

## Definitions and abbreviations

<b>Outcome</b>	Something that follows as a result <sup>1</sup>
<b>Outcomes research</b>	The study of the end result of health services that takes patients' experiences, preferences, and values into account <sup>2</sup>
<b>Questionnaire</b>	A set of questions for obtaining statistically useful or personal information from individuals; a survey made by the use of a questionnaire. It includes standardized questions and response choices. Synonyms are measure, test, tool, survey, or instrument <sup>3</sup>
<b>Item</b>	A single question or statement and its standardized set of responses <sup>4</sup>
<b>Generic measures</b>	A category of health measures that are valued by all types of patients as well as general populations, and that have reliability and validity to measure health in populations with diverse characteristics <sup>3</sup>
<b>Disease-specific measures</b>	Focus on particular complaints attributable to the disease or condition of interest <sup>3, 4</sup>
<b>Observer-administered measures</b>	Questions answered by respondents about themselves by responding to an interviewer's question
<b>Patient-administered measures or Self-administered measures</b>	Respondents read and answer the questions by themselves, without assistance
<b>Test-retest reliability</b>	Stability over time <sup>5</sup>
<b>Internal consistency</b>	Based on the average inter-item correlation and number of items <sup>5</sup>
<b>Face validity</b>	Extent to which a measure "looks like" what it is intended to measure; whether respondents understand a measure's questions and find the answers appropriate <sup>3, 4</sup>
<b>Content validity</b>	The extent to which measures represent functions or items of relevance given the purpose and matter at issue <sup>5</sup>
<b>Criterion validity</b>	The extent to which a measure corresponds to an

---

	accurate or previously validated measure of the same concept <sup>4</sup>
– <b>Concurrent validity</b>	A form of validity in which the measure being tested and the comparison measure are administered at the same point in time <sup>3</sup>
<b>Construct validity</b>	The degree to which an instrument measures the theoretical construct it was designed to measure <sup>5</sup>
– <b>Convergent validity</b>	Seeing whether a measure displays the pattern of converging relationships it should <sup>5</sup>
<b>WOMAC</b>	Western Ontario and MacMaster Universities Osteoarthritis Index
<b>KOOS</b>	Knee injury and Osteoarthritis Outcome Score
<b>SF-36</b>	Short Form 36-item questionnaire derived from the Medical Outcome Study
<b>ACL</b>	Anterior Cruciate Ligament
<b>OA</b>	Osteoarthritis
<b>Post-traumatic OA</b>	Osteoarthritis that follows as a result of injury

---

<sup>1</sup>Webster's New Encyclopedic Dictionary, 1993.

<sup>2</sup>Clancy CM and Eisenberg JM, *Science* 1998; 282: 245–246.

<sup>3</sup>Bungay KM and Ware JE. *Measuring and Monitoring Health-Related Quality of Life*. Kalamazoo, MI: The Upjohn Company; 1993.

<sup>4</sup>Stewart AL and Ware JE, eds. *Measuring Functioning and Well-being: The Medical Outcomes Study Approach*. Durham, NC: Duke University Press; 1992.

<sup>5</sup>Johnston MV et al., *Arch Phys Med Rehabil* 1992; 73: S3–S23.

# Introduction

## Knee injury and knee OA

Knee osteoarthritis is common. Joint disease, causing pain and functional limitations, is the most common chronic disease in the elderly, more common than high blood pressure, heart disease, diabetes, etc (Havlik et al. 1986). The treatment of knee osteoarthritis is focused on pain control, initially by drugs or other modalities, and at later stages by surgical replacement of the affected joint. In Sweden more than 5,000, and in the USA more than 250,000 knee arthroplasties are carried out yearly (Knutson et al. 1994).

Although OA is primarily a disease of the older, knee OA is common at younger age. It can be estimated that more than 5% of subjects between 35 and 54 years old have radiographic signs of knee OA (Tzonchev et al. 1968, Hernborg and Nilsson 1973, HANES 1 1979, Williams et al. 1994). In a Swedish population study, including subjects between 35 and 54 years old without known knee injury, the prevalence of radiographic tibiofemoral knee OA was 1.5 % (Pettersson et al. 1997). Knee injury is a known risk factor of radiographic as well as symptomatic knee OA. Injury to the menisci or the anterior cruciate ligament causes 6/10,000 and 3/10,000 individuals, respectively, to seek medical care every year in Denmark (Hede et al. 1990, Buhl-Nielsen 1991). Approximately half of these patients have radiographic signs of knee OA after 10–15 years (Lohmander and Roos 1994, Roos et al. 1995, Roos et al. 1998b). The majority of subjects under the age of 50 undergoing tibial osteotomy because of symptomatic knee OA, have a previous injury to the anterior cruciate ligament or menisci (Odenbring et al. 1989). A high proportion of the population, in an age group with high demands on physical activity, is at risk of developing symptoms of knee OA. The treatment offered to these relatively young and active patients is symptom modifying. This far, no treatment has been shown to modify the disease process of osteoarthritis in humans.

## Outcome measures

Traditionally, process measures like radiographs, laxity and other clinical findings have been used to evaluate knee injury and knee OA. Clinical trials have used these measures as their primary dependent variables. Seldom have patients' preferences for outcomes been used to evaluate treatment; they have often been perceived as important but subjective and unreliable. However, we need to identify sources of cost without benefit, and the best treatment. Clinicians need to select

effective treatments, and patients want to make informed treatment choices. These concerns have stimulated researchers in the clinical, quantitative, behavioral, and social sciences to expand the methods and metrics used to evaluate the effects of health services. The number of valid patient-centered measures has increased dramatically, and the use of both generic and disease-specific patient-centered measures is recommended in clinical trials of OA (Altman et al. 1996, Bellamy et al. 1997). It is well recognized that process measures and patient-related measures evaluate different aspects of knee injury and knee OA. Weak correlations and frequent discordance is found when comparing process measures like radiographic findings and laxity to patient-relevant outcomes such as pain, function and activity level (Hadler 1992, Lethbridge-Cejku et al. 1995, Cicuttini et al. 1996, Snyder-Mackler et al. 1997).

In summary, patient-relevant outcome measures are now promoted in general health care, orthopedics, and sports medicine, and should be considered the primary outcome measure in clinical trials (Amadio 1993, Johnson 1994, Altman et al. 1996, Bellamy et al. 1997, Clancy and Eisenberg 1998).

# Objectives

## General

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

## Specific

- to evaluate short-term and long-term symptoms and function after meniscectomy,
- to validate the WOMAC Osteoarthritis Index for use in Sweden,
- to determine the sensitivity of the WOMAC Osteoarthritis Index for post-traumatic osteoarthritis,
- to develop a self-administered instrument, the Knee injury and Osteoarthritis Outcome Score (KOOS), to be used in the acute phase and over time, for assessment of patient-relevant outcomes in patients with anterior cruciate deficiency, meniscus injury, and cartilage damage or post-traumatic osteoarthritis,
- to evaluate the KOOS with regard to reliability, validity, and responsiveness,
- to evaluate the KOOS for use with assessment of reconstruction of the anterior cruciate ligament, knee arthroscopy, and physical therapy,
- to validate the KOOS for use in North America and Sweden,
- to compare the KOOS to other instruments used for similar purposes and diagnostic groups.

## List of papers

- I. Roos E, Östenberg A, Roos H, Ekdahl C, Lohmander LS. Long-term results of meniscectomy—symptoms, function, and performance tests in patients with or without radiographic osteoarthritis compared to matched controls. Manuscript.
- II. Roos E, Roos H, Lohmander LS. WOMAC Osteoarthritis Index—additional dimensions for use in post-traumatic osteoarthritis of the knee. *Osteoarthritis and Cartilage* 7(3): 00–00, 1999.
- III. Roos E, Klässbo M, Lohmander LS. WOMAC Osteoarthritis Index—reliability, validity, and responsiveness in patients with arthroscopically assessed osteoarthritis. Submitted.
- IV. Roos E, Roos H, Lohmander LS, Ekdahl C, Beynnon B. Knee injury and Osteoarthritis Outcome Score (KOOS)—development of a self-administered outcome measure. *The Journal of Orthopaedic and Sports Physical Therapy* 78(2): 88–96, 1998.
- V. Roos E, Roos H, Ekdahl C, Lohmander LS. Knee injury and Osteoarthritis Outcome Score (KOOS)—validation of a Swedish version. *Scandinavian Journal of Medicine and Science in Sports* 8: 439–448, 1998.
- VI. Roos E, Roos H, Ryd L, Lohmander LS. Substantial disability three months after arthroscopic partial meniscectomy—a prospective study of patient-relevant outcomes. Submitted.

Some additional data, not previously published, have been included in Results.

Reprints were made with permission from the publishers.

## Previous investigations

### **Impairment, disability, and handicap**

Outcomes can be classified into different measurement levels. Commonly, the International Classification into Impairment, Disability, and Handicap (ICIDH) is used (WHO 1980). Impairment reflects disturbances at the level of the organ, while disability is concerned with abilities that are generally accepted as essential components of everyday life. For a knee trauma patient, the rupture of a ligament would be classified as impairment, while the inability to walk, run or squat would be the resulting disability. Handicap is a disadvantage for a given individual. A knee trauma, resulting in difficulty with fast cutting actions, could be a handicap for a professional athlete or a construction worker, while a sales person or secretary might not experience a handicap from the same impairment or the same disability. An overview of what levels of measurement for different knee scoring scales and other outcome instruments measure is given in Table 3.

### **Assessment of knee injury and knee OA**

Knee injury most often includes damage to the ligaments, the menisci, or the cartilage. These injuries can occur isolated but are frequently combined (Cerabona et al. 1988). Orthopedic interventions aim to restore damaged structures. Consequently, important outcomes of knee injury include laxity measurement and radiographs. However, patients are more concerned with symptoms and functional limitations. These aspects have been assessed by knee scoring scales like the Lysholm knee scoring scale (Tegner and Lysholm 1985), the Noyes or Cincinnati knee ligament rating scale (Noyes et al. 1989), or the IKDC score (Hefti et al. 1993). These scoring scales are developed by orthopedic surgeons and reflect the operating surgeons' perspectives. For each scale symptoms and functional limitations are weighted differently. The scores of the separate items are then aggregated into one total score. Not surprisingly low correlations have been shown between existing scales, indicating that the scales measure different constructs and that studies employing different scales are not comparable (Bollen and Seedholm 1991, Sgaglione et al. 1995, Labs and Paul 1997, Neeb et al. 1997).

During the early 1990's the first patient-administered questionnaire intended for assessment of ACL deficient patients, and tested for validity, was introduced by Flandry and co-workers (Flandry et al. 1991). During 1998, another three questionnaires, meticulously assessed with regard to the developmental process,

reliability, validity and sensitivity to clinical change were published. The Quality of Life Outcome Measure for Chronic Anterior Cruciate Ligament Deficiency (ACL-QOL) (Mohtadi 1998) and The Activities of Daily Living Scale of the Knee Outcome Survey (ADLS) (Irrgang et al. 1998b) are described below. The third instrument is the Knee injury and Osteoarthritis Outcome Score (KOOS) which is described in detail in "Present investigations".

Osteoarthritis is usually assessed by radiographs. The severity of radiographic OA is evaluated according to one of several available classification systems (Kellgren and Lawrence 1963, Ahlbäck 1968, Altman et al. 1995). Clinical examination, symptoms and functional limitations have frequently been summarized by one of several knee scoring systems available for knee OA (Ranawat and Insall 1976, Jónsson 1981, Ewald 1989). Two outcome measures, the WOMAC Osteoarthritis Index and the Lequesne Index, are the most commonly used outcome measures. The similarity is assessment of symptoms and function, and the dissimilarities include format of questions and answer options, administration mode, and weighting and aggregation of scores.

Some knee scoring scales, and other important outcome measures, are described below and in Table 3.

### The Tapper & Hoover grading system

In 1969 Tapper and Hoover introduced an evaluation system to be used by an observer, categorizing outcome after meniscectomy into excellent, good, fair, and poor (Tapper and Hoover 1969), Table 1. Frequently, the total scores of knee scoring scales are categorized into the same four categories.

Table 1. The Tapper & Hoover grading system

Excellent	an effective and completely normal knee.
Good	a knee giving minor symptoms, but no disability, i.e. the knee is functional in all activities including vigorous sports, but with some ache or swelling afterwards.
Fair	a knee giving definite symptoms and some disability, preventing vigorous sports.
Poor	a knee giving symptoms, e.g., aching while kneeling or climbing stairs, which interferes with daily activities. Definite mechanical symptoms, e.g., locking, also indicates a poor grade.

### The Lysholm Knee Scoring Scale

The Lysholm knee scoring scale was introduced by Lysholm and Gillquist (1982). The Lysholm scoring scale is meant to be used by an examiner and no instructions to the patient are provided. In 1985 a revised form was published,

being the version used today (Tegner and Lysholm 1985). For assessment of meniscus lesions, a modified version is available (Lysholm and Gillquist 1982). However, in studies on injury to the menisci, different versions have been used (Hamberg and Gillquist 1984, Hamberg et al. 1984, Ferkel et al. 1985, Hede et al. 1992, Katz et al. 1992, Bolano and Grana 1993, Jaureguito et al. 1995, Rockborn and Gillquist 1995, Maletius and Messner 1996, Schimmer et al. 1998).

When the total Lysholm score is categorized as poor, fair, good or excellent, the cut off for good/excellent result of the first version was 77 points, as recommended by Lysholm & Gillquist (Lysholm and Gillquist 1981). For the updated version the cut off for good/excellent results has been set to 84 points (Rockborn and Gillquist 1995, Maletius and Messner 1996), however both 77 and 84 points have been used as cut-off for good/excellent results for the updated version. The three versions of the Lysholm knee scoring scale are given in Table 2.

Table 2. The three versions of the Lysholm knee scoring scale

	1982	1982 meniscal version	1985
Instability	30	20	25
Pain	30	25	25
Catching, Locking	-	15	15
Swelling	10	10	10
Stairs	10	10	10
Squat	5	5	5
Limp	5	5	5
Support	5	5	5
Thigh atrophy	5	-	-
Total	100	95	100
Cut-off excellent/good	77	77	84

### The Cincinnati Knee Ligament Rating System

The Cincinnati Knee Ligament Rating Form was introduced by Noyes and co-workers in the 1980's (Noyes et al. 1989). The system includes physical examination and instrumented testing as well as a four-part evaluation format to assess symptoms and function. This includes 1) a symptoms scale assessing pain, swelling, and partial and full giving way, depending on six specifically defined activity levels; 2) assessment of function by determining how each patient performs certain activities (including walking, climbing stairs, squats, running, jumping, and pivoting); 3) a sports activity rating scale that stratifies function depending on four levels of sport type and frequency of participation; 4) a final rating system

that provides an overall grade defined by the lowest score in any individual category. The rating system was developed to be used by an observer.

### **The International Knee Documentation Committee Knee Ligament Standard Evaluation Form (IKDC)**

The International Knee Documentation Committee (IKDC) was formed from members of the American Orthopedic Society for Sports Medicine (AOSSM) and the European Society of Sports Traumatology, Knee Surgery and Arthroscopy (ESSKA). The focus of this group was to define the terms that should be used to describe the injured knee and to set standards for the evaluation of knee ligament injuries. The efforts of the IKDC resulted in the development of the Knee Ligament Standard Evaluation Form (Hefti et al. 1993). The IKDC Standard Evaluation, which is developed to be used by an observer, consists of eight groups including: patient's subjective assessment of function, symptoms, range of motion, ligament examination, compartmental findings, harvest site pathology, radiographic findings, and functional tests. However, only the first four groups are included in the final overall IKDC rating. Each group consists of one or more items that are rated as normal, nearly normal, abnormal, or severely abnormal according to established guidelines. The worst rating of any item in a given group determines the overall group rating, and the final overall rating of the knee is based on the worst rating for the categories of patient's subjective assessment of function, symptoms, range of motion, and ligament examination. Therefore, the final overall rating of the knee is limited by the worst rating for any one particular item. For example, if all items receive a rating of normal or nearly normal, except for the Lachman test, which receives a rating of abnormal, then the overall final rating for the knee is abnormal. This should prevent giving the knee a satisfactory rating when a significant problem continues to exist (Irrgang et al. 1998a).

### **The Flandry Questionnaire**

In 1991, Flandry and co-authors from the Hughston Sports Medicine Foundation, Inc. introduced a 26 item self-administered questionnaire for assessment of subjective knee complaints (Flandry et al. 1991). The questionnaire consists of 28 items and uses a visual analog scale response format. An average score from 0 to 100 is calculated. In a clinical study comprising 182 patients with knee complaints, the system was shown to be valid and comparable to other methods while offering several advantages. It brought greater sensitivity and greater statistical power to data collection and analysis by allowing a broader range of responses than did traditional categorical responses. It removed bias that was introduced by

examiner questioning, and it allowed graphic temporal comparisons. Most importantly, patient affinity was higher for this type of subjective evaluation than for other methods. The questionnaire was translated into German in 1995 (Höher et al. 1995).

### **Quality of Life Outcome Measure for Chronic Anterior Cruciate Ligament Deficiency (ACL-QOL)**

The ACL-QOL questionnaire was initially published in a book chapter on outcome assessment after ACL reconstructive surgery in 1993 (Mohtadi 1993). The questionnaire consists of 32 items and uses a visual analog scale response format. The questionnaire assesses symptoms, physical complaints, work-related concerns, recreational activity and sport participation, lifestyle, social and emotional questions. An average total score ranging from 0 to 100 is calculated. In 1998 the developmental process and the validation study was published (Mohtadi 1998). The ACL-QOL was developed to evaluate the patients' view on knee problems. The reliability, validity and sensitivity of the questionnaire for patients with ACL deficiency were found to be good.

### **The Activities of Daily Living Scale (ADLS) of the Knee Outcome Survey**

The Activities of Daily Living Scale (ADLS) of the Knee Outcome Survey is a patient-reported measure of functional limitations imposed by pathological disorders and impairments of the knee during activities of daily living (Irrgang et al. 1998b). The ADLS consists of 17 items and assesses the impact of symptoms and functional limitations on activities of daily living. An average total score ranging from 0 to 100 is calculated. In a clinical study comprising 397 patients referred to physical therapy because of a wide variety of disorders of the knee, the scale was proven to be a reliable, valid and sensitive outcome measure for the assessment of functional limitations due to knee problems.

### **Hospital for Special Surgery Knee Score (HSS)**

The Hospital for Special Surgery Knee Score (HSS) is meant to be used by an observer to summarize symptoms and clinical signs: pain (30 points); function (22 points); range of motion (18 points); muscle strength (10 points); (6) flexion deformity (10 points); and instability (10 points). All items are summoned together into a score ranging from 0 to 100. The overall scores are converted into

the categories excellent (85-100), good (70-84), fair (60-69), and poor (<60) as suggested by Ranawat and Insall (1976).

### **The Knee Society Clinical Rating System (KSS)**

The Knee Society Clinical Rating System was developed in the late 1980's to provide an up-to-date and more stringent evaluation form. The system is subdivided into a knee score that rates only the knee joint itself and a functional score that rates the patient's ability to walk and climb stairs. The dual rating system was proposed since it eliminated the problem of declining knee scores associated with patient infirmity (Insall et al. 1989). The functional part assesses the patient's ability to walk (50 points) and climb stairs (50 points). A maximum functional score is given to a patient who can walk an unlimited distance and go up and down stairs normally.

### **Lequesne Index of Severity – Knee (Lequesne ISK)**

Index of severity for OA of the hip was introduced in 1980 and a modified index for OA of the knee was introduced some years later (Lequesne 1989) The Lequesne Index of severity for OA of the knee assesses Pain or Discomfort (5 questions), Walking (2 questions), and Activities of Daily Living (4 questions). The questions are weighted differently, and all items are aggregated into one total score ranging from 0 to 24 points.

### **WOMAC Osteoarthritis Index**

Western Ontario and MacMaster Universities Osteoarthritis Index (WOMAC), the first patient-administered outcome measure developed for OA, was constructed during the 1980's (Bellamy et al. 1988a, 1988b). The WOMAC measures three separate dimensions: Pain (5 questions), Stiffness (2 questions), and Function (17 questions). The original WOMAC, developed in Canada, is available in two formats, Visual Analog Scales and Likert-boxes, with similar metric properties. The scores are reported separately for the three subscales. For the Likert version, the range of possible subscale scores for pain, stiffness, and function is 0–20, 0–8, and 0–68, respectively. The questionnaire is validated for use in knee and hip osteoarthritis (Bellamy et al. 1988a, 1988b) and is frequently used around the world.

### **12-item questionnaire for total knee replacement**

In 1998 a new questionnaire on perceptions of patients having a total knee replacement was published (Dawson et al. 1998). The twelve questions concern symptoms and functional limitations. Each item is scored from 1 to 5, and combined to produce a single score with a range from 12 (least difficulties) to 60 (most difficulties). In a clinical study of 117 patients the 12-item questionnaire proved to be a practical, reliable, valid and sensitive outcome measure.

### **SF-36—a generic measure**

The Short Form 36-item of the Medical Outcome Study (SF-36) is a widely used generic instrument for assessment of health status. It is patient-administered and comprises 8 subscales assessing physical and mental health to various degrees. The score of each subscale ranges from 0 (poor) to 100 (good). The advantage of using generic questionnaires is that comparisons can be made across diagnoses, and thus be a tool for health care planners. The SF-36 has been used to evaluate patients with anterior cruciate ligament injury (Shapiro et al. 1996), and to determine the outcomes of meniscectomy (Katz et al. 1992) and knee replacement (Hawker et al. 1995). In 1994 it was suggested by a nominal group's process that the SF-36 should be used, in addition to disease-specific measures, when assessing the outcome of arthroscopic meniscectomy (Small et al. 1994). The SF-36 is also recommended to be included in clinical trials on knee OA (Hawker et al. 1995, Altman et al. 1996).

### **Musculoskeletal Functional Assessment Questionnaire (MFA)— a generic measure**

The Musculoskeletal Functional Assessment Questionnaire (MFA), a tool for the evaluation of patients' perception about their physical, psychological, and social well being, was introduced in 1996 (Engelberg et al. 1996, Martin et al. 1996). The questionnaire consists of 101 yes or no questions grouped into ten categories; self-care, sleep and rest, hand and fine motor skills, mobility, housework, employment and work, leisure and recreational activities, family relationships, cognition and thinking, and emotional adjustment, coping, and adaptation. The result can be presented in separate scores or one total score. In a study comparing the MFA to the SF-36, the WOMAC and the Sickness Impact Profile (SIP) (Bergner et al. 1976), all instruments were found to be of good reliability (intra-class correlation coefficients of more than 0.70) (Martin et al. 1997). The study group consisted of 444 patients with traumatic fractures or soft-tissue-injury of the extremities, a history of repetitive-motion disorder, osteoarthritis, or rheuma-

toid arthritis. In this group the MFA demonstrated better content validity with no ceiling or floor effects. In addition, it was more responsive than the SF-36. It was concluded that the MFA could be used to assess health status of patients with a musculoskeletal disorder.

### **The Patient-Specific Index—a measure for total hip replacement**

The Patient-Specific Index was introduced in 1994 and is used to assess the outcome of total hip replacement by evaluating the preferences of the individual patient (Wright et al. 1994, Wright and Young 1997). Although not available for knee OA, the index merits being mentioned because of its approach to assess patient-relevant outcomes. The patients rate the severity and importance of each complaint, these numbers are then added together. An advantage of this approach is that the patient's goals are clearly determined and specified. This can prevent a situation where clinicians rate hip replacements as successful in patients who are disappointed because the complaints that they considered the most important were not alleviated. Another advantage of this approach is statistical, higher effect sizes are seen which indicate fewer patients needed in clinical studies to demonstrate statistically significant differences.

### **Patient-Specific Functional Scale (PFPS)—a generic measure of disability**

The Patient-Specific Functional Scale (PFPS) was introduced in 1995 from McMaster University in Hamilton, Ontario, Canada (Stratford et al. 1995). The PFPS is a measure of disability that can be used across musculoskeletal conditions, and has been validated for use in subjects with low back pain (Stratford et al. 1995), knee dysfunction (Chatman et al. 1997), and neck dysfunction (Westaway et al. 1998). The PFPS is designed to measure at the level of the individual patient rather than the group level, and should complement the findings of generic or disease-specific measures. The PFPS seem to be ideal for quality assurance in clinical settings where a wide variety of conditions are treated, such as a physical therapy practice, and is not primarily designed for assessment of groups in clinical studies.

## **Methodological aspects of outcome measures for knee injury and knee OA**

Methodological aspects of outcome measures for knee injury and knee OA have received increasing attention lately. From the literature on methodological aspects, which is further discussed in the General discussion section, the following conclusions can be suggested:

- Knee symptoms and functional limitations, as measured by the Cincinnati and the Lysholm scores, occur in young athletes considering themselves “knee healthy” (Demirdjian et al. 1998).
- The Lysholm score is less sensitive to ACL injury than to other diagnostic groups (Bengtsson et al. 1996).
- The Lysholm score does not accurately identify problems during strenuous activities (Risberg and Ekeland 1994).
- The Lysholm score yields a better outcome than the Cincinnati score (Bollen and Seedholm 1991, Sgaglione et al. 1995).
- Self-administration of the Lysholm score yields worse outcome than completion by an observer (Höher et al. 1997).
- The IKDC is not a valid measure of *patient-relevant* outcomes (Snyder-Mackler et al. 1997, Irrgang et al. 1998a).
- 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).
- Composite knee scoring systems are exceedingly unreliable (Ryd et al. 1997).
- The reliability for measures of pain, stiffness, and function is reported to be high, while the results for clinical signs are less conclusive (Sun et al. 1997)
- The WOMAC Osteoarthritis Index is preferable to the Lequesne Index (Stucki et al. 1998).
- The SF-36 is sensitive to change in subjects with knee injury (Katz et al. 1992, Shapiro et al. 1996).
- The inclusion of both a disease-specific and a generic instrument is recommended in clinical trials (Small et al. 1994, Hawker et al. 1995).

Table 3. Outcome instruments for knee injury and knee OA categorized into measurement levels, administration mode, and how the result is presented

Scale	Measurement level <sup>1</sup>			Administration method		Assesses measurement levels in separate scores
	Impairment	Disability	Handicap	Patient administered	Observer administered	
<b>Knee injury</b>						
Tapper & Hoover, 1969	●	●			●	
Lysholm, 1982	●	●			●	
Cincinnati, 1984	●	●			●	
IKDC, 1993	●	●			●	
Flandry, 1992	●	●		●		
ACL-QOL, 1998	●	●	●	●		
<b>Knee injury and knee OA</b>						
KOOS, 1998	●	●	●	●		●
ADLS, 1998	●	●		●		
<b>Knee OA</b>						
HSS, 1976	●	●			●	
KSS, 1989	●	●			●	●
Lequesne ISK, 1987	●	●			●	
WOMAC, 1988	●	●		●		●
12-item, 1998	●	●		●		
<b>Generic and other</b>						
SF-36, 1992	●	●	●	●		●
MFA, 1998	●	●	●	●		●
PSI, 1994	●	●	●	●		
PFPS, 1995		●			●	● <sup>2</sup>

<sup>1</sup> According to the international classification of impairments, disabilities, and handicap (WHO 1980).<sup>2</sup> Measure only at the disability level

# Present investigations

## Subjects

Table 4. Number and characteristics of the 390 subjects included in studies I–VI

Study	N	Character	Sex M/F	Mean age at exam (range)
<i>Cohort</i>				
<b>I</b>	159	Meniscectomy 1973 or 1978	124/35	53 (33–78)
	68	Controls	50/18	55 (36–79)
<b>II</b>	41	Meniscectomy 1973 and radiographic OA	29/12	57 (38–76)
	50	Controls, no radiographic OA	38/12	53 (37–79)
<i>ACL reconstruction</i>				
<b>IV</b>	21	ACL reconstruction	9/12	32 (18–46)
<i>Arthroscopy</i>				
<b>III</b>	52	Arthroscopic OA	27/25	48 (20–69)
<b>V</b>	142	Knee arthroscopy (meniscus injury, ACL injury, arthroscopic OA)	89/53	37 (14–75)
<b>VI</b>	95	Arthroscopic partial meniscectomy	62/33	42 (14–75)

## Meniscus cohorts

In 1973 and 1978, all residents of the Lund district who received orthopedic medical advice and care had such provided at the Lund University Hospital. All patients who underwent an open total meniscectomy at this hospital in 1973 or 1978 were identified through the surgical code system that was in use in the hospital, and their current addresses were located through the National Population Records. Exclusion criteria were: a report of death, relocation outside the South Swedish Healthcare Region, a diagnosis of rheumatoid arthritis, major psoriasis, multitrauma, associated cruciate ligament injury diagnosed at the time of surgery or at follow-up examination, radiographic changes indicating knee OA at the time of surgery, knee surgery (other than for OA) before or after the meniscectomy in 1973/1978, or being under age 10 at the time of surgery. 123 patients in 1973 and 94 patients in 1978 fulfilled these criteria. 107 (87%) and 72 (76%) patients, respectively, were re-examined 21 and 18 years after open meniscectomy.

Of these 179 patients, eight patients were excluded in paper I because of general functional disability of other origin ( $n = 4$ ), hip arthroplasty on the operated side ( $n = 2$ ), recurrent Achilles tendon rupture on the operated side ( $n = 1$ ), and

not being able to follow instructions ( $n = 1$ ). 12 did not fill out the Flandry questionnaire regarding symptoms and function, or underwent performance tests, leaving 159 patients (35 women) for the study group (Table 4).

The Flandry questionnaire was used to obtain data regarding symptoms and function for the 107 patients being operated in 1973. KOOS data was obtained by mail approximately 20 months after the index follow-up. The KOOS questionnaire was returned by 87 subjects. Of these, 41 subjects had definite radiographic changes of OA and were used in paper II for comparison of sensitivity between the three WOMAC subscales pain, stiffness and function to the two KOOS subscales sport and recreational function and knee-related quality of life (Table 4).

### *Control group*

Sixty control subjects were regarded a sufficient number to provide statistical power for determination of the relative risk of developing radiographic signs of OA following meniscectomy (Roos et al. 1998b). To ensure 60 subjects, 2 age- and sex-matched controls for each of the 107 patients operated on in 1973 were identified from the National Population Records (same birth year, same sex, and same mail zip code). Each of these persons was contacted by mail and asked to complete a questionnaire. Of the 214 designated control subjects, 46 did not answer the questionnaire, 37 of those who answered did not want to undergo clinical and radiographic examination, and 16 were excluded because of previous meniscectomy or a known cruciate ligament injury. Of the remaining 115 subjects, 40 were excluded because they each represented a "double control" i.e., a second control subject was already matched with that particular patient. Thus, 75 control subjects remained and were invited to the examination. Six of the subjects did not show up despite their acceptance of the invitation, and 1 had side-to-side difference in knee laxity exceeding 3 mm and was therefore excluded, leaving a total of 68 subjects (18 women) in the control group. These 68 subjects were used as controls in study I, and in study II the 50 subjects that did not have any radiographic signs of OA were used as controls, Table 4.

### **Anterior cruciate ligament reconstruction**

The 21 patients used for validation of the American-English version of the KOOS questionnaire were the first 21 patients enrolled in a clinical study of different rehabilitation methods following reconstruction of the anterior cruciate ligament. The subjects were operated at the Department of Orthopedics at University of Vermont in Burlington, Vermont, USA.

### **Knee arthroscopy**

During a six-month period, questionnaires were mailed to 200 consecutive patients on the waiting list for knee arthroscopy at the Department of Orthopaedics at Lund University Hospital, Sweden. Inclusion criteria were undergoing arthroscopy and speaking Swedish as first language. Exclusion criteria were multiple joint affection or having other perioperative diagnoses than the following: meniscal lesion, ACL injury, or cartilage damage of the tibio-femoral joint. The injuries were isolated or combined as shown in figure 1, paper V. A standardized arthroscopy record form was filled out by the operating orthopedic surgeon. One orthopedic surgeon reviewed the record forms and classified the subjects with regard to arthroscopy findings. The ACL was regarded as insufficient if a rupture was seen at arthroscopy and at least two of three clinical tests performed under anesthesia were positive (Lachman sign, anterior drawer sign, and pivot shift). Meniscus tears were regarded as significant if they required surgery. Cartilage damage was defined as open lesion with bone contact or exposed bone. Cartilage damage was found on tibia, femur or both joint surfaces. A total of 153 patients fulfilled the inclusion and exclusion criteria. Of these 11 did not return the pre-operative questionnaires, thus baseline data was available for 142 (93 %) patients. These 142 patients constituted the study group in paper V, Table 4.

The short-term effect of arthroscopic partial meniscectomy was studied in paper VI. Of 106 patients having an arthroscopic partial meniscectomy as *only* intervention baseline data was available for 95 (90 %) patients, Table 4. As shown in Table 1, paper VI, 47 patients had an isolated meniscus tear, 27 had associated cartilage damage, and 21 had an associated injury of the ACL.

## Methods

### **The Flandry Questionnaire**

The Flandry questionnaire, which is described in “Previous investigations”, was used to evaluate symptoms and function in paper I. The test-retest reliability of the score has not been established. Instead of reporting the average for all items aggregated into one score, 21 of the items were grouped into 4 subscales: pain (3 items), symptoms (6 items), activities of daily living (ADL) (7 items), and sport and recreational function (Sport/Rec) (5 items) (Table 5). This is similar to four of the five subscales of the Knee Injury and Osteoarthritis Outcome Score (KOOS). As for the KOOS, an average percentage score from 0-100 was calculated for each of the four subscales pain, symptoms, ADL, and sport and recreation function. 100 indicated no knee-related complaints.

### **WOMAC Osteoarthritis Index**

The Western Ontario and MacMaster Universities Osteoarthritis Index (WOMAC) measures three separate dimensions: Pain (5 questions), Stiffness (2 questions), and Function (17 questions). The original Likert version 3.0, employing 5 Likert boxes, was translated into Swedish. The linguistic validation process was carried out in four steps: translation, back-translation, committee review, and pre-testing, according to published guidelines (Guillemin et al. 1993, Guillemin 1995), as described in Paper III.

#### *Score calculations*

According to the WOMAC User's Guide (Bellamy 1995) the first step is to take the data off the raw questionnaire. Numerical values are assigned to each of the five response categories (0=none, 1=mild, 2=moderate, 3=severe, 4=extreme). For each WOMAC dimension, a subscale score is calculated by simple summation of the assigned values scored on component items. Thus, the ranges of possible subscale scores for the three dimensions are as follows: pain=0-20, stiffness=0-8, physical function=0-68. However, to enable comparisons across subscales and to other outcome instruments, the summoned scores were transformed into a 0 to 100 scale, an approach also employed by others (Stucki et al. 1996, Creamer et al. 1998). To further simplify comparisons, 100 indicated no symptoms or functional disability and 0 indicated extreme symptoms and functional disability, as common in orthopedic scales (Tegner and Lysholm 1985, Windsor et al. 1988, Noyes et al. 1989).

Table 5. 21 of the 26 questions from the questionnaire by Flandry were grouped into the subscales pain, symptoms, ADL, and sports and recreation function. The anchors were never/unable and always/able, respectively

---

*Pain*

- How often does your knee hurt
- Do you have night pain
- Does your knee ache while you are sitting

*Symptoms*

- Do you have swelling in your knee
- Does your knee lock up so you are unable to straighten it
- Does your knee catch or hang up when moving
- Is your knee stiff
- Do you feel grinding when your knee moves
- Do you have stiffness or discomfort when you first start to walk

*Activities of Daily Living*

- Are you able to walk on level ground
- Do you have problems carrying heavy objects because of your knee
- Do you have problems climbing stairs
- Do you have problems going down stairs
- Do you have problems getting in or out of a car
- Do you have problems getting in or out of a chair
- Do you have problems turning over in bed

*Sport and Recreational Function*

- Do you have problems twisting or pivoting on your knee
- Do you have problems running
- Do you have problems jumping
- Do you have problems kneeling
- Do you have problems squatting

*Other Flandry-questions<sup>a</sup>*

- How bad is the pain at its worst
  - Does your knee give way or buckle
  - Are you able to walk on rough ground, inclines, or negotiate curves
  - Do you need crutches, cane, or walker to walk
  - Do you have problems decelerating (slowing down) after running or jogging
  - Do you have problems cutting (changing directions while running by pivoting on affected knee)
  - Do you have problems taking part in competitive sports
- 

<sup>a</sup> These questions were not grouped into any of the four subscales

## **Knee Injury and Osteoarthritis Outcome Score (KOOS)**

### *Development*

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. The concepts of questionnaire construction follow those of Liang & Jette (1981): 1) the instrument need 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. As described in Figure 1, we reviewed the literature, consulted an expert panel including patients, and carried out a pilot study to generate items to be included in the questionnaire.

### *Linguistic validation*

An American English and a Swedish version of the KOOS were developed simultaneously. The linguistic validation of the Swedish version of the KOOS was 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.

### *Validation studies*

Two clinical studies were designed to assess the reliability, validity, and responsiveness of the American English version and the Swedish version of the KOOS. In the American study patients about to undergo ACL reconstruction (N=21) were studied, and in the Swedish study patients about to undergo knee arthroscopy (N=142) were studied.

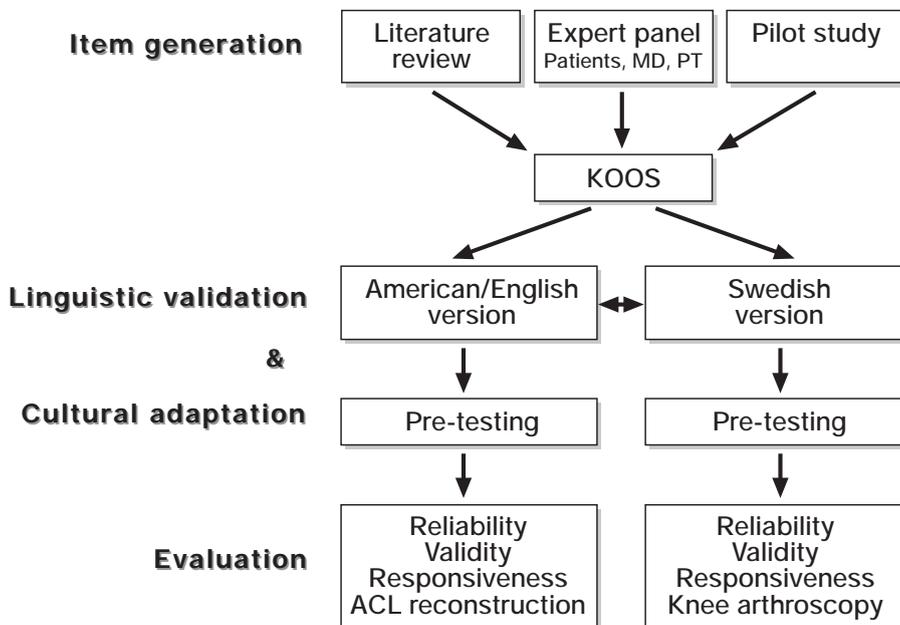


Figure 1. Development and evaluation of the KOOS.

## **SF-36**

The Short Form 36 item of the Medical Outcome Study (SF-36) is a widely used generic measure of health status which comprises eight subscales: Physical Function, Role-Physical, Bodily Pain, General Health, Vitality, Social Function, Role-Emotional, and Mental Health (Ware and Sherbourne 1992). The subscale scores range from 0 to 100, 100 indicating the least health-related problems and 0 the worst health-related problems. The test-retest reliability and the internal consistency of the SF-36 has been found satisfactory (Sullivan and Karlsson 1994). The SF-36 is further described under previous investigations. In paper IV the acute American-English version was used, and in paper III, V and VI the Swedish Acute version 1.0 was used (Sullivan and Karlsson 1994).

## **Lysholm Knee Scoring Scale**

The Lysholm knee scoring scale was used for determination of convergent construct validity in paper V. The test-retest reliability of the scale when administered by phone on three different occasions (day 1, 3, and 14) was 0.75 and 0.69, respectively (Kendalls correlation coefficient) (Bengtsson et al. 1996). The scale is further described under previous investigations.

## **Performance tests**

### *One-leg-rising*

The test was modified from Ekdahl et al. (1989) with the purpose to mainly assess hip-knee extensor strength in a functional position. The subject was sitting on a height adjustable bench, the heel of one foot placed 10 cm in front of the bench on a stool secured to the floor. This way the minimum height possible was 0 cm. The other foot was held in the air. Both arms were held out in front of the body. The subject was asked to rise on one leg without help, neither by swinging the body nor the arms, Figure 1, paper I. The subject chose the beginning height and got three trials. If the subject did not succeed, the bench was raised and three new trials were allowed. The subject continued until he/she could rise from the bench. The height, between the height adjustable bench and the stool attached to the floor, at the lowest height the subject succeeded to rise, was registered in cm. A low number in cm was seen as a better result than a high number. The procedure was repeated with the opposite leg. In a separate pre-study, the intra-tester reliability was assessed one week apart in 40 subjects with knee symptoms (19 patients (16 females) mean age 58 (36–75) and 21 controls (10 females) with a mean age of 44 (36–66) years. The Spearman's test-retest correlation coefficients were 0.78 and 0.84 for the patients and controls, respectively, indicating an ac-

ceptable test-retest reliability (E Roos, unpublished data).

### *Knee-bending*

The purpose of the test was to determine the maximal number of knee bendings a person could perform in 30 seconds to assess mainly the endurance in the hip/knee extensors. The starting position, as well as the performance of the test was standardized. The patient was asked to align the long axis of the foot to a straight line, and place the toes on a perpendicular line. The examiner gave fingertip support to prevent rotation at the pelvis and to provide some balance control. The subject was asked to bend the knee until he/she, without bending forward from the hip, no longer could see the line along the toes. This standardization gave a knee bending of 30-35 degrees, comparable to walking down a stair. The test is visualized in Figure 2, paper I. The test-retest reliability of the test has not been established.

### *Toe-raises*

The purpose of the test was to determine the maximal number of toe raises a person could perform in 20 seconds. The test was included as a test of endurance in the lower extremity that not primarily should be affected by knee pain or other knee symptoms during movement. The starting position, as well as the performance of the test was standardized. The subject was asked to stand on one leg. Fingertip support was allowed to provide some balance control. The subject was asked to rise up on the toes with extended knee and hip. The least allowed distance between floor and heel was approximately 2.5 centimeters. The test-retest reliability of the test has not been established.

## **Measure of recreational physical activity level**

In papers III and VI, current recreational activities were self-reported on a scale from 0 to 6 (modified from Steven Edworthy, Mc Craig Center for Joint Injury & Arthritis Research, Calgary, Alberta, Canada, personal communication, 1995), 0 defined as a minimum of recreational activities and 6 as competitive sports (table 6). Test-retest agreement of the activity scale has been found satisfactory. When administered twice with a mean of  $5.0 \pm 1.8$  days in 63 subjects prior to knee arthroscopy, 56 patients reported the same activity level at both administrations. Four patients reported one level higher on the second administration, while 2 patients reported one level lower. One patient reported three levels higher on the second administration (E Roos, unpublished data).

Table 6. Self-reported current recreational activities

---

*Which description describes your recreational activities the best?  
Please mark one alternative.*

---

- 6 competitive sports: soccer, racquet sports, track & field, skiing, etc.
  - 5 recreational sports: jogging, skiing, racquet sports, etc.
  - 4 golf, dancing, hiking, water aerobics
  - 3 heavy yard work, heavy household work, walking on even ground
  - 2 light yard work, light household work, shopping
  - 1 minimal household work, card games, sewing
  - 0 no household work, TV, reading
- 

## Radiographic evaluation

In papers I and II, all patients and controls had standing radiographs taken of both knees in 15 degrees of flexion, with a Siemens Basic Radiological System (film focus distance 1.4 m; Siemens GmbH, Erlangen, Germany) at 70kV and 10mA. All the radiographs were obtained with the same standardized technique by the same technician. The frontal views of the tibiofemoral joints from both knees of the patients and the controls were classified according to the recommendations of the Osteoarthritis Research Society (Altman et al. 1995). The radiograph of each operated knee was compared with the same-side knee view from the controls. A radiographic atlas was used to evaluate the appearance of the joint spaces and the presence of osteophytes and to grade these features on a scale from 0-3 (Altman et al. 1995). The radiographic data were presented and statistically analyzed in 2 ways, both involving the index knee of the patients and the corresponding (same-side) knee of the controls. The category "Radiographic changes grade A" required the presence of joint space narrowing (JSN) grade 1 or more as the only feature. "Radiographic changes grade B" required the presence of JSN of grade 2 or more, or JSN grade 1 combined with osteophytes. The more advanced changes under the latter category represent a more "stringent" version of Kellgren-Lawrence knee OA grade 2 (Kellgren and Lawrence 1957).

## Statistics

### *Parametric versus non-parametric analysis*

The underlying data obtained from questionnaires like the KOOS are ordinal, which implies the use of non-parametric statistics. However, means and standard deviations are often calculated instead of medians and inter-quartile ranges for this type of questionnaire data (Liang et al. 1985, Tegner and Lysholm 1985, Ware 1988, Bellamy 1995, Wright and Young 1997). In most cases during this study both parametric and non-parametric analyses have been performed. In no

case did the choice of method change the interpretation of the data. In accordance with European traditions, non-parametric statistics have been used throughout all papers.

### *Reliability*

Bland (Bland 1995) suggested that measurement error for questionnaire scales should be presented as the correlation coefficient between pairs of readings. The intraclass correlation coefficient (ICC), a special form that does not take into account the order in which observations were taken, is preferred for this application. The random effects intraclass correlation coefficient (Shrout and Fleiss 1979) was calculated to assess test-retest reliability in paper III–V. In study V percentage agreement of individual items was calculated to identify single items with poor test-retest reliability.

Given concerns regarding the dynamic nature of many chronic diseases, an alternative approach to test-retest reliability is to determine reliability from a single application of the technique using measures of internal consistency. The most frequently employed statistic is Cronbach's alpha (Cronbach 1951). The method penalizes items showing poor inter-item correlation in a questionnaire. If the inventory of a questionnaire is relatively homogenous and unambiguous, then the inter-item correlation will be high. The possible range of values is 0–1.0. A Cronbach's alpha of more than or equal to 0.80 is generally regarded as acceptable (Bellamy 1993).

### *Validity*

Spearman's correlation coefficient ( $r_s$ ) was used to assess construct validity.

### *Responsiveness*

Postoperative change across all times was assessed by Friedman's test, and postoperative change at specific follow-ups was assessed by Wilcoxon's signed rank test. Responsiveness was calculated by effect size, defined as mean score change divided by the standard deviation of the preoperative score (Kazis et al. 1989). An effect size over 0.8 is regarded as high (Cohen 1977). A p-value less than or equal to 0.05 was regarded as significant in all papers but paper V where a p-value less than or equal to 0.01 was regarded as significant. No adjustments were made for multiple comparisons.

### *Differences between groups*

Kruskal-Wallis test was used to determine overall differences between more than two groups. The Mann-Whitney U-test was used to determine differences between two groups.

*Dimensionality*

In paper V, principal component factor analysis on each KOOS subscale was performed to determine if the individual items loaded on a single factor. Failure to load on a single factor suggests that the items in the scale do not all estimate the same aspect. An eigenvalue criterion of 1.0 was used for these factor analyses (Norman and Streiner 1986) and the results are given in terms of the percentage of variance in the scale score explained by the principal factor.

*Influence of potential predictors*

In paper VI, the influence of potential predictor variables on postoperative 'knee-related quality of life' was analyzed by means of linear regression analysis. First, the influence of each potential predictor was assessed in simple regression analyses. Those predictors which implied a p-value less than 0.20 were considered further in a multivariate regression analysis. Backward stepwise elimination was then employed to obtain the significant predictors. Model fit was checked by analyzing residuals (Altman 1991).