BMJ Open

Validation of the Polish version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in osteoarthritis patients undergoing total knee replacement

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-006947
Article Type:	Research
Date Submitted by the Author:	20-Oct-2014
Complete List of Authors:	Paradowski, Przemyslaw; Sunderby Central Hospital of Norrbotten, Department of Orthopaedics; Medical University, Department of Orthopaedics and Traumatology Keska, Rafal; Regional Centre of Orthopaedics and Rehabilitation, Radliński Hospital, Department of Reconstructive Surgery and Arthroscopy of the Knee Joint Witonski, Dariusz; Regional Centre of Orthopaedics and Rehabilitation, Radliński Hospital, Department of Reconstructive Surgery and Arthroscopy of the Knee Joint
Primary Subject Heading:	Medical management
Secondary Subject Heading:	Surgery, Research methods, Patient-centred medicine
Keywords:	Adult orthopaedics < ORTHOPAEDIC & TRAUMA SURGERY, Knee < ORTHOPAEDIC & TRAUMA SURGERY, Musculoskeletal disorders < ORTHOPAEDIC & TRAUMA SURGERY, Orthopaedic & trauma surgery < SURGERY

Title

Validation of the Polish version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in osteoarthritis patients undergoing total knee replacement.

Authors

Przemysław Tomasz Paradowski[1,2], Rafał Kęska[3], Dariusz Witoński[3]

[1]Department of Orthopaedics, Sunderby Central Hospital of Norrbotten, SE-971 80 Luleå, Sweden.

[2]Department of Orthopaedics and Traumatology, Medical University, WAM University Hospital, Żeromskiego 113, PL-90-549 Łódź, Poland.

[3]Department of Reconstructive Surgery and Arthroscopy of the Knee Joint, Regional Center of Orthopaedics and Rehabilitation, Radliński Hospital, Drewnowska 75, PL-91-002 Łódź, Poland.

Corresponding author

Przemyslaw Tomasz Paradowski, M.D., Ph.D., Department of Orthopedics, Sunderby Central Hospital of Norrbotten, SE-971 80 Luleå, Sweden. E-mail: przemyslaw.t.paradowski@gmail.com

ABSTRACT

Objective: To test the clinimetric properties and to evaluate the internal consistency, validity and reliability of the Polish version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in older patients with end-stage knee osteoarthritis undergoing total knee replacement (TKR).

Design and setting: A prospective cohort study performed at the university hospital and the outpatient clinic.

Methods: The patients were asked to complete the KOOS questionnaire and the Short Form 36 Health Survey (SF-36). We evaluated floor/ceiling effects, reliability (using Cronbach's alpha, intraclass correlation coefficients (ICC) and measurement error), structural validity (performing exploratory principal factor analysis), construct validity (with the use of three *a priori* hypotheses) and responsiveness (using data obtained before and after the surgery, and described by Global Perceived Effect, effect size and standardized response mean).

Results: The study consisted of 68 subjects (mean age 68.8, 82% women). The floor effects were found prior to surgery for the subscales Sports and Recreation Function and Quality of Life. There were no ceiling effects in any KOOS subscales neither preoperatively, nor at follow-up. The Cronbach's alpha was between 0.90 to 0.92 for all subscales indicating excellent internal consistency. The test-retest reliability at follow-up was excellent with ICCs ranging from 0.81 to 0.86 for all KOOS subscales. The minimal detectable change ranged from 18.2 to 24.3 on an individual level and from 2.4 to 2.9 on a group level. Responsiveness was confirmed with a statistically significant correlation between all KOOS subscales and the Global Perceived Effect score (ranging from 0.56 to 0.70, $p<0.001$).

Conclusions: The Polish version of KOOS demonstrated good reliability, validity and responsiveness for use in patient groups having had TKR. Since the smallest change

considered clinically relevant cannot reliably be detected in individual subjects, the Polish version of the KOOS is advocated for assessment of groups of patients.

ARTICLE SUMMARY

Strengths and Limitations of the study

- This is the first validation study of any outcome scale to be used in Poland in patients undergoing total knee replacement (TKR).
- We report that the Polish version of the Knee injury and Osteoarthritis Outcome Score (KOOS) demonstrated good reliability, validity and responsiveness for use in patient groups having had TKR.
- The subjects in the present study do not represent the entire spectrum of patients with knee OA but only those with the end-stage disease eligible for TKR. However, since the construct validity is expected to be higher in younger and more active individuals, one can presume that the KOOS scale would be at least equally useful for patients with less severe forms of OA.

Keywords

Patient-reported outcome, validation study, total knee replacement, orthopaedic surgery

Word count

4,090 (excluding title page, abstract, references and tables).

INTRODUCTION

Total knee replacement (TKR) is one of the most common and successful procedures in orthopedic surgery. It provides substantial relief from pain and functional improvement in patients with end-stage knee osteoarthritis (OA).[1] Although most patients undergoing TKR improve their quality of life, there is still an important minority of those who do not improve or even get worse.[2]

Since neither clinical examination nor radiographic imaging correlate with patients' complaints, it is important to assess clinical outcome from the patient's perspective. Cross-culturally adapted and clinically validated patient-reported outcomes (PROs) provide such an approach and describe the function, activity and quality of life avoiding the observer-related bias.[3]

The Knee injury and Osteoarthritis Outcome Score (KOOS)[4,5] is a commonly used PRO, originally prepared in English and Swedish, and currently available in 39 different languages and language variants.[6] KOOS has been found a valid, reliable and responsive self-administered instrument in patients with knee injuries undergoing meniscectomy and anterior cruciate ligament reconstruction (ACLR),[5,7] as well as in patients with knee OA.[8–12] The KOOS scale had already been translated and cross culturally adapted to the Polish language and validated in patients undergoing ACLR[13]. However, there is a need to monitor the outcome of intervention also in elderly patients with OA undergoing total knee replacement. The aim of this study was therefore to test the clinimetric properties and to evaluate the internal consistency, validity and reliability of the Polish version of the KOOS in patients with end-stage knee OA who had undergone TKR.

METHODS

Linguistic and cross-cultural validation process

The cross-cultural adaptation process of the KOOS followed the standard guidelines and was described in detail in the previous study performed in the subjects undergoing ACLR.[13] The Polish KOOS version was pretested in patients with end-stage OA eligible for TKR. All patients who later formed the validation study group were prior to the study asked whether they fully understood the items, whether they found any items ambiguous and whether they had any problems in answering them (see also *Content validity* chapter).

Clinical validation study

The psychometric properties of the KOOS scale were evaluated according to the Consensus-based Standards for the selection of health Measurements Instruments (COSMIN).[14,15] The Polish version of the KOOS questionnaire is available free of charge at http://www.koos.nu.[6]

Patients

All patients who were eligible to take part in the study were native Polish speakers with an intermediate or higher educational level. The patients had met the appropriateness criteria for TKR.[16] The patients had end-stage knee OA diagnosis confirmed[17] and were enrolled for the surgery. Patients were operated on at the Department of Reconstructive Surgery and Arthroscopy of the Knee Joint, Medical University in Łódź between February 2007 and October 2011. The follow up control was carried out between April 2008 and July 2013. The mean follow up time was 1.7 years (0.5–3.1). All subjects had undergone standard total knee replacement with the Genesis II posterior-stabilized (PS) cemented knee prosthesis (Smith and Nephew, Memphis, TN, USA). No patellar replacement was performed. The patients received the same postoperative medical care and were advised to complete individual physical therapy sessions supervised by one therapist.

At time of follow-up, all subjects had returned to their normal activities. Participants were asked to complete the Polish version of the KOOS three times: first preoperatively, then during the routine one to two-year follow-up and finally for test-retest purposes one to two weeks later. Patients filled out the KOOS in the clinic the first two times and at home the third time. Questionnaires were returned by ordinary mail. The one to two week test-retest period is considered appropriate and previously used for the validation of KOOS.[4,5,18] The patients completed the SF-36[19] (license number H1 031207-30347) questionnaire once during the one to two-year postoperative follow-up.

All patients signed their informed consent forms before participating in the study. All self-reported questionnaires, demographics and relevant information were personally administered by one orthopedic surgeon. The study was approved by the ethics committee at the Medical University of Łódź (approval no. RNN/190/07/KB).

Questionnaires

The KOOS is a 42-item self-administered knee-specific questionnaire with five subscales: Pain (9 items), Symptoms (7 items), Activities of Daily Living Function (ADL Function, 17 items), Sports and Recreation Function (5 items) and knee-related Quality of Life (QOL, 4 items).Each item is responded to by marking one of five response options from 0 (best) to 4 (worst) on a Likert scale. Raw scores from 0 (extreme problems) to 100 (no problems at all) is calculated separately for each subscale.

The Short Form 36 (SF-36) Health Survey is a generic self-administered questionnaire that includes 36 items that are combined in eight health domains: Physical Functioning (PF), Role-Physical (RP), Bodily Pain (BP), General Health (GH), Vitality (VT), Social Functioning (SF), Role-Emotional (RE) and Mental Health (MH), and one single item measure of health transition which is not used to score the scales nor in summary measures. A score from 0

(worst possible health status) to 100 (best possible health status) is independently generated for each domain. The SF-36 had already been validated in Polish.[20]

Missing items

According to the 2003 Users Guide for the KOOS questionnaire, two missing items were allowed in each subscale. Missing data were then subsequently imputed with the mean of other values within the same subscale.[6] SF-36 results were calculated using standard scoring procedures whereby missing values were replaced by scale means where valid responses were available for at least half of the scale items.[19]

Floor/ceiling effects

Floor or ceiling effects were assessed pre- and postoperatively. They considered to be present if more than 15% of the participants achieved either the lowest or highest possible scores.[21] Preoperatively floor effects can be expected since experiencing symptoms is an indication for surgery. Post-operatively, ceiling effects can be expected if the intervention has been successful and the patient has returned to his or her normal activities and has no symptoms.

Statistical analysis

Analyses were performed with the use of SPSS for Windows 15.0.0 (SPSS, Chicago, IL, USA). We considered a two-tailed P less than 0.05 to be significant.

Reliability

Reliability is an estimation of the consistency and stability of a measure. It includes analysis of the extent to which a measure is internally consistent and free of measurement error.

Internal consistency

Internal consistency is defined as the degree of the interrelatedness among the items. It was determined by calculating Cronbach's alpha coefficient. Cronbach's alpha was determined at the first one to two-year follow-up assessment. Cronbach's alpha value of more than 0.70 was considered satisfactory.[22]

Test-retest reliability

Test-retest reliability is the extent to which scores for the same patients are unchanched for repeated measurements over time. Test-retest reliability of the KOOS subscales was assessed one to two-year after the TKR twice with one to two-week interval. The test-retest reliability of the KOOS was analyzed using two-way random effect model of the intra-class correlation coefficient (ICC) for absolute agreement and presented with 95% confidence interval (CI). An ICC equal to or greater than 0.80 was considered acceptable for groups and an ICC of more than 0.90 for individual patient use

Measurement error

Measurement error is the systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured. Standard error of measurement (SEM) for absolute agreement of the test–retest reliability estimates how repeated measures of a person on the same instrument tend to be distributed around his or her "true" score. SEM was calculated according to the following formula: SEM= SD $\sqrt{(1-R)}$ where SD represents standard deviation of the sample and R represents the reliability parameter (ICC).[23] Then, in turn, the minimal detectable change (MDC), which is the threshold for determining clinical changes outside measurement error, was calculated using the formula: MDC = SEM × 1.96 × $\sqrt{2}$, where 1.96 derives from the 0.95% confidence interval of no change and $\sqrt{2}$ represents

two measurements evaluating the change.[23,24] The MDC can be modified for group comparison, depending on the size of the group (n = 68), as follows: MDCgroup = MDCindividual/√n.[24] The MDC should preferably be smaller than the minimal important change (MIC). MIC is the smallest change score needed for the effect to be considered clinically relevant.[25] A MIC of 8-10 points was considered to be appropriate for the different KOOS subscales.[18] However, it must be acknowledged that the MIC is dependent on context factors, including patient group, intervention and time to follow-up. Therefore, it is more appropriate to establish the MIC for specific contexts.

Validity

Content validity

Content validity is assessed by making a judgment of relevance and comprehensiveness of the items. All subjects recruited for the study group were asked to assess whether the content covered the items, whether the description of the construct was clear, and whether explanation of the domains was understandable.

Structural validity (exploratory principal factor analysis)

Principal component factor analyses were performed to confirm the previously established subscale structure of the KOOS. Failure to load on a single factor suggests that subscale items do not describe the same aspect. An eigenvalue criterion of 1.0 was used, and the results are given as percentage of variance in the subscale score explained by the principal factor(s).

Hypotheses testing

Construct validity is defined as the degree to which the subscales of the KOOS scale measure the characteristic to be measured. We examined the construct validity of the instruments by

testing an *a priori* set of hypotheses about the expected relationships between the KOOS subscales and the SF-36 scale at baseline. The Spearman's rank correlation was used to assess the association between domains. Correlation coefficients greater than 0.5 were considered strong, correlations between 0.35 and 0.5 moderate and less than 0.35 were considered weak.[26] We expected the highest correlations when comparing the subscales that measure similar constructs. We hypothesized that:

1) since KOOS Pain and SF–36 BP measure a sufficiently similar construct, the correlation between these two measures should be strong and in the same direction,

2) the correlation between KOOS ADL Function and SF–36 PF should be moderate or strong and in the same direction,

3) the correlation between KOOS Sports and Recreation Function and SF–36 PF should be at least moderate and in the same direction,

Responsiveness

Responsiveness is an ability of a measure to detect meaningful clinical change over time in the construct to be measured. It is critical for the use and application of a measure. We have expected to be able to detect clinical change that occurred following TKR. In order to evaluate responsiveness, a Global Perceived Effect (GPE) score was used. Patients were asked to rate at follow-up possible knee condition changes following TKR. They had following answer options: much better (3), better (2), somewhat better (1), no change (0), somewhat worse (-1), worse (-2) and much worse (-3). As with construct validity, we tested the responsiveness by setting *a priori* hypotheses.

We have expected that change in scores in all KOOS subscales between initial examination and follow-up would correlate with GPE score and that a correlation would be at least 0.5 for

all subscales. We also calculated the effect size (ES) defined as score change in all KOOS subscales divided by baseline SD.[27] In addition to ES, responsiveness was also presented as standardized response mean (SRM). SRM was calculated by dividing the mean score change with the standard deviation of that score change.[28]

We also hypothesized that SRM and ES should be higher for patients who reported their condition to be somewhat better, better or much better than in patients reporting much worse, worse, somewhat worse or no change in the GPE score.

To compare KOOS scores before TKR and at follow-up, the Wilcoxon signed rank test was used.

RESULTS

Linguistic and cross-cultural translation process

The Polish version of the KOOS questionnaire was well-accepted by OA patients. All questions and response options were considered satisfactory and understandable by the subjects. Thus, we used the same KOOS questionnaire as was previously validated in younger patients with ACL injury who underwent ACLR.[13]

Clinical validation study

Sample characteristics

In total, 68 subjects, 59 women and 9 men, were enrolled in the study. Patient characteristics is given in Table 1.

Table 1. Characteristics of patients after primary total knee replacement (TKR).

Characteristics	
N (% women)	68 (82)
Age at surgery, mean (SD) years	68.8 (7.8)
Time to follow-up after TKR, mean (SD) years	1.7 (0.8)

Missing items

For the KOOS scale at baseline, a total of four items out of the possible 42 (number of items) x 68 (number of patients), or 0.14% were missing. At follow-up, three items (0.1%) were missing. For SF-36, the number of missing items at follow-up was five (0.2%) out of possible 36 (items) x 68 (number of patients).

Floor/ceiling effects

Preoperatively, there were neither ceiling effects, nor any patients with best possible scores in any of the KOOS subscales. The floor effects (indicating worst possible status) were found prior to surgery for the subscales Sports and Recreation Function (56%) and QOL (19%). The worst possible scores were reported by 3% of patients for the subscales Pain and Symptoms and 4% by the subscale ADL.

At the follow-up, there were no ceiling effects in any KOOS subscales. The best possible scores were reported by 13% of patients for the subscale Pain, 3% for the subscales Symptoms, ADL Function and Sports and Recreation Function and 2% for the subscale QOL. As expected, at follow-up, floor effects were reported only for the subscale Sports and Recreation Function (16%). There were no worst possible scores found after surgery for the other KOOS subscales.

Reliability

Median number of days from test to retest was 6 (ranging from 4 to 13).

Internal consistency

Cronbach alpha ranged from 0.90 to 0.92 indicating excellent internal consistency of all subscales both pre- and postoperatively (Table 2).

Table 2.

Test-retest reliability

The reliability of all KOOS subscales was excellent with ICCs ranging from 0.81 to 0.86 (Table 2).

Minimal detectable change

At the individual level, the MDC was lowest (18.2) for KOOS ADL Function and highest (24.3) for the KOOS subscales Sports and Recreation Function and QOL. At group level, MDC ranged from 2.4 to 2.9 (Table 2).

Validity

Content validity

All KOOS items were estimated to be relevant. The content covered all items, the description of the domains was assessed to be understandable and the construct appeared to be clearly described. Thus, the items were assessed to be comprehensive.

Structural validity (exploratory principal factor analysis)

On the basis of an exploratory principal component analysis in all subscales, we demonstrated that the items from subscales Pain, Sports and Recreation Function and Quality of Life loaded on one major factor (58%, 69% and 51% explained by the principal factor, respectively). The items in the subscale Symptoms loaded on two factors (53 and 14% explained by the principal components) while items in the subscale ADL Function loaded on three factors (54%, 8% and 7% explained by the principal components).

Hypothese testing

All *a priori* established hypotheses were supported. We confirmed strong correlation between KOOS Pain and SF–36 BP (rs = 0.57), KOOS ADL Function and SF–36 PF (rs= 0.53) (hypothesis 1 and 2, respectively), and moderate correlation between KOOS Sports and Recreation Function and SF–36 PF (rs = 0.42) (hypothesis 3) (Table 3).

Table 3. Construct validity, given as Spearman's correlations of the five KOOS subscales and the eight SF-36 subscales in subjects following primary total knee replacement (TKR) (n = 68).

		KOOS subscales				
		Pain	Symptoms	ADL	Sports/Rec	QOL
SF-36 subscales	PF	0.34	0.32	**0.53**	**0.42**	0.43
	RP	0.36	0.29	0.45	0.31	0.30
	BP	**0.57**	0.41	0.46	0.28	0.50
	GH	0.21	0.16	0.44	0.23	0.25
	VT	0.26	0.23	0.44	0.06	0.26
	SF	0.45	0.33	0.52	0.21	0.32
	RE	0.36	0.29	0.42	0.24	0.30
	MH	0.24	0.29	0.41	0.19	0.29

Abbreviations: ADL Activities of Daily Living, Sports/Rec Sports and Recreation Function, QOL Quality of Life, PF Physical Functioning, RP Role-Physical, BP Bodily Pain, GH General Health, VT Vitality, SF Social Functioning, RE Role-Emotional, MH Mental Health.
* As hypothesized, expected correlations were above 0.35 for a priori hypotheses 1-3.

Responsiveness

As hypothesized, change in all five subscales of the KOOS correlated at least at 0.35 with GPE score. The weakest correlation was observed in the KOOS subscale Symptoms (0.56) and the strongest for the subscale ADL Function (0.70). ES and SRM were lower for patients reporting “much worse”, “worse” “somewhat worse” or “no change” than patients reporting “much better”, “better” and “somewhat better” for all five KOOS subscales (Table 4).

Table 4.

DISCUSSION

The present study was performed according to guidelines recommended for validation processes.[29]

The results of our study show that the Polish version of KOOS questionnaire has a good internal consistency and that the questionnaire items are relevant for elderly patients who have undergone TKR due to OA.

In this validation we observed an excellent internal consistency with Cronbach’s alphas ranging from 0.90 to 0.92. These values are higher than in previous KOOS validation studies[7-9,11] but slightly lower than in our previous study performed in subjects undergoing ACLR.[13] Cronbach’s alpha coefficients were generally reported to be lowest. This tendency was not observed in the present study. The Cronbach’s alphas in the KOOS subscales Pain and Symptoms were about 0.2 higher than those described in the studies of Xie et al.[8] and de Groot et al.[10] but only slightly higher than in the two studies of Salavati et al.[7] One possible explanation of such a good consistency is a relative homogeneity of the groups examined postoperatively as compared to patients with OA awaiting surgery.

We have found that the test-retest reliability was excellent with ICCs ranging from 0.81 to 0.86. It proved to have a satisfactory stability and reproducibility of all the KOOS subscales over time in examined subjects. The ICCs with values comparable to ours were observed in previous methodological studies performed in patients with OA awaiting joint replacement[9,12,18] and patients with mild OA after previous ACLR[10]. The ICCs in our group were, however, slightly lower than those reported in subjects with moderate OA following high tibial osteotomy, but higher than in patients eligible for revision knee arthroplasty.[10] This can be explained by the fact that our study group was less homogeneous than patients who had undergone osteotomy but more consistent than revision patients. Since the patients examined in our study had the highest ICC in the KOOS subscale Sports and Recreation Function, we conclude that in those who had undergone TKR the questions about sport were less relevant than the questions in other KOOS domains. We found that ICCs for subscale Sports and Recreation Function were identical to the values we previously observed in subjects undergoing ACLR. It suggests a similar reliability of these subscales in different patients and other subscales.

The MDC value of 3 points or less for the group level indicates that the Polish version of KOOS scale has an ability to detect a difference of 3 points between the measurements. The change of KOOS outcome of 8-10 points (that suggested a minimal clinical important change of each subscale)[31] could thus be easily detected at a group level. Since greater changes are needed to be detected at an individual level (MDC value 18.2-24.3 points for different subscales), the Polish version of KOOS is advocated for use in groups of patients.

Since the content validity of the Polish version of the KOOS had so far been tested only in young individuals who had undergone ACLR,[13] we decided to assess it also in older patients with end-stage OA undergoing TKR. In our study we confirmed the relevance and comprehensiveness of the KOOS items.

With respect to the dimensional structure of the KOOS scale according to confirmatory factor analysis, we found that items from subscales Pain, Sport and Recreation Function and QOL loaded on one major factor, items in the subscale Symptoms loaded on two factors and items in the subscale ADL Function loaded on three factors. The lack of previous reports of structural validity of the KOOS in elderly patients from the TKR group prevented a comparison to other studies. We were able, however, to perform an additional factor analysis retrospectively (which has not been published before) in subjects undergoing ACLR who participated in our previous study.[13] This assessment revealed that the structural validity was similar for all the KOOS subscales except for the subscale ADL Function which loaded on two factors for the younger patients and three factors for the elderly patients undergoing TKR. It was no surprise to confirm that the items of the subscale Symptoms loaded on two factors considering the concomitant structural damage often seen in patients having TKR (and ACLR). The loading on three factors makes the subscale ADL Function even more heterogeneous in older patients than in younger ones.

The construct validity of the KOOS questionnaire was determined by comparing the KOOS subscales with the subscales of the SF–36. As expected, we found strong correlations between KOOS subscales and those subscales of SF–36 that measured corresponding constructs. In our study, the highest correlations were observed between SF-36 subscale Bodily Pain and KOOS subscale Pain and between SF–36 subscale Physical Functioning and KOOS subscales ADL and Sports and Recreation Function. All *a priori* hypotheses were thus confirmed.

The construct validity for the patients in our study was lower than those observed in subjects who had undergone ACLR.[13] This observation was, however, expected, since the KOOS was preliminary designed for use in younger and physically active patients who are more sensitive especially for questions in the subscale Sports and Recreation Function. Similarly, the correlation coefficients reported in our study were about 0.1 lower from those obtained by

Roos et al.[7] and Goncalves et al.[11] who performed their studies in subjects with less severe forms of OA. Our findings are thus more in line with previous results in elderly patients undergoing TKR.[8,18]

Since the outcome in TKR is not specific to the joint but to overall impact on health, we have expected that the correlations between the KOOS subscales and SF-36 subscales representing Physical Function are lower than in patients undergoing ACLR and that there is no big discrepancy between correlations of the KOOS and SF-36 subscales representing Physical Function and Mental Health. As has been shown in one study, the KOOS subscale Sports and Recreation Function holds items being of great importance for all young knee patients but only for about half of elderly patients having TKR.[10] Consequently, our observations and findings reported by others confirm a closer relationship between mental and physical aspects in elderly patients with degenerative disease than in younger patients with knee injury,[30] and suggest different construct validity of the KOOS in younger and older age groups.[5,10]

In our study, to determine KOOS' ability to detect whether patients undergo clinically relevant changes, we assessed GPE. As hypothesized, change in all five subscales of the KOOS correlated at least at 0.35 with GPE score. The results of this assessment showed that the Polish version of KOOS was able to recognize clinical changes over time.

We would like to point out some important limitations of the study. First, the subjects in the present study do not represent the entire spectrum of patients with knee OA but only those with the end-stage disease eligible for TKR. However, since the construct validity is expected to be higher in younger and more active individuals, one can presume that the KOOS scale would be at least equally useful for patients with less severe forms of OA.

In the present study we assessed a relatively small amount of patients. Though the sample was big enough to evaluate reliability, responsiveness and construct validity of the KOOS, it is questionable whether it was big enough to assess its structural validity. Two different

approaches for researchers using exploratory factor analysis were taken suggesting either a minimum total sample size, or a ratio of subjects to variables. However, both recommendations present scarce evidence in practical studies and are not comprehensive enough to be definitive.[32] It has been suggested that the sample size below one hundred gives poor relevance of the results.[33] However, different studies recommend a sample size from an N=50[34] to N=400[35] and a ratio of subjects to variables not less than 2:1.[36,37] Thus, we decided to perform the analysis of structural validity of the KOOS on a group consisting of 68 patients with a ratio of subjects to variables between 4 (in the subscale ADL Function) and 17 (in the subscale QOL).

In our study, women constituted 82% of the study population. Since the prevalence of symptomatic knee OA in women had been reported to be two to three times higher than in men[38], female patients were overrepresented in our study group. However, women often develop more severe symptoms of OA and therefore they accounted for a remarkable majority of TKR.[38] The rate of TKR in women in our study was almost five-fold higher than that for men. Nonetheless, it reflected the gender distribution of patients with end-stage OA in our department over time. The female-to-men ratio of TKR in our study group was higher than in Scandinavia[39] and USA[40] but lower than in the South Korea.[41]

Conclusion

In conclusion, the Polish version of the KOOS demonstrated good reliability, validity and responsiveness for use in patient groups having had TKR. Since the smallest change considered clinically relevant cannot reliably be detected in individual subjects, the Polish version of the KOOS is advocated for assessment of groups of patients. The Polish version of KOOS provides a valuable basis for national and international clinical projects focusing on patient-based assessments in knee OA.

ACKNOWLEDGEMENTS

The authors would like to thank Robert Foltyn, MSc, for the assistance in preparing the manuscript.

CONTRIBUTORS

PTP and DW were responsible for the conceptualization of this project. PTP was the lead writer and was responsible for instrument development, statistical procedures and interpretation of data. DW was the co-principal investigator and performed the surgeries. RK was responsible for creation of datasets and drafted the questionnaires. All authors critically revised the successive drafts and approved the final version of the manuscript.

FUNDING

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

COMPETING INTERESTS

The authors declare no conflict of interest related to this manuscript.

ETHICS APPROVAL

Ethics committee at the Medical University of Łódź (approval no. RNN/190/07/KB).

PROVENANCE AND PEER REVIEW

Not commissioned; externally peer reviewed.

DATA SHARING STATETEMENT

No additional data are available.

REFERENCES

1 Kane RL, Saleh KJ, Wilt TJ, Bershadsky B. The functional outcomes of total knee arthroplasty. J Bone Joint Surg Am 2005;87:1719–24.

2 Bourne RB, Chesworth BM, Davis AM, Mahomed NN, Charron KD. Patient satisfaction after total knee arthroplasty: who is satisfied and who is not? Clin Orthop Relat Res 2010;468:57–63.

3 Speight J, Barendse SM. FDA guidance on patient reported outcomes. BMJ 2010;340:c2921.

4 Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee injury and osteoarthritis outcome score (KOOS) – development of a self–administered outcome measure. J Orthop Sports Phys Ther 1998;28:88–96.

5 Roos EM, Roos HP, Ekdahl C, Lohmander LS. Knee injury and Osteoarthritis Outcome Score (KOOS) – validation of a Swedish version. Scand J Med Sci Sports 1998;8:439–48.

6 Knee injury and Osteoarthritis Outcome Score. http://www.koos.nu/index.html.

7 Salavati M, Mazaheri M, Negahban H, et al. Validation of a Persian-version of Knee injury and Osteoarthritis Outcome Score (KOOS) in Iranians with knee injuries. Osteoarthritis Cartilage 2008;16:1178–82.

8 Xie F, Li SC, Roos EM, et al. Crosscultural adaptation and validation of Singapore English and Chinese versions of the Knee injury and Osteoarthritis Outcome Score (KOOS) in Asians with knee osteoarthritis in Singapore. Osteoarthritis Cartilage 2006;14:1098–103.

9 Ornetti P, Parratte S, Gossec L, et al. Cross-cultural adaptation and validation of the French version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in knee osteoarthritis patients. Osteoarthritis Cartilage 2008;16:423–8.

10 de Groot IB, Favejee MM, Reijman M, Verhaar JAN, Terwee CB. The Dutch version of the Knee injury and Osteoarthritis Outcome Score: a validation study. Health Qual Life Outcomes 2008;6:16.

11 Gonçalves RS, Cabri J, Pinheiro JP, Ferreira PL. Crosscultural adaptation and validation of the Portuguese version of the Knee injury and Osteoarthritis Outcome Score (KOOS). Osteoarthritis Cartilage 2009;17:1156–62.

12 Nakamura N, Takeuchi R, Sawaguchi T, et al. Cross-cultural adaptation and validation of the Japanese Knee Injury and Osteoarthritis Outcome Score (KOOS). J Orthop Sci 2011;16:516–23.

13 Paradowski PT, Witoński D, Kęska R, Roos EM. Cross-cultural translation and measurement properties of the Polish version of the Knee injury and Osteoarthritis Outcome Score (KOOS) following anterior cruciate ligament reconstruction. Health Qual Life Outcomes 2013;11:107.

14 Consensus-based standards for the selection of health measurement instruments. http://www.cosmin.nl/index.php.

15 Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 2010;10:22.

16 Quintana JM, Escobar A, Arostegui I, et al. Health-related quality of life and appropriateness of knee or hip joint replacement. Arch Intern Med 2006;166:220–6.

17 Paradowski PT. Osteoarthritis of the Knee: Assessing the Disease. Health Care Curr Rev 2014;2:e103.

18 Roos EM, Toksvig-Larsen S. Knee injury and Osteoarthritis Outcome Score (KOOS) – validation and comparison to the WOMAC in total knee replacement. Health Qual Life Outcomes 2003;1:17.

19 Ware JE, Kosinski M, Keller SD. SF-36 physical and mental health summary scales: a user's manual. Boston: The Health Institute, New England Medical Center 1988.

20 Żołnierczyk-Zreda D. The Polish version of the SF-36v2 questionnaire for the quality of life assessment (article in Polish). Przegl Lek 2010;67:1302–7.

21 McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 1995;4:293–307.

22 Nunnally JC, Bernstein IR. Psychometric Theory. 3rd edition. New York: McGraw-Hill 1994.

23 Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 1998;26:217–38.

24 de Vet HC, Bouter LM, Bezemer PD, Beurskens AJ. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. Int J Technol Assess Health Care 2001;17:479–87.

25 de Vet HC, Ostelo RW, Terwee CB, et al. Minimally important change developed by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res 2007;16:131–42.

26 Juniper EF, Guyatt GH, Jaeschke R. How to develop and validate a new health-related quality of life instrument. In: Spilker B, editor. Quality of Life and Pharmacoeconomics in Clinical Trials. Philadelphia: Lippincott-Raven 1996:49–56.

27 Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27(3 Suppl):S178–89.

28 Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Control Clin Trials 1991;12:S142–58.

29 Angst F. The new COSMIN guidelines confront traditional concepts of responsiveness. BMC Med Res Methodol 2011;11:152.

30 Terwee CB, van der Slikke RM, van Lummel RC, et al. Self-reported physical functioning was more influenced by pain than performance-based physical functioning in knee-osteoarthritis patients. J Clin Epidemiol 2006;59:724–31.

31 Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. Health Qual Life Outcomes 2003;1:64.

32 Guadagnoli E, Velicer WF. Relation of sample size to the stability of component patterns. Psychol Bull 1988;103:265–75.

33 Comrey AL, Lee HB. A first course of factor analysis. 2nd edition. New York: Psychology Press. Taylor and Francis Group 1992.

34 Barrett PT, Kline P. The observation to variable ratio in factor analysis. Pers Study Group Behav 1981;1:23–33.

35 Aleamoni LM. The relation of sample size to the number of variables in using factor analysis techniques. Educ Psychol Meas 1976;36: 879–83.

36 Henson RK., Roberts JK. Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. Educ Psychol Meas 2006;66:393–416.

37 Costello AB, Osborne JW. Best practice in exploratory factor analysis: four recommendations for getting the most from your analysis. Pract Assess Res Eval 2005;10:1–9.

38 Zhang Y, Xu L, Nevitt MC et al. Comparison of the prevalence of knee osteoarthritis between the elderly Chinese population in Beijing and whites in the United States: The Beijing Osteoarthritis Study. Arthritis Rheum 2001;44:2065–71.

39 Robertsson O, Bizjajeva S, Fenstad AM, et al. Knee arthroplasty in Denmark, Norway and Sweden. A pilot study from the Nordic Arthroplasty Register Association. Acta Orthop 2010;81:82–9.

40 Jain NB, Higgins LD, Ozumba D, et al. Trends in epidemiology of knee arthroplasty in the United States, 1990–2000. Arthritis Rheum 2005;52:3928–33.

41 Kim H-A, Kim S, Seo YI, et al. The epidemiology of total knee replacement in South Korea: national registry data. Rheumatology 2008;47:88–91.

Table 2. Mean KOOS scores (0 to 100, worst to best scale) at test and retest assessment one to two weeks apart, test-retest reliability, internal consistency and minimal detectable change of KOOS subscales for individuals and groups 1.7 years after primary total knee replacement (TKR).

KOOS subscales (number of items)	Mean KOOS score (SD)		ICC (95% CI)	Cronbach's alpha coefficients	SEM	Minimal detectable change (95% CI) in individuals	Minimal detectable change (95% CI) in groups
	First follow-up assessment	Second follow-up assessment					
TKR, n=68							
Pain (9)	78.7 (17.4)	81.1 (15.9)	0.83 (0.74–0.89)	0.91	7.2 (5.8-8.9)	19.9 (16.0-24.6)	2.4 (1.9-3.0)
Symptoms (7)	76.3 (17.8)	80.2 (16.6)	0.81 (0.71–0.88)	0.90	7.8 (6.2-9.6)	21.6 (17.1-26.5)	2.6 (2.1-3.2)
ADL (17)	78.1 (16.0)	79.0 (14.7)	0.83 (0.73–0.89)	0.91	6.6 (5.3-8.3)	18.2 (14.6-22.9)	2.2 (1.8-2.8)
Sports/Recreation (5)	24.6 (23.5)	29.9 (27.3)	0.86 (0.78–0.91)	0.92	8.8 (7.1-11.0)	24.3 (19.6-30.4)	2.9 (2.4-3.7)
QOL (4)	53.7 (21.3)	57.3 (19.4)	0.83 (0.74–0.89)	0.91	8.8 (7.1-10.9)	24.3 (19.6-30.1)	2.9 (2.4-3.7)

Abbreviations: ICC Intraclass correlation coefficient, SEM Standard error of measurement, ADL Activities of Daily Living, QOL Quality of Life.

Table 4. Mean KOOS scores (0 to 100, worst to best scale) in subjects (n = 68) prior to primary total knee replacement (TKR) and 1.7 years after the surgery. Responsiveness given as Spearman's correlations of the five KOOS subscales and GPE score. Standardized effect size (SES) and standardized response mean (SRM) in subjects who scored "somewhat better", "better" and "much better" (n=54) and in those who scored "much worse", "worse", "somewhat worse" and "no change" (n=14).

KOOS subscales	Mean score (SD)		P	GPE score	"Somewhat better", "better" or "much better", n = 54		"Much worse", "worse", "somewhat worse" or "no change", n = 14	
	Before surgery	At follow-up		Spearman r	SES	SRM	SES	SRM
Pain	35.7 (17.3)	78.7 (17.4)	< 0.001	0.58	3.07	2.37	1.37	0.96
Symptoms	35.3 (22.6)	76.3 (17.8)	< 0.001	0.56	2.51	2.20	0.50	0.68
ADL	33.0 (17.1)	78.1 (16.0)	< 0.001	0.70	3.50	3.05	1.63	1.29
Sports/Recreation	7.2 (13.6)	24.6 (23.5)	< 0.001	0.62	1.55	0.94	0.04	0.04
QOL	16.8 (13.3)	53.7 (21.3)	< 0.001	0.61	3.30	1.91	1.31	0.81

Abbreviations: ICC Intraclass correlation coefficient, ADL Activities of Daily Living, QOL Quality of Life; ES, effect size; GPE, Global Perceived Effect; SRM, standardized response mean.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No.	Recommendation	Page No.	Relevant text from manuscript
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1, 2	Title is including the study's design as follows: "Validation of the Polish version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in osteoarthritis patients undergoing total knee replacement."
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2, 3	Summary of results provided in abstract, the Results chapter.
Introduction				
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4	"The KOOS scale had already been translated and cross culturally adapted to the Polish language and validated in patients undergoing anterior cruciate ligament reconstruction. (…) there is a need to monitor the outcome of intervention also in elderly patients with osteoarthritis undergoing total knee replacement."
Objectives	3	State specific objectives, including any prespecified hypotheses	4	"(…) to test the clinimetric properties and to evaluate the internal consistency, validity and reliability of the Polish version of the KOOS in patients with end-stage knee OA who had undergone total knee replacement."
Methods				
Study design	4	Present key elements of study design early in the paper	5	Description in details in the Methods chapter, in Linguistic and cross-cultural validation

				process and Patients sections.
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5, 6	Description in details in the Patients section.
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up.	5, 6	Description in details in the Patients section.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7-11	"The psychometric properties of the KOOS scale were evaluated according to the Consensus-based Standards for the selection of health Measurements Instruments (COSMIN)." Definition of all methods is described.
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	7-11	Methods of assessment described as mentioned earlier.
Bias	9	Describe any efforts to address potential sources of bias		It is described throughout the Methods chapter in different sections.
Study size	10	Explain how the study size was arrived at		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-11	See earlier
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7-11	Statistical methods are described throughout the whole Methods chapter.
		(*b*) Describe any methods used to examine subgroups and interactions		Due to the character of the study (validation study) we analysed only one group of patients
		(*c*) Explain how missing data were addressed	7	The missing items were addressed According to the users guide for the KOOS and the SF-36 questionnaires. See details in the Missing items section.
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed		Due to the character of the study, only patients who were followed-up were eligible.
		(*e*) Describe any sensitivity analyses	7-11	

Results				
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	11-12	See details in the Clinical validation study section.
		(b) Give reasons for non-participation at each stage		
		(c) Consider use of a flow diagram		
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	11-12	
		(b) Indicate number of participants with missing data for each variable of interest	12	See description in Missing items section, Sample characteristics subsection.
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	12	See Table 1.
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	12-15	Described throughout the whole Results chapter.
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	12-15	See details in Results chapter.
		(*b*) Report category boundaries when continuous variables were categorized		-
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period		-
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	12-15	
Discussion				
Key results	18	Summarise key results with reference to study objectives	15-18	Due to the character of the study, the key results are discussed throughout the most of the Discussion chapter.
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	18	
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	15-18	Interpretation of data can be found throughout the Discussion chapter.

Generalisability	21	Discuss the generalisability (external validity) of the study results	19	
Other information				
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	20	This research received no specific grant from any funding agency, neither public, nor private..

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Validation of the Polish version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in osteoarthritis patients undergoing total knee replacement

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-006947.R1
Article Type:	Research
Date Submitted by the Author:	13-Apr-2015
Complete List of Authors:	Paradowski, Przemyslaw; Sunderby Central Hospital of Norrbotten, Department of Orthopaedics; Medical University, Department of Orthopaedics and Traumatology Keska, Rafal; Regional Centre of Orthopaedics and Rehabilitation, Radliński Hospital, Department of Reconstructive Surgery and Arthroscopy of the Knee Joint Witonski, Dariusz; Regional Centre of Orthopaedics and Rehabilitation, Radliński Hospital, Department of Reconstructive Surgery and Arthroscopy of the Knee Joint
Primary Subject Heading:	Medical management
Secondary Subject Heading:	Surgery, Research methods, Patient-centred medicine
Keywords:	Adult orthopaedics < ORTHOPAEDIC & TRAUMA SURGERY, Knee < ORTHOPAEDIC & TRAUMA SURGERY, Musculoskeletal disorders < ORTHOPAEDIC & TRAUMA SURGERY, Orthopaedic & trauma surgery < SURGERY

Title

Validation of the Polish version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in osteoarthritis patients undergoing total knee replacement.

Authors

Przemysław Tomasz Paradowski[1,2], Rafał Kęska[3], Dariusz Witoński[3]

[1]Department of Orthopaedics, Sunderby Central Hospital of Norrbotten, SE-971 80 Luleå, Sweden.

[2]Department of Orthopaedics and Traumatology, Medical University, WAM University Hospital, Żeromskiego 113, PL-90-549 Łódź, Poland.

[3]Department of Reconstructive Surgery and Arthroscopy of the Knee Joint, Regional Center of Orthopaedics and Rehabilitation, Radliński Hospital, Drewnowska 75, PL-91-002 Łódź, Poland.

Corresponding author

Przemyslaw Tomasz Paradowski, M.D., Ph.D., Department of Orthopedics, Sunderby Central Hospital of Norrbotten, SE-971 80 Luleå, Sweden. E-mail: przemyslaw.t.paradowski@gmail.com

ABSTRACT

Objective To test the clinimetric properties and to evaluate the internal consistency, validity and reliability of the Polish version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in older patients with end-stage knee osteoarthritis undergoing total knee replacement (TKR).

Design and setting A prospective cohort study performed at the university hospital and the outpatient clinic.

Methods The patients were asked to complete the KOOS questionnaire and the Short Form 36 Health Survey (SF-36). We evaluated floor/ceiling effects, reliability (using Cronbach's alpha, intraclass correlation coefficients (ICC) and measurement error), structural validity (performing exploratory principal factor analysis), construct validity (with the use of three *a priori* hypotheses) and responsiveness (using data obtained before and after the surgery, and described by Global Perceived Effect, effect size and standardized response mean).

Results The study consisted of 68 subjects (mean age 68.8, 82% women). The floor effects were found prior to surgery for the subscales Sports and Recreation Function and Quality of Life. The Cronbach's alpha was from 0.90 to 0.92 for all subscales indicating excellent internal consistency. The test-retest reliability at follow-up was excellent, with ICCs ranging from 0.81 to 0.86 for all KOOS subscales. The minimal detectable change ranged from 18.2 to 24.3 on an individual level and from 2.4 to 2.9 on a group level. All KOOS items were relevant and all *a priori* established hypotheses were supported. Responsiveness was confirmed with a statistically significant correlation between all KOOS subscales and the Global Perceived Effect score (ranging from 0.56 to 0.70, $p<0.001$).

Conclusions The Polish version of the KOOS demonstrated good reliability, validity and responsiveness for use in patient groups having had TKR. Since the smallest change

considered clinically relevant cannot reliably be detected in individual subjects, the Polish version of the KOOS is advocated for assessment of groups of patients.

ARTICLE SUMMARY

Strengths and Limitations of the study

- This is the first validation study of any outcome scale to be used in Poland in patients undergoing total knee replacement (TKR).
- We report that the Polish version of the Knee injury and Osteoarthritis Outcome Score (KOOS) demonstrated good reliability, validity and responsiveness for use in patient groups having had TKR.
- The subjects in the present study do not represent the entire spectrum of patients with knee OA but only those with the end-stage disease eligible for TKR. However, since the construct validity is expected to be higher in younger and more active individuals, one can presume that the KOOS scale would be at least equally useful for patients with less severe forms of OA.

Keywords

Patient-reported outcome, validation study, total knee replacement, orthopaedic surgery

Word count

4,948 (excluding title page, abstract, references and tables).

INTRODUCTION

Total knee replacement (TKR) is one of the most common and successful procedures in orthopedic surgery. It provides substantial relief from pain and functional improvement in patients with end-stage knee osteoarthritis (OA).[1] Although most patients undergoing TKR improve their quality of life, there is still an important minority of those who do not improve or even get worse.[2]

Since neither clinical examination nor radiographic imaging correlate with patients' complaints, it is important to assess clinical outcome from the patient's perspective. Cross-culturally adapted and clinically validated patient-reported outcomes (PROs) provide such an approach and describe the function, activity and quality of life, avoiding the observer-related bias.[3]

The Knee injury and Osteoarthritis Outcome Score (KOOS)[4,5] is a commonly used PRO, originally prepared in English and Swedish, and currently available in 39 different languages and language variants.[6] The KOOS has been found a valid, reliable and responsive self-administered instrument in patients with knee injuries undergoing meniscectomy and anterior cruciate ligament reconstruction (ACLR),[5,7] as well as in patients with knee OA.[8–12] The KOOS scale had already been translated and cross culturally adapted to the Polish language and validated in patients undergoing ACLR[13]. However, there is a need to monitor the outcome of intervention also in elderly patients with OA undergoing TKR. The aim of this study was therefore to test the clinimetric properties and to evaluate the internal consistency, validity and reliability of the Polish version of the KOOS in patients with end-stage knee OA who had undergone TKR.

METHODS

Linguistic and cross-cultural validation process

The cross-cultural adaptation process of the KOOS followed the standard guidelines and was described in detail in the previous study performed in subjects undergoing ACLR.[13]

The Polish version of KOOS was pretested in patients with end-stage OA eligible for TKR. All patients who later formed the validation study group were prior to the study asked whether they fully understood the questions (items), whether they found any items ambiguous and whether they had any problems in answering them (see also *Content validity* chapter).

Clinical validation study

The psychometric properties of the KOOS scale were evaluated according to the Consensus-based Standards for the selection of health Measurements Instruments (COSMIN).[14,15] The Polish version of the KOOS questionnaire is available free of charge at http://www.koos.nu.[6]

Patients

All patients recruited in the study had met the appropriateness criteria for TKR.[16] One hundred and fifty-seven patients had end-stage knee OA diagnosis confirmed[17] and were enrolled for the surgery. Patients were operated on at the Department of Reconstructive Surgery and Arthroscopy of the Knee Joint, Medical University in Łódź between February 2007 and October 2011. The follow up control was carried out between April 2008 and July 2013. The mean follow up time was 1.7 years (0.5–3.1). All subjects had undergone standard total knee replacement with the Genesis II posterior-stabilized (PS) cemented knee prosthesis (Smith and Nephew, Memphis, TN, USA). No patellar replacement was performed. The patients received the same postoperative medical care and were advised to complete individual physical therapy sessions supervised by one therapist.

At the time of follow-up, all subjects had returned to their normal activities. Participants were asked to complete the Polish version of the KOOS three times: first preoperatively, then during the routine one to two-year follow-up and finally for test-retest purposes one to two weeks later. Patients filled out the first two KOOS questionnaires in the clinic, while the third one was completed at home. Questionnaires were returned by ordinary mail. The one to two week test-retest period is considered appropriate and previously used for the validation of KOOS.[4,5,18] The patients completed the SF-36[19] (license number H1 031207-30347) questionnaire once during the one to two-year postoperative follow-up.

All patients signed and dated personally their informed consent forms at the admission into hospital, before participating in the study. All self-reported questionnaires, demographics and relevant information were personally administered by one orthopedic surgeon. The study was approved by the ethics committee at the Medical University of Łódź (approval no. RNN/190/07/KB).

Questionnaires

The KOOS is a 42-item self-administered knee-specific questionnaire with five subscales: Pain (9 items), Symptoms (7 items), Activities of Daily Living Function (ADL Function, 17 items), Sports and Recreation Function (5 items) and knee-related Quality of Life (QOL, 4 items). Each item is responded to by marking one of five response options from 0 (best) to 4 (worst) on a Likert scale. Raw scores from 0 (extreme problems) to 100 (no problems at all) is calculated separately for each subscale.

The Short Form 36 (SF-36) Health Survey is a generic self-administered questionnaire that includes 36 items that are combined in eight health domains: Physical Functioning (PF), Role-Physical (RP), Bodily Pain (BP), General Health (GH), Vitality (VT), Social Functioning (SF), Role-Emotional (RE) and Mental Health (MH), and one single item measure of health

transition which is not used to score the scales nor in summary measures. A score from 0 (worst possible health status) to 100 (best possible health status) is independently generated for each domain. The SF-36 had already been validated in Polish.[20]

Missing items

According to the 2003 Users Guide for the KOOS questionnaire, two missing items were allowed in each subscale. Missing data were then subsequently imputed with the mean of other values within the same subscale.[6] SF-36 results were calculated using standard scoring procedures whereby missing values were replaced by scale means where valid responses were available for at least half of the scale items.[19]

Floor/ceiling effects

Floor or ceiling effects were assessed pre- and postoperatively. They were considered to be present if more than 15% of the participants achieved either the lowest or highest possible scores.[21] Preoperatively, floor effects can be expected since experiencing symptoms is an indication for surgery. Post-operatively, ceiling effects can be expected if the intervention has been successful and the patient has returned to his or her normal activities and has no symptoms. Comparisons of proportions for men and women with the lowest and the highest possible scores were evaluated with the McNemar's test.

Statistical analysis

Analyses were performed with the use of SPSS for Windows 15.0.0 (SPSS, Chicago, IL, USA). We considered a two-tailed P less than 0.05 to be significant.

Reliability

Reliability is an estimation of the consistency and stability of a measure. It includes analysis of the extent to which a measure is internally consistent and free of measurement error.

Internal consistency

Internal consistency is defined as the degree of the interrelatedness among the items. It was determined by calculating Cronbach's alpha coefficient. Cronbach's alpha was determined at the first one to two-year follow-up assessment. Cronbach's alpha value of more than 0.70 was considered satisfactory.[22]

Test-retest reliability

Test-retest reliability is the extent to which scores for the same patients remain unchanged for repeated measurements over time. Test-retest reliability of the KOOS subscales was assessed one to two-year after the TKR twice, with one to two-week interval. Test-retest reliability of the KOOS was analyzed using two-way random effect model of the intra-class correlation coefficient (ICC) for absolute agreement and presented with 95% confidence interval (CI). An ICC equal to or greater than 0.80 was considered acceptable for groups and an ICC of more than 0.90 for individual patient use.

Measurement error

Measurement error is the systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured. Standard error of measurement (SEM) for absolute agreement of the test–retest reliability estimates how repeated measures of a person on the same instrument tend to be distributed around his or her "true" score. SEM was calculated according to the following formula: SEM= SD $\sqrt{}$(1-R), where SD represents standard deviation of the sample and R represents the reliability parameter (ICC).[23] Then,

in turn, the minimal detectable change (MDC), which is the threshold for determining clinical changes outside measurement error, was calculated using the formula: MDC = SEM × 1.96 × √2, where 1.96 derives from the 0.95% confidence interval of no change and √2 represents two measurements evaluating the change.[23,24] The MDC can be modified for group comparison, depending on the size of the group (n = 68), as follows: MDCgroup = MDCindividual/√n.[24] The MDC should preferably be smaller than the minimal important change (MIC). MIC is the smallest change score needed for the effect to be considered clinically relevant.[25] A MIC of 8-10 points was considered to be appropriate for the different KOOS subscales.[18] However, it must be acknowledged that the MIC is dependent on context factors, including patient group, intervention and time to follow-up. Therefore, it is more appropriate to establish the MIC for specific contexts.

Validity

Content validity

Content validity is assessed by making a judgment of relevance and comprehensiveness of the items. All subjects recruited for the study group were asked to assess whether the content covered the items, whether the description of the construct was clear, and whether explanation of the domains was understandable.

Structural validity (exploratory principal factor analysis)

The factor analysis is a method designed to determine if the observed variables (items) could be explained by a smaller number of latent variables (called factors). Due to the sample size of 68, we performed an exploratory factor analysis. Investigations were conducted on all items of the KOOS scale with use of principal component analyses with the orthogonal rotation procedure (Varimax). According the Kaiser's criterion,[26] factors with eigenvalue

greater than 1 were extracted. The scree plot of the correlation matrix of all items was drawn. Factors that appeared over the point where the curve bends ("elbow") were considered to be meaningful.[27] An analysis of the factor structure and loading was made. Factor loading of 0.4 or above was defined as substantial loading and desirable for an item to be significant. The subscale item that had a substantial loading on more than one factors (cross-loading), was considered to be "complex", meaning that it had an affinity to two or more of the derived factors and it did not describe the same aspect. The results are given as percentage of variance in the subscale score explained by the principal factor(s).

Hypotheses testing

Construct validity is defined as the degree to which the subscales of the KOOS scale measure the characteristic to be measured. We examined the construct validity of the instruments by testing an *a priori* set of hypotheses about the expected relationships between the KOOS subscales and the SF-36 scale at baseline. The Spearman's rank correlation was used to assess the association between domains. Correlation coefficients greater than 0.5 were considered strong, correlations between 0.35 and 0.5 moderate and less than 0.35 were considered weak.[28] We expected the highest correlations when comparing the subscales that measure similar constructs. We hypothesized that:

1) since KOOS Pain and SF–36 BP measure a sufficiently similar construct, the correlation between these two measures should be strong and in the same direction,

2) the correlation between KOOS ADL Function and SF–36 PF should be moderate or strong and in the same direction,

3) the correlation between KOOS Sports and Recreation Function and SF–36 PF should be at least moderate and in the same direction,

Responsiveness

Responsiveness is an ability of a measure to detect meaningful clinical change over time in the construct to be measured. It is critical for the use and application of a measure. We have expected to be able to detect clinical change that occurred following TKR. In order to evaluate responsiveness, a Global Perceived Effect (GPE) score was used. At follow-up, patients were asked to rate knee condition changes, if any, following TKR. They had the following answer options: much better (3), better (2), somewhat better (1), no change (0), somewhat worse (-1), worse (-2) and much worse (-3). As with construct validity, we tested the responsiveness by setting *a priori* hypotheses.

We have expected that the change in scores in all KOOS subscales between initial examination and follow-up would correlate with the GPE score and that a correlation would be at least 0.5 for all subscales. We also calculated the effect size (ES) defined as a score change in all KOOS subscales divided by baseline SD.[29] In addition to ES, responsiveness was also presented as standardized response mean (SRM). SRM was calculated by dividing the mean score change by the standard deviation of that score change.[30]

We also hypothesized that SRM and ES should be higher for patients who reported their condition to be somewhat better, better or much better than in patients reporting much worse, worse, somewhat worse or no change in the GPE score.

To compare KOOS scores before TKR and at follow-up, the Wilcoxon signed rank test was used.

RESULTS

Linguistic and cross-cultural translation process

The Polish version of the KOOS questionnaire was well-accepted by OA patients. All questions and response options were considered appropriate and understandable by the subjects. Thus, we used the same KOOS questionnaire as was previously validated in younger patients with ACL injury who had undergone ACLR.[13]

Clinical validation study

Sample characteristics

Sixty-eight out of 157 (43%) patients who were enrolled in the study returned fully completed sets of questionnaires and formed the study sample. Of them, 59 were women and 9 men. All patients who were eligible to take part in the study were native Polish speakers with secondary or higher education. To evaluate a possible inclusion bias, the subjects who participated in the study, and those who did not respond, were analyzed with regard to age and gender. We found no significant differences in these characteristics (data not shown). The patient characteristics is given in Table 1.

Table 1. Characteristics of patients after primary total knee replacement (TKR).

Characteristics	
N (% women)	68 (82)
Age at surgery, mean (SD) years	68.8 (7.8)
Time to follow-up after TKR, mean (SD) years	1.7 (0.8)

Missing items

For the KOOS scale at baseline, a total of four items out of the possible 42 (number of items) x 68 (number of patients), or 0.14% were missing. At follow-up, three items (0.1%) were

missing. For SF-36, the number of missing items at follow-up was five (0.2%) out of possible 36 (items) x 68 (number of patients).

Floor/ceiling effects

Preoperatively, there were neither ceiling effects, nor any patients with best possible scores in any of the KOOS subscales. The floor effects (indicating worst possible status) were found prior to surgery for the subscales Sports and Recreation Function (56%) and QOL (19%). The worst possible scores were reported by 3% of patients for the subscales Pain and Symptoms and 4% for the subscale ADL.

At follow-up, there were no ceiling effects in any KOOS subscales. The best possible scores were reported by 13% of patients for the subscale Pain, 3% for the subscales Symptoms, ADL Function and Sports and Recreation Function and 2% for the subscale QOL.

As expected, at follow-up, floor effects were reported only for the subscale Sports and Recreation Function (16%). There were no worst possible scores found after surgery for the other KOOS subscales. No differences in the number of patients having the worst or best possible scores related to gender were observed.

Reliability

The median number of days from test to retest was 6 (ranging from 4 to 13).

Internal consistency

Cronbach's alpha ranged from 0.90 to 0.92, indicating an excellent internal consistency of all subscales (Table 2).

Table 2.

Test-retest reliability

The reliability of all KOOS subscales was excellent, with ICCs ranging from 0.81 to 0.86 (Table 2).

Minimal detectable change

At the individual level, the MDC was lowest (18.2) for KOOS ADL Function and highest (24.3) for the KOOS subscales Sports and Recreation Function and QOL. At the group level, MDC ranged from 2.4 to 2.9 (Table 2).

Validity

Content validity

All KOOS items were estimated to be relevant. The content covered all items, the description of the domains was assessed to be understandable and the construct appeared to be clearly described. Thus, the items were assessed to be comprehensive.

Structural validity (exploratory principal factor analysis)

The Kaiser-Meyer-Olkin measure of sampling adequacy was middling (0.79), but close to good (≥0.8), which suggested the sample was adequate for an exploratory factor analysis. The scree plot confirmed the retention of the first five factors. Thus, five factors were sufficient to describe the data. This solution accounted for 63.3% of the total variance for the Polish version of the KOOS questionnaire (with eigenvalues of 16.6, 3.5, 2.4, 2.3 and 1.9 for respective factors).

Items S1 and S3-S5 from the subscale Symptoms loaded substantially on the third factor (ranging 0.45 to 0.78). The S2 item had a substantial loading of the fifth factor. In the case of

items S6 and S7, a cross-loading of both the third (0.56 and 0.54 respectively) and the fifth factor (0.52 in both items) was observed.

Seven out of 17 items from the subscale ADL Function had a substantial loading on only the first factor (ranging between 0.42 and 0.76). Items A1 and A2 had a substantial loading on only the second factor (ranging between 0.69 and 0.77 respectively) and item A8 on only the third factor (0.44). In all other items, the cross-loading of different combination of factors was observed. Items A6 and A7 loaded on both the first and the second factor, whereas a cross-loading of the first and the third factor was observed in items A3 and A9-A11. Item A5 cross-loaded on the second and the third factor. Item A3 (rising from sitting) loaded on three factors: the first, the third and the fifth one (0.40, 0.42 and 0.43 respectively).

All items from the subscale Sports and Recreation Function loaded highly on the fourth factor (ranged from 0.63 to 0.84). In the SP5 item, a cross-loading of the fifth and the fourth factor (0.63 and 0.41 respectively) was observed.

In the subscale Quality of Life, items QOL1 and QOL4 loaded on the fifth factor (0.65 and 0.62 respectively), item QOL2 loaded on the third (0.42) and QOL3 the second factor (0.68) (data not shown).

Hypothese testing

All *a priori* established hypotheses were supported. We confirmed a strong correlation between KOOS Pain and SF–36 BP (rs = 0.57), KOOS ADL Function and SF–36 PF (rs= 0.53) (hypothesis 1 and 2, respectively), and a moderate correlation between KOOS Sports and Recreation Function and SF–36 PF (rs = 0.42) (hypothesis 3) (Table 3).

Table 3. Construct validity, given as Spearman's correlations of the five KOOS subscales and the eight SF-36 subscales in subjects following primary total knee replacement (TKR) (n = 68).

		KOOS subscales				
		Pain	Symptoms	ADL	Sports/Rec	QOL
SF-36 subscales	PF	0.34	0.32	**0.53**	**0.42**	0.43
	RP	0.36	0.29	0.45	0.31	0.30
	BP	**0.57**	0.41	0.46	0.28	0.50
	GH	0.21	0.16	0.44	0.23	0.25
	VT	0.26	0.23	0.44	0.06	0.26
	SF	0.45	0.33	0.52	0.21	0.32
	RE	0.36	0.29	0.42	0.24	0.30
	MH	0.24	0.29	0.41	0.19	0.29

Abbreviations: ADL Activities of Daily Living, Sports/Rec Sports and Recreation Function, QOL Quality of Life, PF Physical Functioning, RP Role-Physical, BP Bodily Pain, GH General Health, VT Vitality, SF Social Functioning, RE Role-Emotional, MH Mental Health.
* As hypothesized, expected correlations were above 0.35 for a priori hypotheses 1-3.

Responsiveness

As hypothesized, the change in all five subscales of the KOOS correlated at least at 0.35 with GPE score. The weakest correlation was observed in the KOOS subscale Symptoms (0.56) and the strongest for the subscale ADL Function (0.70). ES and SRM were lower for patients reporting "much worse", "worse" "somewhat worse" or "no change" than patients reporting "much better", "better" and "somewhat better" for all five KOOS subscales (Table 4). No correlation between ES and SRM and the duration of the follow-up period was observed.

Table 4.

DISCUSSION

The present study was performed according to the guidelines recommended for validation processes.[31]

The results of our study show that the Polish version of the KOOS questionnaire has a good internal consistency and that the questionnaire items are relevant for elderly patients who have undergone TKR due to OA.

In this validation, we observed an excellent internal consistency with Cronbach's alphas ranging from 0.90 to 0.92. These values are higher than in previous KOOS validation studies[7-9,11], but slightly lower than in our previous study performed in subjects undergoing ACLR.[13] Cronbach's alpha coefficients were generally reported to be lowest. This tendency was not observed in the present study. The Cronbach's alphas in the KOOS subscales Pain and Symptoms were about 0.2 higher than those described in the studies of Xie et al.[8] and de Groot et al.[10], but only slightly higher than in the two studies of Salavati et al.[7] One possible explanation of such a good consistency is a relative homogeneity of the groups examined postoperatively as compared to patients with OA awaiting surgery.

We have found that the test-retest reliability was excellent, with ICCs ranging from 0.81 to 0.86. It proved to have a satisfactory stability and reproducibility of all the KOOS subscales over time in examined subjects. The ICCs with values comparable to ours were observed in previous methodological studies performed in patients with OA awaiting joint replacement[9,12,18] and patients with mild OA after previous ACLR.[10] The ICCs in our group were, however, slightly lower than those reported in subjects with moderate OA following high tibial osteotomy, but higher than in patients eligible for revision knee arthroplasty.[10] This can be explained by the fact that our study group was less homogeneous than patients who had undergone osteotomy, but more consistent than revision patients. Since the patients examined in our study had the highest ICC in the KOOS subscale Sports and Recreation Function, we conclude that in those who had undergone TKR the questions about sport were less relevant than the questions in other KOOS domains. We found that ICCs for subscale Sports and Recreation Function were identical to the values we

previously observed in subjects undergoing ACLR. It suggests a similar reliability of these subscales in different patients and other subscales.

The MDC value of 3 points or less for the group level indicates that the Polish version of the KOOS scale has an ability to detect a difference of 3 points between the measurements. The change of KOOS outcome of 8-10 points (that suggested a minimal clinical important change of each subscale)[32] could thus be easily detected at the group level. Since greater changes are needed to be detected at the individual level (MDC value 18.2-24.3 points for different subscales), the Polish version of KOOS is advocated for use in groups of patients.

Since the content validity of the Polish version of the KOOS had so far been tested only in young individuals who had undergone ACLR,[13] we decided to assess it also in older patients with end-stage OA undergoing TKR. In our study, we confirmed the relevance and comprehensiveness of the KOOS items.

With respect to the dimensional structure of the KOOS scale according to confirmatory factor analysis, we found that the Polish version of the KOOS contains five principal factors. This observation is in line with that of Roos et al.,[5] who found that the Swedish version of the KOOS loaded on five factors. All items of the Polish version of the KOOS questionnaire had a substantial loading of at least one factor. A large first eigenvalue (16.6) and much smaller subsequent eigenvalues (3.5 and lower) suggested a leading global factor. Indeed, the first factor dominated in 17 items in the subscales Pain and ADL Function. While some items loaded on a single factor, other items had association to two, and in the case of the A3 item, even three factors, providing evidence of the complex nature of some of the questions. In addition, we noticed that the pairs of items that addressed the same activities related to pain (in the subscale Pain) and function (subscale ADL Function) such as "walking on flat surface" (P5 and A6), "going up and down stairs" (P6 and A1-A2), "sitting or lying" (P8 and A14) and "standing upright" (P9 and A4), loaded on the same principal factor. In fact, we observed that

some patients might have had difficulty to distinct between pain and physical functioning in ADL. Apparently, the KOOS subscales Symptoms and Sport and Recreation Function are much more homogenous than ADL Function and QOL.

The lack of previous reports of structural validity of the KOOS in elderly patients from the TKR group prevented a comparison to other studies. We were able, however, to perform an additional factor analysis retrospectively (which has not been published before) in subjects undergoing ACLR who participated in our previous study.[13] This assessment revealed that the KOOS contained four principal factors. The number of items that had an association to more than one factor was even higher than in our present study. However, if we ignore the complexity and assume that each item belongs to the factor on which they have the highest loading, we recognize that each subscale of the Polish version of the KOOS has its dominant factor in both younger subjects undergoing ACLR and elderly patients after TKR.

The construct validity of the KOOS questionnaire was determined by comparing the KOOS subscales with the subscales of the SF–36. The SF-36 measures the general health status and contains domains that make it possible to assess the correlations between KOOS subscales and SF-36 subscales representing both mental and physical health. As expected, we found strong correlations between KOOS subscales and those subscales of SF–36 that measured corresponding constructs. In our study, the highest correlations were observed between SF-36 subscale Bodily Pain and KOOS subscale Pain and between the SF–36 subscale Physical Functioning and the KOOS subscales ADL and Sports and Recreation Function. All *a priori* hypotheses were thus confirmed.

The construct validity for the patients in our study was lower than those observed in subjects who had undergone ACLR.[13] This observation was, however, expected, since the KOOS was preliminary designed for use in younger and physically active patients who are more sensitive especially for questions in the subscale Sports and Recreation Function. Similarly,

the correlation coefficients reported in our study were about 0.1 lower from those obtained by Roos et al.[7] and Goncalves et al.[11] who performed their studies in subjects with less severe forms of OA. Our findings are thus more in line with the previous results in elderly patients undergoing TKR.[8,18]

Since the outcome in TKR is not specific to the joint but to overall impact on health, we have expected that the correlations between the KOOS subscales and SF-36 subscales representing Physical Function are lower than in patients undergoing ACLR and that there is no big discrepancy between correlations of the KOOS and SF-36 subscales representing Physical Function and Mental Health. As has been shown in one study, the KOOS subscale Sports and Recreation Function holds items of great importance for all young knee patients but only for about half of elderly patients having TKR.[10] Consequently, our observations and findings reported by others confirm a closer relationship between mental and physical aspects in elderly patients with degenerative disease than in younger patients with knee injury,[33] and suggest different construct validity of the KOOS in younger and older age groups.[5,10]

In our study, to determine KOOS' ability to detect whether patients undergo clinically relevant changes, we assessed GPE. As hypothesized, change in all five subscales of the KOOS correlated at least at 0.35 with GPE score. Part of the patients examined had a relatively long follow-up period that hypothetically could have affected the responsiveness of the KOOS. We did not notice, however, that responsiveness depended on the duration of the follow-up time. The results of this assessment showed that the Polish version of KOOS was able to recognize clinical changes over time.

We would like to point out some important limitations of the study. First, the subjects in the present study do not represent the entire spectrum of patients with knee OA but only those with the end-stage disease eligible for TKR. However, since the construct validity is expected

to be higher in younger and more active individuals, one can presume that the KOOS scale would be at least equally useful for patients with less severe forms of OA.

In the present study we assessed a relatively small amount of patients. Though the sample was big enough to evaluate reliability, responsiveness and construct validity of the KOOS, it is questionable whether it was big enough to assess its structural validity. In earlier studies two different approaches for researchers using exploratory factor analysis have been taken, suggesting either a minimum total sample size, or a ratio of subjects to variables. However, both recommendations present scarce evidence in practical studies and are not comprehensive enough to be definitive.[34] It has been suggested that the sample size below one hundred gives poor relevance of the results.[35] However, different studies recommend a sample size from N=50[36] to N=400[37] and a ratio of subjects to variables not less than 2:1.[38,39] Thus, we decided to perform the analysis of structural validity of the KOOS on a group consisting of 68 patients, with a ratio of subjects to variables between 4 (in the subscale ADL Function) and 17 (in the subscale QOL).

In our study, women constituted 82% of the study population. Since the prevalence of symptomatic knee OA in women had been reported to be two to three times higher than in men,[40] female patients were overrepresented in our study group. However, women often develop more severe symptoms of OA and that fact accounted for a remarkable majority of TKR.[40] The rate of TKR in women in our study was almost five-fold higher than that for men. Nonetheless, it reflected the gender distribution of patients with end-stage OA in our department over time. The female-to-men ratio of TKR in our study group was higher than in Scandinavia[41] and USA[42] but lower than in the South Korea.[43]

As we examined a relatively small group of patients which was skewed towards a female population, we could expect that it affected the presence of floor and/or ceiling effects in the most sensitive domains like the KOOS subscales Sport and Recreation Function and QOL.

However, in our study we did not observe gender-related differences in proportion of patients having reported the worst and best possible scores. In order to assess reliably if such differences exist, an analysis in a study sample of at least 500 participants is required.

Conclusion

In conclusion, the Polish version of the KOOS demonstrated good reliability, validity and responsiveness for use in patient groups having had TKR. Since the smallest change considered clinically relevant cannot reliably be detected in individual subjects, the Polish version of the KOOS is advocated for assessment of groups of patients. The KOOS may be useful in national and international projects focusing on patient-based assessment of clinical outcome in therapeutic interventions due to knee OA.

ACKNOWLEDGEMENTS

The authors would like to thank Robert Foltyn for the assistance in preparing the manuscript.

CONTRIBUTORS

PTP and DW were responsible for the conceptualization of this project. PTP was the lead writer and was responsible for instrument development, statistical procedures and interpretation of data. DW was the co-principal investigator and performed the surgeries. RK was responsible for the creation of datasets and drafted the questionnaires. All authors critically revised the successive drafts and approved the final version of the manuscript.

FUNDING

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

COMPETING INTERESTS

The authors declare no conflict of interest related to this manuscript.

ETHICS APPROVAL

Ethics committee at the Medical University of Łódź (approval no. RNN/190/07/KB).

PROVENANCE AND PEER REVIEW

Not commissioned; externally peer reviewed.

DATA SHARING STATETEMENT

No additional data are available.

REFERENCES

1 Kane RL, Saleh KJ, Wilt TJ, Bershadsky B. The functional outcomes of total knee arthroplasty. J Bone Joint Surg Am 2005;87:1719–24.

2 Bourne RB, Chesworth BM, Davis AM, Mahomed NN, Charron KD. Patient satisfaction after total knee arthroplasty: who is satisfied and who is not? Clin Orthop Relat Res 2010;468:57–63. doi: 10.1007/s11999-009-1119-9.

3 Speight J, Barendse SM. FDA guidance on patient reported outcomes. BMJ 2010;340:c2921. doi: 10.1136/bmj.c2921.

4 Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee injury and osteoarthritis outcome score (KOOS) – development of a self–administered outcome measure. J Orthop Sports Phys Ther 1998;28:88–96.

5 Roos EM, Roos HP, Ekdahl C, Lohmander LS. Knee injury and Osteoarthritis Outcome Score (KOOS) – validation of a Swedish version. Scand J Med Sci Sports 1998;8:439–48.

6 Knee injury and Osteoarthritis Outcome Score. http://www.koos.nu/index.html.

7 Salavati M, Mazaheri M, Negahban H, et al. Validation of a Persian-version of Knee injury and Osteoarthritis Outcome Score (KOOS) in Iranians with knee injuries. Osteoarthritis Cartilage 2008;16:1178–82. doi: 10.1016/j.joca.2008.03.004. [Published online first: 14 April 2008].

8 Xie F, Li SC, Roos EM, et al. Crosscultural adaptation and validation of Singapore English and Chinese versions of the Knee injury and Osteoarthritis Outcome Score (KOOS) in Asians with knee osteoarthritis in Singapore. Osteoarthritis Cartilage 2006;14:1098–103. [Published online first: 30 June 2006].

9 Ornetti P, Parratte S, Gossec L, et al. Cross-cultural adaptation and validation of the French version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in knee osteoarthritis patients. Osteoarthritis Cartilage 2008;16:423–8. [Published online first: 1 October 2007].

10 de Groot IB, Favejee MM, Reijman M, Verhaar JAN, Terwee CB. The Dutch version of the Knee injury and Osteoarthritis Outcome Score: a validation study. Health Qual Life Outcomes 2008;6:16. doi: 10.1186/1477-7525-6-16

11 Gonçalves RS, Cabri J, Pinheiro JP, Ferreira PL. Crosscultural adaptation and validation of the Portuguese version of the Knee injury and Osteoarthritis Outcome Score (KOOS). Osteoarthritis Cartilage 2009;17:1156–62. doi: 10.1016/j.joca.2009.01.009. [Published online first: 11 March 2009].

12 Nakamura N, Takeuchi R, Sawaguchi T, et al. Cross-cultural adaptation and validation of the Japanese Knee Injury and Osteoarthritis Outcome Score (KOOS). J Orthop Sci

2011;16:516–23. doi: 10.1007/s00776-011-0112-9. [Published online first: 16 July 2011].

13 Paradowski PT, Witoński D, Kęska R, Roos EM. Cross-cultural translation and measurement properties of the Polish version of the Knee injury and Osteoarthritis Outcome Score (KOOS) following anterior cruciate ligament reconstruction. Health Qual Life Outcomes 2013;11:107. doi: 10.1186/1477-7525-11-107.

14 Consensus-based standards for the selection of health measurement instruments. http://www.cosmin.nl/index.php.

15 Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 2010;10:22. doi: 10.1186/1471-2288-10-22.

16 Quintana JM, Escobar A, Arostegui I, et al. Health-related quality of life and appropriateness of knee or hip joint replacement. Arch Intern Med 2006;166:220–6.

17 Paradowski PT. Osteoarthritis of the Knee: Assessing the Disease. Health Care Curr Rev 2014;2:e103. doi: 10.4172/hccr.1000e103.

18 Roos EM, Toksvig-Larsen S. Knee injury and Osteoarthritis Outcome Score (KOOS) – validation and comparison to the WOMAC in total knee replacement. Health Qual Life Outcomes 2003;1:17.

19 Ware JE, Kosinski M, Keller SD. SF-36 physical and mental health summary scales: a user's manual. Boston: The Health Institute, New England Medical Center 1988.

20 Żołnierczyk-Zreda D. The Polish version of the SF-36v2 questionnaire for the quality of life assessment (article in Polish). Przegl Lek 2010;67:1302–7.

21 McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 1995;4:293–307.

22 Nunnally JC, Bernstein IR. Psychometric Theory. 3rd edition. New York: McGraw-Hill 1994.

23 Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 1998;26:217–38.

24 de Vet HC, Bouter LM, Bezemer PD, Beurskens AJ. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. Int J Technol Assess Health Care 2001;17:479–87.

25 de Vet HC, Ostelo RW, Terwee CB, et al. Minimally important change developed by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res 2007;16:131–42. [Published online first: 11 October 2006].

26 Kaiser HF. The application of electronic computers to factor analysis. Educ Psychol Meas 1960;20:141–51.

27 Cattell RB. The scree test for the number of factors. Multivar Behav Res 1966;1:245–76.

28 Juniper EF, Guyatt GH, Jaeschke R. How to develop and validate a new health-related quality of life instrument. In: Spilker B, editor. Quality of Life and Pharmacoeconomics in Clinical Trials. Philadelphia: Lippincott-Raven 1996:49–56.

29 Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27(3 Suppl):S178–89.

30 Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Control Clin Trials 1991;12:S142–58.

31 Angst F. The new COSMIN guidelines confront traditional concepts of responsiveness. BMC Med Res Methodol 2011;11:152. doi: 10.1186/1471-2288-11-152.

32 Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. Health Qual Life Outcomes 2003;1:64.

33 Terwee CB, van der Slikke RM, van Lummel RC, et al. Self-reported physical functioning was more influenced by pain than performance-based physical functioning in knee-osteoarthritis patients. J Clin Epidemiol 2006;59:724–31.

34 Guadagnoli E, Velicer WF. Relation of sample size to the stability of component patterns. Psychol Bull 1988;103:265–75.

35 Comrey AL, Lee HB. A first course of factor analysis. 2nd edition. New York: Psychology Press. Taylor and Francis Group 1992.

36 Barrett PT, Kline P. The observation to variable ratio in factor analysis. Pers Study Group Behav 1981;1:23–33.

37 Aleamoni LM. The relation of sample size to the number of variables in using factor analysis techniques. Educ Psychol Meas 1976;36: 879–83.

38 Henson RK., Roberts JK. Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. Educ Psychol Meas 2006;66:393–416.

39 Costello AB, Osborne JW. Best practice in exploratory factor analysis: four recommendations for getting the most from your analysis. Pract Assess Res Eval 2005;10:1–9.

40 Zhang Y, Xu L, Nevitt MC et al. Comparison of the prevalence of knee osteoarthritis between the elderly Chinese population in Beijing and whites in the United States: The Beijing Osteoarthritis Study. Arthritis Rheum 2001;44:2065–71.

41 Robertsson O, Bizjajeva S, Fenstad AM, et al. Knee arthroplasty in Denmark, Norway and Sweden. A pilot study from the Nordic Arthroplasty Register Association. Acta Orthop 2010;81:82–9. doi: 10.3109/17453671003685442.

42 Jain NB, Higgins LD, Ozumba D, et al. Trends in epidemiology of knee arthroplasty in the United States, 1990–2000. Arthritis Rheum 2005;52:3928–33.

43 Kim H-A, Kim S, Seo YI, et al. The epidemiology of total knee replacement in South Korea: national registry data. Rheumatology 2008;47:88–91.

Table 2. Mean KOOS scores (0 to 100, worst to best scale) at test and retest assessment one to two weeks apart, test-retest reliability, internal consistency and minimal detectable change of KOOS subscales for individuals and groups 1.7 years after primary total knee replacement (TKR).

KOOS subscales (number of items)	Mean KOOS score (SD)		ICC (95% CI)	Cronbach's alpha coefficients	SEM	Minimal detectable change (95% CI) in individuals	Minimal detectable change (95% CI) in groups
	First follow-up assessment	Second follow-up assessment					
TKR, n=68							
Pain (9)	78.7 (17.4)	81.1 (15.9)	0.83 (0.74–0.89)	0.91	7.2 (5.8-8.9)	19.9 (16.0-24.6)	2.4 (1.9-3.0)
Symptoms (7)	76.3 (17.8)	80.2 (16.6)	0.81 (0.71–0.88)	0.90	7.8 (6.2-9.6)	21.6 (17.1-26.5)	2.6 (2.1-3.2)
ADL (17)	78.1 (16.0)	79.0 (14.7)	0.83 (0.73–0.89)	0.91	6.6 (5.3-8.3)	18.2 (14.6-22.9)	2.2 (1.8-2.8)
Sports/Recreation (5)	24.6 (23.5)	29.9 (27.3)	0.86 (0.78–0.91)	0.92	8.8 (7.1-11.0)	24.3 (19.6-30.4)	2.9 (2.4-3.7)
QOL (4)	53.7 (21.3)	57.3 (19.4)	0.83 (0.74–0.89)	0.91	8.8 (7.1-10.9)	24.3 (19.6-30.1)	2.9 (2.4-3.7)

Abbreviations: ICC Intraclass correlation coefficient, SEM Standard error of measurement, ADL Activities of Daily Living, QOL Quality of Life.

Table 4. Mean KOOS scores (0 to 100, worst to best scale) in subjects (n = 68) prior to primary total knee replacement (TKR) and 1.7 years after the surgery. Responsiveness given as Spearman's correlations of the five KOOS subscales and GPE score. Standardized effect size (SES) and standardized response mean (SRM) in subjects who scored "somewhat better", "better" and "much better" (n=54) and in those who scored "much worse", "worse", "somewhat worse" and "no change" (n=14).

KOOS subscales	Mean score (SD)		P	GPE score	"Somewhat better", "better" or "much better", n = 54		"Much worse", "worse", "somewhat worse" or "no change", n = 14	
	Before surgery	At follow-up		Spearman r	SES	SRM	SES	SRM
Pain	35.7 (17.3)	78.7 (17.4)	< 0.001	0.58	3.07	2.37	1.37	0.96
Symptoms	35.3 (22.6)	76.3 (17.8)	< 0.001	0.56	2.51	2.20	0.50	0.68
ADL	33.0 (17.1)	78.1 (16.0)	< 0.001	0.70	3.50	3.05	1.63	1.29
Sports/Recreation	7.2 (13.6)	24.6 (23.5)	< 0.001	0.62	1.55	0.94	0.04	0.04
QOL	16.8 (13.3)	53.7 (21.3)	< 0.001	0.61	3.30	1.91	1.31	0.81

Abbreviations: ICC Intraclass correlation coefficient, ADL Activities of Daily Living, QOL Quality of Life; ES, effect size; GPE, Global Perceived Effect; SRM, standardized response mean.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No.	Recommendation	Page No.	Relevant text from manuscript
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1, 2	Title is including the study's design as follows: "Validation of the Polish version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in osteoarthritis patients undergoing total knee replacement."
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2, 3	Summary of results provided in abstract, the Results chapter.
Introduction				
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4	"The KOOS scale had already been translated and cross culturally adapted to the Polish language and validated in patients undergoing anterior cruciate ligament reconstruction. (…) there is a need to monitor the outcome of intervention also in elderly patients with osteoarthritis undergoing total knee replacement."
Objectives	3	State specific objectives, including any prespecified hypotheses	4	"(…) to test the clinimetric properties and to evaluate the internal consistency, validity and reliability of the Polish version of the KOOS in patients with end-stage knee OA who had undergone total knee replacement."
Methods				
Study design	4	Present key elements of study design early in the paper	5	Description in details in the Methods chapter, in Linguistic and cross-cultural validation

				process and Patients sections.
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5, 6	Description in details in the Patients section.
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up.	5, 6	Description in details in the Patients section.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7-11	"The psychometric properties of the KOOS scale were evaluated according to the Consensus-based Standards for the selection of health Measurements Instruments (COSMIN)." Definition of all methods is described.
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	7-11	Methods of assessment described as mentioned earlier.
Bias	9	Describe any efforts to address potential sources of bias		It is described throughout the Methods chapter in different sections.
Study size	10	Explain how the study size was arrived at		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-11	See earlier
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7-11	Statistical methods are described throughout the whole Methods chapter.
		(*b*) Describe any methods used to examine subgroups and interactions		Due to the character of the study (validation study) we analysed only one group of patients
		(*c*) Explain how missing data were addressed	7	The missing items were addressed According to the users guide for the KOOS and the SF-36 questionnaires. See details in the Missing items section.
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed		Due to the character of the study, only patients who were followed-up were eligible.
		(*e*) Describe any sensitivity analyses	7-11	

Results				
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	11-12	See details in the Clinical validation study section.
		(b) Give reasons for non-participation at each stage		
		(c) Consider use of a flow diagram		
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	11-12	
		(b) Indicate number of participants with missing data for each variable of interest	12	See description in Missing items section, Sample characteristics subsection.
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	12	See Table 1.
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	12-15	Described throughout the whole Results chapter.
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	12-15	See details in Results chapter.
		(*b*) Report category boundaries when continuous variables were categorized		-
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period		-
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	12-15	
Discussion				
Key results	18	Summarise key results with reference to study objectives	15-18	Due to the character of the study, the key results are discussed throughout the most of the Discussion chapter.
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	18	
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	15-18	Interpretation of data can be found throughout the Discussion chapter.

Generalisability	21	Discuss the generalisability (external validity) of the study results	19	
Other information				
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	20	This research received no specific grant from any funding agency, neither public, nor private..

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.