Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index as an Assessment Tool to Indicate Total Knee Arthroplasty in Patients with Primary Knee Osteoarthritis

Ab Rahman S[a], Narhari P[a,b], Sharifudin, MA [a,c], Shokri AA[a]

[a]Department of Orthopedics, School of Medical Sciences, Universiti Sains Malaysia, Kelantan, Malaysia
[b]Department of Orthopaedics, Penang General Hospital, Penang, Malaysia
[c]Faculty of Medicine, Universiti Sultan Zainal Abidin, Medical Campus, Jalan Sultan Mahmud, Terengganu, Malaysia

ABSTRACT

Introduction: The WOMAC questionnaire is used extensively in the assessment of knee osteoarthritis but not as an indicator of total knee arthroplasty (TKA). We aimed to assess if the WOMAC score can be used as an assessment tool to indicate TKA in primary knee osteoarthritis. **Materials and Methods:** Patients with primary knee osteoarthritis Kellgren and Lawrence grades 3 or 4 were assessed separately by either one of two arthroplasty surgeons for indications to undergo TKA and evaluated using the WOMAC score. Patients who had received any form of intra-articular knee injection within six months before assessment were excluded. Patients were divided into two groups; indicated for TKA, and for non-surgical treatment. The mean WOMAC score of patients from both groups was compared using Independent t-test. The accuracy at different cut-off points in the score indicated for TKA was determined using the receiver operating characteristic (ROC) curve plotted on sensitivity and 1-specificity graph. **Results:** The study included 74 patients with a mean age of 62.5 years (SD=8.26). There was a significant difference in the total WOMAC score (p<0.001) and its components between patients of the two groups. At the cut-off point of 30, the specificity and sensitivity of the score indicated for TKA was 91.18% and 95.00% respectively. The positive predictive value was 93.24%. **Conclusion:** We recommend all primary knee osteoarthritis Kellgren and Lawrence grades 3 or 4 with a score of 30 and above should be referred to an arthroplasty surgeon for the possibility of requiring TKA.

KEYWORDS: knee joint, osteoarthritis, joint replacement, receiver operating characteristic curve, WOMAC questionnaire

INTRODUCTION

The decision-making concerning total knee arthroplasty (TKA) is extremely complex from the patients' perspective. It is equally difficult from the surgeons' point of view as well.[1] The only parameter that has good consensus for the indication of TKA is pain not responsive to pharmacological therapy.[2] Several authors and institutions have come up with indications for TKA.[3,4] Generally, the criteria used previously were; general guidelines, non-specific and demand the surgeon to make his or her decision based on their own experience.[5,6] It was reported that there is disagreement on opinion for TKA among orthopaedic surgeons. From 34 different patient characteristics, only 14 characteristics had an agreement between surgeons for the indication of TKA.[2] This possibly explains why the rate of TKA differs largely between institutions.[2,7]

Western Ontario and McMaster Osteoarthritis Index (WOMAC) is a disease-specific questionnaire extensively used in patients with knee and hip arthritis. It was initially designed to evaluate clinically important, patient-relevant changes in health status and therapeutic benefits following

Corresponding Author:
Assoc. Prof. Dr Mohd Ariff Sharifudin
Faculty of Medicine,
Universiti Sultan Zainal Abidin,
Medical Campus, Jalan Sultan Mahmud,
20400 Kuala Terengganu, Terengganu, Malaysia
Tel :+609-627 5630
Email : dr.ariff.s@gmail.com

interventions.[8] WOMAC has been translated into more than 65 different languages worldwide. It was originally meant to be a self-administered tool but it has been validated for usage via telephone interviews, multimedia with audio-visual presentations, verbally spoken computer programs, and mobile phone applications.[9-11] Minimal guide is required to complete the questionnaire and interpret the score. Studies have shown that it has greater or at least comparable responsiveness to changes compared to other scoring tools, with excellent reliability, construct, content and criteria validity.[8,9]

Despite having a numerical scale, there is no formal range for severity in WOMAC.[8] It is interesting to know if the WOMAC score can be used as an indicator score to guide in decision making for TKA in primary knee osteoarthritis. We hypothesized that there is a significant difference in the total WOMAC score and its components between patients with primary knee osteoarthritis Kellgren and Lawrence grades 3 or 4 indicated for TKA and indicated for non-surgical treatment. The present study aimed to compare the mean WOMAC score between patients of the two groups and to determine accuracy at different cut off points in the WOMAC score to decide on the need for TKA in primary knee osteoarthritis.

MATERIALS AND METHODS

The study was designed as a comparative cross-sectional study involving patients with primary knee osteoarthritis. The mean WOMAC score between patients with primary knee osteoarthritis indicated for TKA and those who were indicated for non-surgical treatment were compared and evaluated to determine the cut-off point for the indication to perform TKA. Participants were selected among patients aged 50 years and above with primary knee osteoarthritis Kellgren and Lawrence grades 3 or 4 presenting to the arthroplasty clinic of a tertiary institution. Patients who had received any form of intra-articular knee injection within six months before assessment were excluded from the study. Those with post-traumatic osteoarthritis or other types of arthritis including rheumatoid arthritis, gouty arthritis, psoriatic arthritis, and post-septic arthritis were also excluded. Other exclusion criteria were patients whose assessed knee had underwent TKA or with a history of major psychiatric illness, dementia or medical illness unfit for operative procedures.

Using the Power and Sample Size (PS) software to calculate the sample size for both objectives, a minimum of 30 participants per group was required for significant statistical analysis. From our pilot study, the response within each subject group was normally distributed with a standard deviation of 17.8. If the true difference in means of both groups is 27.7, a minimum of eight patients for each group was required to be able to reject the null hypothesis that the means of both groups are equal with probability (power) 0.8. The type-I error probability associated with this two-sided test of the null hypothesis is 0.05.

Data collection was commenced after approval by the Human Research Ethics Committee of the institution. The study was conducted in three phases. Phase A was an assessment using the WOMAC questionnaire. Phase B involved independent assessment by arthroplasty surgeon on the need for TKA. For both phases, the assessments were performed in either manner; Phase A followed by Phase B or vice versa. Phase C was measuring patients' weight and height, and body mass index (BMI) calculation.

In Phase A, a single general orthopaedic surgeon was designated to document participants' basic demographics and detailed history of previous medical conditions. Participants were evaluated using the WOMAC questionnaire by the investigator. As some of the elderly patients were illiterate or could not comprehend the written form, WOMAC was used as a physician-filled tool. The investigator verbally asked the questions and participants' responses were documented. This was to standardize the data collection method. Phase A was conducted in a designated area in the arthroplasty clinic and neither of the two arthroplasty surgeons was present at the time of completion of the WOMAC score to eliminate any form of bias. The total WOMAC score was calculated and converted into percentage.

Upon completion of the WOMAC questionnaire, participants were brought to a separate consultation room where they were assessed separately by either one of two consultant arthroplasty surgeons on the need for arthroplasty (Phase B). During the entire process of assessment, the arthroplasty surgeons were blinded to the results of the WOMAC score. In deciding for the need for arthroplasty, patients' agreement was discarded, as it did not affect the relationship between the WOMAC score and

assessment by the surgeons on the need for TKA. The selection of which surgeon performed the assessment was done randomly on the availability of the surgeons.

Since the study used two independent consultant arthroplasty surgeons to decide regarding the need for TKA, a pilot study was conducted before the actual study to evaluate the reliability and consistency of both participating surgeons in decision -making on a given participant using the Cohen's Kappa Statistic. Fifteen percent of the total sample size was required for a good validity of Kappa Statistics.

Data was entered and analysed using Predictive Analytics Software (PASW) Statistics version 18.0. The mean WOMAC score of patients indicated for TKA and patients for non-surgical treatment were compared using Independent t-test. The accuracy at different cut-off points in the score indicated for TKA was determined using the receiver operating characteristic (ROC) curve plotted on sensitivity and 1-specificity graph. STATA 10.1 (Stata Corporation, College Station, USA) was used to analyse Kappa Statistics and ROC curve.

RESULTS

A total of 75 participants with a mean age of 62.47 (SD=8.26) years (range 50 to 86 years) fulfilled the inclusion criteria. Most of the participants (n=63, 84.0%) were in the 50-69 years old age group. The majority of the participants (n=55, 73.3%) were female. The cohort represents a multi-ethnic distribution (Figure 1).
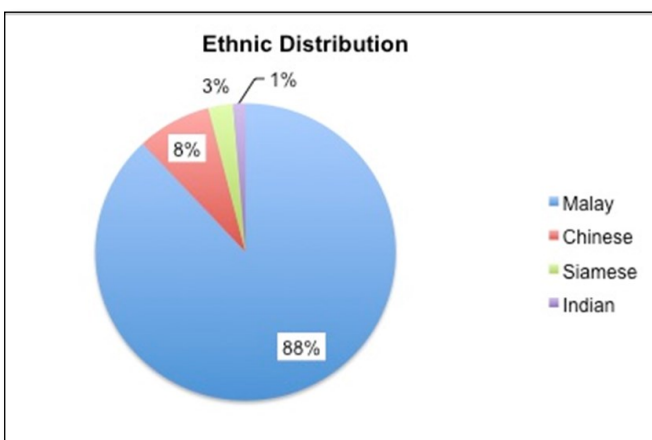

Figure 1: Pie chart depicting the ethnic distribution of the patients (n=75)

Out of the total 75 participants, 12 (16.0%) were evaluated by both assessing surgeons independently during the pilot study. Both surgeons came to an agreement on the need for TKA in 11 out of 12 participants. Kappa Statistics showed a value of 0.833 (p = 0.002) with an excellent agreement rate of 91.7%, indicating an almost perfect agreement between the two surgeons. The remaining 63 participants were assessed by either of the two surgeons. Surgeon 1 assessed a total of 43 participants and indicated TKA in 24 (55.8%) of them, whereas TKA was indicated in 19 out of 44 (43.2%) participants assessed by Surgeon 2. As one of the participants who was assessed by both surgeons during the pilot study had a discrepancy in the need for TKA, the participant was omitted from the total statistic to avoid false results.

Based on the evaluation of both surgeons, TKA was indicated in 34 (45.9%) out of 74 participants. The mean age of participants indicated for TKA was 65.35 years (SD=8.65) compared to participants who did not need the surgery (mean age of 60.03 (SD=7.14)). They also had a higher mean BMI of 31.59 (SD=4.46) compared to participants who do not require TKA (Table I). Table I also summarizes the means of all three components and the total WOMAC score for both groups of participants. Participants indicated for TKA scored higher in the total WOMAC score and all three components compared to participants who were not. The differences between the mean scores were statistically significant (p < 0.001 in all four parameters).

The ROC curve on the indication for TKA showed that below the WOMAC score of 19, there was a 100.0% sensitivity that the patients do not need the surgery (Table II). On the other hand, a WOMAC score of 33 and more had a 100.0% specificity on the need for TKA. Participants were correctly classified at best (93.2%) when the score of 30 or 31 was taken as the cut-off point. Specificity was highest (97.5%) at a WOMAC score of 32, but with a lower rate of sensitivity (82.4%) with only 90.5% of participants were correctly classified according to their need for TKA. The cut-off point with the highest sensitivity and specificity rate (91.2% and 95.0%, respectively) with the best rate of participants correctly classified was the score of 30. The area under the curve was 0.982 (Figure 2).

Table I: Descriptive statistics for overall participants and comparison of means of all three components and the total WOMAC score based on indication for TKA.

Variables	Mean (SD*) of overall participants (n=74)	Mean (SD*) based on indication for TKA		Mean difference (95% CI†)	t-statistics (df§)	p- value**
		Yes (n=34)	No (n=40)			
Age	62.47 (8.26)	65.35 (8.65)	60.03 (7.14)			
BMI	29.95 (4.26)	31.59 (4.46)	28.55 (3.57)			
Pain score	5.07 (4.42)	8.29 (4.17)	2.33 (2.23)	-5.97 (-7.57, 4.36)	-7.477 (49)	<0.001
Stiffness score	2.93 (2.01)	4.12 (1.81)	1.93 (1.59)	-2.19 (-2.98, -1.41)	-5.553 (72)	<0.001
Physical function score	19.08 (13.19)	30.21 (9.93)	9.63 (6.55)	-20.58 (-24.57,-16.59)	-10.328 (55)	<0.001
Total WOMAC score	27.08 (17.82)	42.62 (12.70)	13.88 (8.27)	-28.74 (-33.84,-23.65)	-11.310 (55)	<0.001

*SD = standard deviation, †CI = confidence interval, §df = degree of freedom
**Independent t test

DISCUSSION

In cases of end-stage knee osteoarthritis with severe knee pain and disability, it is quite easy to decide on TKA.[5,12] However, there seem to be arguments on the need for TKA in patients who have moderate arthritis with disability. These arguments have been intensified, as there are other newer methods besides TKA to treat mild to moderate knee osteoarthritis.[5,12] Most of these patients are seen by general practitioners at primary care or district health institutions.[13] There is an on-going dilemma for the general practitioners to refer cases for consultation on TKA as who need the surgical intervention and who do not.[5]
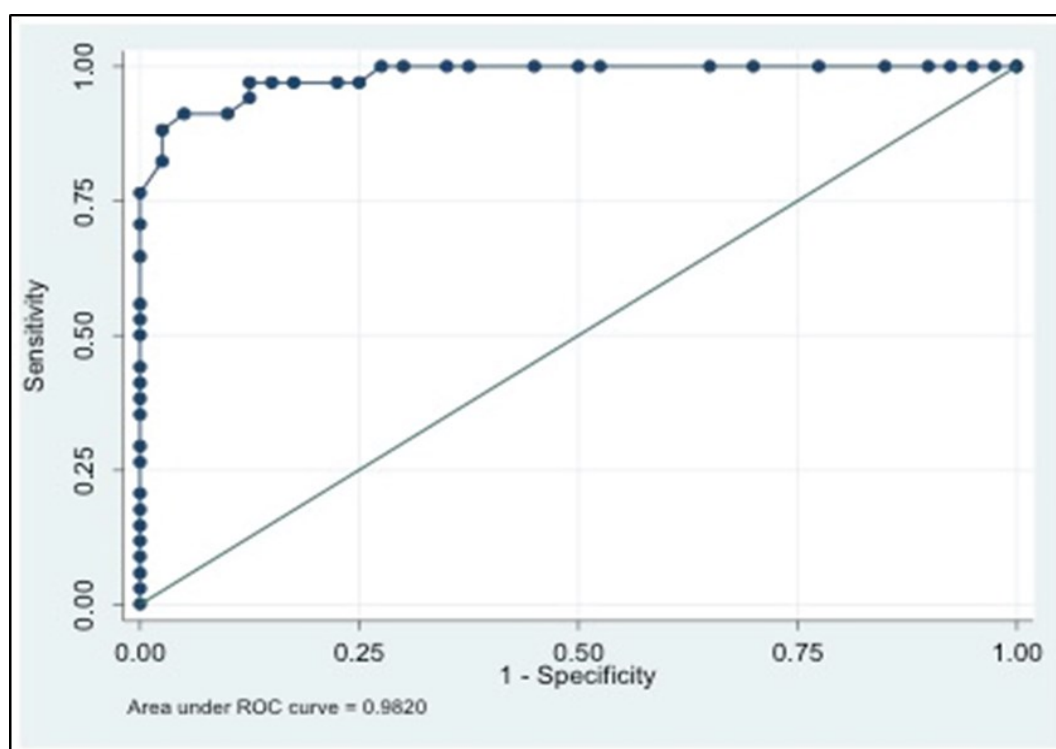

Figure 2: Receiver operating characteristic (ROC) curve on the indication for TKA

Table II: The sensitivity and specificity at various cut-off points of WOMAC score to determine indication for TKA

Cut point	Sensitivity	Specificity	Correctly Classified	LR+*	LR-*
≥ 1	100.00%	0.00%	45.95%	1.0000	
≥ 2	100.00%	2.50%	47.30%	1.0256	0.0000
≥ 4	100.00%	5.00%	48.65%	1.0526	0.0000
≥ 17	100.00%	65.00%	81.08%	2.8571	0.0000
≥ 18	100.00%	70.00%	83.78%	3.3333	0.0000
≥ 19	100.00%	72.50%	85.14%	3.6364	0.0000
≥ 20	97.06%	75.00%	85.14%	3.8824	0.0392
≥ 21	97.06%	77.50%	86.49%	4.3137	0.0380
≥ 28	94.12%	87.50%	90.54%	7.5294	0.0672
≥ 29	91.18%	90.00%	90.54%	9.1176	0.0980
≥ 30	91.18%	95.00%	93.24%	18.2353	0.0929
≥ 31	88.24%	97.50%	93.24%	35.2941	0.1207
≥ 32	82.35%	97.50%	90.54%	32.9412	0.1810
≥ 33	76.47%	100.00%	89.19%		0.2353
≥ 65	5.88%	100.00%	56.76%		0.9412
≥ 67	2.94%	100.00%	55.41%		0.9706
> 67	100.00%	100.00%	54.05%		1.0000

*LR = likelihood ratio

National Institutes of Health (NIH) United States guidelines for TKA indicate replacement surgery if the patient has radiographic evidence of joint damage, moderate to severe persistent pain that is not adequately relieved by nonsurgical management and clinically significant functional limitation resulting in diminished quality of life.[14] South West London Effective Commissioning Initiative (SWLECI) criteria were developed to consider eligible patients to receive funds for TKA.[15] Like the previous guideline, the latter is also very subjective and non-specific.

The more elaborated tools for indications on TKA, on the other hand, are too complex and not suitable for clinical use. An example would be the classification tree by Escobar et al.,[4] which needs proper assessment of pain severity, age group of patient, proper radiographic assessment and classifying the radiological findings using the Ahlback classification,[8] type of symptoms, localization of pain, and ability to ambulate. To use these tools, x-rays are needed to look for the severity of arthritis. In a small health institution where an x-ray service is not available, this tool cannot be readily used. An ideal tool for the guideline should be simple, easy to use, readily available, objective and preferably can be classified numerically, can be done with no or minimal investigations, and should have a relatively good track record in terms of consistency, reliability, and validity.

The WOMAC score is thought to be a suitable tool to decide on TKA as the parameters that make up the score are the main areas that are assessed in a patient with knee osteoarthritis who are planned for TKA. The score has a total of 24 questions covering almost all possible aspects of knee osteoarthritis presentations and severity.[8] Pain and functional disability are consistently identified as the most important indication for TKA.[7,16] Twenty-two out of 24 questions in the WOMAC score assess these two components. The remaining two questions cover the aspect of stiffness, which is also important in deciding on TKA. The score has a numerical value ranging from 0 to 100; 0 being the best possible score with no disability and 100 being the worst. It has good to excellent validity, consistency, reliability, and responsiveness.[9] It is easy to use and takes approximately 10 minutes to be completed. X-rays are not needed to complete the score, which makes its usage almost universal at any health institution, especially the smaller ones that do not have imaging facilities.

A study by Hawker et al. used the WOMAC score to predict time to TKA.[17] They used any score on the

WOMAC scale above 39 as an indication for TKA. The cut-off score of 39 was used as it represents the 25th percentile of scores for patients undergoing arthroplasty in Ontario. We disagree with this oversimplification technique to determine the cut-off point. As it is a 25th percentile range, it essentially means that 75% of patients undergoing TKA has a score of 39 and above. However, it also reflects that 25% of all TKA patients might have a score below 39 points on the WOMAC score. Although the idea of using the WOMAC score was similar to the current study, the statistical method was completely different. We believe that the statistical tool used in the previous study did not produce the sensitivity and specificity that we are attempting to address in the current study. We are interested to see the exact cut-off point with its given specificity and sensitivity, which can be extrapolated on an ROC curve.

There are abundant works of literature that show inconsistency between indications for TKA among health care providers.[2,7,16] We performed a Kappa analysis to establish the consistency of the two arthroplasty surgeons involved in the study. Contrary to most literature that showed inconsistency between arthroplasty surgeons, we noted excellent agreement between the two surgeons in this study. The Kappa value above 0.81 is considered as almost perfect.[18] This shows that the consistency in deciding for TKA is very high between the two assessing surgeons. This consistency might not be universal to all arthroplasty surgeons as it relies on the surgeons' experience and their previous working experience. Area variation has also shown to affect a surgeon's decision.[7] As both the surgeons are from the same practicing area, this difference is most probably eliminated.

We can observe that not only the WOMAC score as a whole shows a significant difference between the two groups, but each of the component scores also showed the difference (Table I). From our analyses, a total WOMAC score of above 19 had 100% sensitivity for indication on TKA. This means that if a person is indicated for TKA, the lowest possible score on the WOMAC scale is 19. However, if the score of 19 is used as the cut-off point, the specificity is only 72.5% and 85.14% of times the cut-off point will be correctly classified. In other words, there is a 14.86% chance that the score of above 19 on the WOMAC scale did not require TKA as assessed by the two arthroplasty surgeons.

On the other hand, there is a specificity of 100% on

the indication for TKA when a score of above 33 is taken, but the sensitivity drops to 76.47%. 89.19% will be correctly classified if the score of 33 is taken. We noted that the cut-off point of 30 gives the best possible sensitivity, specificity, and positive predictive value. The total WOMAC score of 30, when used as a cut-off point for indicating TKA gives a sensitivity of 91.18%, specificity of 95% and 93.24% of times it will be correctly classified. This shows that the WOMAC score can be used as an excellent tool to indicate TKA in our population when it has a total score of 30 and above (Table II). From the ROC curve, it is noted that the area under the curve had an almost perfect score of 0.98 (Figure 1). This again proves that there is a strong correlation between increasing WOMAC score and indication for TKA. The higher the WOMAC score, the more likely is the patient indicated for TKA.

The current study has several limitations. In the majority of set-ups, the operating arthroplasty surgeon is the one who assesses if TKA is indicated in a patient. Besides the surgeon's decision, the patients also play a crucial role in deciding on undergoing TKA. The decision-making process regarding the surgery is extremely complex, as patients have to weigh up numerous considerations before they can decide on surgery.[1] In the current study, we put aside the fact that patients' will is an important factor in deciding the need for TKA since our study questions the ability of the WOMAC score to determine the need for TKA when compared to experts' opinions. Several previous studies used case scenarios to establish criteria for TKA.[4,7]

These studies used the Delphi method, which is based on an assessment of scenarios created by panels of experts.[4] The Delphi method is in essence, a series of sequential questionnaires or 'rounds', interspersed by controlled feedback, that seek to gain the most reliable consensus of a group of experts.[19] Despite being a "democratic" way of attaining consensus from an expert panel, it does not use real patients in developing the criteria. It is a difficult method to be performed well and requires a large number of experts.[19,20] Nevertheless, the accuracy of the Delphi method in producing a consensus has been questioned probably to be as good as a single expert opinion.[21]

CONCLUSION

With established appropriate WOMAC score as a guideline for the need for TKA, we could encourage first-line physicians or general practitioners to use the WOMAC score as a tool to assess if a patient is suitable to be referred for surgical intervention at appropriate timing. The study recommends that all primary knee osteoarthritis Kellgren and Lawrence grades 3 or 4 with a score of 30 and above should be referred to an arthroplasty surgeon for the possibility of requiring TKA.

REFERENCES

1. O'Neill T, Jinks C, Ong BN. Decision-making regarding total knee replacement surgery: a qualitative meta-synthesis. BMC Health Serv Res 2007; 7:52.
2. Cross WW 3rd, Saleh KJ, Wilt TJ, Kane RL. Agreement about indications for total knee arthroplasty. Clin Orthop Relat Res 2006; 446:34 -9.
3. Naylor CD, Williams JI. Primary hip and knee replacement surgery: Ontario criteria for case selection and surgical priority. Qual Health Care 1996; 5:20-30.
4. Escobar A, Quintana JM, Arostegui I, et al. Development of explicit criteria for total knee replacement. Int J Technol Assess Health Care 2003; 19:57-70.
5. Van Manen MD, Nace J, Mont MA. Management of primary knee osteoarthritis and indications for total knee arthroplasty for general practitioners. J Am Osteopath Assoc 2012; 112:709-15.
6. Toronto Arthroplasty Research Group Writing Committee, Wright JG, Hawker GA, et al. Variability in physician opinions about the indications for knee arthroplasty. J Arthroplasty 2011; 26:569-75.e1.
7. Wright JG, Coyte P, Hawker G, et al. Variation in orthopedic surgeons' perceptions of the indications for and outcomes of knee replacement. CMAJ 1995; 152:687-97.
8. Hernandez-Vaquero D, Fernandez-Carreira JM. Relationship between radiological grading and clinical status in knee osteoarthritis. A multicentric study. BMC Musculoskelet Disord 2012; 13:194.
9. McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. Arthritis Rheum 2001; 45:453-61.
10. Bischoff-Ferrari HA, Vondechend M, Bellamy N, Theiler R. Validation and patient acceptance of a computer touch screen version of the WOMAC 3.1 osteoarthritis index. Ann Rheum Dis 2005; 64:80-4.
11. Bellamy N, Wilson C, Hendrikz J, et al. Osteoarthritis Index delivered by mobile phone (m-WOMAC) is valid, reliable, and responsive. J Clin Epidemiol 2011; 64:182-90.
12. Feeley BT, Gallo RA, Sherman S, Williams RJ. Management of osteoarthritis of the knee in the active patient. J Am Acad Orthop Surg 2010; 18:406-16.
13. Zakaria ZF, Bakar AA, Hasmoni HM, Rani FA, Kadir SA. Health-related quality of life in patients with knee osteoarthritis attending two primary care clinics in Malaysia: a cross-sectional study. Asia Pac Fam Med 2009; 8:10.
14. NIH Consensus Panel. NIH Consensus Statement on total knee replacement. J Bone Joint Surg Am 2004; 86:1328-35.
15. Judge A, Arden NK, Price A, et al. Assessing patients for joint replacement: can pre-operative Oxford hip and knee scores be used to predict patient satisfaction following joint replacement surgery and to guide patient selection? J Bone Joint Surg Br 2011; 93:1660-4.
16. Cobos R, Latorre A, Aizpuru F, et al. Variability of indication criteria in knee and hip replacement: an observational study. BMC Musculoskelet Disord 2010; 11:249.
17. Hawker GA, Wright JG, Coyte PC, et al. Determining the need for hip and knee arthroplasty: the role of clinical severity and patients' preferences. Med Care 2001; 39:206-16.
18. Kundel HL, Polansky M. Measurement of observer agreement. Radiology 2003; 228:303-8.
19. Powell C. The Delphi technique: myths and realities. J Adv Nurs 2003; 41:376-82.
20. Gordon M, Baker P, Catchpole K, Darbyshire D, Schocken D. Devising a consensus definition and framework for non-technical skills in healthcare to support educational design: A modified Delphi study. Med Teach 2015; 37:572-7.
21. Juni P, Dieppe P, Donovan J, et al. Population requirement for primary knee replacement surgery: a cross-sectional study. Rheumatology (Oxford). 2003; 42:516-21.