Results

Development and validation of the questionnaires (studies III–V)

Content validity

The objective of paper IV was to develop a patient-relevant outcome measure for patients with knee injury and/or early osteoarthritis that could be used from the time of injury to development of OA. When reviewing the literature, we were guided by the principles by Liang and Jette (Liang and Jette 1981): 1) the instrument needs to ask specific questions, 2) the data collection procedure should be specified 3) the instrument should allow for quantification 4) the instrument should be reliable, valid, and sensitive to change. We found no outcome measure for knee injury that fulfilled these requirements. The WOMAC Osteoarthritis Index, developed for elderly with more advanced knee OA, fulfilled the metric requirements but the validity for younger subjects with knee injury or early OA was not tested. In addition, we consulted an expert panel comprising patients referred to physical therapy because of knee problems, orthopedic surgeons, and physical therapists from Sweden and the USA, to find out what areas needed to be included. Seven factors were identified by the panel: pain, early disease-specific symptoms, late disease-specific symptoms (e.g., symptoms of osteoarthritis), function, quality of life, activity level, and satisfaction. In addition we used data from paper I to determine the most relevant factors among patients with post-traumatic OA. Questions that most frequently received high responses and thus were considered to reflect the most predominant symptoms included those relating to pain, swelling, stiffness, and the ability to run, jump, kneel, and squat.

The KOOS was constructed on the basis of the literature review, expert panel and pilot study described above. Five of the seven identified factors were included, pain, other symptoms, function related to activities of daily living, function related to sports and recreation, and knee-related quality of life. To ensure content validity for the older population with OA, the questions from WOMAC Osteoarthritis Index were included in their full and original form in the KOOS questionnaire (with permission, Nicholas Bellamy, 1995). The KOOS subscale activities of daily living is equivalent to that of function in the WOMAC Osteoarthritis Index. Questions included in the subscales sport and recreational function and knee-related quality of life, were adopted as they were originally written or with some modification from other outcome measures used to assess ACL injury (Flandry et al. 1991, Mohtadi 1993). The standardized and user-friendly format of the WOMAC Osteoarthritis Index with five-item Likert scales was chosen. The KOOS questionnaire is self-explanatory and takes about ten minutes to complete.

Linguistic validation

An American English and a Swedish version of the KOOS were developed simultaneously. The linguistic validation of the Swedish versions of the KOOS and the WOMAC were carried out according to the guidelines by Guillemin (Guillemin et al. 1993, Guillemin 1995): translation, back-translation, committee reviewing, and pre-testing. The validation process of the Swedish version of KOOS is reported in paper V and the Swedish version of the WOMAC is reported in paper III.

Swedish version of KOOS

Independent back-translations from Swedish to English were performed. A committee of bilingual physicians, physical therapists, and knee patients reviewed the existing versions, agreed on a mutual Swedish version and ensured that the translations were fully comprehensible. The Swedish version of the KOOS was tested with regard to clearness of the language, ambiguities, and ability of subjects to complete the questionnaire without assistance. During this process, one adaptation was made to Swedish conditions; the question regarding difficulties when taking a bath was changed to bath/shower.

Swedish version of WOMAC

The WOMAC items were independently translated into Swedish and back translated into English in three separate processes. The first process is described above, the second was initiated by a Swedish physical therapist with interest in hip OA, and the third by a health related outcomes specialist. These three versions of WOMAC were used to evaluate subjects with hip and knee OA at different sites in Sweden. When learning of the existing versions of the WOMAC, it was decided that a committee should review the existing versions and agree upon a common version. When comparing the preliminary versions of WOMAC, few differences were noted and a common Swedish version was agreed upon. The reviewed version was tested for clearness of the language, ambiguities, and ability of subjects with hip or knee OA to complete the questionnaire without assistance. When pre-testing the KOOS it became obvious that many Swedes prefer to shower and do not take regular baths. This was especially true for subjects with hip or knee OA, who frequently had had their bathtub replaced by a shower. Therefore the modification of the question regarding bath to bath/shower undertaken in the KOOS questionnaire was maintained. An alternative change of item description considered was to get on a bike, which is a common task in Sweden requiring similar range of motion in hip and knee as entering a bathtub. However, according to the WOMAC User's Guide, and the originator Nicholas Bellamy, the intended aspect is the hygienic aspect.

KOOS score calculation

The five dimensions of KOOS are scored separately: pain (nine items); symptoms (seven items); activities of daily life function (17 items); sport and recreation function (five items); quality of life (four items). All items are scored from 0 to 4, and each of the five scores is calculated as the sum of the items included, in accordance with score calculations of WOMAC Osteoarthritis Index (Bellamy et al. 1988b). Scores are then transformed to a 0 - 100 scale, with zero representing extreme knee problems and 100 representing no knee problems, as common in orthopedic scales (Tegner and Lysholm 1985, Windsor et al. 1988, Noyes et al. 1989). Scores between 0 and 100 represent the percentage of total possible score achieved. Table 7 provides a transformation formula and information necessary to apply the formula to each scale. The separate scores of the five dimensions can be visualized as a profile, Figure 2. An aggregate score is not calculated since it is regarded desirable to analyze and interpret the five dimensions separately.

If a mark is placed outside a box, the closest box is used. If two boxes are marked, the box that indicated the more severe problems is chosen. Missing data are treated in agreement with SF-36 (Ware 1988); substituting missing values with the average value for the dimension. If more than two items are omitted, the response is considered invalid.

Table 7. Formulas for scoring and transforming KOOS subscales

Scale	raw score = sum of values of the following items [a]	possible raw score range
Pain	P1–P9	36
Symptoms	S1–S7	28
Activities of Daily Living	A1–A17	68
Sport and Recreation Function	SP1–SP5	20
Quality of Life	Q1–Q4	16

[a] the KOOS items are found in Appendix B–D

Formula and example for transformation of raw scale scores to a zero to 100 scale (0 = extreme knee problems, 100 = no knee problems)

$$\text{Transformed scale} = 100 - \frac{\text{actual raw score x } 100}{\text{possible raw score range}}$$

Example: A Pain raw score of 16 would be transformed as follows:

$$100 - \frac{(16 \text{x} 100)}{36} = 56$$

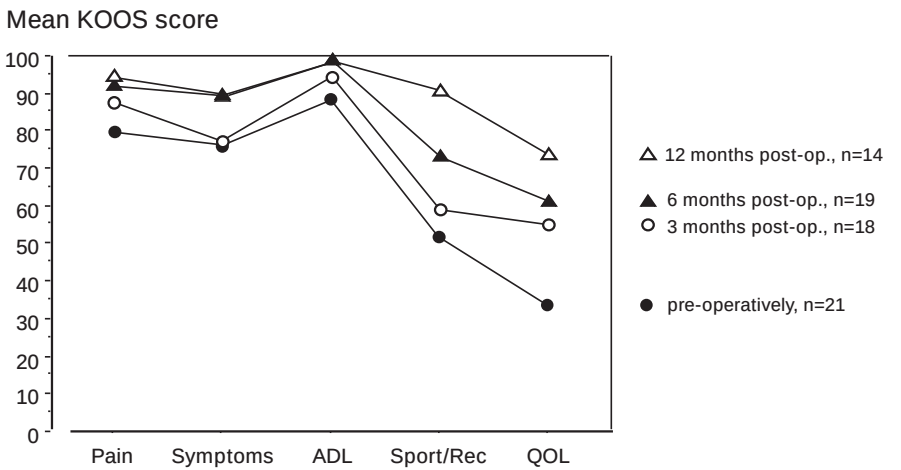

Figure 2. **KOOS Profile.** Mean KOOS scores preoperatively, three, six, and twelve months after reconstruction of the ACL (From paper IV).

Reliability

The clinimetric properties of the American English version of the KOOS, the Swedish version of the KOOS, and the Swedish version of the WOMAC are reported in papers III, IV, and V, respectively.

The test-retest reliability of the three questionnaires was found satisfactory. The intraclass correlation coefficients (ICC) were 0.74 or higher for all subscales of all versions except for the subscale Stiffness for the Swedish version of WOMAC (Table 8). Intraclass correlation coefficients of 0.75 or more are regarded as high (Rosner 1995).

Table 8. Intraclass correlation coefficients (ICC) for the American English and the Swedish versions of KOOS and the Swedish version of WOMAC

	Pain	Symptoms/ stiffness	ADL/ function	Sport/Rec.	QOL
KOOS US (IV)	0.85	0.93	0.75	0.81	0.86
KOOS Swe (V)	0.86	0.84	0.91	0.78	0.83
WOMAC Swe (III)	0.74	0.58	0.92		

Internal consistency, or inter-item correlation, an additional measure of reliability was calculated by Cronbach's alpha for the Swedish versions of the KOOS and the WOMAC. The alphas ranged from 0.71 to 0.96, being similar for the

corresponding subscales of the two questionnaires. Alphas of 0.8 or more are generally considered acceptable (Bellamy 1993).

Construct validity

Comparisons to the SF-36 were made to determine convergent and divergent construct validity. Generally, we expected higher correlations to the SF-36 scales with a high ability to measures physical health (convergent validity) and lower correlations to SF-36 scales with a high ability to measure mental health (divergent validity). Mean correlations of the five KOOS subscales of 0.50 and 0.36, respectively, were obtained in the Swedish (paper V) and the American (paper IV) validation studies when comparing to the SF-36 subscale Physical Function. Corresponding mean correlations when comparing to the SF-36 subscale Mental Health were 0.20 and 0.12, supporting our *a priori* assumption. As expected the highest correlations in both validation studies occurred between the scales that are intended to measure the same or similar constructs (KOOS Pain vs. SF-36 Bodily Pain, KOOS ADL vs. SF-36 Physical Function, KOOS Sport/Rec vs. SF-36 Physical Function), Table 9. The correlations obtained were generally higher in paper V, but when interpreting the pattern of correlations the conclusions were similar for four of the five KOOS subscales. For the KOOS subscale QOL quite opposite results were found. In the American validation study low correlations were found when comparing the KOOS subscale QOL to the subscales Physical Function and Bodily Pain of the SF-36 (0.19, 0.02), and in the Swedish validation study higher correlations were found when comparing QOL to the same subscales of SF-36 (0.45, 0.54), Table 9.

Table 9. Correlation coefficients (r_s) of the KOOS subscales to the SF-36 determined in the Swedish (paper V) and the American English (paper IV) validation studies. Significant correlations ($p<0.05$) in bold figures

KOOS subscales	SF-36 Physical Function	SF-36 Bodily Pain	SF-36 Mental Health
Pain Swe (V)	**0.49**	**0.65**	0.23
Pain US (IV)	0.29	**0.46**	0.09
Symptoms Swe (V)	**0.32**	0.29	0.04
Symptoms US (IV)	0.29	0.08	0.13
ADL Swe (V)	**0.68**	**0.65**	0.16
ADL US (IV)	**0.57**	0.35	0.22
Sport/Rec Swe (V)	**0.57**	**0.43**	0.12
Sport/Rec US (IV)	**0.47**	0.27	0.23
QOL Swe (V)	**0.45**	**0.54**	0.06
QOL US (IV)	0.19	0.02	0.33

Responsiveness

In papers III, IV, and V effect sizes (mean score difference divided by the preoperative standard deviation (Kazis et al. 1989)) were calculated as a measure of responsiveness. Six months after reconstruction of the ACL (paper IV) and three months after meniscectomy (paper V) the effect sizes ranged from 0.67 to 1.65 and could all be regarded as high (> 0.8) with the exception of the subscale ADL three months after meniscectomy. Knee related quality of life was the most responsive subscale in both studies. For the WOMAC (paper III) the effect sizes ranged from 0.51 to 0.71.

Sensitivity and responsiveness of the KOOS versus the WOMAC (studies II–V)

In Table 10 descriptive statistics of the KOOS data and the WOMAC data for the papers II–V are given. Generally, the mean scores of the KOOS were lower than the mean scores of the WOMAC for comparable dimensions (pain vs. pain and symptoms vs. stiffness). The KOOS subscale ADL is equivalent to the WOMAC subscale function, and thus the scores are the same. The differences in mean score were mostly small, indicating no clinically relevant difference in sensitivity between the instruments. Pain constituted an exception however. The ACL reconstructed patients (paper IV) had a KOOS pain score that was nearly ten points lower than the corresponding WOMAC pain score. A high mean score for the question regarding pain when twisting/pivoting on the knee increased the sensitivity of the KOOS pain scale.

In paper II, the control group reported substantially more symptoms (KOOS) than stiffness (WOMAC), indicating symptoms like grinding, catching and lack of full knee flexion being more prevalent than stiffness in subjects without a diagnosis of ACL injury, meniscus injury, or radiographic OA. In the group with radiographic OA (paper II) the mean score difference between the KOOS subscale symptoms and the WOMAC subscale stiffness was negligible due to the fact that the by far highest mean scores for symptom items were given to items related to stiffness.

In paper III, patients with arthroscopic OA reported more symptoms (KOOS) than stiffness (WOMAC) both preoperatively and postoperatively. In 75 % of the patients a partial meniscectomy was performed, indicating probable mechanical problems, which are assessed in the KOOS subscale symptoms but not in the WOMAC subscale stiffness.

In paper III, it was concluded that WOMAC was a valid outcome measure for patients with arthroscopic OA undergoing meniscectomy. A question that was not addressed was if the KOOS would have been a more sensitive or responsive outcome measure for this group of patients. When comparing the mean scores,

the differences were small for the comparable subscales of the KOOS and the WOMAC. However, because of the additional subscales of the KOOS, sport and recreation function and knee-related quality of life, the KOOS was the more responsive instrument.

Table 10. Mean, median, and standard deviations of KOOS and WOMAC scores in papers II, III, IV, and V

	KOOS Pain	WOMAC AC Pain	KOOS Symp-toms	WOMAC AC Stiff-ness	KOOS ADL	WOMAC AC Func-tion	KOOS Sport/ Rec	KOOS QOL
II								
OA, n=41								
Mean	81	84	78	76	79	79	52	59
Median	100	100	82	75	87	87	65	56
SD	25	25	19	25	24	24	32	27
Controls, n=50								
Mean	95	98	84	96	92	92	88	91
Median	100	100	100	100	100	100	100	100
SD	13	8	20	11	14	14	19	15
III								
OA, preop., n=52								
Mean	54	51	50	61	65	65	25	31
Median	53	50	55	62	66	66	22	31
SD	16	21	20	24	19	19	23	14
OA, postop., n=40								
Mean	66	65	63	71	74	75	36	42
Median	64	62	64	75	77	78	28	38
SD	23	28	24	24	21	21	28	24
IV [a]								
ACL, preop., n=26								
Mean	80	89	76	75	88	88	52	35
Median	80	95	75	75	90	90	55	34
SD	13	12	15	20	11	11	18	16
ACL, 3 mo., n=22								
Mean	87	95	77	74	93	93	58	52
Median	89	95	78	75	95	95	58	56
SD	8	6	10	16	7	7	23	13
ACL, 6 mo., n=24								
Mean	91	98	87	85	96	96	70	58
Median	92	100	89	88	98	98	72	62
SD	8	4	9	16	4	4	15	10
V								
Preop., n=142								
Mean	61	63	58	70	73	73	31	34
Median	61	62	60	75	75	76	30	31
SD	17	23	19	24	18	18	24	16

[a] Due to continued data collection, the n is higher in this table than in paper IV.

Comparison of the KOOS subscales and the KOOS profiles between study groups (studies II–V)

In all symptomatic study groups, the scores of the subscales sport and recreation function and knee-related quality of life were markedly lower than the scores of the subscales pain, symptoms and activities of daily living. The differences between the highest and the lowest mean scores ranged from 29 to 53 points. Most commonly the largest differences were seen between ADL (highest score) and sport and recreation function or knee-related quality of life (lowest score). The only exception was the control group where the difference between the highest (pain) and the lowest (symptoms) mean scores was 11 points (Figure 3). As previously described in "subjects, control group", these subjects had no known previous injury, no clinical signs of current injury to the ACL or menisci, and no radiographic signs of OA.

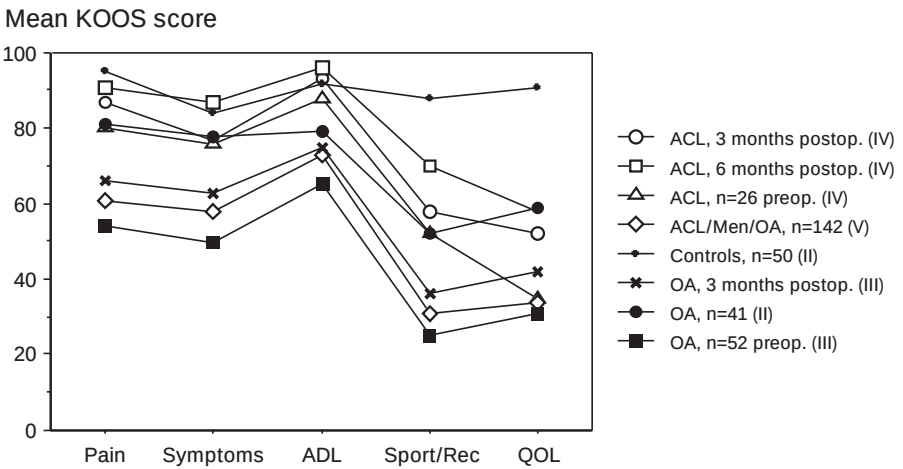

Figure 3. **KOOS Profiles** for papers II-V. Generally, the mean scores of the subscales sport and recreation function and knee-related quality of life are lower than the mean scores of the subscales pain, symptoms, and activities of daily living.

Six months after reconstruction of the ACL (paper IV) the subscores for symptoms and ADL were better than for the control group. Most likely this is an age effect, the mean age of the ACL patients was 32 years compared to 53 years for the control group. Preoperatively ACL patients reported less pain, symptoms and difficulty with ADL and sport and recreation function than the other groups. However, their knee-related quality of life was low, just a few points from being the lowest of all groups. In addition, quality of life was the subscore that improved the most over time in the group with ACL reconstruction. The ACL

group was also the only group to have a mean score of QOL that was lower than their mean score of sport and recreational function.

The study group with post-traumatic OA (paper II) had a profile similar to patients about to undergo reconstruction of the ACL, with the exception of knee-related quality of life. The older patients with post-traumatic OA reported a much better quality of life.

Short-term results of meniscectomy (study VI)

The objective of paper VI was to prospectively assess patient-relevant outcomes in patients undergoing arthroscopic partial meniscectomy.

Patient-relevant outcomes were evaluated at a mean of 14 weeks after arthroscopic partial meniscectomy. Significant improvement was found in all KOOS subscales, the Lysholm scoring scale, and the relevant subscales of the SF-36. However, despite reporting only minor pain and other symptoms postoperatively, significant physical disability and handicap (as measured by the KOOS subscales sport and recreation function and knee-related quality of life and the SF-36 subscales physical function, role-physical and bodily pain) was reported compared to reference groups. The KOOS data is shown in Figure 4.

Postoperatively 32% were active in sports, compared to 68% pre-injury. A sedentary life-style was reported by 38%, compared to 7% pre-injury as seen in

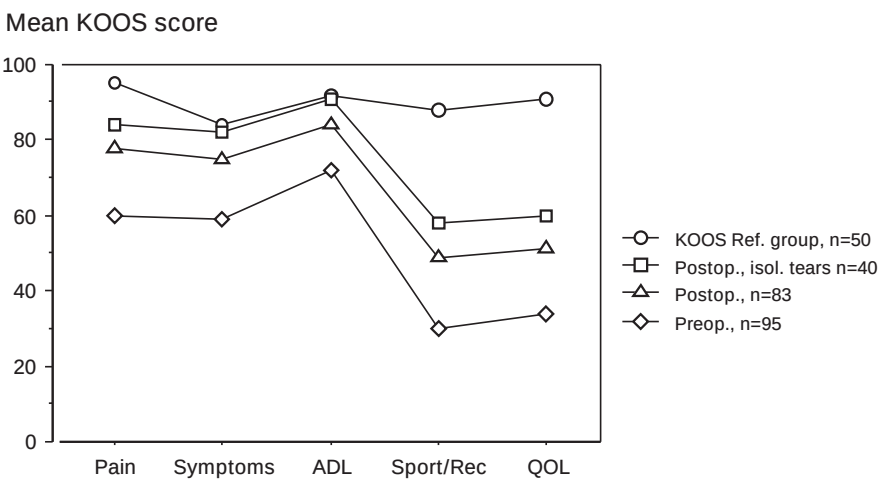

Figure 4. Preoperative and postoperative KOOS data for the total group having arthroscopic partial meniscectomy. Postoperative data is also given for the subgroup with isolated meniscus tears. The controls are identical with the control group in paper II.

Table 4, paper VI. Similar results were seen for the diagnostic subgroup with isolated meniscus tears. Associated cartilage damage explained 6%, and an associated ACL injury explained 2 %, of the variation in postoperative quality of life.

Long-term results of meniscectomy (study I)

The objective of paper I was to describe the long-term influence of meniscectomy, radiographic OA, gender, and age on symptoms, function, and performance tests.

Self-reported symptoms and knee-related function was obtained, performance tests were carried out, and radiographs were taken at a mean of 19 (17-22) years after meniscectomy. Age- and gender-matched controls were examined likewise. The data was analyzed in two steps. Firstly, subjects with meniscectomy (n = 159) were compared to the controls (n = 68), and subgroup analyses were carried out with regard to radiographic OA, gender, and age. Secondly, comparisons were carried out within the meniscectomized group with regard to radiographic OA, gender, and age.

Meniscectomized subjects reported significantly ($p < 0.001$) more symptoms and functional limitations than the controls did. The mean score difference ranged from 9 to 13 points. Significant differences of 8 to 9 points were also found when operated subjects without OA were compared to controls without OA. In the meniscectomized group, more severe radiographic OA was associated with more pain, symptoms, limitations in activities of daily living, and limitations in sport and recreation function. However, 41% of the patients with more severe radiographic OA did not report more pain than the controls did, confirming a clinically weak relationship between pain and radiographic knee OA. Meniscectomized women reported worse symptoms and functional limitations than meniscectomized men, and compared to female controls. Older age was not associated with more pain or more functional limitations. The results, including the performance tests, are summarized in Table 11.

Table 11. Summary of long-term patient-related results of meniscectomy

	Self-reported symptoms and function		Performance tests	
	Meniscectomy group	Control group	Meniscectomy group	Control group
Meniscectomy	⇓		(⇓)	
OA	⇓		–	
Female gender	⇓	–	⇓	–
Age	–	–	⇓	⇓

⇓ = Worse outcome
– = No effect on outcome

General discussion

Development of the KOOS

When developing the KOOS, we searched the literature, conducted a pilot-study, and consulted an expert panel. The literature revealed three important factors; symptoms, functional status, and satisfaction. The expert panel, comprised of patients referred to physical therapy because of knee injuries, orthopedic surgeons, and physical therapists from both Sweden and the United States, identified in addition to the above mentioned factors, also quality of life and activity level. The pilot study, carried out in subjects with post-traumatic OA, informed us that the most predominant symptoms included those relating to pain, swelling, stiffness, and the ability to run, jump, kneel, and squat. The KOOS questionnaire would not have contained the same items if any of these steps had been left out. It seems difficult to omit any step to simplify the validation process. As pointed out by Kirkley et al. (1998), if a measurement tool is developed with appropriate methodology it should be valid and have adequate reliability and responsiveness. Still, it is important to formally document these characteristics in different settings over different time periods.

It was decided not to include activity level and satisfaction into the KOOS questionnaire. These factors are separate constructs that do not necessarily correlate to function. The most commonly used instruments for assessment of activity level in patients with anterior cruciate ligament injury are the Tegner activity scale (Tegner and Lysholm 1985) and the Sports Activity Rating Scale (SARS) by Noyes et al. (1989). The Tegner activity scale grades physical activities on a scale from 0 (sick leave or disability pension because of knee problems) to 10 (competitive sports, soccer on national or international level). The SARS first categorizes the subject according to frequency of sports activity (4–7 days/week, 1–3 days/week, 1–3 times/month, or no sports possible). Secondly, a classification is done depending on how knee demanding the sports activity is. The result is a score from 0 to 100, where 0 indicates no sports activity possible and 100 knee provoking activities like soccer or basketball 4–7 times a week. No consensus has been reached on what approach to use to assess activity level in patients with ACL injury, and neither so for patients with other knee injuries or knee OA. For patients with ACL injury, stability-demanding activities like soccer are included at the highest activity level. If patients with meniscus injury were at focus, it is possible that soccer would be accompanied by weight lifting, which requires full knee flexion, at the highest level. To resolve these issues requires a solid validation work, and it was considered beyond the scope of this thesis.

Satisfaction was the other factor, considered important by patients, which was

not included in the KOOS questionnaire. It seemed impossible to agree upon wording that would be applicable in all situations. Additionally, the differences between health care systems in different countries are consequential, in some countries surgical procedures and/or physical therapy are financially restrained by insurance companies which can influence on reported satisfaction. Validation for each country would be appropriate. In paper IV (American subjects undergoing reconstruction of the ACL) we asked how satisfied the patients were with the result of their treatment using 5 Likert-boxes, ranging from not at all satisfied to extremely satisfied, as answer options. We collected data at several points after treatment, and found the data to be very inconsistent, supporting the theory that satisfaction is a composite construct that can not be readily assessed by one simple question.

Likert versus visual analog scale response format

The issue of Likert-scales versus visual analog scales as response format was not addressed in this study. When developing the KOOS, we compared the instructions needed for patients to complete a visual analog scale version and a Likert-version. We found the Likert-version requiring fewer instructions, and thus preferable, although the visual analog format was reported to be slightly more responsive in both major validation studies of the WOMAC (Bellamy et al. 1988a, 1988b). We regarded patient-friendliness being the more important issue. There have been many debates in the literature as to which of these two scales is to be preferred. Generally, high levels of correlations have been observed between scores made concurrently on Likert and visual analog scales (Bellamy et al. 1988a, 1988b).

Parametric versus non-parametric statistics

The first requirement for using parametric statistics is that the data is on an interval/ratio level of measurement (Hicks 1995). The underlying data from questionnaires such as the KOOS are ordinal and therefore non-parametric statistics are implemented and have been used throughout all papers. The use of parametric or non-parametric tests is a matter of debate (Campbell and Machin 1993, Bland 1995). Most calculations of the questionnaire data in this study have been carried out with both parametric and non-parametric statistics. In no case did the type of calculation change the interpretation of the results, indicating that either analysis may be used. Bellamy also noted this when comparing non-parametric and parametric treatment of WOMAC data (Bellamy et al. 1988b).

For questionnaire data means and standard deviations are often calculated instead of medians and inter-quartile range (Liang et al. 1985, Tegner and Lysholm 1985, Ware 1988, Bellamy 1995, Wright and Young 1997). If a set of data is normally distributed the mean and median should be similar. This has been the case except in paper II, where medians were given to help interpret the data.

Construct validity

Validation of a disease-specific measure is an ongoing process that no single experiment can provide. Since there is no "gold standard" with which the tool can be compared, investigators use strategies borrowed from psychologists who have developed methods for deciding whether questionnaires examining emotional function, intelligence, and attitudes are really valid. The most rigorous of these strategies is construct validity. A general understanding of the disease process and what investigators are trying to measure allows predictions to be made about how the tool will relate to other measures. If the predictions are confirmed in the population of interest then this strengthens the evidence for validity. If the instrument performs as expected over time in varied settings then one can be more confident that it is valid (Kirkley et al. 1998).

In this study, the KOOS correlated the best with the Lysholm scoring scale and the subscales of SF-36 with a high ability to measure physical health. Just as important, low correlations were found when comparing the KOOS to the SF-36 subscales with a high ability to measure mental health. The reason for choosing the SF-36 for validation purposes was that the SF-36 is well defined as to what constructs it measure, and appropriate for use in subjects with knee-related problems (Katz et al. 1992, Shapiro et al. 1996). The results were repeated in two validation studies of different populations from different countries undergoing different treatments. The subscale knee-related quality of life behaved differently in the Swedish and the American validation study. The reasons for this finding could include differences in diagnoses and diverse cultures. This finding warrants further research. Recently, a disease-specific measure of quality of life for ACL injury has become available (Mohtadi 1998). Comparing the KOOS subscale QOL to both a disease-specific and a generic measure assessing quality of life might add new information.

Comparison of the KOOS to the WOMAC

The KOOS contains the full and original version of the WOMAC, and thus WOMAC scores can be calculated from the KOOS questionnaire. When comparing the mean scores of comparable KOOS and WOMAC subscales (pain versus pain and symptoms versus stiffness) the KOOS generally had somewhat lower mean scores. The differences between the comparable subscales of the KOOS and the WOMAC were greater for patients with ACL injury and meniscus injury than for patients with OA. This was as expected, since the questions added in the KOOS were meant to increase the sensitivity for subjects with knee injury.

In paper II, the sensitivity of the WOMAC was compared to the KOOS subscales sport and recreation function and knee-related quality of life. When comparing the mean scores of the WOMAC and KOOS subscales, the difference

between the mean scores of the controls and the subjects with post-traumatic OA was twice for the KOOS subscales compared to the WOMAC subscales. Thus, when adding the subscales sport and recreation function and knee-related quality of life to the WOMAC, the sensitivity and the validity was increased for these "young subjects with old knees".

The KOOS is more sensitive and responsive than the WOMAC, largely due to the added subscales sport and recreation function and knee-related quality of life. Therefore the use of the KOOS is recommended in the populations investigated. For older populations with more severe OA however, WOMAC is still the questionnaire of choice. To determine the relevancy, and thus content validity, of the KOOS subscales sport and recreation function and knee-related quality of life in an older population with more severe OA is the objective of future studies.

In paper III, the preoperative and postoperative WOMAC mean pain scores were lower than the corresponding KOOS mean pain scores. This is the only instance in Table 10 when a WOMAC mean score is lower than the comparable KOOS mean score. However, when calculating effect sizes (mean score difference divided by preoperative standard deviation) we found the KOOS subscales to have higher effect sizes than the corresponding WOMAC subscale, thus indicating the KOOS being the more responsive instrument. The effect size has implications for the number of patients needed to detect significant differences. When using the data for the KOOS and WOMAC subscales pain (paper III), it can be calculated that 55 patients would be needed for the KOOS subscale pain to detect a significant difference (at the 0.05 level with a power of 80%) compared to 70 patients for the WOMAC subscale pain. For the corresponding subscales symptom and stiffness KOOS was also the more responsive instrument. The third WOMAC subscale, function, is identical with the KOOS subscale ADL, and identical effect sizes were calculated. The effect sizes calculated in paper III could not be regarded as high. This was expected however, since arthroscopy is not an intervention associated with major positive effects in patients with OA.

Comparison of KOOS subscales and KOOS profiles between study groups

In all study groups, the scores of the subscales sport and recreation function and knee-related quality of life were markedly lower than the scores of the subscales pain, symptoms and activities of daily living. When looking at raw data, symptomatic patients often report moderate to severe problems with items related to sport and recreation function and knee-related quality of life (i.e. disability and handicap) while reporting only mild to moderate problems with items related to impairment. This phenomenon could be a coping strategy. When interviewing patients with a chronic disease, such as OA of the knee, some patients tell that

they have learned to live with the pain and avoid thinking of it to be able to get on with their lives. However, the consequences of the impairment are more difficult to learn to live with and thus receive worse scores. Another explanation could be that symptomatic patients at their best avoid situations that cause pain or other symptoms, but still experience the disability and handicap. Further exploration of this area should incorporate the use of a valid measure of physical activity and qualitative research strategies.

When comparing to the profiles of the other symptomatic groups, the patients about to undergo ACL reconstruction experienced relatively large problems with knee-related quality of life in comparison to the other four KOOS subscales. In the USA, where the patients included in paper IV were treated, a high percentage of subjects with ACL injury are reconstructed. It could be speculated that early decision regarding operative treatment is the reason for the relatively minor symptoms and functional problems reported. Quite on the contrary, it could be the low quality of life that brings the patient to the operating room (Mohtadi 1993).

The study group with post-traumatic OA (paper II) had a profile similar to patients about to undergo reconstruction of the ACL, with the exception of knee-related quality of life. The older patients with post-traumatic OA reported a much better quality of life. Age is the most probable reason for this discrepancy in reported quality of life, but adaptation of physical activity level in the group with post-traumatic OA because of long-term knee problems is another possible explanation.

Minimum significant level of change

Few recommendations are found for established outcome measures regarding the minimum significant level of clinical change. Statistical methods can be used to calculate the minimum level of *detectable* change (Stratford et al. 1996), but these two levels do not necessarily agree. We compared the KOOS data available after ACL reconstruction to the clinical knowledge of rehabilitation phases following ACL reconstruction. Three months postoperatively, the patients experienced some pain, swelling, and restriction of range of motion and had not pushed their knee during sporting activities. This was reflected by (statistically non-significant) changes of 1 to 7 units in pain, symptoms, and sport and recreation function over this time interval, compared to preoperative values. Six months postoperatively, the patients were back at more vigorous activities including sport and had few symptoms, reflected by (statistically significant) changes of 8 to 23 units in all subscores. Thus, it seems that a change in score of 8 units or more may represent a clinically significant change following ACL reconstruction. To fully explore this difficult question, additional studies including different treatments should be undertaken.

Comparison of the KOOS to established, observer-administered knee scoring scales

Most scores for evaluation of knee injury assess similar constructs (i.e. symptoms and function). However, they differ significantly with respect to how qualitative data is being transformed into numbers and finally reported. Previously, most scores relied on weighting of each item according to the developers' opinion. Today, the approach is to let the patient take part in the developing process, and then let the patients themselves assess the severity of each item on a visual analog scale or with Likert-boxes (no problems, mild problems, moderate problems, severe problems, and extreme problems). However, regardless of what approach is used to score the individual item, most scores still aggregate all included items into one total score, making it difficult to interpret the result and to tell what construct the score actually is a valid measure of. The KOOS differs from the established knee scores, and from the other modern disease-specific and patient-relevant instruments, in this aspect. The KOOS assesses pain, other symptoms, activities of daily living, sport and recreation function, and knee-related quality of life in five separate scores.

Measurement level

The International Knee Documentation Committee (IKDC) uses four categories for an overall evaluation of knee injury; normal, nearly normal, abnormal, and severely abnormal. The overall IKDC rating is based on patient subjective assessment, symptoms, range of motion, and ligament examination, i.e. measures of impairment and disability are aggregated into one final rating. The final rating is equal to the worst individual group rating. Irrgang et al. (1998a) evaluated 133 patients with ACL reconstruction with the purpose to provide evidence for the validity of the guidelines for rating knee ligament injuries established by the IKDC. Construct validity was determined by establishing the relationship between the final IKDC rating and each patients' global rating of knee function. Concurrent validity was established by determining the relationship between the final IKDC rating and the modified Cincinnati knee score, an instrument commonly used to assess outcome following knee ligament surgery. In addition, the contribution of each individual group rating to the overall final IKDC rating was assessed. Irrgang et al. found that the average subjective rating was lower for patients with a worse IKDC rating, but the difference was only significant for subjects who were rated severely abnormal compared to better ratings, and for subjects rated as abnormal compared to those with better ratings. Additionally, 62 % of the variability in the final IKDC rating was accounted for by symptoms and laxity (measures of impairment), while range of motion (impairment) and patient's subjective assessment (disability) together explained 8 % of the variation in the final IKDC score. Thus, the IKDC cannot be considered a valid measure of patient-relevant outcomes. This statement is supported by the findings of Sny-

der-Mackler et al. (1997) who compared two groups, consisting of 10 patients each, with ACL deficiency. One group was able to return to sports (copers) and the other group was unable to return to sports (non-copers). However, there was no difference in measured laxity. All patients in the coper group received IKDC ratings of abnormal or severely abnormal, which did not correlate with the patient's self-report of knee function.

Score categorization

Tapper and Hoover, who in 1969 introduced their system for evaluation of symptoms and function following meniscectomy, categorized the outcome into four categories; excellent, good, fair, and poor. This approach is appealing and the raw scores of established knee scoring scales are frequently categorized into the same four categories, using arbitrarily chosen cut-off scores. However, categorizing can be criticized for its inexactness and introduction of bias. Sgaglione et al. (1995) found that categorizing the raw scores of ligament rating scales tended to inflate the result, and that the interpretation of categorical data depends on the content of the particular rating scale and the relative weight given to each component that is summoned to arrive at the final score. Individuals rated as excellent or good on one scale may therefore be rated as fair on another scale. They concluded that avoidance of data generalization remained the optimal method for studying the outcome of knee injury. Categorizing KOOS scores is not recommended.

Validity

The Lysholm knee scoring scale is probably the most commonly used system for evaluation of knee injury. It was developed to assess outcome following ACL injury and/or surgery, but Bengtsson et al. (1996) found that the Lysholm scores were on average higher in patients following ACL injury compared to those who had a meniscal tear, patellofemoral pain, or a lateral ankle sprain. Thus, the sensitivity of the Lysholm knee scoring scale to detect functional limitations in those for which the scale was designed is questionable, limiting its validity. One explanation could be that symptoms of instability (which constitutes 25 points of the Lysholm score) are not specific to ACL injury. In paper V, each item of the KOOS score was checked for reported severity. The groups with ACL deficiency (isolated or combined) did not experience more functional instability when assessed by the KOOS items 'do you have difficulty turning/twisting on your injured knee?', 'how much troubled are you with lack of confidence in your knee?', or the Lysholm scale item 'instability', indicating symptoms of instability not being specific to ACL injury. From the patients' perspective severe instability is an experienced functional limitation of ACL injury, as well as of meniscus injury and cartilage damage.

Risberg & Ekeland (1994) further questioned the validity of the Lysholm score. They evaluated functional tests to be used after anterior cruciate ligament

surgery. Evaluative factors were the Lysholm score, thigh atrophy, and knee instability. It was concluded that the Lysholm score was inaccurate in identifying functional problems during strenuous activities. This lack of patient-relevancy, and thus validity, is not really surprising since the Lysholm scoring scale was developed by orthopedic surgeons as a tool to aid the doctor in his/her assessment of symptoms and function related to ACL injury.

Comparison of established knee scoring scales

As pointed out by several authors, the results of different knee scoring scales are not comparable (Bollen and Seedholm 1991, Sgaglione et al. 1995, Labs and Paul 1997, Neeb et al. 1997). The most plausible explanations are differences in outcomes assessed and the weight assigned to each outcome. Bollen & Seedholm (1991) compared the Lysholm and Cincinnati knee scoring questionnaires in a prospective study of 41 patients with unilateral anterior cruciate deficiency. They found that patients score consistently higher on the Lysholm scale than on the Cincinnati scale, and concluded that it was difficult to compare the results of these two commonly used knee scoring scales. Sgaglione et al. (1995) compared the Lysholm, Cincinnati and Hospital for Special Surgery (HSS) knee scales in 65 patients whom underwent anterior cruciate ligament reconstruction. The Cincinnati scores were lower than the Lysholm and HSS scores. The HSS and the Lysholm did not correlate highly with the Cincinnati final rating, but they did correlate with each other. The Cincinnati score correlated more highly with individual grading and most precisely defined outcome in athletically active patients. Labs and Paul (1997) compared eight established scoring systems (including Lysholm, HSS, and IKDC) in a prospective study of 56 patients who received a Leeds-Keio ligament as an anterior cruciate ligament replacement. The systems were compared based upon their total scores, as well as their subjective, objective and functional criteria. It was concluded that the diversity of both the results and criteria within the systems made valid comparisons impossible. Neeb et al. (1997) compared nine measuring systems (4 knee scoring scales, 3 clinical tests, and 2 functional tests) used in patients with ACL injury. They found low correlations between all systems, indicating that all scores and tests seem to be related to different aspects of the ACL injured patient. It was concluded that evaluating systems measuring both at impairment and disability level are needed to gain insight in patients with an injured ACL.

When comparing patient-relevant outcome measures, the correlations should theoretically be higher than those obtained when comparing traditional knee scoring scales, provided that the patients included in the different developing processes experience the same symptoms and encounter the same functional limitations. The recently available patient-relevant measures were developed for similar but not identical populations, and yet no comparative studies are available.

Time frame dependency

An alternative explanation to questionnaire content as the reason for low correlations between established knee scoring scales, is the instructions provided, including the time frame considered by the patient when answering the questions. For the established knee scoring scales no instructions are given, i.e. it is not known if the patient considers his/her status at the actual time of assessment, the last week, or at an average from time of treatment to the time of evaluation. When an observer asks the questions, differences specific to each observer might occur. For the KOOS, thorough instructions are given and the patient is asked to consider the last week when answering the items. The last week was considered suitable since it was supposed that most functions included in the activities of daily living and sport and recreational function subscales had been performed during this period, and that the answers given should be more stable compared to a shorter time frame. A longer time frame was considered inappropriate since follow-ups one week after treatment of knee injury might be wanted. Little evidence is found as to the importance of the time frame used. Griffiths et al. (1993) compared three time frames (previous 24 hours, previous 48 hours, and previous 2 weeks) for the WOMAC when administered to 19 patients entered into a clinical trial of non-steroid anti-inflammatory drugs. No time frame dependency was observed, and the authors felt justified in varying the time frame of the WOMAC from 1 to 14 days. Although the time frame can be varied, it is important to give a time frame to avoid differences within and between groups of patients.

Comparison of the KOOS to other self-administered measures for knee injury

Administration mode

The Flandry questionnaire was the first self-administered questionnaire for assessment of knee injury and constituted the first step towards patient-relevant questionnaires. Having the patients themselves answer the questions avoided the significant observer bias which is introduced both by doctors and independent observers (Lieberman et al. 1996, McGrory et al. 1996, Höher et al. 1997, Roos et al. 1998a). Höher et al. (Höher et al. 1997) compared two methods of data collection on the result one year after anterior cruciate ligament reconstruction. Two Lysholm scores were obtained for 61 patients at the same clinical visit. First, the patients completed a self-administered questionnaire, and second, the Lysholm score form was completed by the investigator in the course of the patient interview. A comparison revealed that the mean score was significantly lower with self-administration. The assignment to different categories (excellent, good, fair, and poor) was also significantly altered by the manner of data collection. It

was concluded that to avoid observer bias a standardized self-administered questionnaire should be used.

With the attempt to avoid bias, the KOOS was developed to be self-explanatory and to be administered in a waiting room or by mail. This requires thorough linguistic validation to avoid ambiguities. In paper V, where the KOOS was administered by mail to 142 patients waiting for knee arthroscopy, it was found that very few (0.8 %) individual items of the KOOS were missing. Thus, scores could be calculated for all patients and all subscales but one subscale for one patient. The percentage missing items was comparable to the SF-36; a questionnaire also developed to be self-explanatory. However, for the self-administered version of the Lysholm scoring scale substantially more items were missing, probably due to the fact that the Lysholm scoring scale was not developed to be self-explanatory, lacks instructions, and holds many ambiguities. Until now the KOOS has been administered to more than a thousand patients, both by mail and in waiting rooms, and only few and minor questions have been noted, confirming the KOOS being suitable for self-administration.

Score aggregation

The quality of life outcome measure for chronic anterior cruciate ligament deficiency (ACL-QOL) (Mohtadi 1993, 1998) was the first outcome measure introduced for knee injury that can be said to be truly patient-relevant. The questionnaire was developed according to set guidelines, included patient input throughout the process, and had satisfactory metric properties. However, one total average score is calculated for all the 32 items. This procedure tends to flatten the results and makes interpretation more difficult. When developing the KOOS, initially a total score was calculated in addition to the five subscores, the reason being that one score (the total score) is easier to deal with than five subscores. However, when comparing the information derived from the total score, which told if the patient had any problems whatsoever, to the information derived from the five subscores, which gave a much more detailed picture as to what areas were causing difficulties, we decided that the use of five subscores was preferable.

The items included in the Activities of Daily Living Subscale (ADLS) (Irrgang et al. 1998b) were derived through literature search of knee scales for injury and OA. The items were selected to reflect problems reported by patients with ligamentous or meniscal injuries, patellofemoral pain, or osteoarthritis. This initial list of items was reviewed by twelve physical therapists, which resulted in addition or modification of several items. Patients were not interviewed, which in theory is questionable for a questionnaire supposed to be patient-relevant. The ADLS includes items assessed in the KOOS subscales pain, symptoms, and ADL. For the ADLS, as well as for the ACL-QOL, items assessing the impact of both impairment (such as pain, grinding, stiffness, and swelling) and disability (such as walking, stairs, standing, kneeling, and squatting) are aggregated into

one total score. The total score of the ADLS measures the level of function that the pathological condition or impairment of the knee imposes during activities of daily living.

Responsiveness

The responsiveness of the ACL-QOL has not yet been thoroughly investigated. In the validation study (Mohtadi 1998), twenty-five consecutive patients with documented chronic ACL deficiency were asked to complete the questionnaire in a prospective fashion on at least two separate visits over a six-month period. At the time of the second visit, patients indicated whether they had improved, deteriorated, or had no change in their clinical status. This was compared with the direction of change in score on the ACL-QOL questionnaire. Twenty-one of the 25 patients had appropriate overall scores on the repeat administration of the ACL-QOL based on the clinical change that had occurred. This information strengthens the validity of the questionnaire, but provides no information usable for comparison to other questionnaires, or determination of number of patients needed in clinical trials.

The validation study of the ADLS (Irrgang et al. 1998b) included 397 patients referred to physical therapy because of knee-related complaints. The primary diagnoses were ligamentous or meniscus injury, patellofemoral pain, and osteoarthritis. Fifty-seven percent were referred for therapy after operative treatment. Reliability, validity, and responsiveness were assessed and found satisfactory. No attempt was made to standardize the treatment, as the study was not designed to assess the effectiveness of any particular type of physical therapy. Responsiveness of the ADLS was evaluated by determining the change in score between the time of initial administration and the one, four, and eight-week administrations. Effect sizes of 0.44, 0.94, and 1.26 were calculated at one, four, and eight weeks, respectively. The effect size determined eight weeks after initiation of physical therapy is high, and comparable to the effect size calculated for the KOOS subscale sport and recreation function six months after ACL reconstruction.

Comparison of the KOOS to outcome measures for knee OA

Reliability and validity

In a literature review, Sun et al. (1997) studied the reliability and validity of commonly used clinical scores for assessment of hip and knee osteoarthritis. Reliability studies were only reported for 6 of the 45 identified clinical rating scores. It was concluded that relatively high reliability was reported for most measures of pain, stiffness, and physical function, while results were less conclusive for clinical signs. These results were supported by the study of Ryd et al. (1997) where 10 experienced orthopedic surgeons assessed 15 patients with knee OA using 3

commonly used composite scoring systems (HSS, KSS and the Venn diagram scoring system (Jónsson 1981)). It was shown that clinical measurements are not robust and that composite knee scores are exceedingly unreliable. The test-retest reliability and the internal consistency of the KOOS has been found high, and is comparable to the reliability of other self-administered patient-relevant outcome measures.

Validity studies were reported for 15 scores, and had been more comprehensively studied for the Lequesne Index, the WOMAC, and the ILAS (Lower Extremity Assistance Scale of the University of Iowa (Shields et al. 1995)), and these scores have shown satisfactory responsiveness to different treatment effects. The 12-item questionnaire for patients having total knee replacement is a new questionnaire assessing pain, symptoms and function in one score (Dawson et al. 1998). The authors claim the advantages of the questionnaire being that it is short, simple, and validated. However, the validity and sensitivity was proven in comparison with SF-36, the HAQ, and the American Knee Society Score. The qualities of the 12-item questionnaire in comparison to the WOMAC, the currently most used outcome measure for knee OA, remains unknown.

Comparison of the Lequesne Index and the WOMAC

The Lequesne index of severity for knee OA is comparable to the Lysholm score for knee injury with regard to type of questions asked, administration mode, and format. Hence, the same disadvantages with doubtful patient-relevancy, observer bias, and score aggregation apply to the Lequesne index as to the Lysholm score.

The WOMAC differs from the Lequesne index in several aspects, one of them being the format used for questions and answer options. How the questions are phrased, and which answer options that are given, affect the patient-relevancy of a questionnaire. Typically in observer-administered scores, like the Lequesne, the questions and the grading of the answer options are based on how knee function and reduction in knee function are defined by the constructor of the questionnaires. However, depending on actual and desired activity level of a specific subject, the same increased knee stiffness or decreased walking ability might be regarded as more or less relevant by the patient. By asking the patient how much difficulty they experience with each task, as in WOMAC, instead of using answer options like " more than 1 but less than 15 minutes of stiffness" or "maximum distance walked from 500 to 900 meters", the answers become patient-relevant.

The Lequesne Index and the WOMAC are also used for hip OA. Stucki et al. (1998) compared the clinimetric properties of the Lequesne Index and the WOMAC in 51 patients before hip arthroplasty. Both instruments had sufficient test-retest reliability, and the WOMAC proved to be valid. The Lequesne Index was however, not internally consistent (indicating the items measuring several aspects) and only weakly correlated to radiographic OA and range of motion. It was concluded that the Lequesne Index should be used cautiously before addi-

tional testing of its metric properties and its validity. Although hip and knee OA are two disparate diseases, the Lequesne Index for hip OA and the Lequesne Index for knee OA are developed in similar processes. Thus the same cautious use of the Lequesne Index for knee OA can be recommended.

Score format

Normalized scores ranging from 0 (worst score) to 100 are calculated for the KOOS. For the 12-item questionnaire the score ranges from 12 (worst score) to 60, and for the original version of WOMAC the scores range from 0 (best score) to 20, 0 to 8, and 0 to 68, for the three subscales pain, stiffness, and function. This implies that a score of 25 should be interpreted quite differently in these three scoring systems. The two reasons for transforming the KOOS raw scores into percentage scores were that interpretation should be enhanced for the clinician or researcher, and uniformity with other knee scoring scales. Studying the scoring format used for the WOMAC and the 12-item questionnaire supports our decision. For the WOMAC, comparisons between subscales are not meaningful, and for the 12-item questionnaire a best possible score of 12 seem not characteristic. In recent studies employing the WOMAC, scores have been normalized to a 0 to 100 scale (Creamer et al. 1998, Stucki et al. 1998).

Comparison of the KOOS to generic outcome measures

Responsiveness

The SF-36 is a generic instrument assessing health status. The disadvantage of generic instruments is the generally lower responsiveness for specific conditions. When used for ACL injury, the SF-36 showed changes, however non-significant, with treatment (surgical and non-surgical) over time (Shapiro et al. 1996). Since only score changes and no standard deviations were given in this paper, effect sizes could not be calculated. In the KOOS study of ACL reconstruction (paper IV), the effect size for the SF-36 subscale bodily pain was comparable to the KOOS subscale pain 6 months after ACL reconstruction. For the SF-36 subscale physical function, a more relevant outcome than pain following ACL injury, the effect size 6 months after ACL reconstruction was 0.84 compared to 0.94 for the KOOS subscale ADL and 1.16 for the KOOS subscale sport and recreational function. The most responsive subscale 6 months after ACL reconstruction was the KOOS subscale knee-related quality of life with an effect size of 1.65.

The SF-36 has also been used in a retrospective study on meniscus injury (Katz et al. 1992). The SF-36 was administered over the phone at an average 18.5 months after arthroscopic partial meniscectomy, and an effect size of 1.36 could be calculated for the SF-36 subscale physical function. Three months after ar-

throscopic partial meniscectomy (paper VI), effect sizes of 0.57, 0.66, and 0.80 were found for the SF-36 subscales physical function, role-physical, and bodily pain. These effect sizes should be compared to 1.11 for the KOOS subscale pain, 0.67 for the subscale ADL, and 0.90 for the subscale sport and recreation function. Also in this study the highest effect size, 1.15, was seen for the KOOS subscale knee-related quality of life. Generally the disease-specific KOOS was more responsive than the generic SF-36. However generic measures can be used across conditions and assess outcomes related to other constructs, and thus the recommendation to include both generic and disease-specific outcome measures in clinical trials seem adequate (Small et al. 1994, Hawker et al. 1995, Altman et al. 1996, Bellamy et al. 1997).

Assessment of groups and individuals

Outcome measures can be used for research where conclusions will be drawn based on mean scores, or to monitor progress of an individual. The acceptable degree of reliability for these different circumstances is not definite. One suggestion for acceptable test-retest reliability for assessment of an individual is intraclass correlation coefficients of 0.85 (Weiner and Stewart 1984). In studies on the Patient-Specific Functional Scale (PSFS), intraclass correlation coefficients of 0.87 and 0.91 were found satisfactory for monitoring individuals with knee dysfunction and neck dysfunction (Chatman et al. 1997, Westaway et al. 1998). The Patient-Specific Functional Scale (PSFS) is primarily designed for monitoring individuals, and seem to be ideal for quality assurance in clinical settings where a wide variety of conditions are treated, such as a physical therapy practice. When comparing groups, a lower intraclass correlation coefficient is likely acceptable and a limit of 0.75 has been suggested (Rosner 1995). The intraclass correlation coefficients of the KOOS ranged from 0.75 to 0.93 in the American English validation study (ACL reconstruction) and from 0.78 to 0.91 in the Swedish validation study (arthroscopy). For all subscales, but sport and recreation function, the reliability coefficient in either of the validation studies exceeded the suggested limit of 0.85 for assessment of individuals. The subscale ADL had the lowest reliability in the American validation study and the highest reliability in the Swedish validation study. The difference in reliability between the two studies is most likely due to the different populations studied. The subscale sport and recreational function had intraclass correlation coefficients determined in both validation studies below the suggested minimum reliability for assessment on an individual level. The obtained intraclass correlation coefficients were 0.81 and 0.78. Despite this finding for one of the five subscales, we suggest that the KOOS can be used to monitor individuals, as well as groups.

Clinical applications

Short-term results of meniscectomy

Much effort was placed into the development and evaluation of the KOOS. However, to complete the process, clinical data should be obtained. To our knowledge, this was the first prospective study of patient-relevant outcomes, using a validated score, following arthroscopic partial meniscectomy.

Meniscectomy is considered a successful knee operation with total recovery expected within 4-6 weeks (Patel et al. 1982, Hamberg and Gillquist 1984, Zarins et al. 1985, Vander Schilten 1990, St Pierre 1995). However, this depends on what outcomes that are measured. Pain and mechanical symptoms were alleviated to a high extent after 3 months, but the patients still experienced major difficulties with knee demanding functions and knee-related quality of life. This is well known from experience in physical therapy, but would not have been possible to prove scientifically without outcome instruments measuring at disability and handicap level.

Additionally, factors that influence postoperative knee-related QOL were investigated. None of the demographic variables examined (age, BMI, gender, duration of problems, or pre-injury work level) influenced the reported postoperative knee-related QOL. Involvement of the lateral meniscus was associated with more residual symptoms, but this did not affect function or quality of life negatively. The only factors, of those previously reported, that predicted postoperative quality of life were associated intra-articular injuries. However, these measures of impairment accounted for only 8 % of the variance. This confirms that objective measures of impairment are not closely related to subjective measures of disability and handicap.

Patient-relevant outcomes provide additional information and should be assessed after arthroscopic partial meniscectomy. It is suggested that the preoperative information to the meniscectomy patient should include the realistic expected functional outcome. Additionally, it could be speculated whether or not patients with more realistic expectations on patient-relevant outcomes will be more satisfied with the outcome.

Long-term results of meniscectomy

20 years after meniscectomy, the former patients reported more pain and functional limitations than did controls. This was also true when patients who had developed radiographic signs of OA were removed from both the patient group and the control group, indicating that meniscectomy in itself is associated with pain, other symptoms, and functional limitations. It has to be considered however, that radiography is an insensitive measure of cartilage damage of the knee. Early changes of OA, which are not detected by radiography, can be seen at arthroscopy (Lysholm et al. 1987).

Despite the rate of radiographic OA not being higher among women than men in the study group, meniscectomized women reported worse outcome than

meniscectomized men did. This is in concordance with Hede et al. (1992) who reported a significantly lower Lysholm score for women with total meniscectomy than for men with total meniscectomy. It could be speculated that women, regardless of knee injury, would report worse knee-related outcome because of anatomical reasons, more valgus alignment, lesser muscle strength, or increased laxity. However, among the controls no differences were seen because of gender, indicating meniscectomy having a more adverse effect on women than on men.

While self-reported symptoms and functional limitations were not affected by age, worse outcome of the performance tests was seen with increasing age. Older subjects performed worse, both in the meniscectomized group and in the control group. Worse physical performance of other, more general, physical tests and worse health status was seen with increasing age in a Swedish study investigating life style, physical performance, and health status in more than 2 000 Swedish men and women in the ages from 20 to 65 (Engström et al. 1993). Together, these findings indicate that separate reference values for physical performance tests are needed for different age groups, both in a general population and in follow-up studies of knee injury. Before recommending the three performance tests we have used, further evaluation is necessary.

In this study the Flandry questionnaire was used to obtain self-reported symptoms and functional limitations. When data from the Flandry questionnaire is presented in four subscales (pain, symptoms, activities of daily living, and sport and recreation function) it can be interpreted similar to KOOS data. When the KOOS questionnaire became available, approximately 20 months after the index follow-up, we obtained KOOS data by mail for 87 of the 97 subjects operated in 1973. When comparing the mean Flandry scores to the mean KOOS scores for the 87 subjects assessed by both questionnaires, reasonable agreement was found. The difference in mean scores ranged from 4 to 8 points. The Flandry pain score was lower than the KOOS pain score, while the Flandry scores for symptoms, ADL, and sport where higher than the KOOS scores, indicating the additional items included in the KOOS causing the difference rather than deterioration in status over the 20 month time period.

The results of our study confirmed the previously reported weak relationship between radiographic OA and pain (Hadler 1992, Lethbridge-Cejku et al. 1995, Cicuttini et al. 1996), indicating the importance of using outcome instruments measuring at different levels.

The meniscectomies carried out in the seventies were total or subtotal. The loss of a meniscus is associated not only with less protection for the cartilage but also with increased laxity. In addition, the postoperative care during the 70's included immobilization and reduced weight bearing for weeks. All these factors could contribute to more pain and functional limitations, suggesting that a more favorable outcome after arthroscopic partial meniscectomy may be possible.

The control group

Exclusion criteria of the control group were diagnoses of meniscal injury, ACL injury, or radiographic OA of the tibio-femoral joint. However, subjects with a diagnosis of patellofemoral problems or other repetitive-motion disorders were not excluded, yielding a group where knee symptoms and functional limitations would be present. In addition, some selection bias among the control subjects may have occurred, since a considerable proportion of control subjects (21 %) declined to answer the initial questionnaire. We suggest that control subjects who had knee complaints would have been more motivated to participate in the study than those without, possibly generating a falsely high proportion of control subjects with knee complaints. As shown in paper I, the reference group had mean KOOS scores ranging from 88 to 98, all medians being 100. It seems like minor symptoms and functional limitations related to the knee should be considered normal. Demirdjian et al. (Demirdjian et al. 1998) administered the Cincinnati and the Lysholm knee scoring questionnaires to over 400 young and knee healthy athletes and found scores ranging from 68 to 100. The mean scores were 99 for men and 97 for women. The 95 % confidence interval computed for either questionnaire did not contain the maximal value of 100. We have used the KOOS reference group for comparison to groups of patients and found significant differences. If all subjects reporting a history of knee problems were excluded from the reference group, these differences would have been even greater.

What outcomes to measure, and what instrument to choose?

What measure to use to assess the outcome of knee injury and knee OA depends on the goal of the investigator. If the goal is to evaluate operating technique, laxity is an important outcome. If the goal is to evaluate cartilage transplants, magnetic resonance imaging, radiographs, and arthroscopic findings are important outcomes. However, in these cases, as in all other clinical situations, the knee is a part of an individual who experiences disturbing symptoms and functional limitations and thus seeks medical care. Patient-relevant outcomes should, therefore, be considered paramount in assessing response to surgery, physical therapy, or other treatment.

We suggest that such an outcome measure should:

- be developed in a process including patient interviews to ensure content validity,
- be self-administered to avoid bias,
- report the outcomes related to impairment, disability, and handicap separately.

High content validity ensures patient-relevancy and high sensitivity. An instrument able to track severe difficulties holds the possibility of larger changes with treatment. In the current study, outcomes related to disability and handicap were especially sensitive and responsive. Inclusion of measures reporting the

scores of outcome related to impairment, disability, and handicap separately helps interpretation and increases responsiveness, thus indicating fewer patients needed in clinical studies to detect significant changes.

The self-administered design of measures such as the KOOS makes them suitable for collection of data in large studies, such as registers of ACL reconstruction, knee arthroplasty, etc. Collection of data by mail is associated with markedly lower costs compared to clinical examination. Another advantage of using patient-relevant questionnaires compared to clinical examination is their value in predicting outcome. Britton et al. (Britton et al. 1997) in a study of total hip arthroplasty, found pain level to be the best predictor of revision. In addition, when using pain level as end-point instead of revision, they found differences between the implants used that were not detected else.

In clinical studies, the disease-specific outcome measure should be used together with a generic measure. This allows comparison across diagnoses. However, the disease-specific instrument can not be replaced by a generic measure since the sensitivity, validity, and responsiveness of generic measures is lower compared to disease-specific instruments.

Perspectives for the future include testing the relevancy and the metric properties of the KOOS for other diagnoses or groups of patients, such as patients qualified for total knee replacement and total hip replacement, or patients with instability of the ankle joint.

Questionnaires such as the KOOS adds information and help interpret the outcome of clinical trials. Currently, three questionnaires developed and evaluated according to set guidelines are available for assessment of patients with knee injury and early OA. Time will show if any of these measures will be the gold standard for assessment of patient-relevant outcomes. It is obvious, however, that patient-relevant outcome measures always should be included, and be considered the primary outcome measure, in clinical trials of knee injury and knee OA.

Conclusions

- The KOOS proved to be a patient-relevant outcome measure, valid for subjects with ACL injury, meniscus injury, and post-traumatic OA.
- The KOOS was test-retest reliable and internally consistent, and can be used for assessment of groups or monitoring individuals.
- The KOOS was responsive to ACL reconstruction, physical therapy, and arthroscopic partial meniscectomy.
- The KOOS is available for use in Sweden and United States.
- The WOMAC is available in a validated Swedish version.
- The KOOS was more responsive than the WOMAC in the investigated study groups.
- Three months after arthroscopic partial meniscectomy significant improvement was seen in patient-related outcomes, but substantial disability and handicap was still reported compared to reference groups.
- Twenty years after meniscectomy symptoms and functional limitations were reported, both in subjects with and without radiographic OA.
- Meniscectomized women reported worse symptoms and functional limitations than meniscectomized men, and compared to female controls
- The relation between pain and radiographic OA was poor, 40% of subjects with more severe radiographic OA did not report more pain than controls did.

Abstract

The overall purpose was to evaluate patient-relevant outcomes in patients with knee injury and post-traumatic osteoarthritis of the knee.

A self-administered questionnaire, The Knee injury and Osteoarthritis Outcome Score (KOOS) was developed by literature search, consulting an expert panel (including patients referred to physical therapy because of knee-related problems, orthopedic surgeons, and physical therapists both from Sweden and the USA), and a pilot study. Two validation studies were carried out, 21 patients were studied following reconstruction of the anterior cruciate ligament, and 142 patients were studied following knee arthroscopy. The KOOS proved to be a user-friendly and patient-relevant outcome measure with satisfactory metric properties. The KOOS was found reliable for assessment of groups and monitoring individuals. The KOOS was responsive to anterior cruciate ligament (ACL) reconstruction, physical therapy, and arthroscopic partial meniscectomy. The KOOS is available in two validated versions, for use in Sweden and the USA.

The Western Ontario and MacMaster Universities Osteoarthritis Index (WOMAC) is included in the KOOS. The WOMAC was translated into Swedish in three separate processes. A committee agreed on a mutual version that was tested for reliability and responsiveness in 52 subjects with arthroscopic knee osteoarthritis (OA) undergoing arthroscopy. The reliability was found satisfactory and comparable to the original version. Significant improvement was seen three months after arthroscopy.

The KOOS was more responsive than the WOMAC when evaluating subjects 20 years after meniscectomy. It is recommended that the KOOS subscales sport and recreation function and knee-related quality of life be added to the WOMAC when assessing post-traumatic OA.

Three months after arthroscopic partial meniscectomy significant improvement was seen in patient-related outcomes, but substantial disability and handicap was still reported compared to reference groups.

Twenty years after meniscectomy symptoms and functional limitations were reported, both in subjects with and without radiographic OA. Meniscectomized women reported worse symptoms and functional limitations than meniscectomized men, and compared to female controls. The relation between pain and radiographic OA was poor, 40 % of subjects with more severe OA did not report more pain than controls did.

It is concluded that patient-relevant questionnaires, such as the KOOS, adds information and should be used, and be considered the primary outcome measure, in clinical trials of knee injury and knee OA.

Summary in Swedish

KNÄSKADA OCH KNÄLEDSFÖRSLITNING – ur patientens perspektiv

Knäledsförslitning är vanligt. Ledsjukdom, med smärta och nedsatt funktion som följd, är den vanligaste kroniska sjukdomen hos äldre, vanligare än högt blodtryck, hjärtsjukdom och diabetes. Behandlingen av knäledsförslitning är fokuserad på smärtlindring. Till en början används vanligt mediciner mot smärtan, men senare i sjukdomsförloppet byts den smärtande leden ofta ut mot en knäprotes. I Sverige sätts årligen 5 000 knäproteser in, och i USA är siffran 250 000 per år.

Knäledsförslitning är vanligt också i yngre åldrar. I en svensk studie visade sig en och en halv procent av personer mellan 35 och 54 år ha förslitning som man kunde se på röntgen. I denna undersökning ingick inte personer som tidigare skadat sina knän. Knäskada är en känd riskfaktor för tidig knäledsförslitning. Ungefär hälften av de som skadar korsband eller menisker i knät har tecken på förslitning i knäleden efter 10-15 år. Eftersom knäskador är vanliga i yngre åldrar kan man beräkna att totalt sett har ca 5 % av befolkningen mellan 35 och 54 år förslitning i knäleden som syns på röntgen. Med stor sannolikhet utvecklar dessa personer förr eller senare symptom av röntgenförändringarna. Behandlingen som erbjuds dessa "unga personer med gamla knän" är huvudsakligen inriktad på smärtlindring, eftersom ingen behandling hittills visat sig kunna påverka sjukdomsförloppet vid knäledsförslitning. Forskning pågår i många olika riktningar för att kunna erbjuda dessa personer bättre behandling. Lovande resultat med minskad smärta och ökad funktion har setts efter muskelträning och konditionsträning. För att på ett pålitligt, giltigt och känsligt sätt utvärdera patientens syn på symptom och funktion i behandlingsstudier behövs relevanta utvärderingsinstrument.

Syfte

Att med hjälp av väl standardiserade utvärderingsinstrument med god pålitlighet, giltighet och känslighet för förändring, utvärdera patientens syn på symptom, funktion och livskvalitet efter främre korsbandsskada, meniskskada och vid knäledsförslitning tidigt i livet.

Resultat

För att täcka in de områden som är viktiga för patienter med knäskada och knäledsförslitning utvecklades ett frågeformulär, Knee Injury and Osteoarthritis Outcome Score (KOOS) som separat mäter smärta, övriga symptom, dagliga livets aktiviteter, sport- och fritidsfunktion och knä-relaterad livskvalitet. KOOS

är en utvidgning av WOMAC Osteoarthritis Index, ett frågeformulär som mäter smärta, stelhet och påverkan på dagliga livets göromål hos äldre med knäledsförslitning. Både WOMAC och KOOS, som finns i svensk, engelsk och dansk version, har testats och befunnits pålitliga såtillvida att man får samma svar om formuläret fylls i vid två tillfällen med en veckas mellanrum. Som jämförelse är pålitligheten bättre än för tolkning av förslitning på röntgenbilder. I jämförelse med andra frågeformulär har det visats att KOOS fem delskalor verkligen mäter de begrepp de avser mäta. Då patienter som genomgick korsbandsrekonstruktion studerades, speglade gruppens medelpoäng för de fem delskalorna mycket väl de olika rehabiliteringsfaserna. Först efter 2 år uppnådde någon patient full poäng i alla delskalor.

Vid meniskkirurgi, som är ett rutiningrepp med kort sjukskrivningstid och kort förväntad rehabiliteringstid, visade det sig att även om patienterna bara hade ringa smärta, symptom och problem med dagliga livets aktiviteter 3 månader efter ingreppet så hade de avsevärda problem med sport- och fritidsfunktion och den livskvalitet som är beroende av knäfunktion. Som en följd av detta rapporterades också en rejäl sänkning av aktivitetsnivån, 3 månader efter ingreppet var endast 33 % idrottsaktiva jämfört med 70 % före skadan. 38 % rapporterade en fysiskt helt inaktiv livsstil jämfört med 7 % före skadan. Korttidsresultaten vid meniskkirurgi stämmer inte överens med sjukvårdens gängse uppfattning.

Personer som opererats för en meniskskada för 20 år sen rapporterar mer smärta, symptom och funktionsnedsättningar än vad andra personer av samma kön och ålder gör. Detta gäller oavsett om personerna utvecklat förslitning som syntes på röntgen eller ej. Opererade kvinnor rapporterade mer symptom och funktionsnedsättningar än opererade män och jämfört med friska kvinnliga kontroller. Knappt hälften av de som hade förslitning som syntes på röntgen hade mer ont än vad kontroller som inte opererats eller hade förslitning hade, dvs sambandet mellan förslitning som syns på röntgen och smärta är svagt.

Patientens syn på smärta, funktion och livskvalitet stämmer dåligt överens med objektiva mått på knäskada. Studier av symptom, funktion och livskvalitet bör leda till en mer realistisk patient-information inför operativa ingrepp, att ytterligare studier genomförs för att undersöka om knäbehandling kan förbättras utifrån patientens perspektiv, och till att patientens åsikt tillmäts större betydelse vid tolkning av resultaten i knästudier.

Acknowledgements

I wish to express my sincere gratitude to the following persons:

Stefan Lohmander, my mentor, for always being supportive, always being possible to reach (except when going to Antarctica), always giving constructive criticism, always asking the right questions, sharing his knowledge, and for accepting me as his graduate student, the second from the Roos family.

Charlotte Ekdahl, my mentor, for invaluable support, stimulating discussions, and good laughs.

Ulrich Moritz, professor emeritus, for introducing me into science.

Nils Westlin, sports medicine legend, who willingly let me auscultate every Friday in the late 1970's and introduced me into the field of knee injury and sports medicine.

All co-workers at Sports Medicine Center, Helsingborg, for supporting my interest in research instead of manual treatment of patients, and especially to Anna Östenberg, co-author, and Annette von Porat for stimulating lunch discussions and for, bit by bit, sharing my interest in research.

Bruce Beynnon, co-author and director of research at Department of Orthopaedics and Rehabilitation, University of Vermont, who without hesitation let me take part in the ongoing research activities during our year in Vermont and also let me share a biomechanists view on patient-related outcomes.

Per Renström, professor of Sports Medicine in Stockholm, who lived in Vermont for ten years and besides being the reason for us going there, together with his wife Lena made our life brighter.

Maria Klässbo, co-author, physical therapist and graduate student in Säffle, for sharing my interest in patient-relevant outcomes and for stimulating discussions of research and life over the phone and by e-mail.

Nina Beyer, research physical therapist at Bispebjerg Hospital in Copenhagen, Denmark for enthusiastically translating KOOS into Danish and discussions of the differences between Danish and Swedish everyday tasks.

All graduate students at the Departments of Physical Therapy and Orthopaedics, for stimulating discussions, revisions of manuscripts, and sharing the struggle for grants and publications.

Leif Ryd, co-author, Department of Orthopedics for stimulating discussions on what to assess and how to do it.

All personnel at the Departments of Orthopedics and Physical Therapy in Lund.

Ulf Strömberg, Dept. of Environmental Medicine, for excellent statistical advice and for always knowing what I wanted, even when I wasn't sure myself.

Board-members of the Swedish Society of Sports Medicine. Once being one of you has provided me with a network of people always willing to share their enthusiasm and knowledge.

Harald, Karl, and Ebba, my family, for EVERYTHING!

Clarie Ek, my mother, for being my biggest supporter and for always being there when the children or I need you.

All other relatives, friends and neighbors who have been supportive and taken good care of Karl and Ebba when I, physically or mentally, were elsewhere.

References

Ahlbäck S. Osteoarthritis of the knee: a radiographic investigation. Acta Radiol 1968; (Suppl 277): 7-72.

Altman D G. *Practical statistics for medical research*. Chapman & Hall, London 1991.

Altman R, Brandt K, Hochberg M, Moskowitz R. Design and conduct of clinical trials in patients with osteoarthritis: Recommendations from a task force of the Osteoarthritis Research Society. Osteoarthritis Cartilage 1996; 4: 217-243.

Altman R D, Hochberg M, Murphy W A, Wolfe F, Lequesne M. Atlas of individual radiographic features in osteoarthritis. Osteoarthritis Cartilage 1995; 3(Suppl A): 3-70.

Amadio P C. Editorial, Outcome Measurements. J Bone Joint Surg (Am) 1993; 75-A(11): 1583-1584.

Barber S, Noyes F, Mangine R, McCloskey J, Hartman W. Quantitative assessment of functional limitations in normal and anterior cruciate deficient knees. Clin Orthop 1990; 255: 204-214.

Bellamy N. *Musculoskeletal Clinical Metrology*. Kluwer Academic Publishers, Dordrecht 1993.

Bellamy N. *WOMAC Osteoarthritis User´s Guide*. Victoria Hospital, London, Ontario, Canada 1995.

Bellamy N, Buchanan W, Goldsmith C. Validation study of WOMAC: A health status instrument for measuring clinically important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. Orthop Rheumat 1988a; 1: 95-108.

Bellamy N, Kirwan J, Boers M, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997; 24(4): 799-802.

Bellamy N, Watson Buchanan W, Goldsmith C H, Campbell J, Stitt L W. 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 the knee. J Rheumatol 1988b; 15: 1833-1840.

Bengtsson J, Möllborg J, Werner S. A study for testing the sensitivity and reliability of the Lysholm knee scoring scale. Knee Surgery, Sports Traumatology and Arthroscopy 1996; 4: 27-31.

Bergner M, Bobbitt R, Pollard W, Martin D, Gilson B. The sickness impact profile: validation of a health status measure. Med Care 1976; 14: 57-67.

Bland M. *An introduction to medical statistics*. Oxford University Press, Oxford 1995.

Bolano L E, Grana W A. Isolated arthroscopic partial meniscectomy. Functional radiographic evaluation at five years. Am J Sports Med 1993; 21(3): 432-437.

Bollen S, Seedholm B B. A comparison of the Lysholm and Cincinnati knee scoring questionnaires. Am J Sports Med 1991; 19(2): 189-190.

Britton A R, Murray D W, Bulstrode C J, McPherson K, Denham R A. Pain levels after total hip replacement. Their use as endpoints for survival analysis. J Bone Joint Surg (Br) 1997; 79-B(1): 93-98.

Buhl-Nielsen A. The epidemiology of anterior cruciate ligament injuries in athletes. Acta Orthop Scand 1991; 62(Suppl 243): 13.

Campbell M J, Machin D. *Medical statistics: a commonsense approach*. John Wiley & Sons, Chichester 1993.

Cerabona F, Sherman M F, Bonamo J R, Sklar J. Patterns of meniscal injury with acute anterior cruciate ligament tears. Am J Sports Med 1988; 16(6): 603-609.

Chatman A B, Hyams S P, Neel J M, Stratford P W, Schomberg A, Stabler M. The patient-specific functional scale: measurement properties in patients with knee dysfunction. Phys Ther 1997; 88(8): 820-829.

Cicuttini F M, Baker J, Hart D J, Spector T D. Association of pain with radiological changes in different compartments and views of the knee joint. Osteoarthritis Cartilage 1996; 4: 143-147.

Clancy C M, Eisenberg J M. Outcomes research: Measuring the end results of health care. Science 1998; 282: 245-246.

Cohen J. *Statistical Power Analysis for the Behavioural Sciences*. Academic Press, New York 1977.

Creamer P, Lethbridge-Cejku M, Hochberg M C. Where does it hurt? Pain localization in osteoarthritis of the knee. Osteoarthritis Cartilage 1998; 6: 318-323.

Cronbach L. Coefficient alpha and the internal structure of tests. Psychometrica 1951; 16: 297-334.

Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. J Bone Joint Surg (Br) 1998; 80-B(1): 63-69.

Demirdjian A, Petrie S, Guanche C, Thomas K. The outcomes of two knee scoring questionnaires in a normal population. Am J Sports Med 1998; 26(1): 46-51.

Ekdahl C, Andersson S I, Svensson B. Muscle function of the lower extremities in rheumatoid arthritis and osteoarthritis. A descriptive study of patients in a primary health care district. J Clin Epidemiol 1989; 42: 947-954.

Engelberg R, Martin D P, Agel J, Obremsky W, Coronado G, Swiontkowski M F. Musculoskeletal Function Assessment Instrument: criterion and construct validity. J Orthop Res 1996; 14: 182-192.

Engström L-M, Ekblom B, Forsberg A, von Koch M, Seger J. *Livstil, prestation, hälsa. LIV 90. Rapport 1*. Folksam, Högskolan för lärarutbildning, Idrottshögskolan, Karolinska institutet, Korpen, Riksidrottsförbundet, Stockholm 1993 [in Swedish].

Ewald F C. The knee society total knee arthroplasty roentgenographic evaluation and scoring system. Clin Orthop 1989; 248: 9-12.

Ferkel R D, Davis J R, Friedman M J, et al. Arthroscopic partial meniscectomy: an analysis of unsatisfactory results. Arthroscopy 1985; 1(1): 44-52.

Flandry F, Hunt J P, Terry G C, Hughston J C. Analysis of subjective knee complaints using visual analog scales. Am J Sports Med 1991; 19(2): 112-118.

Griffiths G, Bellamy N, Kean W, Campbell J, Gerez-Simon E. A study on the time frame dependency of responses to the WOMAC osteoarthritis index. Inflammopharmocology 1993; 2: 85-87.

Guillemin F. Cross-cultural adaptation and validation of health status measures. Scand J Rheumatol 1995; (24): 61-63.

Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptations of health related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993; 46: 1417-1432.

Hadler N M. Knee pain is the malady-not osteoarthritis (ed.). Ann Intern Med 1992; 116: 598-599.

Hamberg P, Gillquist J. Knee function after arthroscopic meniscectomy. Acta Orthop Scand 1984; 55: 172-175.

Hamberg P, Gillquist J, Lysholm J. A comparison between arthroscopic meniscectomy and modified open meniscectomy. J Bone Joint Surg (Br) 1984; 66-B(2): 189-192.

HANES 1. Basic data on osteoarthritis of the knee, hip and sacroiliac joints in adults aged 25-74 years. USA 1971-75. Series 11(213). National Center for Health Statistics, 1979.

Havlik R, et al. Health statistics on older persons. Vital and health statistics. Series 3, No. 25, Table 17. National Center for Health Statistics, USA, 1986.

Hawker G, Melfi C, Paul J, Green R, Bombardier C. Comparison of a generic (SF-36) and a disease specific (WOMAC) instrument in the measurement of outcomes after knee replacement surgery. J Rheumatol 1995; 22: 1193-1196.

Hede A, Jensen D, Blyme P, Sonne-Holm S. Epidemiology of meniscal lesions in the knee. 1.215 open operations in Copenhagen. Acta Orthop Scand 1990; 61: 435-437.

Hede A, Larsen E, Sandberg H. Partial versus total meniscectomy. A prospective, randomised study with long-term follow-up. J Bone Joint Surg (Br) 1992; 74-B(1): 118-121.

Hefti F, Muller W, Jakob R P, Stubli H U. Evaluation of knee ligament injuries with the IKDC form. Knee Surg Sports Traumatol Arthroscopy 1993; 1(3-4): 226-34.

Hernborg J S, Nilsson B E. Age and sex incidence of osteophytes in the knee joint. Acta Orthop Belg 1973; 44: 66-68.

Hicks C M. *Research for physiotherapists: project design and analysis*. Churchill Livingstone, London 1995.

Höher J, Bach T, Münster A, Bouillon B, Tiling T. Does the mode of data collection change results in a subjective knee score? Am J Sports Med 1997; 25(5): 642-647.

Höher J, Munster A, Klein J, Eypasch E, Tiling T. Validation and application of a subjective knee questionnaire. Knee Surg Sports Traumatol Arthroscopy 1995; 3: 26-33.

Insall J N, Dorr L D, Scott R D, Scott W N. Rationale of the knee society clinical rating system. Clin Orthop 1989; 248: 13-14.

Irrgang J J, Ho H, Harner C D, Fu F H. Use of the International Knee Documentation Committee guidelines to assess outcome following anterior cruciate ligament reconstruction. Knee Surg Sports Traumatol Arthroscopy 1998a; 6: 107-114.

Irrgang J J, Snyder-Mackler L, Wannier R, Fu F, Harner C. Development of a patient-reported measure of function of the knee. J Bone Joint Surg (Am) 1998b; 80(8): 1132-1145.

Jaureguito J, Elliot J, Lietner T, Dixon L, Reider B. The effects of arthroscopic partial meniscectomy in an otherwise normal knee: a retrospective review of functional, clinical and radiographic results. Arthroscopy 1995; 11(1): 29-36.

Johnson R J. Outcomes research in AOSSM. Am J Sports Med 1994; 22(6): 734-738.

Jónsson G T. Compartmental arthroplasty for gonarthrosis. Acta Orthop Scand 1981; 52(Suppl 193): 3-109.

Katz J N, Harris T M, Larson M G, et al. Predictors of functional outcomes after arthroscopic partial meniscectomy. J Rheumatol 1992; 19(12): 1938-1942.

Kazis L, Anderson J, Meenan R. Effect sizes for interpreting changes in health status. Med Care 1989; 27(suppl): S178-S189.

Kellgren J, Lawrence J. Radiological assessment of osteoarthrosis. Ann Rheum Dis 1957; 16: 494-502.

Kellgren J H, Lawrence J S. *The Epidemiology of Chronic Rheumatism. Atlas of Standard Radiographs*. Blackwell Scientific, Oxford 1963.

Kirkley A, Griffin S, McLintock H, Ng L. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability. The Western Ontario Shoulder Instability Index (WOSI). Am J Sports Med 1998; 26(6): 764-772.

Knutson K, Lewold S, Robertsson O, Lidgren L. The Swedish knee arthroplasty register. A nation-wide study of 30 003 knees 1976-1992. Acta Orthop Scand 1994; 65: 375-386.

Labs K, Paul B. To compare and contrast the various evaluation scoring systems after anterior cruciate ligament reconstruction. Arch Orthop Trauma Surg 1997; 116: 92-96.

Lequesne M. Informational indices. Validation of criteria and tests. Scand J Rheumatol 1989; Suppl 80: 17-27.

Lethbridge-Cejku M, Scott W W, Reichle R, et al. Association of radiographic features of osteoarthritis of the knee with knee pain: Data from the Baltimore longitudinal study of aging. Arthritis Care and Research 1995; 8(3): 182-189.

Liang M, Jette A M. Measuring functional ability in chronic arthritis. Arthritis Rheum 1981; 24(1): 80-86.

Liang M, Larson M, Cullen K, Schwartz J. Comparative measurement of efficiency and sensitivity of five health status instruments for arthritis research. Arthritis Rheum 1985; 28(5): 542-547.

Lieberman J R, Dorey F, Shekelle P, et al. Differences between patients' and physicians' evaluation of outcome after total hip arthroplasty. J Bone Joint Surg (Am) 1996; 78-A(6): 835-838.

Lohmander L S, Roos H P. Knee injury, surgery and osteoarthrosis. Acta Orthop Scand 1994; 65: 605-609.

Lysholm J, Gillquist J. Endoscopic meniscectomy. Int Orthop 1981; 5: 265-270.

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: 150-154.

Lysholm J, Hamberg P, Gillquist J. The correlation between osteoarthrosis as seen on radiographs and on arthroscopy. Arthroscopy 1987; 3(3): 161-165.

Maletius W, Messner K. The effect of partial meniscectomy on the long-term prognosis of knees with localized, severe chondral damage. Am J Sports Med 1996; 24(3): 258-262.

Martin D P, Engelberg R, Agel J, Snapp D, Swiontkowski M F. Development of a musculoskeletal extremity health status instrument: the musculoskeletal function assessment instrument. J Orthop Res 1996; 14: 173-181.

Martin D P, Engelberg R, Agel J, Swiontkowski M F. Comparison of the Musculoskeletal Function Assessment Questionnaire with the Short-Form 36, The Western Ontario and McMaster Universities Osteoarthritis Index, and the Sickness Impact Profile health-status measures. J Bone Joint Surg (Am) 1997; 79-A(9): 1323-1335.

McGrory B J, Morrey B F, Rand J A, Ilstrup D M. Correlation of patient questionnaire responses and physician history in grading clinical outcome following hip and knee arthroplasty. The Journal of Arthroplasty 1996; 11(1): 47-57.

Mohtadi N. Quality of life assessment as an outcome in anterior cruciate ligament reconstructive surgery. In: *The Anterior Cruciate Ligament Current and Future Concepts*. (Ed. Jackson D) Raven Press, New York 1993: 439-444.

Mohtadi N. Development and validation of the quality of life outcome measure (questionnaire) for chronic anterior cruciate ligament deficiency. Am J Sports Med 1998; 26(3): 350-359.

Neeb T B, Aufdemkampe G, Wagener J H D, Mastenbroek L. Assessing anterior cruciate ligament injuries: The association and differential value of questionnaires, clinical tests, and functional tests. J Orthop Sports Phys Ther 1997; 26(6): 324-331.

Norman G R, Streiner D L. *PDQ Statistics*. BC Decker, Hamilton, Ontario 1986.

Noyes F R, Barber S D, Mooar L A. A rationale for assessing sports activity levels and limitations in knee disorders. Clin Orthop 1989; (246): 238-249.

Odenbring S, Tjörnstrand B, Egund N, et al. Function after tibial osteotomy for medial gonarthrosis below aged 50 years. Acta Orthop Scand 1989; 60: 527-531.

Patel D, Fahmy N, Sakayan A. Isokinetic and functional evaluation of the knee following arthroscopic surgery. Clin Orthop 1982; 167: 84-91.

Petersson I F, Boegård T, Saxne T, Silman A J, Svensson B. Radiographic osteoarthritis of the knee classified by the Ahlbäck and Kellgren & Lawrence systems for the tibiofemoral joint in people aged 35-54 years with chronic knee pain. Ann Rheum Dis 1997; 56: 493-496.

Ranawat C S, Insall J N. Duocondylar knee arthroplasty. Clin Orthop 1976; 120: 76-92.

Risberg M A, Ekeland A. Assessment of functional tests after anterior cruciate ligament surgery. J Orthop Sports Phys Ther 1994; 19(4): 212-217.

Rockborn P, Gillquist J. Outcome of arthroscopic meniscectomy. Acta Orthop Scand 1995; 66(2): 113-117.

Roos E M, Hansson M, Järnvall P. Hur mår korsbandsrekonstruerade patienter? -Jämförande skattning av patient, läkare och sjukgymnast. In: Proceedings of Svenska Läkaresällskapets Riksstämma. Göteborg:1998a: 201 [in Swedish].

Roos H, Adalberth T, Dahlberg L, Lohmander L S. Osteoarthrosis of the knee after injury to the cruciate ligament or meniscus: the influence of time and age. Osteoarthritis Cartilage 1995; 3: 261-267.

Roos H, Laurén M, Adalberth T, Roos E M, Jonsson K, Lohmander L S. Knee osteoarthritis after meniscectomy. The prevalence of radiographic changes after 21 years, compared to matched controls. Arthritis Rheum 1998b; 41(4): 687-693.

Rosner B. *Fundamentals of Biostatistics*. Duxbury Press, Belmont, California 1995.

Ryd L, Kärrholm J, Ahlvin P. Knee scoring systems in gonarthrosis. Evaluation of interobserver variability and the envelope of bias. Acta Orthop Scand 1997; 68(1): 41-45.

Schimmer R, Brulhart K, Duff C, Glinz W. Arthroscopic partial meniscectomy: a 12-year and two-step evaluation of the long-term course. Arthroscopy 1998; 14(2): 136-142.

Sgaglione N A, Del Pizzo W, Fox J M, Friedman M J. Critical analysis of knee ligament rating systems. Am J Sports Med 1995; 23(6): 660-667.

Shapiro E T, Richmond J C, Rockett S E, McGrath M M, Donaldson W R. The use of a generic, patient-based health assessment (SF-36) for evaluation of patients with anterior cruciate ligament injuries. Am J Sports Med 1996; 24(2): 196-200.

Shields R, Enloe L, Evans R, Smith K, Steckel S. Reliability, validity, and responsiveness of functional tests in patients with total joint replacement. Phys Ther 1995; 7(3): 169-179.

Shrout P, Fleiss J. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 1979; 86(2): 420-428.

Small N C, Sledge C B, Katz J N. A conceptual framework for outcome research in arthroscopic meniscectomy: Results of a nominal group process. Arthroscopy 1994; 10(5): 486-492.

Snyder-Mackler L, Fitzgerald G K, Bartolozzi A R, Ciccotti M G. The relationship between passive joint laxity and functional outcome after anterior cruciate injury. Am J Sports Med 1997; 25(2): 191-195.

St Pierre D. Rehabilitation following arthroscopic meniscectomy. Sports Med 1995; 20(5): 338-347.

Stratford P, Binkley J, Solomon P. Defining the minimum level of detectable change for the Roland-Morris Questionnaire. Phys Ther 1996; 76: 359-365.

Stratford P W, Gill C, Westaway M, Binkley J M. Assessing disability and change on individual patients: a report of a patient-specific measure. Physiotherapy Canada 1995; 47: 258-263.

Stucki G, Meier D, Stucki S, et al. Evaluation of a German version of the WOMAC osteoarthritis index. Zeitschrift für Rheumatologie 1996; 55(1): 40-49.

Stucki G, Sangha O, Stucki S, et al. Comparison of the WOMAC (Western Ontario and McMaster Universities) osteoarthritis index and a self-report format of the self-administered Lequesne-Algofunctional index in patients with knee and hip osteoarthritis. Osteoarthritis Cartilage 1998; 6(2): 79-86.

Sullivan M, Karlsson J. *SF-36 Hälsoenkät: Swedish Manual and Interpretation Guide*. Health Care Unit, Sahlgrenska Hospital, Gothenburg, Sweden 1994 [in Swedish].

Sun Y, Sturmer T, Gunther K, Brenner H. Reliability and validity of clinical outcome measurements of osteoarthritis of the hip and knee -a review of the literature. Clin Rheum 1997; 16(2): 185-198.

Tapper E, Hoover N. Late results after meniscectomy. J Bone Joint Surg (Am) 1969; 51-A: 517-526.

Tegner Y, Lysholm J. Rating systems in the evaluation of knee ligament injuries. Clin Orthop 1985; 198: 43-49.

Tzonchev V T, Pilossoff T, Kanev K. *Prevalence of osteoarthritis in Bulgaria in population studies of the rheumatic diseases*. Excerpta Medica Foundation, Amsterdam 1968.

Vander Schilten J L. Improvements in rehabilitation of the postmeniscectomized or meniscal-repaired patient. Clin Orthop 1990; 252: 73-79.

Ware J. *How to score the revised MOS short form health scale (SF-36)*. The Health Institute, New England Medical Center Hospitals, Boston, MA 1988.

Ware J E, Sherbourne C D. The MOS 36-item short form health survey (SF-36).I. Conceptual framework and item selection. Med Care 1992; 30: 472-483.

Weiner E, Stewart B. *Assessing individuals: Psychologic and educational tests and measurements*. Little Brown, Boston 1984.

Westaway M D, Stratford P W, Binkley J M. The Patient-specific functional scale: Validation of its use in persons with neck dysfunction. J Orthop Sports Phys Ther 1998; 27(5): 331-338.

WHO. *International Classification of Impairments, Disabilities, and Handicaps*. World Health Organization, Geneva 1980.

Williams M, Frankel S, Nanchahal K, Coast J, Donovan J. DHA project: Epidemiologically based needs assessment: Total knee replacement, 1994.

Windsor R, Insall J, Warren R. The Hospital for Special Surgery knee ligament rating form. Am J Knee Surg 1988; 1: 140-145.

Wright J, Rudicel S, Feinstein A. Ask patients what they want. Evaluation of individual complaints before total hip replacement. J Bone Joint Surg (Br) 1994; 76-B(2): 229-234.

Wright J G, Young N L. The patient-specific index: asking patients what they want. J Bone Joint Surg (Am) 1997; 79-A(7): 974-983.

Zarins B, Boyle J, Harris B A. Knee rehabilitation following arthroscopic meniscectomy. Clin Orthop 1985; 198: 36-42.

Appendices

A. KOOS Users Guide (including reference data and scoring manual)

B. Swedish version of KOOS

C. Danish version of KOOS

D. AmericanEnglish version of KOOS

E. Swedish version of WOMAC

When copying the Users Guide and the questionnaires:

Use a magnifying factor of 141 % to receive a European size A4 (equivalent to American letter size). European size A4 is suitable for patient administration.

The Swedish version of WOMAC is to be copied in landscape view on both sides of one sheet. Fold to get a booklet with 4 pages.