much narrower 95% prediction interval and lower true variance among the PCS scores. As a result, we believe their impact to be minimal, but they remain a potential limitation notwithstanding.

CONCLUSION

Patients undergoing TKA with a preoperative functional status similar to that of their counterparts in the past. Changes in functional status, therefore, do not explain the increasing volume of TKA procedures performed annually. Patient age, sex and location do not influence the functional status at which patients are considered to be candidates for surgery. Further studies are needed to determine the optimal functional status at which patients should undergo surgery. Acknowledgements: The authors thank Dr. Magdalena Tarchala, Dr. Susan Ge and Dr. Majid Khademolhosseini for their assistance in translating articles into English from Polish, Chinese and Persian, respectively.

Affiliations: From the Division of Orthopaedic Surgery, McGill University, Montréal, Que. (Dust, Huk, Zukor, Antoniou, Bergeron); the Department of Orthopaedic Surgery, Hôpital du Suroît, Salaberry-de-Valleyfield, Que. (Krijuit); and the Department of Orthopaedic Surgery, General Hospital of Nikaia Piraeus “Ag. Panteleimon,” Nikaia, Piraeus, Greece (Stavropoulos).

Competing interests: None declared.

Contributors: O. Huk, D. Zukor, J. Antoniou and S. Bergeron designed the study. P. Dust, J. Krijuit and J. Stavropoulos acquired the data, which P. Dust and S. Bergeron analyzed. P. Dust wrote the article, which J. Krijuit, J. Stavropoulos, O. Huk, D. Zukor, J. Antoniou and S. Bergeron critically revised. All authors gave final approval of the version for publication.

Content licence: This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY-NC-ND 4.0) licence, which permits use, distribution and reproduction in any medium, provided that the original publication is properly cited, the use is noncommercial (i.e., research or educational use), and no modifications or adaptations are made. See: https://creativecommons.org/licenses/by-nc-nd/4.0/

References


