

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

Przemysław Tomasz Paradowski,<sup>1,2</sup> Rafał Kęska,<sup>3</sup> Dariusz Witoński<sup>3</sup>

**To cite:** Paradowski PT, Kęska R, Witoński D. Validation of the Polish version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in patients with osteoarthritis undergoing total knee replacement. *BMJ Open* 2015;**5**:e006947. doi:10.1136/bmjopen-2014-006947

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2014-006947>).

Received 20 October 2014  
Revised 13 April 2015  
Accepted 29 April 2015



CrossMark

<sup>1</sup>Department of Orthopaedics, Sunderby Central Hospital of Norrbotten, Luleå, Sweden

<sup>2</sup>Department of Orthopaedics and Traumatology, Medical University, WAM University Hospital, Łódź, Poland

<sup>3</sup>Department of Reconstructive Surgery and Arthroscopy of the Knee Joint, Regional Center of Orthopaedics and Rehabilitation, Radliński Hospital, Łódź, Poland

**Correspondence to**  
Dr Przemysław Tomasz Paradowski;  
[przemyslaw.t.paradowski@gmail.com](mailto: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. 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 3 a priori hypotheses) and responsiveness (using data obtained before and after the surgery, and described by Global Perceived Effect, effect size and standardised response mean).

**Results:** The study consisted of 68 participants (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 KOOS demonstrated good reliability, validity and responsiveness for use in patient groups that had undergone TKR. Since the smallest change considered clinically relevant cannot reliably be detected in individual cases, the Polish version of KOOS is advocated for assessment of groups of patients.

## Strengths and limitations of this 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 that had undergone TKR.
- The participants in the present study do not represent the entire spectrum of patients with knee osteoarthritis (OA) but only those with 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.

## INTRODUCTION

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

Since neither clinical examination nor radiographic imaging correlate with patients' symptoms, 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.<sup>3</sup>

The Knee injury and Osteoarthritis Outcome Score (KOOS)<sup>4 5</sup> is a commonly

used PRO, originally prepared in English and Swedish, and currently available in 39 different languages and language variants.<sup>6</sup> KOOS has been found to be a valid, reliable and responsive self-administered instrument in patients with knee injuries undergoing meniscectomy and anterior cruciate ligament reconstruction (ACLR),<sup>5 7</sup> as well as in patients with knee OA.<sup>8–12</sup> The KOOS scale has already been translated and cross-culturally adapted to the Polish language and validated in patients undergoing ACLR.<sup>13</sup> However, there is also a need to monitor the outcome of intervention 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 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 KOOS followed the standard guidelines and was described in detail in the previous study performed in patients undergoing ACLR.<sup>13</sup>

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 asked, prior to the study, whether they fully understood the questions (items), whether they found any items ambiguous and whether they had any problems in answering them (see also the chapter titled ‘Content validity’).

### 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).<sup>14 15</sup> The Polish version of the KOOS questionnaire is available free of charge at <http://www.koos.nu>.<sup>6</sup>

### Patients

All patients recruited in the study had met the appropriateness criteria for TKR.<sup>16</sup> One hundred and fifty-seven patients had end-stage knee OA diagnosis confirmed<sup>17</sup> and were enrolled for the surgery. Patients were operated on at the Department of Reconstructive Surgery and Arthroscopy of the Knee Joint, at the 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 patients had undergone standard TKR with the Genesis II posterior-stabilised cemented knee prosthesis (Smith and Nephew, Memphis, Tennessee, 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 participants had returned to their normal activities. Participants were asked to complete the Polish version of KOOS three times: first preoperatively, then during the routine 1–2 years of follow-up, and, finally, for test–retest purposes 1–2 weeks later. Patients filled out the first two KOOS questionnaires in the clinic, while the third was completed at home. Questionnaires were returned by ordinary mail. The 1–2 week test–retest period is considered appropriate and has previously been used for the validation of KOOS.<sup>4 5 18</sup> The patients completed the Short Form 36 (SF-36) Health Survey<sup>19</sup> (license number H1 031207-30347) questionnaire once during the 1–2 years of post-operative follow-up.

All patients signed and personally dated informed consent forms during the admission into hospital, before participating in the study. All self-reported questionnaires, demographics and relevant information were personally administered by one orthopaedic surgeon.

### Questionnaires

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) are calculated separately for each subscale.

The SF-36 Health Survey is a generic self-administered questionnaire that includes 36 items, combined in eight health domains: Physical Functioning (PF), Role-Physical, Bodily Pain (BP), General Health (GH), Vitality, Social Functioning, Role-Emotional and Mental Health, 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 has already been validated in Polish.<sup>20</sup>

### 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.<sup>6</sup> 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 the scale items.<sup>19</sup>

### Floor/ceiling effects

Floor or ceiling effects were assessed preoperatively and postoperatively. They were considered to be present if more than 15% of the participants achieved either the lowest or the highest possible scores.<sup>21</sup> Preoperatively,

floor effects can be expected since experiencing symptoms is an indication for surgery. Postoperatively, 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 V.15.0.0 (SPSS, Chicago, Illinois, USA). We considered a two-tailed *p* value 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 inter-relatedness among the items. It was determined by calculating Cronbach's  $\alpha$  coefficient. Cronbach's  $\alpha$  was determined at the first 1–2 years of follow-up assessment. Cronbach's  $\alpha$  value of more than 0.70 was considered satisfactory.<sup>22</sup>

#### 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 1–2 years after the TKR twice, with 1–2 weeks interval. Test–retest reliability of KOOS was analysed using two-way random effect model of the intra-class correlation coefficient (ICC) for absolute agreement, and presented with 95% CI. An  $ICC \geq 0.80$  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. 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 SD of the sample and R the reliability parameter (ICC).<sup>23</sup> Then, in turn, the minimal detectable change (MDC), which is the threshold for determining clinical changes outside measurement error, was calculated using the formula:  $MDC = SEM \times 1.96 \times \sqrt{2}$ , where 1.96 derives from the 0.95% CI of no change and  $\sqrt{2}$  represents two measurements evaluating the change.<sup>23 24</sup> The MDC can be modified for group comparison, depending on the size of the

group ( $n=68$ ), as follows:  $MDC_{group} = MDC_{individual} / \sqrt{n}$ .<sup>24</sup> 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.<sup>25</sup> An MIC of 8–10 points was considered to be appropriate for the different KOOS subscales.<sup>18</sup> 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 judgement of relevance and comprehensiveness of the items. All participants 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). Owing to the sample size of 68, we performed an exploratory factor analysis. Investigations were conducted on all items of the KOOS scale with the use of principal component analyses with the orthogonal rotation procedure (Varimax). According to the Kaiser's criterion,<sup>26</sup> factors with an 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.<sup>27</sup> 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 factor (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 characteristics 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.<sup>28</sup> We expected the highest correlations when comparing the subscales that measure similar constructs. We hypothesised that:

- ▶ 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.
- ▶ The correlation between KOOS ADL Function and SF-36 PF should be moderate or strong and in the same direction.
- ▶ 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.<sup>29</sup> In addition to ES, responsiveness was also presented as standardised response mean (SRM). SRM was calculated by dividing the mean score change by the SD of that score change.<sup>30</sup>

We also hypothesised 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 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 patients with OA. All questions and response options were considered appropriate and understandable by the patients. Thus, we used the same KOOS questionnaire as was previously validated in younger patients with ACL injury who had undergone ACLR.<sup>13</sup>

### Clinical validation study

#### Sample characteristics

Sixty-eight 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 were men. All patients who were eligible to take

**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)

part in the study were native Polish speakers with secondary or higher education. To evaluate a possible inclusion bias, the patients who participated in the study, and those who did not respond, were analysed with regard to age and gender. We found no significant differences in these characteristics (data not shown). The patient characteristics are given in [table 1](#).

### Missing items

For the KOOS scale at baseline, a total of four items of the possible 42 (number of items)×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 5 (0.2%) of a possible 36 (items)×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](#)).

### Test-retest reliability

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

**Table 2** Mean KOOS (0–100, worst to best scale) at test and retest assessment 1–2 weeks apart, test–retest reliability, internal consistency and minimal detectable change of KOOS subscales for individuals and groups 1.7 years after primary TKR

KOOS subscales (number of items)	Mean KOOS (SD)		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)	0.91	7.2 (5.8 to 8.9)	19.9 (16.0 to 24.6)	2.4 (1.9 to 3.0)
Symptoms (7)	76.3 (17.8)	80.2 (16.6)	0.90	7.8 (6.2 to 9.6)	21.6 (17.1 to 26.5)	2.6 (2.1 to 3.2)
ADL (17)	78.1 (16.0)	79.0 (14.7)	0.91	6.6 (5.3 to 8.3)	18.2 (14.6 to 22.9)	2.2 (1.8 to 2.8)
Sports/recreation (5)	24.6 (23.5)	29.9 (27.3)	0.92	8.8 (7.1 to 11.0)	24.3 (19.6 to 30.4)	2.9 (2.4 to 3.7)
QOL (4)	53.7 (21.3)	57.3 (19.4)	0.91	8.8 (7.1 to 10.9)	24.3 (19.6 to 30.1)	2.9 (2.4 to 3.7)

ADL, activities of daily living; ICC, intraclass correlation coefficient; KOOS, Knee injury and Osteoarthritis Outcome Score; QOL, quality of life; TKR, total knee replacement.

### 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 ( $\geq 0.8$ ), 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 0.45–0.78). 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 factors (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 combinations of factors was observed. Items A6 and A7 loaded on both the first and the second factors, whereas a cross-loading of the first and the third factors was observed in items A3 and A9–A11. Item A5 cross-loaded on the second and the third factors. Item A3 (rising from sitting) loaded on three factors: the first, the third and the fifth (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 factors (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 on the third (0.42) and QOL3 loaded on the second factor (0.68) (data not shown).

#### Hypotheses testing

All a priori-established hypotheses were supported. We confirmed a strong correlation between KOOS Pain and SF-36 BP ( $r_s=0.57$ ), KOOS ADL Function and SF-36 PF

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

	KOOS subscales				
	Pain	Symptoms	ADL	Sports/Rec	QOL
SF-36 subscales					
PF	0.34	0.32	<b>0.53</b>	<b>0.42</b>	0.43
RP	0.36	0.29	0.45	0.31	0.30
BP	<b>0.57</b>	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

As hypothesised, expected correlations were above 0.35 for a priori hypotheses 1–3 (as shown in bold typeface). ADL, activities of daily living; BP, bodily pain; GH, general health; KOOS, Knee injury and Osteoarthritis Outcome Score; MH, mental health; PF, physical functioning; QOL, quality of life; RE, role-emotional; RP, role-physical; SF, social functioning; SF-36, Short Form 36; Sports/Rec, sports and recreation function; TKR, total knee replacement; VT, vitality.

(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).

### Responsiveness

As hypothesised, the change in all five subscales of KOOS correlated at least at 0.35 with the 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.

### DISCUSSION

The present study was performed according to the guidelines recommended for validation processes.<sup>31</sup>

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  $\alpha$ s ranging from 0.90 to 0.92. These values are higher than in previous KOOS validation studies,<sup>7–9 11</sup> but slightly lower than in our previous study performed in patients undergoing ACLR.<sup>13</sup> Cronbach's  $\alpha$  coefficients were generally reported to be lowest. This tendency was not observed in the present study. The Cronbach's  $\alpha$ s in the KOOS subscales Pain and Symptoms were about 0.2 higher than those described in the studies of Xie *et al*<sup>8</sup> and de Groot *et al*,<sup>10</sup> but only slightly higher than in the two studies of Salavati *et al*.<sup>7</sup> One possible explanation for such a good consistency is a relative homogeneity of the groups examined postoperatively as compared with 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 patients who were examined. The ICCs with values comparable to ours were observed in previous methodological studies performed in patients with OA awaiting joint replacement,<sup>9 12 18</sup> and patients with mild OA after previous ACLR.<sup>10</sup> The ICCs in our group were, however, slightly lower than those reported in patients with moderate OA following high tibial osteotomy, but higher than in patients eligible for revision knee arthroplasty.<sup>10</sup> 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

**Table 4** Mean KOOS (0–100, worst to best scale) in patients (n=68) prior to primary TKR, and 1.7 years after the surgery

KOOS subscales	Mean score (SD)		p Value	GPE score Spearman's r	'Somewhat better', 'better' or 'much better', n=54		'Much worse', 'worse', 'somewhat worse' or 'no change', n=14	
	Before surgery	At follow-up			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

Responsiveness given as Spearman's correlations of the five KOOS subscales and GPE score. SES and SRM in patients 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).

ADL, activities of daily living; GPE, global perceived effect; KOOS, Knee injury and Osteoarthritis Outcome Score; QOL, quality of life; SES, standardised effect size; SRM, standardised response mean; TKR, total knee replacement.

questions about sport were less relevant than the questions in other KOOS domains. We found that ICCs for the subscale Sports and Recreation Function were identical to the values we previously observed in patients 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 clinically important change of each subscale)<sup>32</sup> 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 KOOS had so far been tested only in young individuals who had undergone ACLR,<sup>13</sup> 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 KOOS items.

With respect to the dimensional structure of the KOOS scale according to confirmatory factor analysis, we found that the Polish version of KOOS contains five principal factors. This observation is in line with that of Roos *et al.*,<sup>5</sup> who found that the Swedish version of 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 with 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 (in the 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 in distinguishing between pain and PF in ADL. Apparently, the KOOS subscales Symptoms and Sport and Recreation Function, are much more homogeneous than ADL Function and QOL.

The lack of previous reports of structural validity of KOOS in elderly patients from the TKR group prevented a comparison with other studies. However, we were able to perform an additional factor analysis retrospectively (which has not been published before) in patients undergoing ACLR who participated in our previous study.<sup>13</sup> This assessment revealed that 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 it has the highest loading, we recognise that each subscale of the Polish version of KOOS has its dominant factor in younger patients undergoing ACLR as well as in 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 GH status and contains domains that make it possible to assess the correlations between the KOOS subscales and SF-36 subscales, representing both mental and physical health. As expected, we found strong correlations between KOOS subscales and those of SF-36 that measured corresponding constructs. In our study, the highest correlations were observed between SF-36 subscale BP 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 that observed in patients who had undergone ACLR.<sup>13</sup> This observation was, however, expected, since KOOS was primarily 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 than those obtained by Roos *et al.*<sup>7</sup> and Goncalves *et al.*,<sup>11</sup> who performed their studies in participants with less severe forms of OA. Our findings are thus more in line with the previous results in elderly patients undergoing TKR.<sup>8 18</sup>

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 significant discrepancy between correlations of KOOS and SF-36 subscales representing Physical Function and MH. 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 the elderly patients having TKR.<sup>10</sup> Consequently, our observations and findings, also 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,<sup>33</sup> and suggest different construct validity of KOOS in younger and older age groups.<sup>5 10</sup>

In our study, to determine KOOS' ability to detect whether patients undergo clinically relevant changes, we assessed GPE. As hypothesised, change in all five subscales of KOOS correlated at least at 0.35 with GPE score. Some of the patients examined had a relatively long follow-up period, which, hypothetically, could have affected the responsiveness of 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 recognise clinical changes over time.

We would like to point out some important limitations of the study. First, the participants in the present study do not represent the entire spectrum of patients with knee OA but only those with 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 number of patients. Though the sample was big enough to evaluate reliability, responsiveness and construct validity of 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 were taken, suggesting either a minimum total sample size, or a ratio of participants to variables. However, both recommendations present scarce evidence in practical studies and are not comprehensive enough to be definitive.<sup>34</sup> It has been suggested that a sample size less than 100 gives poor relevance of the results.<sup>35</sup> However, different studies recommend a sample size from N=50<sup>36</sup> to N=400,<sup>37</sup> and a ratio of participants to variables not less than 2:1.<sup>38–39</sup> Thus, we decided to perform the analysis of structural validity of KOOS on a group consisting of 68 patients, with a ratio of participants 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 has been reported to be two to three times higher than in men,<sup>40</sup> female patients were over-represented in our study group. However, women often develop more severe symptoms of OA, and that fact accounted for a remarkable majority of TKRs.<sup>40</sup> The rate of TKRs in women in our study was almost fivefold 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-male ratio of TKR in our study group was higher than in Scandinavia<sup>41</sup> and the USA,<sup>42</sup> but lower than in South Korea.<sup>43</sup>

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 to 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 KOOS demonstrated good reliability, validity and responsiveness for use in patient groups that had undergone TKR. Since the smallest change considered clinically relevant cannot reliably

be detected in individual patients, the Polish version of KOOS is advocated for assessment of groups of patients. 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 his assistance in preparing the manuscript.

**Contributors** PTP and DW were responsible for the conceptualisation of this project. PTP was responsible for instrument development, statistical procedures and interpretation of the data. DW performed the surgeries. RK was responsible for the creation of datasets and drafted the questionnaires. All the 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** None declared.

**Ethics approval** Ethics committee at the Medical University of Łódź (approval number RNN/190/07/KB).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** No additional data are available.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

## REFERENCES

1. Kane RL, Saleh KJ, Wilt TJ, *et al.* The functional outcomes of total knee arthroplasty. *J Bone Joint Surg Am* 2005;87:1719–24.
2. Bourne RB, Chesworth BM, Davis AM, *et al.* 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, *et al.* 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, Ek Dahl C, *et al.* 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, *et al.* 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, *et al.* 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, *et al.* 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. Zolnierczyk-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 edn. 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, *et al.* 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. 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, ed. *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.
  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 edn. 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.
  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.