Function and activity in patients with knee arthroplasty: validity and reliability of a German version of the Lysholm Score and the Tegner Activity Scale

Jaap Swanenburg[a,b,c], Peter P. Koch[d], Nadja Meier[a,e], Brigitte Wirth[e]

[a] Department of Physiotherapy, Balgrist University Hospital, Zurich, Switzerland

[b] Physiotherapy and Occupational Therapy Research, Directorate of Research and Education, University Hospital Zurich, Switzerland

[c] Department of Chiropractic Medicine, Faculty of Medicine, Balgrist University Hospital, University of Zürich, Switzerland

[d] Department of Orthopaedics, Balgrist University Hospital, Zurich, Switzerland

[e] Institute of Human Movement Sciences and Sport, ETH Zurich, Switzerland

Summary

OBJECTIVE: The Lysholm score and the Tegner activity scale are widely used for assessing knee function and activity level after knee ligament injuries. This study aimed to asses validity and reliability of the German versions of the Lysholm score (Lysholm-G) and the Tegner activity scale (Tegner-G) in patients after total knee arthroplasty (TKA).

METHODS: Two patient groups (<3 months postoperative; 3 to 12 months postoperative) and a healthy control group completed both questionnaires twice within 3 to 7 days. On the first occasion, the SF-12 questionnaire was also administered. The instruments' absolute and relative reliability, internal consistency, construct validity (discriminative validity and exploratory factor analysis), criterion validity, and floor/ceiling effects were determined.

RESULTS: Both instruments showed acceptable relative reliability (Lysholm-G: ICC2,1 = 0.87; Tegner-G: ICC2,1 = 0.79), and the Lysholm-G had a Cronbach's alpha of 0.93. Kruskal-Wallis test and post-hoc Mann-Whitney U tests showed significant differences in both scores among the groups. Factor analysis of Lysholm-G yielded three factors that together explained 76% of the variance. Correlations between the Lysholm-G/Tegner-G and the physical component of the SF-12 were moderate (ρ = 0.60/0.59). Item analysis of the Lysholm-G revealed a low relative reliability for the item "instability" and low discriminative validity for the items "locking" and "instability".

CONCLUSIONS: This study demonstrated acceptable psychometric performances for the Lysholm-G and the Tegner-G scales as outcome measures for patients with TKA. A short version of the Lysholm score without the "locking" and "instability" items might be more appropriate for TKA patients.

Key words: activity; function; reliability; total knee arthroplasty; validity

Objective

In today's society, degenerative joint diseases are one of the main causes of pain and disability. The knee joint is among the most often affected joints in the human body [1]. Pain and limited knee function resulting from degenerative joints can considerably affect a patient's quality of life [2]. Therapies include a wide range of conservative methods, but in cases of advanced degenerative osteoarthritis, a total knee arthroplasty (TKA) can relieve a patient's pain [3]. This pain reduction leads to an increase in activity, which is an important contributor to quality of life [4, 5]. Thus, measuring the patient's activity level in combination with knee function [6] could provide an important outcome measurement for TKA surgery [7]. Numerous assessment tools, such as the WOMAC, the SF-36 and the Knee Society score, are currently used to evaluate TKA outcomes [8–11]. However, there is a need for a simple and concise standardised questionnaire to assess the outcomes of TKA surgery [11, 12].

The Lysholm score and the Tegner activity scale, which measure the patient's perceptions of function and activity [13], are commonly used to document outcomes after anterior cruciate ligament injuries [14, 15]. The Lysholm score ranges from 0 to 100 points and is based on eight domains: limp, locking, pain, stair-climbing, support, instabil-

Abbreviations

ACL	Anterior cruciate ligament
TKA	Total knee arthroplasty
HRQOL	Health-related quality of life
GPs	General practitioners
ICC	Intraclass correlation coefficient
SEM	Standard error of measurement
SDC	Smallest detectable change
SF-12	Short-Form 12 HRQOL
PCS	Physical component summary of the SF-12
MCS	Mental component summary of the SF-12
FV	First visit
SV	Second visit

ity, swelling, and squatting. A score of 95 to 100 is considered to be excellent, 84 to 94 is good, 65 to 83 is fair, and <65 is poor [16]. The Tegner activity scale is a numerical scale ranging from an activity level of 0 (sick leave or disability pension due to knee problems) to 10 (competitive sports on a very high level). Both tools are easy to administer, short in length, and have been validated as clinician- and patient-administered instruments for a variety of knee conditions, primarily anterior cruciate ligament (ACL) and other ligament injuries, but also including chondral disorders, meniscal injuries, and patellar dislocation [13, 14, 17]. However, in contrast to ACL surgery, TKA surgery is not sports-related [18]; rather, this procedure is a consequence of osteoarthritis [3] and is primarily performed on patients older than 65 years [19]. Thus, the applicability of the Lysholm score and the Tegner activity scale for this patient group has yet to be investigated. To date, the Tegner activity scale has been shown to increase during rehabilitation in TKA patients [20], and the scale's reliability was reported to be excellent, with no ceiling effects [21]. To our knowledge, however, no complete validation of the two scores has been conducted in patients with TKA [22, 23]. Therefore, the purpose of the present study was to determine the psychometric properties of the Lysholm score and the Tegner activity scale in patients with a TKA. As both instruments have been translated into German (Lysholm-G and Tegner-G) and those versions have been validated in ACL patients [24, 25], the German versions were used in this study.

Methods

Participants
The patients were recruited consecutively during a post-TKA surgery check-up appointment at the knee unit of the orthopaedic department at Balgrist University Hospital, Zurich. They were divided into two groups: the first-visit group (FV; 0 to 3 months post-surgery) and the second-visit group (SV; 3 to 12 months post-surgery). Additionally, an age matched control group (CG) of healthy participants was recruited within the medical and therapeutic staff of the Balgrist University Hospital and students of ETH Zurich. All participants were not familiar with the questionnaires. The exclusion criteria for all participants were arthritis (inflammatory disease), loco-regional tumour or metastasis, inability to speak and read German fluently and severe psychiatric disorders.

Validation process
In one session, demographic and clinical characteristics were collected, and both the Lysholm-G and Tegner-G were completed with instructions of a study team member. The instructions given by the research assistant were the same as those on the questionnaire itself. Additionally, the patients were asked to complete the health status questionnaire SF-12 [26]. The participants were then instructed to complete the Lysholm-G and Tegner-G again at home 3 to 7 days later and return the forms by mail. If the questionnaires were not returned within four days, the patient received a reminder telephone call.

Data analysis and statistics
Descriptive statistics were used to describe the participants' characteristics. Differences between participants' characteristics were tested with a Mann-Whitney U test. Missing values and dropouts were noted and excluded from further analyses. Parametric statistical tests were chosen to calculate the relative reliability (intraclass correlation coefficient ($ICC_{2,1}$) and its 95% confidence interval) [24, 25]. ICC results greater than 0.70 were considered to be acceptable [27]. Cronbach's alpha was used to assess the internal consistency of the Lysholm-G but was not determinable for the Tegner-G, which consists of a single item. The acceptable range of coefficient alpha values is 0.70–0.95 [28]. To assess the absolute reliability, the standard error of measurement (SEM) and the smallest detectable change (SDC) were calculated. The SEM represents the standard deviation of repeated measures in the same patient. The SDC represents the minimal change that a patient must achieve on the scale to ensure that the observed change is real and not just measurement error. The SEM was calculated using the formula SEM = the pooled standard deviation of the first and second assessment*$\sqrt{(1-ICC)}$ [29]. The SDC was calculated as 1.96 x $\sqrt{2}$ x SEM [27]. For construct validity, the ability to differentiate between FV, SV, and CG (the discriminative validity) was tested using the Kruskal-Wallis and Mann-Whitney U tests as *post-hoc* tests, and an exploratory factor analysis was completed for the Lysholm-G. In the absence of a gold standard, criterion validity was established by correlating the total scores of the Lysholm-G and Tegner-G with the physical component summary (PCS) and the mental component summary (MCS) of the SF-12 [25]. Spearman's coefficient values were interpreted as indicating a relationship that was excellent (>0.9), good (0.7–0.9), moderate (0.5–0.69), fair (0.2–0.5), or minimal to absent (0.0–0.2) [30]. Floor and ceiling effects were considered to be present if more than 15% of the respondents achieved the lowest or highest possible total score [14]. In addition to the analysis of the Tegner-G and Lysholm-G total scores, the single items of the Lysholm-G were analysed identically. All of the analyses were conducted using the IBM-SPSS 20 statistical software (SPSS, Inc., Chicago, IL). The significance level alpha was set at 0.05 and at 0.05/3 = 0.017 for the *post hoc* tests.

Ethics committee
All of the participants gave their informed written consent. The study was approved by the ethics commission of the Canton of Zurich, Switzerland (KEK-ZH-Nr: 2011–0303).

Results

A total of 56 participants were recruited, 52 of whom completed all of the measurements. Due to braces used for after surgery, four patients did not answer the question about instability, and one did not answer the question about locking. Consequently, their four data sets were removed from all of the analyses. The participants' characteristics are shown in table 1. Four of the Tegner-G score retest forms and two of the Lysholm-G retest forms had inconclusive answers (more than one answer was marked), which were clarified by phone. There was no difference in age.

$ICC_{2,1}$ showed good reliability for both measures (0.87 for the Lysholm-G and 0.79 for the Tegner-G) in all of the TKA patients (table 2), and Cronbach's alpha for the Lysholm-G was alpha = 0.93. The SEM was 7.84 for the Lysholm-G and 0.44 for the Tegner-G. The SDC was 22 (21.72) for the Lysholm-G and 2 (1.16) points for the Tegner-G (table 2 and table 3).

The Kruskal-Wallis test showed an overall significant difference among the three groups for both scores (Lysholm-G: $\chi^2 = 29.16$, $p <0.001$ and Tegner-G: $\chi^2 = 29.46$, $p <0.001$). The *post-hoc* Mann-Whitney U tests showed significant differences among all of the groups; $p \leq 0.001$. The exploratory factor analysis yielded three factors: Factor 1, with an eigenvalue of 3.54, explained 44.20% of the variance; the addition of Factor 2, with an eigenvalue of 1.46, explained 62.44% of the variance; and the addition of Factor 3, with an eigenvalue of 1.05, explained 75.59% of the

Table 1: The participant characteristics and mean scores for all groups.

	CG Control group (n = 18)	FV First-visit group (n = 14)	Mann-Whitney U *p*	SV Second-visit group (n = 20)	Mann-Whitney U *p*
Female / Male	10 / 8	7 / 7	–	10 / 10	–
Age, yrs (SD) Range	65 (8) 39/81	65 (8) 51/76	0.790	66 (8) 47/81	0.988
Weight, kg (SD)	72 (15)	83 (14)	0.071	80 (14)	0.052
Height, cm (SD)	168 (9)	168 (7)	0.568	168 (9)	0.640
Lysholm-G (SD)	96 (7)	57 (20)	<0.001*	73 (18)	<0.001*
Tegner-G (SD)	4 (1)	1 (1)	<0.001*	2 (1)	<0.001*
SF-12 PCS (SD)	52 (6)	31 (11)	<0.001*	38 (12)	<0.001*
SF-12 MCS (SD)	52 (10)	57 (5)	0.201	53 (10)	0.795
Days post-op(SD)	–	6 (5)	–	34 (16)	–

SD = Standard deviation; PCS = SF-12 physical component summary; MCS = SF-12 mental component summary, U = Mann-Whitney U Test, *p** = sig <0.05.

Table 2: The test-retest reliability parameters of the Lysholm-G in all patients.

	1st measurement mean (SD)	2nd measurement mean (SD)	ICC (95% CI)	Cronbach's alpha	SDC	SEM
Lysholm-G (Max. 100 points)	66.38 (20.83)	65.71 (22.84)	0.89 (0.80/0.95)	0.93	21.72	7.84
Lysholm-G short version (Max. 60 points)	34.94 (14.63)	35.03 (15.69)	0.87 (0.76/0.94)	0.93	15.14	5.47

SD = standard deviation; 95% CI = 95% confidence interval; SDC = smallest detectable change; SEM = error of measurement

Table 3: The test-retest reliability parameters of the Tegner-G scores in all of the TKA patients.

	1st measurement mean (SD)	1st measurement median	2nd measurement mean (SD)	2nd measurement median	ICC (95% CI)	SDC	SEM
Tegner-G	1.97 (1.40)	2.00	2.18 (1.45)	2.00	0.79 (0.63/0.89)	1.80	0.65

SD = standard deviation; 95% CI = 95% confidence interval; SDC = smallest detectable change; SEM = error of measurement

Table 4: The test-retest reliability, validity and agreement parameters of the Lysholm-G items in TKA patients.

	1st measurement mean (SD)	2nd measurement mean (SD)	ICC (95% CI)	SDC	SEM	Kruskal-Wallis χ2 (p)
Item 1: limp (Max. 5 points)	2.41 (1.73)	2.82 (1.69)	0.72 0.51/0.85	2.51	0.91	35.48 (p <0.001)
Item 2: support (Max. 5 points)	3.68 (1.51)	3.62 (1.61)	0.86 0.75/0.93	1.62	0.59	17.47 (p <0.001)
Item 3: locking (Max. 15 points)	12.18 (4.25)	11.68 (4.74)	0.73 0.53/0.87	6.47	2.34	3.13 (p = 0.209)
Item 4: instability (Max. 25 points)	19.26 (7.60)	17.65 (8.63)	0.50 0.20/0.71	15.90	5.74	4.52 (p = 0.104)
Item 5: pain (Max. 25 points)	15.88 (7.53)	15.41 (8.84)	0.71 0.49/0.84	12.22	4.41	20.25 (p <0.001)
Item 6: swelling (Max. 10 points)	4.94 (4.49)	4.65 (4.38)	0.61 0.36/0.79	7.66	2.77	18.31 (p <0.001)
Item 7: stair-climbing (Max. 10 points)	5.53 (3.34)	6.15 (3.07)	0.70 0.49/0.84	4.86	1.75	29.22 (p <0.001)
Item 8: squatting (Max. 5 points)	2.50 (1.79)	2.38 (1.83)	0.66 0.42/0.82	2.93	1.06	31.15 (p <0.001)

SD = standard deviation; 95% CI = 95% confidence interval; SDC = smallest detectable change; SEM = error of measurement

variance. The varimax-rotated factor matrix is presented in table 5.

Spearman's ρ coefficient values showed moderate correlations between the Lysholm-G and Tegner-G and the SF-12 PCS (Lysholm-G: ρ = 0.60, p <0.001; Tegner-G: ρ = 0.69, p <0.001) and no correlation with the SF-12 MCS (ρ = −0.02 and 0.06; not significant). Within the FV group, the Tegner-G score exhibited a floor effect of 50%, with 7 patients scoring the lowest possible total score (0). The item analysis of the Lysholm score showed that all of the items except Items 3 ("locking") and 4 ("instability") were able to discriminate among the three groups. This analysis further revealed a low ICC of 0.50 for Item 4 and a large SDC (table 4).

Based on these results, the validity and reliability of a short version of the Lysholm-G (without the items 3 and 4) were determined. This analysis resulted in the same ICC (0.87) and Cronbach's alpha (0.93) that were determined for the original version of the Lysholm score; however, the SDC and SEM were proportionally higher, allowing discrimination among the three groups (Kruskal-Wallis test: χ2 = 32.90, p <0.001; post-hoc: FV versus SV: p = 0.03, FV and SV versus CG subjects: p <0.001). The factor analysis revealed two factors that explained 72.68% of the variance (table 6).

Discussion

In this study, the Lysholm-G score and Tegner-G activity scale demonstrated generally acceptable validity and reliability, justifying their use as outcome measures for patients with TKA. Thus, the Lysholm-G combined with the Tegner-G might be a simple and concise assessment tool to assess the outcomes of TKA [11].

Five items (7%) of the Lysholm-G were not completed because of a post-surgery protocol that required these parti-

Table 5: The varimax-rotated factor matrix of the Lysholm score (factor loads <0.3 suppressed).

		Factors		
		1	2	3
Item 1	Limp	.813		
Item 2	Support	.791		.390
Item 3	Locking		.830	
Item 4	Instability		.832	
Item 5	Pain		.419	.760
Item 6	Swelling			.749
Item 7	Stair-climbing	.837		
Item 8	Squatting	.854		

Table 6: The varimax-rotated factor matrix of the short version of the Lysholm score (factor loads <0.3 suppressed).

		Factors	
		1	2
Item 1	Limp	0.798	
Item 2	Support	0.762	
Item 5	Pain		0.850
Item 6	Swelling		0.716
Item 7	Stair-climbing	0.852	
Item 8	Squatting	0.884	

cipants to wear braces. In 11% of the retest assessments, the participants gave more than one answer to a question, which implies that the patients were uncertain about how to correctly answer the questionnaire at home. It can be concluded that to address this problem, this patient population should receive extra instruction before the self-administration of these instruments.

Overall, the Lysholm-G exhibited test-retest reliability in TKA patients that was adequate and comparable with studies of other populations [16, 31, 32]. Furthermore, the internal consistency was better than those reported in studies of patients with meniscal injury or ACL injuries [16, 24, 31, 33]. The SDC of the Lysholm-G score (22 points) was much larger than those reported in studies of patients with meniscal injury (SDC = 9 points) or ACL injuries (SDC = 11 to 14.6 points) [16, 24, 31, 33]. The large range of Lysholm scores (minimum = 15 / maximal = 100) results could be the reason for this large SDC. Future studies should investigate the SDC in more homogeneous groups with respect to days post surgery.

However, we found that the SDC of the Tegner-G was 2 points, which differs from the SDC of 1 point reported in ACL injury populations [16, 31]. Nevertheless, because the difference between the CG (mean = 4 points) and FV (mean = 1 point) groups was larger than the SDC, the Tegner-G can still be used in a clinical TKA setting. The good discriminative validity of the Lysholm-G and Tegner-G that was demonstrated in this study indicates that the Lysholm-G and Tegner-G can differentiate between the functional status of TKA patients at various rehabilitation stages and that of healthy controls.

In the absence of a real gold standard, the Lysholm-G and Tegner-G have been recommended as a gold standard [34]. Criterion validity was established by correlating the Lysholm-G and Tegner-G with the SF-12 [35], which has undergone vigorous testing and has been used previously to measure outcomes after arthroscopic knee surgery [36, 37]. In this study, the Lysholm-G and Tegner-G correlated moderately with the SF-12 PCS, but not with the average mental component score. These results were similar to those found in ACL patients [31].

As only one patient (3%) achieved the maximum score on the Lysholm-G (consistent with the reported data for ACL patients (2%) [31]), there was no ceiling effect for either score. The floor effect of the Tegner-G that was observed in the FV group was based on nine patients within the first postoperative week, which implies that these patients will achieve higher scores with increased postsurgical time. Thus, this floor effect is not of clinical relevance.

The factor analysis of the Lysholm-G revealed three factors that could be interpreted related to 1) transfer and changing body position (the items "limp", "support", "stair-climbing", and "squatting"); 2) joint function (the items "locking" and "instability"); and 3) signs of inflammation (the items "pain" and "swelling"). However, given that instability and locking are problems more specific to patients with ACL or meniscal injuries [38, 39], it appears plausible that the item analysis of the Lysholm-G revealed low discriminative validity for the items "locking" and "instability" and low reliability for the item "instability". These were also the items that could not be answered by the four pa-

tients who were wearing braces. The analysis of the psychometric properties of a short version of the Lysholm-G without these two items showed validity and reliability similar to those of the original version, but proportionally higher SEM and SDC, indicating that it might be more difficult for an individual to achieve a clinically relevant change.

One limitation of the present study is the small sample size which could have impacted the psychometric properties values. Another limitation is that the present study did not assess pre-operative data and responsiveness, which is necessary for the complete coverage of psychometric properties using the Lysholm-G and Tegner-G [40]. Future research should therefore assess these instruments' sensitivity to changes in activity levels over time after surgical or rehabilitative interventions for TKA with a larger sample size.

Furthermore, further studies should investigate the relationship between the Lysholm-G (and its short version) and Tegner-G and common tests used with TKA, such as patient-based questionnaires (WOMAC and the Knee Society score) and performance-based tests (sit-to-stand, maximal isometric contraction and timed up-and-go).

Conclusion

Overall, the Lysholm-G score and the Tegner-G scale demonstrated acceptable reliability and validity performances as outcome measures for patients with TKA. A short version of the Lysholm score without the items "locking" and "instability" might be more appropriate for TKA patients; however, this possibility must be investigated in further studies. Future research should also assess responsiveness of both questionnaires in TKA patients. Nevertheless, the results of this study may form a basis for further research that examines, for example, the effects of physical exercise on patients after TKA surgery.

Correspondence: Jaap Swanenburg, PhD, Department of Physiotherapy, Balgrist University Hospital, Forchstrasse 340, CH-8008 Zurich, Switzerland, jaap.swanenburg[at]balgrist.ch

References

1 Gawel J, Fibiger W, Starowicz A, Szwarczyk W. Early assessment of knee function and quality of life in patients after total knee replacement. Ortop Traumatol Rehabil. 2010;12(4):329–37.

2 Hirvonen J, Blom M, Tuominen U, Seitsalo S, Lehto M, Paavolainen P, et al. Health-related quality of life in patients waiting for major joint replacement. A comparison between patients and population controls. Health Qual Life Outcomes. 2006;4:3.

3 Crockarell JR, Guyton JL. Arthroplasty of the knee. In: Campbell's Operative Orthopaedics. edn. Edited by Canale ST, Beaty JH. Philadelphia: Mosby; 2008; 241–299.

4 Mont MA, Marker DR, Seyler TM, Gordon N, Hungerford DS, Jones LC. Knee arthroplasties have similar results in high- and low-activity patients. Clin Orthop Relat Res. 2007;460:165–73.

5 Spinarelli A, Petrera M, Vicenti G, Pesce V, Patella V. Total knee arthroplasty in elderly osteoporotic patients. Aging Clin Exp Res. 2011;23(2 Suppl):78–80.

6 Lysholm J, Tegner Y. Knee injury rating scales. Acta Orthop. 2007;78(4):445–53.

7 Hayes DA, Watts MC, Anderson LJ, Walsh WR. Knee arthroplasty: a cross-sectional study assessing energy expenditure and activity. ANZ J Surg. 2011;81(5):371–4.

8 Mahomed NN, Liang MH, Cook EF, Daltroy LH, Fortin PR, Fossel AH, et al. The importance of patient expectations in predicting functional outcomes after total joint arthroplasty. J Rheumatol. 2002;29(6):1273–9.

9 Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol. 1988;15(12):1833–40.

10 Katz JN, Phillips CB, Baron JA, Fossel AH, Mahomed NN, Barrett J, et al. Association of hospital and surgeon volume of total hip replacement with functional status and satisfaction three years following surgery. Arthritis Rheum. 2003;48(2):560–8.

11 Ghanem E, Pawasarat I, Lindsay A, May L, Azzam K, Joshi A, et al. Limitations of the Knee Society Score in evaluating outcomes following revision total knee arthroplasty. J Bone Joint Surg Am. 2010;92(14):2445–251.

12 Wang D, Jones MH, Khair MM, Miniaci A. Patient-reported outcome measures for the knee. J Knee Surg. 2010;23(3):137–51.

13 Briggs KK, Lysholm J, Tegner Y, Rodkey WG, Kocher MS, Steadman JR. The Reliability, Validity, and Responsiveness of the Lysholm Score and Tegner Activity Scale for Anterior Cruciate Ligament Injuries of the Knee. Am J Sports Med. 2009;37(5):890–7.

14 Tegner Y, Lysholm J. Rating Systems in the Evaluation of Knee Ligament Injuries. Clin Orthop Relat Res. 1985;198:43–9.

15 Lysholm J, Gillquist J. Evaluation of knee ligament surgery results with special emphasis on use of a scoring scale. Am J Sports Med. 1982;10(3):150–4.

16 Briggs KK, Kocher MS, Rodkey WG, Steadman JR. Reliability, validity, and responsiveness of the Lysholm knee score and Tegner activity scale for patients with meniscal injury of the knee. J Bone Joint Surg Am. 2006;88(4):698–705.

17 Lysholm J, Gillquist J. Evaluation of Knee Ligament Surgery Results with Special Emphasis on Use of a Scoring Scale. Am J Sports Med. 1982;10(3):150–4.

18 Gianotti SM, Marshall SW, Hume PA, Bunt L. Incidence of anterior cruciate ligament injury and other knee ligament injuries: a national population-based study. J Sci Med Sport. 2009;12(6):622–7.

19 Kurtz S, Mowat F, Ong K, Chan N, Lau E, Halpern M. Prevalence of primary and revision total hip and knee arthroplasty in the United States from 1990 through 2002. J Bone Joint Surg Am. 2005;87(7):1487–97.

20 Diduch DR, Insall JN, Scott WN, Scuderi GR, Font-Rodriguez D. Total knee replacement in young, active patients. Long-term follow-up and functional outcome. J Bone Joint Surg Am. 1997;79(4):575–82.

21 Naal FD, Impellizzeri FM, Leunig M. Which is the best activity rating scale for patients undergoing total joint arthroplasty? Clin Orthop Relat Res. 2009;467(4):958–65.

22 Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42.

23 Briggs KK, Steadman JR, Hay CJ, Hines SL. Lysholm Score and Tegner Activity Level in Individuals With Normal Knees. Am J Sports Med. 2009;37(5):898–901.

24 Wirth B, Liffert F, de Bruin ED. Development and evaluation of a German version of the Lysholm score for measuring outcome after anterior cruciate ligament injuries. Sportverletz Sportschaden. 2011;25(1):37–43.

25 Wirth B, Meier N, Koch PP, Swanenburg J. Development and evaluation of a german version of the tegner activity scale for measuring outcome after anterior cruciate ligament injury. Sportverletz Sportschaden. 2013;27(1):21–7.

26 Ware J, Jr., Kosinski M, Keller SD. A 12–Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34(3):220–33.

27 Kane R. Outcomes measures. In: Understanding Health Care Outcomes Research. edn. Gaithersburg, MD: Aspen; 1997:17–18.

28 Irrgang JJ, Snyder-Mackler L, Wainner RS, Fu FH, Harner CD. Development of a patient-reported measure of function of the knee. J Bone Joint Surg Am. 1998;80(8):1132–45.

29 de Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol. 2006;59(10):1033–9.

30 Siegel S, Castellan JNJ. Nonparametric Statistics for the Behavioral Sciences 2nd edn. Boston Massachusetts: McGraw Hill, Inc; 1888.

31 Briggs KK, Lysholm J, Tegner Y, Rodkey WG, Kocher MS, Steadman JR. The reliability, validity, and responsiveness of the Lysholm score and Tegner activity scale for anterior cruciate ligament injuries of the knee: 25 years later. Am J Sports Med. 2009;37(5):890–7.

32 Negahban H, Mostafaee N, Sohani SM, Mazaheri M, Goharpey S, Salavati M, et al. Reliability and validity of the Tegner and Marx activity rating scales in Iranian patients with anterior cruciate ligament injury. Disabil Rehabil. 2011;33(23–24):2305–10.

33 Paxton EW, Fithian DC, Stone ML, Silva P. The reliability and validity of knee-specific and general health instruments in assessing acute patellar dislocation outcomes. Am J Sports Med. 2003;31(4):487–92.

34 Johnson DS, Smith RB. Outcome measurement in the ACL deficient knee – what's the score? Knee. 2001;8(1):51–7.

35 Briggs KK, Steadman JR, Hay CJ, Hines SL. Lysholm score and Tegner activity level in individuals with normal knees. Am J Sports Med. 2009;37(5):898–901.

36 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(2):88–96.

37 Katz JN, Harris TM, Larson MG, Krushell RJ, Brown CH, Fossel AH, et al. Predictors of functional outcomes after arthroscopic partial meniscectomy. J Rheumatol. 1992;19(12):1938–42.

38 Torg JS, Conrad W, Kalen V. Clinical diagnosis of anterior cruciate ligament instability in the athlete. Am J Sports Med. 1976;4(2):84–93.

39 Schraeder TL, Terek RM, Smith CC. Clinical evaluation of the knee. N Engl J Med. 2010;363(4):e5.

40 Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick 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.