FACULTY OF HEALTH AND MEDICAL SCIENCES
UNIVERSITY OF COPENHAGEN

PhD thesis

Jonathan D. Comins

Measuring Symptoms, Function, and Psychosocial Consequences in Patients with Anterior Cruciate Ligament Rupture

Development and Validation of a Patient-related Outcome Questionnaire

Submitted July 2012

Accepted for defence October 2012

PhD thesis

Measuring Symptoms, Function, and Psychosocial Consequences in Patients with Anterior Cruciate Ligament Rupture

Development and Validation of a Patient-related Outcome Questionnaire

Jonathan David Comins, RPT, MSc

Section for Sports Traumatology M51, Department of Orthopedic Surgery, Bispebjerg Hospital
Institute for Orthopedics and Internal Medicine, Faculty Of Health and Medical Sciences
University of Copenhagen

and

Department of General Practice
Institute of Public Health, Faculty of Health and Medical Sciences
University of Copenhagen

Academic advisors:

John Brodersen, MD, PhD
Associate Professor
Research Unit and Department of General Practice
Institute of Public Health, Faculty of Health and Medical Sciences
University of Copenhagen

Michael Krogsgaard, MD, PhD
Professor
Section for Sports Traumatology M51, Department of Orthopedic Surgery, Bispebjerg Hospital
Institute for Orthopedics and Internal Medicine, Faculty Of Health and Medical Sciences
University of Copenhagen

Measuring Symptoms, Function, and Psychosocial Consequences in Patients with Anterior Cruciate Ligament Rupture - Development and Validation of a Patient-related Outcome Questionnaire

Ph.D. Dissertation

Correspondence:

Jonathan Comins
Department of General Practice
Institute of Public Health
Center for Health and Society
University of Copenhagen
Øster Farimagsgade 5, P.O. Box 2099
1014 Copenhagen
Denmark
email: jocom@sund.ku.dk

Cover photo: Jean Schweitzer

1. Preface and Acknowledgements

The *Guidelines for the PhD Programme* of the Faculty of Health Science, University of Copenhagen[1] stipulate that when a PhD a thesis incorporates more than one scientific paper submitted for publication, the written structure of the articles should be accompanied by an "extended summary" of approximately 30 pages. The summary is normally structured as follows:

- A brief, general presentation of the research hypotheses presented in the included articles;
- A brief presentation of the results achieved with an assessment of the methods applied and a critical review of the conclusions that can be drawn from the results;
- A comparison with and assessment of other researchers' published results to the extent that this is relevant to the presentation of the author's contribution to the analysis of the research hypothesis;
- A summary conclusion

The Guidelines also state that the scientific articles/manuscripts can be included as chapters instead of the brief presentation of the methods applied and the main findings.

The form of the present thesis adheres to the above-described structure. The thesis is based on three scientific papers, which have been submitted for peer-review, and an accompanying summary consisting of an introduction, background, and objectives, followed by the three manuscripts submitted for publication. Each paper is followed by a discussion of the major findings, assessment of methods, justification of conclusions, and contribution to current knowledge.

The studies performed in this thesis were carried out from 2009 to 2012 in a collaborative effort between the Department of Sports Traumatology, Institute of Orthopedics and Internal Medicine, Faculty of Health Science, at the University of Copenhagen, and the Department and Research Unit of General Practice, Institute of Public Health, Faculty of Health Science, University of Copenhagen.

Financial support

Funding was provided by the Danish Agency for Science, Technology, and Innovation, and Sahva A/S, a Danish prosthetics and orthotics company.

Origins of project

It was by chance that I met my supervisor, John Brodersen, at a private dinner party in 2004. John, an MD and specialist in general practice, was approaching the end of his own PhD study concerning the psychosocial consequences of false-positive breast cancer screening. He had developed a questionnaire to measure these consequences, and had used an advanced statistical method called Rasch analysis to validate the instrument. I had heard of Rasch analysis, and I was aware that some of the world's leading experts, such as Svend Kreiner, were at the University of Copenhagen. I was working closely at that time with my other supervisor, Chief Surgeon Michael Krogsgaard, in rehabilitation of patients with knee injuries. It was a common occurrence that my patients complained of difficulty answering some of the questions in our standardized questionnaires. Therefore, John, Michael, Senior Scientist and fellow physical therapist Nina Beyer, and I planned to investigate the measurement performance of the questionnaires using Rasch analysis. We published our results in 2008.[2] The present PhD is a direct consequence of that study.

Originally, this PhD project consisted of two major stages. Phase 1 was the development and validation of a condition-specific patient-related outcome (PRO) questionnaire for patients with anterior cruciate ligament (ACL) deficiency pre- and post-ACL reconstruction. The instrument would be subjected to psychometric validation using the Rasch model of item-response theory (IRT).

Phase 2 was to involve computer-generated three-dimensional biomechanical analysis of a cohort of these patients, and the results were subsequently to be compared in a regression analysis using structural equation or log linear chain-graph modeling techniques.

The research process did not progress exactly as planned. The biomechanical analyses are not included in this thesis for two reasons: 1) the timeframe of the development and validation of the condition-specific PRO questionnaire extended beyond the allotted three-year period of the PhD program, and 2) the biomechanical test laboratory, where the testing was to take place, had a technical permanent malfunction of its infrared optical tracking cameras. This rendered viable motion capture impossible.

Acknowledgements

I am grateful for my collaboration with Tine Alkjær and Erik Simonson from the Department of Biomechanics, Panum Institute - University of Copenhagen. From Bispebjerg Hospital, I want to thank Jes Bruun Lauritzen, Nina Beyer, Monika Bayer, Christian Couppé, Pia Andersen, and all the staff and the surgeons at the Department of Sports Traumatology. Also thanks to all my colleagues at Sahva, in particular Christer Levin.

I want to thank my dear friends and colleagues at the Research Unit and Section of General Practice, particularly in the UPPS group.

Special thanks to Hanne Thorsen (for getting this ball rolling, long before I knew it existed, and then passing it on to John Brodersen) and Klaus Witt. Also thanks to Olivia Spalletta and David Stodolsky for editorial assistance and scientific discussions.

Very special thanks to Svend Kreiner and Volkert Siersma. I do not think humans can get much brighter or kinder (imperviously kind!), as is the case with most everyone in the Matilda Bay Club, which is a group of statisticians and scientists, led by David Andrich, committed to using Rasch models for various measurement applications.

Finally, I want most specially to thank my supervisors Michael Krogsgaard and John Brodersen. What can I say about these two fellows? Great minds, great hearts, great clinicians. It is a privilege simply to be associated with them. While Michael is the best surgeon I have had the fortune to work with, John's insight into the human condition, and the challenges we face as clinicians in the justification of our actions, is unsurpassed. In addition, his ability to grasp the fundamental concepts of Rasch analysis, and even more important, employ these together with qualitative concepts to generate useful metrics is nothing short of brilliant. Words cannot express my gratitude.

My intention with this project has been to be as efficient as possible; however, as my friend Timothy Dunne, a mathematics and statistics professor from South Africa, so poignantly stated, "Efficiency without compassion approaches abuse".[1]

This thesis is dedicated to my wife Eva, my sons Niklas and Max, and my family in the United States (my mother, brother, sisters, and particularly to my father, the late Dr. Arnold C. Comins).

[1] Tim Dunne spontaneously uttered this phrase in response to a comment by John Brodersen on the current trend to streamline, at all costs, healthcare systems in Western societies. This was while attending the Fifth International Conference on Probabilistic Models for Measurement in Perth Australia in January 2012, eating breakfast in the 40-C° summer heat (and no air conditioning).

2. Summary

The objective of this PhD study was to construct and validate a questionnaire that can be used to measure the effect of treatment in patients with ACL deficiency. This process was divided into separate studies described in the three articles presented in this thesis.

Study 1 encompassed a literature search to find all questionnaires used to assess outcome in the targeted patient group. The objective was to identify item content deemed suitable for these patients by clinical experts; that is, items that possess face validity. The next step was to translate all “non-Danish” items into Danish and consolidate items with redundant content. Different questionnaires ask many of the same questions. Thus, item reduction was performed retaining only items with unique content. The literature search included 31 PROs, which yielded 539 items in four languages. Because the majority of items were not in Danish, translation was carried out by extracting just the meaningful content of the item. For example, an item such as “In the past week, I have had difficulty walking down a flight of stairs” would be truncated to “difficulty walking down stairs”. These truncated items were then translated to Danish and assessed for content redundancy. The final number of truncated items with unique content was 157. This process was the substantive part of Paper 1. The article was submitted for publication and is now under revision.

The second study involved focus group interviews with patients with ACL rupture, pre- and postoperatively. Each item from the literature search was discussed on an item-by-item basis to ascertain the content relevance for these patients. Thirty-eight items from the initial item pool, five modified items, and twelve items with new content were confirmed to be relevant by the patients in three focus groups and seven single interviews. The result was a 55-item pilot questionnaire with six proposed functional measurement domains. This process is described in Paper 2, which has been submitted for publication and is in review.

In the third study, 242 patients consisting of patients prior to and subsequent to ACL reconstruction were recruited from the ACL registry list at Bispebjerg Hospital. The patients completed the 55-item draft questionnaire. The subjects consisted of three groups: 62 subjects in the pre-operative group, 87 subjects in the first post-operative group (four to 16 months post-op), and 93 subjects in the third post-operative group (at least 28 months post-operative). The responses were analyzed using Partial-Credit and Graphical Loglinear Rasch models. Forty-one items exhibited fit to the Rasch model and thus possess unidimensional measurement characteristics as applied to patients, pre- and post- ACL reconstruction. The items were

distributed across seven constructs and not six as proposed *a priori*. This was because one proposed domain was found to consist of two separate constructs. The seven scales comprise the newly formed condition-specific PRO questionnaire entitled the *Knee Numeric Entity Evaluation Score – ACL (KNEES-ACL).*[2] This scale validation study is described in Article 3, which is included in this thesis. The paper has been finalized and is in the process of being submitted for peer-review.

[2] Definition of acronym: KNEES – ACL

When an item fits a Rasch model, and thus can be used in numeric comparisons, the output score can be considered a numeric-entity, representing the individual's level of the attribute. The acronym Knee-Numeric-Entity-Evaluation-Score came about through a philosophical discussion of measurement theory with David Stodolsky, PhD. He is thanked for his input in this regard.

3. Resumé (Danish Summary)

Formålet med dette projekt var at konstruere et spørgeskema til måling af behandlingseffekt på patienter med forreste korsbåndsinsufficiens. Processen bestod af 3 studier, som er beskrevne i denne afhandling.

Studie 1 omfattede en systematisk litteratursøgning, som skulle finde frem til alle potentielt relevante spørgeskemaer, der anvendes til at vurdere effekten af behandling af den pågældende patientgruppe. Dette var for at identificere de items og begreber, som kliniske eksperter har skønnet, er egnet til disse patienter. Det næste skridt var at oversætte alle de "ikke-danske" items til dansk. Dette var for at kunne vurdere, hvorvidt indholdet af items overlappede (item redundans). Derefter skulle relevansen af items undersøges i forhold til målgruppen gennem fokusgruppe-interviews for at identificere, hvilke elementer der var kvalitativt relevante i forhold til disse patienter. I litteratursøgningen identificerede vi 31 spørgeskemaer anvendt på patienter med sygdom i knæet, der opfyldte inklusionskriterierne. Det blev til i alt 539 items fordelt over fire sprog. Meget få af disse items var oversat til dansk, og derfor blev den planlagte 2-panels oversættelse af logistisk årsager ikke mulig at gennemføre. Derfor valgte mine vejledere og jeg at udføre oversættelsesprocessen på en anden måde. Vi fjernede sætningsstrukturen i hvert item og isolerede det meningsfyldte indhold. For eksempel, et item som "I den forløbne uge, har jeg haft svært ved at gå ned ad en trappe" ville blive til "svært ved at gå ned ad trappen". Derefter oversatte vi disse forkortede items til dansk og vurderede dem for indholdsredundans, hvilket medførte, at det endelige antal trunkerede items blev reduceret til 157. Denne proces er beskrevet i en videnskabelig artikel og indsendt med henblik på publicering i et videnskabeligt tidsskrift.

I Studie 2 blev fokusgruppe-interviews med præ- og postoperative patienter med ACL-ruptur/ACL-rekonstruktion gennemført. Hvert item fra litteratursøgningen blev vurderet i en én efter én analyse for at fastslå indholdsrelevansen for vores målgruppe. Ud af de 157 items blev 38 items direkte anerkendt som relevante for deltagerne, 5 items blev modificerede og 12 items med nyt indhold blev bekræftet som relevante af interviewdeltagerne. Dette resulterede i sammensætningen af et 55-item testspørgeskema, som skulle felt-testes på en kohorte af de samme typer patienter. De kvalitative interviews og processen vedrørende spørgeskemaets konstruktion er beskrevet i artikel II, som er blevet indsendt med henblik på videnskabelig bedømmelse.

I det tredje studie blev 242 patienter fra Bispebjerg Hospitals database over præoperative og postoperative ACL-rekonstruktioner bedt om at udfylde det 55-item testspørgeskema udviklet i

forbindelse med Studie I og Studie II. Deltagerne bestod af tre grupper: 63 i den præoperative gruppe, 87 i den første postoperative gruppe (4 til 16 måneder efter operation), og 94 i den tredje postoperative gruppe (mindst 28 måneder efter operation). Svarene blev analyseret statistisk ved hjælp af loglinear Rasch-modellen, og 41 items fordelt over 7 skalaer fittede en både en Partial-Credit og en Grafisk Loglinear Rasch model. Disse 7 skalaer udgør det nye sygdomsspecifikke spørgeskema, der benævnes *the Knee Numeric Entity Evaluation Score – ACL (KNEES-ACL).* Manuskriptet er under udarbejdelse mhp. publicering, og bliver indsendt til *peer-review* inden den 30. juni, 2012.

4. Introduction

4.1 Who needs surgery?

What is the most effective strategy for the treatment of anterior cruciate ligament (ACL) deficiency, and what do patients with ACL-deficiency experience as the most significant problems due to their knee injury? Is ACL reconstructive surgery the best way to prevent osteoarthritis in these patients? Alternatively, is non-surgical physical therapy sufficient, or even superior? After forty years of highly sophisticated surgical techniques to reconstruct the ligament using augmentation techniques, a variety of autografts and allografts, a multitude of physical therapeutic modalities, and conflicting evidence that between 14 percent [3] and 70 percent [4] of these patients can regain pre-injury levels of function through non-surgical treatment; the evidence is still not clear.[5-7] A randomized clinical trial showed there was no difference between patients treated surgically and those treated non-surgically at two years follow-up.[8] However, these results have been questioned due to doubts as to the validity of the outcome measure.[6, 9, 10] A fundamental requirement in order to answer such questions is the use of valid outcome measures. With appropriate measurement methods, it should indeed be possible to identify which patients will most likely benefit from surgical versus non-surgical treatment, and thus which of these modalities is most efficacious for the patient with ACL deficiency.

4.2 Anterior Cruciate Ligament Deficiency

4.2.1 Mechanisms of injury

The primary function of the ACL is to restrain the forward translation of the tibia relative to the femur as well as to restrict axial rotation between these two long bones during weight bearing. Multiple mechanisms and risk factors for anterior ACL deficiency have been described and as the focus on general participation in sports has increased in modern society, the mechanisms responsible for ACL deficiency have been the subject of intense scrutiny.[11-19] The majority of lesions to the ACL are incurred during running, cutting (twisting), jumping, and landing maneuvers in connection with sports where the knee is subjected to excessive valgus and axial rotational moments near extension.[11, 20-22] While there is a distinction between *non-contact* and *contact* injuries, the vast majority of ruptures occur in conjunction with some form of contact with the ground. Thus, the *ground reaction force* (GRF) is ultimately the external force responsible for rupturing the ACL, due to increased inter-segmental valgus/rotation torques. More specifically, the lateral femoral chondyle twists and glides on the lateral tibial plateau, which creates a fulcrum between the femur and tibia that cannot be mechanically constrained by the ACL. Numerous factors increase the risk of ACL deficiency.[14, 23, 24] Extrinsic factors include weather conditions, type of contact surface, shoe type, and interactions between these factors. Intrinsic factors include anthropometric aspects (such as femoral and tibial length, increased Q-angle, and interchondylar notch width), gender, hypermobility, body mass index, hormonal, and neuromuscular mechanisms.[5, 19, 23-26]

Figure 1 shows an ACL rupture.

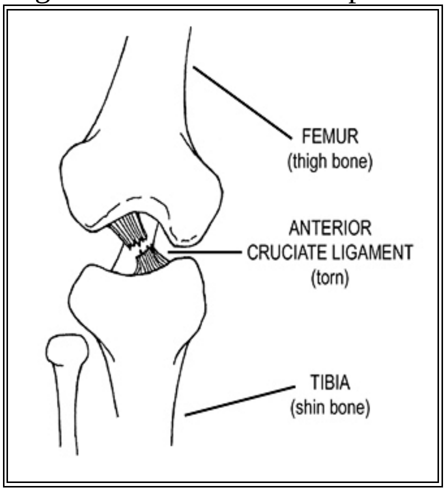
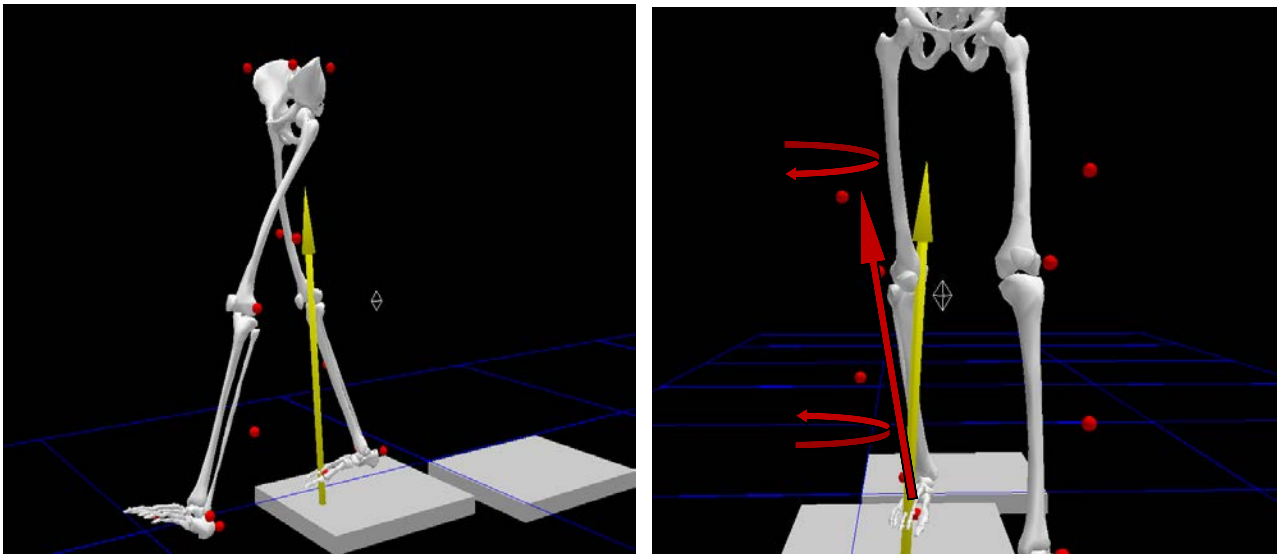


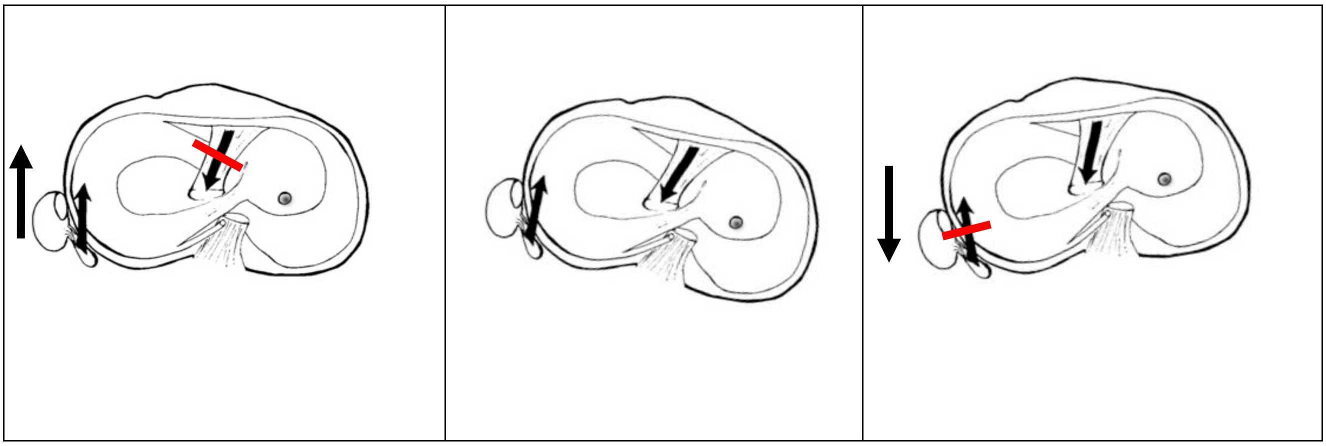
Figure 1 shows the relationship between the femur, tibia, and the ACL. Figure 2 illustrates how the longitudinal rotational axis of the knee passes through the medial compartment and the mechanisms of rotational instability. Secondary constraints to tibio-femoral excursion include capsular and neuromuscular contributions from the popliteus and hamstrings musculotendinous structures.[27, 28] Thus, the causal mechanisms contributing to functional instability are highly multifactorial, depending on anatomical, biomechanical, neuromuscular, as well as psychological input.

Figure 2.



The figures show normal GRF vectors (yellow) relative to the knee joint center of rotation. For ACL rupture to occur, the GRF will be lateral to the center of knee rotation in the coronal plane in combination with oppositely directed axial tibio-femoral rotations at near extension (red arrows).

Figure 3.



The rotation of the lateral compartment is controlled by the ACL, the popliteofibular ligaments, and the popliteus tendon. ACL deficiency results in internal tibial rotation (left). Rupture of the popliteofibular ligaments and the popliteus tendon results in external tibial rotation (right). One or both of these lesions can result in functional rotational instability and subluxation of the lateral femoral condyle on the lateral tibial plateau.[29] (Illustration from Krogsgaard, 2007)[29]

4.2.2 Incidence and socioeconomic aspects of ACL deficiency

Trauma of the ACL is one of the most common injuries to the knee joint in industrialized society.[5, 29-31] Incidence rates per 100,000 person-years are reported to be around 1200,[32] which corresponds to 1.2 percent per year. These numbers seem to be consistent in Western countries, in general.[33] Incidence rates in high-risk activities and sports, such as soccer and football, are increased by as much as a factor of ten.[34, 35] Between 3000 and 3500 ACL reconstructions are performed annually in Denmark, a country of five million inhabitants.[36, 37] The economic impact of ACL deficiency has not been determined in Denmark. However, in the United States, approximately 200,000 ACL reconstructions are performed annually, costing around $3 billion.[18] ACL deficiency is known to increase the risk of knee osteoarthritis [38-40] and is present in around twenty-three percent of these patients.[39] The prevalence of knee osteoarthritis in industrialized countries is ten percent and affects roughly thirty percent of persons aged seventy.[40-42] The socioeconomic impact (total direct and indirect costs) of osteoarthritis is estimated at 0.3% of the gross national product,[42] which corresponds annually to nearly $8 billion in Denmark.

4.2.3 Diagnosis of ACL deficiency

The diagnosis of ACL deficiency is dependent on the history of injury, clinical evaluation, and paraclinical confirmatory analyses. The patient presenting with an ACL deficiency will describe injury events, which correspond with the mechanisms touched upon in Figures 1 and 2: Buckling of the knee during directional change of motion, most often in a sporting situation, and usually with an external valgus moment imparted on the knee. The patient reports there was immediate swelling of the knee, inability to continue the activity, and possibly an “audible pop.” Depending on the acuteness of the injury, the level of activity and the degree of mechanical instability, the patient may exhibit continued swelling and giving way during weight-bearing tasks. Rotational instability is the most common cause of symptoms during functional activities.[29] A clinical diagnosis is confirmed by Lachman’s test, Anterior Drawer test, and Pivot Shift test. Lachman’s test is the Gold Standard of clinical tests with high sensitivity and specificity (eighty-five and ninety-two percent respectively). The Pivot shift test, with a sensitivity of twenty-four percent, is not adequate to reveal ACL deficiency; however with specificity of ninety-eight percent, the test combined with Lachmans is very useful.[43] Magnetic resonance imaging (MRI) can also be used to confirm diagnosis; however, sensitivity and specificity is not enhanced relative to clinical examination. MRI is clearly warranted for assessment of multiple injuries to concomitant knee structures such as meniscus or cartilage lesions.[44, 45]

4.2.4 Quantifying diagnostic efficiency - sensitivity, specificity, and predictive values

Sensitivity, specificity, and the positive and negative predictive values of diagnostic tests are important from a socioeconomic perspective. They are the foundation for choice of action, that is, whether and how to treat. The objective of diagnostic tests is to identify pathology in patients who present with symptoms that might indicate the presence of pathology (sensitivity) or exclude patients without the condition (specificity). For example, persons who might at some point have experienced a swollen knee after some traumatic event, and who test positive for having ACL deficiency, may in fact have an intact ACL (false positive), and persons with normal test results may in fact have a rupture (false negative). See table 1.

Table 1. Diagnostic tests.

	Persons with rupture	Persons without rupture	Total persons
Positive test	A (TP)	B (FP)	A + B
Negative test	C (FN)	D (TN)	C + D
Total tests	A + C	B + D	A+B+C+D

Sensitivity: A/(A+C);
Specificity: D/(B+D)
Positive predictive value (PPV): A/(A+B)
Negative predictive value (NPV): D/(C+D)
TP: True Positive finding, FP: False Positive finding
FN: False Negative finding, TN: True Negative finding

Possible outcomes of tests: “A” is the number of persons with ACL deficiency identified by the test, the true positives. In contrast, “C” is the number of persons with ACL deficiency that the test misses (i.e., the false negatives). “B” is the number of normally functioning persons with an abnormal test result, the false positives. “D” is number of normally function persons with a negative test result, the so-called true negatives.

The positive predictive value (PPV) of a test is simply the proportion of persons with a positive test who, in fact, have the condition relative to the total number of positive tests. This will depend on the particular group to which the test is being applied, as the prevalence of ACL deficiency will invariably be higher in soccer and basketball players than in a group of rowers and in the general population at large. The prevalence is the total number of persons in a population who have the condition. Thus, the PPV of the test will be significantly greater for field and indoor court sportsmen than rowers and cyclists. Conversely, the negative predictive value (NPV) is the number of individuals who in fact do not have the condition relative to the total number of negative tests. The NPV will thus decrease in groups of patients with increased prevalence, such as soccer and basketball players.

In practical terms for the clinician, these concepts can be used to assess the probability of a patient actually having a certain condition, such as ACL deficiency. Using Fagan’s Nomogram of Bayesian statistics,[46] a clinician who knows the base rate of a condition in a certain population (e.g., the incidence of ACL deficiency in female basketball players) can use the prevalence (or alternatively, the likelihood ratio in Fagan’s Nomogram)[46] to calculate the probability of the person having the condition.

4.2.5 *Copers verus Non-copers*

Individuals classified as having ACL deficiency have been categorized as Copers and Non-copers.[47-51] Copers achieve asymptomatic pre-injury levels of activity despite ACL deficiency and represent between 23 and 33 percent of these patients.[18, 52] The treatment for Copers is strictly non-surgical, primarily consisting of proprioceptive-, progressive resistive strengthening-, and agility exercises. Non-copers will most often receive the same non-surgical treatment regimen; however, as they experience knee instability even during basic functional tasks, this usually warrants surgical intervention.[29] The importance of the terms Copers and Non-copers becomes apparent when deciding when a diagnosis of ACL deficiency is clinically significant, that is, represents a condition, which is meaningful for the patient. Non-copers are, as the name implies, unable to cope with the condition of their knee. This indicates that they are unable to function at a level that is satisfactory for them. This fact is important to consider when applying diagnostic tests and screening for patients who may be deemed to benefit from a specific intervention, such as surgery or physical therapy, since the objective of such treatment is to ensure a satisfactory level of function. The concept of satisfactory function, and exactly what this entails, is important. As Copers function normally, one could claim that they are functionally without disease, which may imply that the concepts of ACL deficiency and Non-copers are synonymous. However, compounding the problem is the fact that there are three overlapping diagnostic constructs involved: ACL deficiency, Non-copers, and the anatomic ACL rupture. The degree to which these converge is not very clear.

4.3 Treatment Strategies for ACL Deficiency

In Denmark, the general practitioner, when treating a patient who presents with a knee injury, may choose a “wait and see” strategy, refer the patient to physical therapy, or refer the patient directly to an orthopaedic specialist. Individuals with acutely severe symptoms or patients with persisting symptoms despite physical therapy (Non-copers) are at some point referred to an orthopaedic specialist. These Non-copers receive two options in the attempt to move into the “healthy” category: Non-surgical treatment, consisting of physical therapy, or surgical reconstruction of the ACL. The decision to reconstruct or not is primarily based on the degree of functional impairment experienced by the patient and the expected or desired level of functional activity.[33, 53] Moknes et al.[4] found that in a group of patients with ACL deficiency classified as Non-copers, 70 percent of these patients could be reclassified as Copers, after an aggressive physical therapy regimen and thus were able to forego ACL reconstruction. The patients included in the study were diagnosed using *Magnetic Resonance Imaging* (MRI), had at least 3 mm anterior tibial translation in a

Lachmans test, participated regularly in high-level activities and sports, and had history of a knee injury. However, risk factors for ACL deficiency in different sports are quite heterogeneous [34] and depending on which MRI criteria are used, the sensitivity, specificity, and accuracy of MRI to confirm ACL deficiency (and thus Non-copers) are quite divergent.[44, 45]

4.3.1 Non-surgical treatment

The course of treatment after the acute knee injury in Denmark is:

- The patient is most often seen by a general practitioner (GP) or at the emergency room where the general diagnosis knee sprain is given. The patient can receive a knee brace, possibly crutches, and is instructed to apply the principles of RICE (Rest, Ice, Compression, and Elevation), and paracetamol and/or non-steroid anti-inflammatory drugs can be prescribed.

- If pain and swelling persist more than two weeks, the patient can be referred by the GP to an orthopaedic specialist for knee assessment. Diagnostic tests of ligament instability will be performed and the patient is often referred to physical therapy.

- In physical therapy, a progressive rehabilitation regimen is instigated. The objectives of rehabilitation are to manage pain, reduce swelling, re-establish normal range of motion, normalize strength and balance, and above all restore the sensation of a normally functioning knee, specifically considering the patient's habitual level of activity and goals for treatment.

4.3.2 Surgical intervention (ACL reconstruction)

Treatment intervention will consist of ACL reconstruction for patients classified as Non-copers, that is, patients with positive diagnostic test results and non-responders to physical therapy. Most often autologous graft material harvested from the patellar tendon, the quadriceps tendon, or the hamstring tendons (semitendosis-gracilis) of the patient are used, although allografts are also commonly used for multi-ligament injuries or revision surgery. Reconstruction can be anatomic or non-anatomic single or double bundle depending on surgical preferences and a number of other factors. Postoperative rehabilitation regimens vary to a degree between hospitals and between the private and public sectors in Denmark. Most patients complete a two to four month rehabilitation trajectory under the supervision of physical therapists. Accelerated rehabilitation ad modem Shelbourne is the most prevalent method of progression, where aggressive range of motion, early weight bearing, progressive resistive strengthening, and functional exercises are emphasized.[54]

4.4 Measuring Outcome in Patients with ACL Deficiency

Valid outcome measurement is paramount for assessment of treatment effect. The goal is accurate evaluation of changes in end-point variables compared with the baseline status of the patient. When looking into the effect of intervention for ACL deficiency, the success criterion must be the degree of "normalization of function" (or absence of dysfunction). Thus, the concept of function must be operationalized in order to measure the degree of success of a given intervention.

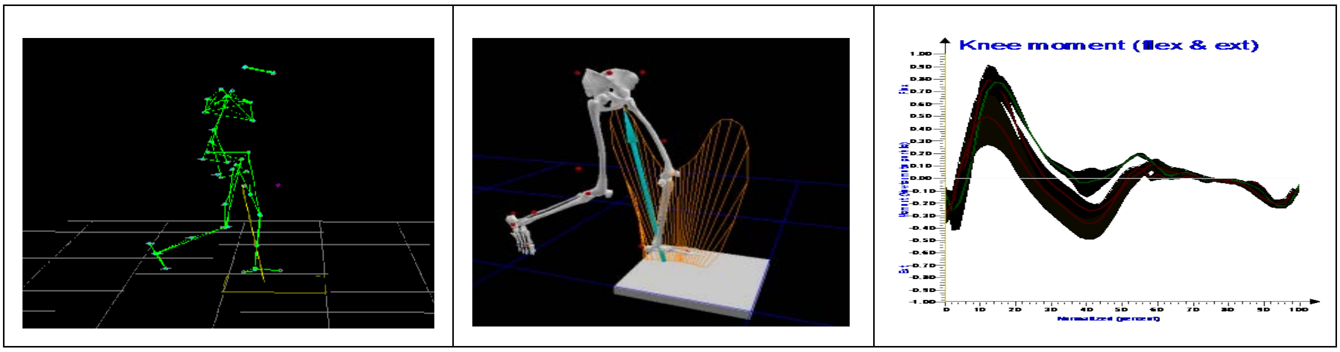
4.4.1 Physical function

General outcome assessment of ACL deficiency includes the diagnostic tests as mentioned in section 4.2.4. Typically, the physical therapist will apply a "return-to-sport" functional criterion assessment.[55] A general rule has been that if the isokinetic or isometric thigh strength, rate of force development, various balance or proprioceptive tasks, running, cutting, and hop tests for distance are at least 80 percent, as compared with the uninvolved side, then the patient is ready for re-entry into pre-injury participation in sporting activities.[22, 53, 55, 56] The cut-off ratio of 80 percent is apparently arbitrary, but seems useful for practical purposes despite recent criticism of the concept.[57] Together, these tests can help guide the patient back towards normalized function and activity level. However, the tests yield rather gross estimates of function and are non-specific in terms of objectively quantifying knee function.

4.4.2 Biomechanical measures of function

Increasingly, video-based instrumented 3-dimensional (3-D) motion capture analysis is being applied clinically to quantify biomechanical knee function and is commonly used in clinical research.[58-62] The method has been used in healthy subjects to identify loading patterns, which can place individuals at risk for ACL injury,[22, 63] and it has been suggested for clinical assessment of patients with ACL deficiency and ACL reconstruction attempting to re-enter competitive level sports activities.[22] This method makes it possible to quantify physically detailed components of knee function.[60] The technique involves placing retro-reflective markers on specific anatomical landmarks of the body. With the use of special infrared cameras, computer software, and synchronisation algorithms, the 2-dimensional coordinates of the markers are reconstructed as 3-D coordinates in space. The 3-D motion data combined with anthropometric parameters from the major body segments is used to construct a biomechanical link-segment model (Figure 3).[62] Using raw data derived from 3-D force plates and inverse dynamic equations of motion,[62] net forces and moments of force about each joint can be calculated.[47, 62, 64, 65] Advantages of the method include the extraction of parameters such as peak knee moment of force at a specific instant in time during the gait cycle, and it has been used to quantify biomechanical deficits during functional tasks in ACL deficiency patients.[47, 49-52] Non-copers and patients with ACL reconstruction were found to have diminished peak knee flexion and significantly reduced extension moments during gait.[49, 66] Inverse dynamic analysis of forward lunge has also revealed temporal and kinetic differences between Copers and Non-copers.[47] These methods are promising; however, there are some drawbacks in relation to the clinical setting. The equipment is costly. Time-consuming aspects include placement of the retro-reflective markers, the acquisition of motion capture data, and the treatment and analysis of the data. Furthermore, there is a degree of kinematic artefact due to marker movement, depending on the type of functional activity that is recorded, as well as considerations concerning biomechanical modelling algorithms and tri-planar axes of joint rotation.[67-70] However, the method does yield quantitative data of functional activities and substantial advances have been made to rectify some of the above-mentioned disadvantages.[69, 70]

Figure 3.



Three-dimensional biomechanical analysis of functional tasks can quantify functional anomalies. The figure to the left shows the 3-D link segment model and the curves to the right are internal knee moments.

4.4.3 Patient-related outcome (PRO)

While physical and biomechanical measures of outcome can help quantify a patient’s level of function from the perspective of the clinician, they do not take into consideration how patients experience, feel, and interpret their own level of functional capacity, or success of the treatment. PRO questionnaires have become an integral component of functional evaluation to determine the effects of surgical intervention, and PROs have attained a level of importance comparable to physical examination.[71] Most PROs are reported to have been validated in some manner,[72, 73] which implies that they can quantify the degree of functional change. Numerous PROs have been developed to evaluate health outcome from the perspective of the patient for a variety of conditions.[73] Some instruments combine individual question-response scores to produce a single over-all index (a sum score), while others yield multiple sub-scores for separate functional domains.[74] Comparisons of different scoring instruments reveal discrepancies in how each self-assessed parameter is reflected in the score.[75-80]

PRO questionnaires can be defined as "any report coming directly from subjects without interpretation of the physician or others about how they function overall or feel in relation to a condition and its therapy."[81] Thus, PROs address the patient's own perception of function, treatment satisfaction, and other aspects deemed relevant by the patient in relation to a certain condition.[82-84] PRO data are most often collected via standardized questionnaires designed to assess underlying constructs not directly measurable, such as pain (or other symptoms), or the ability to carry out functional tasks. Such underlying constructs are referred to in psychometric theory as *latent traits* or *latent variables*. Individual items (an item is a question and its response categories) are grouped into one or more domains, depending on the concept they represent. Table 2 exemplifies items and their response options in a small hypothetical activity scale. A scale in this context is defined as the sum of at least two item-response scores. The item statement is the qualitative content or theme of the item and the response options are the categorical choices representing the endorsement of the item theme. Items with dichotomous response options, such as "Agree/Disagree" or Yes/No", address only whether the item is endorsed or not. If the item has a polytomous response structure as seen in Table 2, then the amount of that endorsement is also addressed. Each option represents a number (e.g., "Not at all" is scored as 0, "A bit" as 1, and so on). This item-response score signifies the amount of item content the person possesses. These response scores are added together and the summary score represents the level of activity for the person who completes the questionnaire. Conceptually, this is the basis of what is known as a *total score* or *sum score*. Thus, the range of possible sum scores in the scale in Table 2 is zero through nine, that is, from least to most functionally impaired.

Table 2.

Item statement	Response options			
	Not at all	A bit	Quite a bit	A lot
1. I have difficulty walking.	☐	☐	☐	☐
2. I have difficulty running.	☐	☐	☐	☐
3. I have difficulty jumping.	☐	☐	☐	☐

A hypothetical 3-item activity scale as an example of a PRO

4.4.4 PROs versus "surrogate" measures

The current rise in popularity of PROs in the health sciences may be attributed to the American Food and Drug Administration (FDA), which emphasizes the use of PROs stating that "physical examination and performance testing is not sufficient to measure what may be most important from the patients perspective."[85] Editor in Chief of the British Medical Journal, Fiona Godlee, expresses that endpoints stemming from the patient, such as visual impairments or quality of life, are "hard outcomes that matter to the patient" and should be prioritized over "surrogate end-points", such as biomarkers. Surrogates will often show much larger and faster responses to treatment and thus are preferred by the biomedical industry.[86] Clinical and surrogate measures are important to confirm diagnoses and establish treatment strategies. However, the most clinically relevant outcomes must be those, which address the patient's perceived response to the condition or the treatment. The clinician can record an objectively perfect result; however, in order to utilize this positive outcome, the patient must endorse the perceived result. Thomée and colleagues [57] stress that despite achieving acceptable levels of physical function as dictated by common return-to-sport criteria, many athletes never reach the goal of actually returning to sport. The authors argue that patient-reported themes such as fear of re-injury should be specifically emphasized, and that physical measures are insufficient to address the demands of high-level activity. Furthermore, physical performance measures do not necessarily correlate well with PROs, as has been demonstrated in numerous studies.[53, 87-89] For example, a recent study compared various hop tests with the *International Knee Documentation Committee* (IKDC) and the *Knee-injury and Osteoarthritis Outcome Score* (KOOS) questionnaires and found low associations in three of the five subscales of KOOS.[90] Interestingly, these same subscales were found to lack unidimensionality in a Rasch model (see section 4.6.12 for more on Rasch analysis) when applied to patients 4-6 months after ACL reconstruction.[2] Conversely, Fuchs et al. found significant correlations between the *Knee Society Score* and sagittal plane knee range of motion and coronal plane knee moments derived from inverse dynamics,[91] and Grindem et al. recorded that a single hop test significantly predicted self-reported knee function in patients with ACL deficiency.[92] A methodologically strong study by Mizner et al. comparing physical performance test outcomes with widely used patient-related measures found that the outcome results did not reflect each other after total knee replacement.[93]

4.5 Measuring Outcome in Danish Patients with ACL Deficiency

4.5.1 Physical performance assessment

The Danish National Registry for Cruciate Ligament Injury has stipulated that certain tests should be used to measure outcome in ACL reconstruction. As a routine protocol at the Copenhagen University Hospital – Bispebjerg, patients scheduled for ACL reconstruction complete a series of tests to assess physical performance pre- and postoperatively. These tests include single-leg hop tests for distance, instrumented thigh strength, measure of thigh girth, and other balance–related measures. The tests are well defined, widely used worldwide, and are considered a gold standard.[51, 53, 94-97] One advantage with physical tests is that they are convenient to perform and standardize, and the data are easy to acquire, process, and tabulate. They are also easily understood by the patient. The downside is that the resultant score does not necessarily reflect the spectrum of relevant outcome from the perspective of the patient.[81]

4.5.2 PRO assessment

The Danish ACL registry requires that all hospitals and private clinics in Denmark, where ACL reconstructive surgery is performed, report the results of the *Tegner Activity Scale* and the KOOS questionnaire preoperatively and at 1, 5, and 10 years post-surgery.[98] Prior to 2008, physical therapists at Bispebjerg were manually responsible for the administration and scoring of these PROs. A common occurrence among physical therapists at Bispebjerg Hospital was that patients complained of not understanding the content of certain questions, for example, in KOOS, or patients could be confused as to which response option to choose.[2]

4.5.3 Ambiguous items

Responses to items are more likely to yield spurious reflections of the patient's condition when the respondents are confused by the item content.[83, 99-102] An example of an item patients with ACL deficiency found difficult to complete in KOOS was: "Do you feel grinding, hear clicking or any other type of noise when your knee moves?" The response options are "Never", "Rarely", "Sometimes", "Often", and "Always". What is the correct response if the patient experiences clicking but not grinding, or other noises, but not clicking? There is only one response option available to answer the question. Items of this type are referred to as *double-barreled* or *ambiguous items* and should be avoided, or at least rephrased and/or split into separate items in the questionnaire development phase.[2, 102] Physical therapists at Bispebjerg were instructed to tell the patient to complete the response that "fits best" when the patient was ambivalent as to an item. Another problematic item in KOOS was "Can you straighten your knee fully?" with the same five response categories as above. Patients responded often that they either could or could not straighten the knee fully, so they had difficulty choosing the appropriate response category. Intuitively, these items should be dealt with in a qualitative manner to ease their use for the specific patients they are intended for, particularly because they are an integral part of a measurement tool. However, the ambivalence patients experience when responding to such ambiguous items is also quantitatively detectable using certain statistical methods.[103, 104]

4.5.4 Analysis of problem items

A closer look at the origin of KOOS can explain why such items were included in the instrument. From its inception, KOOS was designed to measure patient-related function in individuals with an acute knee injury as well as patients with degenerative knee disease.[105] This was in order to capture functional deficits in the acute phase of injury and to track the progression of functional problems in patients with chronic development of osteoarthritis of the knee. The strategy for creating KOOS involved the inclusion of the entire 33-item *Western Ontario and McMasters Osteoarthritis Index* (WOMAC)[3] and the addition of nine new items, which addressed issues pertaining to sport/recreational activities and quality of life. Thus, KOOS is a generic instrument generated from items targeting chronic degenerative osteoarthritis, acute injury, and multiple joints. An investigation into the psychometric properties of KOOS revealed that three of the five proposed scales in the instrument did not fulfill the requirements of a measurement scale.[2] The study showed that the WOMAC subscales of *Knee Symptoms, Pain,* and *Activities of Daily Living* were not appropriate measurement constructs for patients who were 20 weeks post ACL reconstruction. Table 3 shows the results of the overall fit to the partial-credit Rasch model for polytomous items as described in Comins et al. (2008).[2] The asterisks demarcate the WOMAC domains. The concepts in Table 3 are explained more thoroughly in section 4.7 of this thesis.

Table 3.

Dimensions (Number of items)	WP-χ^2	DF	*p*	PSI	α
Knee Symptoms (7)	24.65	14	0.038*	0.70	0.63
Pain (9)	36.11	18	0.007*	0.81	0.80
ADL (17)	67.16	34	0.001*	0.87	0.91
Sport/Rec (5)	9.17	10	0.516	0.81	0.80
QoL (4)	2.91	8	0.940	0.75	0.75

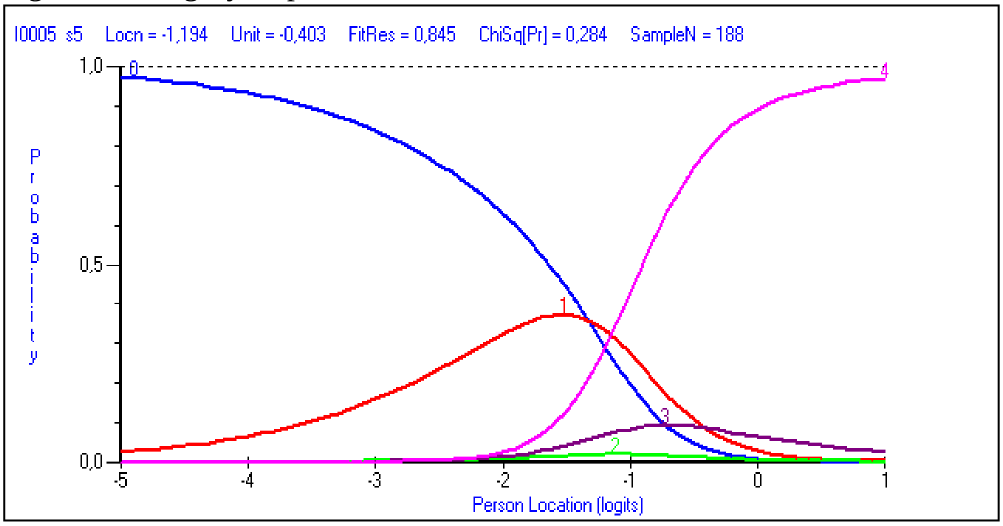
Wright-Panchapakesan (WP) fit statistics, Person Separation Index, and the Cronbach's alpha of five dimensions in the KOOS.[2] The asterisk* indicates significant chi-square misfit.

Rasch analysis revealed that the ambiguous items for patients in the clinical setting were also statistically problematic. For example, the response options of "Sometimes" and "Rarely" in the item: "Can you straighten your knee fully?" were found to be insufficient for patients with ACL reconstruction (see Figure 4). These are the two flattened-out curves at the bottom of the graph. Furthermore, the response category "Often" (the red curve "1" in Figure 4) was qualitatively

[3] The WOMAC is one of the most widely used questionnaires for assessment of osteoarthritis of the knee and hip joints.[106]

problematic, for how "often" can a person be expected to be able to straighten the knee "fully?" Full extension would most expectedly be either possible or not possible for ACL deficient individuals, and therefore a question of this type might be more appropriate if the response options were "Yes" and "No."

Figure 4. Category response curves of an item.



Category Probability Curves of item S5 "Can you bend your knee fully?" The scale from -4 to +1 symbolises the latent trait of knee symptoms, with the severity of symptoms increasing towards the right. Response categories: 0 "Always", 1 "Often", 2 "Sometimes", 3 "Rarely" and 4 "Never".

4.6 Constructing PRO Questionnaires

Many aspects must be considered when developing questionnaires for use as measurement scales.[100] Most importantly, the specific patient group targeted for assessment and the specific condition to be measured must be thoroughly contemplated.

4.6.1 Condition-specific versus generic questionnaires

PROs can be developed to address specific conditions, or more general health concerns, so-called *condition-specific* and *generic* questionnaires.[107] Proponents of generic PROs cite the fact that they are reported to have well-established validity and, therefore, often are used as criterion measures for validation of new instruments.[108] Moreover, they are promoted as a means of comparing the impact of disease and treatment across populations, or to compare acquired patient data with normative values.[100] However, this practice has been criticized, as the use of generic questionnaires for cross-disease comparisons is scientifically questionable because items can have different meanings in different patient groups.[100, 109] Furthermore, a number of studies show that the results are inconsistent when such scales are applied across patient populations.[2, 110-112] Generic

PROs are not designed to capture issues of concern for the specific patient groups and, thus, are more likely to inquire about irrelevant themes. This can increase the likelihood of respondent alienation and missing data, particularly in more severely affected patients.[2, 113, 114]

PRO questionnaires developed for generic applications are often used in condition-specific studies. Such instruments as the Short Form 36 (SF-36)[115] and the EuroQoL (Eq-5D)[116] can be used in any number of ways. Pitfalls associated with the application of generic PROs have led to an increase in the development and use of condition-specific questionnaires.[100, 101, 117-119] Condition-specific PROs are developed to assess aspects of outcome that are important for specific patient populations. Condition-specificity entails qualitatively deriving the content of items directly from the targeted patient populations (see section 4.5.6). Condition-specific PROs have the advantage that they provide more detailed information than generic questionnaires and are thus more likely to be sensitive to disease- or treatment-specific effects.[100, 119]

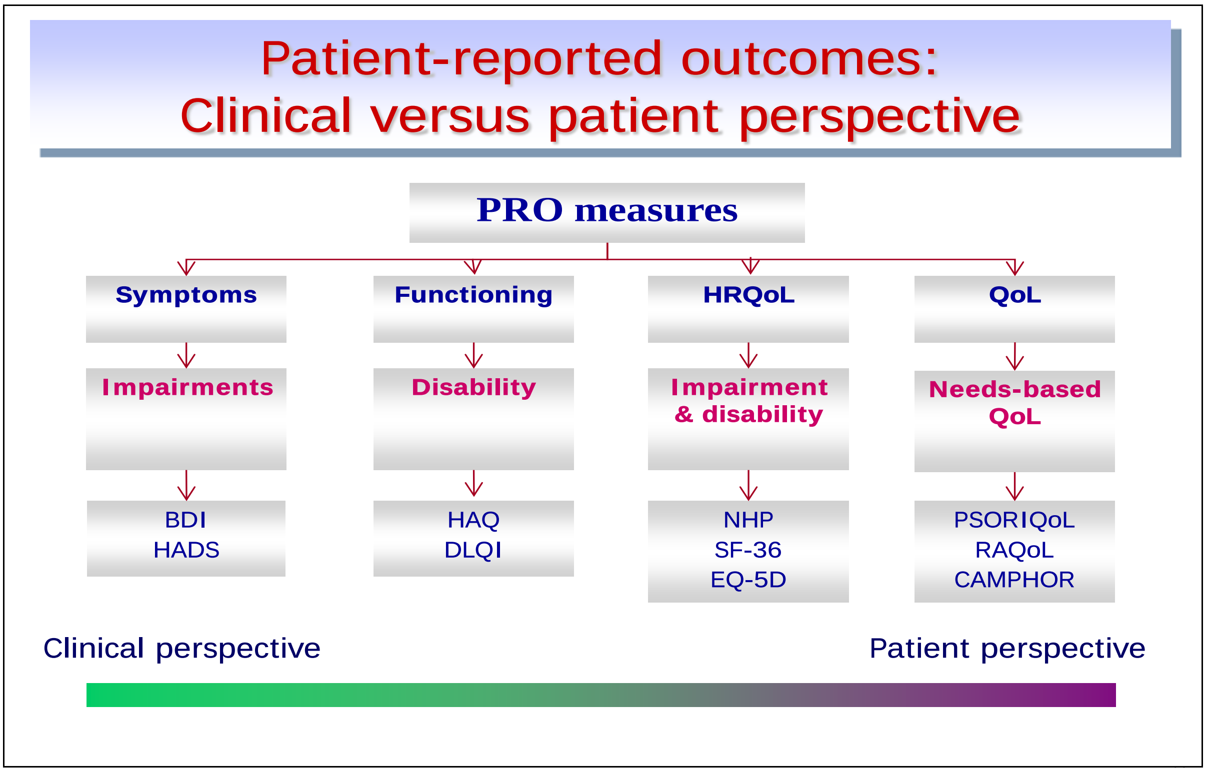
4.6.2 Assembling the questionnaire

Depending on which constructs are deemed relevant, a theoretical model to address the appropriate item content in relation to the targeted patient population must be chosen in order to establish a viable frame of reference to achieve measurement. A myriad of PROs exist for different types of clinical assessment. The most common constructs are measures of:

- Impairment (symptoms)
- Activity limitations (functioning)
- Participation restrictions
- Health status / Health-related Quality of Life (HRQL)
- Quality of Life
- Treatment satisfaction

The constructs of impairments (symptoms), activity (limitations) and participation (restrictions) have been clearly defined in the International Classification of Functioning, Disability and Health (ICF; WHO, 2001) and are widely accepted. These constructs are particularly relevant for PROs used in clinical research, where measures targeting bodily function, activity level, and participation in life's situations are particularly useful. Figure 5 shows the continuum between at the one end clinically related conditions, where patients with musculoskeletal pathology such as ACL deficiency are more appropriately targeted within the ICF categorizations (Impairment, Activity, and Participation). At the other end of the spectrum, the Needs-based model may be better suited to address constructs such as Quality of Life or satisfaction with life's general conditions, for example, in connection with life-threatening diseases.

Figure 5.



Conceptual model of function and disability (with permission from Brodersen)[100]

4.6.3 *Item content generation*

Item content must be generated once the frame of reference and the model have been established. There are essentially two sources of item content: a) the clinician (or clinical expert), and b) the patient.

4.6.4 *Face validity*

Item content derived from the clinical expert will according to Mosier[120, 121] possess *face validity*, in that the instrument will be "considered to be valid, if the sample of items appears to the subject matter experts to represent adequately the total universe of appropriate test questions." This item content can be derived through individual- and focus group- interviews of clinicians or from assembling items from existing PROs created and used by clinical experts.[100, 122-125] The main purpose of collecting item content from the perspective of the clinician is to identify possibly relevant constructs for the patient, based on the fact that the clinical expert has extensive diagnostic and treatment experience relative to the particular patient group.

4.6.5 Content validity

Content derived from the clinician may or may not be relevant for the patient. There is only one method to confirm content relevance, and that is through direct patient confrontation.[84, 100, 118, 126] Thus, after item generation from the clinician, the next step is to confront the specific patient groups with the assembled items and ascertain which items can be endorsed in order to ensure well-functioning items that will be perceived as relevant and comprehendible by the target group.[100, 126] This can be accomplished in single or focus group interviews where content relevance and content coverage can be assessed. Content relevance is addressed through item endorsement. Content coverage, also termed comprehensiveness, addresses the emergence of new items, themes, and constructs that were not previously identified by, say, previous analysis of pre-existing instruments. In practical terms, this really involves the assessment of the saturation of themes patients find relevant, where verbatim comments from the participants can be used to define any new themes and constructs that may emerge.[100, 127, 128] Content relevance and content coverage are the cornerstones of content validity.[83, 102] Confirmation of which items to include in which groups of items is also instrumental in the establishment of relevant condition-specific constructs. Furthermore, item reduction based on feedback from the patients is critical in to order confirm that item content considered directly irrelevant by the particular patient group is discarded. Finally, in order to ensure valid responses to items, the questions must be worded such that they are easily understood by all potential respondents, regardless of age, socio-demographic background, or other personal factors, and appropriate response categories must be established. This can be attained within focus groups and through field-testing in follow-up single interviews.[100, 102, 118, 129]

4.6.6 Construct validity

The last step in creating a valid questionnaire for measurement is to establish the psychometric properties of the instrument. A scale can be defined as the operationalization of magnitudes to construct a variable: “When a variable has been constructed, magnitudes of the properties in entities, which are restricted to persons in this volume, can be measured” (Andrich p. 9)[130] These entities are the items derived from qualitative processes, such as those described above. However, it cannot be assumed that items numerically quantify anything of importance to the persons being measured, without an assessment of the dimensionality of the scale/measure; that is, the degree to which the sum-score actually reflects the level of a person’s ability. Denny Borsboom questions whether the scores are “no more than relatively arbitrary summations of item responses” (Borsboom p. 2).[131]

Scale validation is an immense topic in measurement theory, and there are myriad ways of interpreting, describing, and establishing the validity of measures, depending of course on the objective of the measurement. Samuel Messick contends that the concept of construct validity engulfs all aspects of criterion, concurrent, and in fact all aspects of validity: "The essence of unified validity is that the appropriateness, meaningfulness, and usefulness of score-based inferences are inseparable and that the unifying force behind this integration is the trustworthiness of empirically grounded score interpretation, that is, construct validity."[132] Denny Borsboom goes on to say that validity is not at all complex; the instrument is valid if it measures what it purports to measure: "… a test is valid for measuring an attribute if and only if a) the attribute exists, and b) variations in the attribute *causally* produce variations in the outcomes of the measurement procedure." [Italics added] (p. 150)[131] He writes further:

> Somewhere in the chain of events that occur between item administration and item response, the measured attribute must play *a causal role in determining what value the measurements outcomes will take; otherwise, the test cannot be valid for measuring the attribute.* Importantly, this implies that the problem of validity cannot be solved by psychometric techniques or models alone. On the contrary, it must be addressed by *substantive theory*. Validity is the one problem in testing that psychology cannot contract out to methodology [emphasis added]. (Borsboom, p. 151)[131]

Substantive theory might for example be biomechanical or physical corroboration of PROs used to address musculoskeletal pathologies in case-specific populations.

4.7 Classical Test Theory versus Item Response Theory

In questionnaire development, there are essentially two schools of thought: Traditional methods, known as Classical Test Theory (CTT), and modern test methods known as latent-trait theory or item-response theory (IRT).[133, 134] Conventionally, when CTT methods are used to assess dimensionality, methods such as exploratory factor analysis, correlation, and Cronbachs alpha are employed.[135] Modern test theorists maintain that these methods are insufficient for establishing dimensionality and reliability in data with a categorical response structure.[104, 133, 136-141] However, once the dimensionality of the variable can be established, confirmatory CTT techniques are warranted and commonly used, for example, in conjunction with assessment of interactions between items, and between person factors and items.[142] (See also section 4.5.15 on sources of misfit).

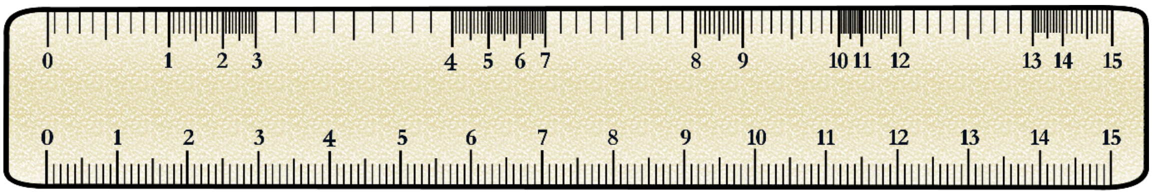
4.7.1 Validating scales with Classical Test Theory and the total score

When questionnaires are validated using CTT, the item-response scores are summed into a total score, which then is used for comparisons and calculations in parametric statistical tests, such as T-tests, Analysis of Variance, and Cronbach's alpha. Parametric techniques assume parametric data structure, which means continuous interval level data, normal distribution of data, relatively homogenous variance in the groups being compared (homogeneity), and equal variance of the residuals (homoscedasticity). At the most basic level, this means that the entities of measurement consist of numeric values that can be added, subtracted, multiplied, and divided as equally weighted units and that the properties represented by these numbers are equal in type and magnitude.[130] Thus, the summary or total score of these numeric entities represents the amount of whatever is being measured. When trying to capture non-physical latent traits, this will always be "more" or "less" of a construct or attribute, such as happiness or pain.[104, 130, 143, 144] As an example, one can consider a situation where ten patients complete a ten-item activity questionnaire. As in Figure 4, the items are polytomous with five response categories, which can be scored from zero to four (i.e., 0, 1, 2, 3, and 4). This means that for ten items the maximal score for a person can be forty. If, say, two persons have a score of twenty-six, this would imply that the persons are equally affected in terms of the latent trait (i.e., they have the same activity levels). However, this assumes that all the items exhibit the same quantity of the underlying construct being measured. One person may have answered three on six items and two on the remaining items. The other person might have responded with a score of three on six other items, four on one item, and two on the last two items. Both combinations add up to twenty-six; however, if the questionnaire consists of items concerning activities such as walking on level ground, stair climbing, and running, then the degree of difficulty associated with the different tasks must of course vary. One person may have scored higher on the most demanding activities, yet both persons appear equally active. In this case, the total score can be misleading because it is based solely on the number of confirmed items and not which items are endorsed. A total score in CTT does not take into consideration which items are more or less difficult or which persons have more or less ability.

Another weakness of CTT is that techniques such as exploratory factor analysis are sometimes used to assess dimensionality for the purpose of item reduction. For example, items that do not load on the first principle component (e.g., eigenvalue less than one)[145, 146] may indicate that the item does not belong to the dimension, and thus might be discarded regardless of content.[108, 147-149] This *data-driven* item reduction method can be problematic in that there must have been a reason for the item being present in the item pool in the first place, and correspondingly, there should be a solid qualitative reason for removal of the item. Good *theory-driven* item generation and item reduction should also include qualitative assessment of the content of all items going in and out of an instrument.[100, 127, 128]

Yet another issue with CTT is that the ordinal item-response structure does not fulfill the assumption of continuous interval data, which in itself warrants the use of other analytic models.[150] The scale must be stable in order to yield valid and repeatable measurement, regardless of which person is being measured, and which part of the scale is being used (the easier or the harder end of the scale or the easier or more difficult items). This means simply that the scale must be *invariant* relative to that which is being measured. Fundamentally, this requires equidistant measurement points along the entire scale.[151, 152] The ruler in Figure 6 illustrates the difference between an ordinal and an interval ruler (scale). Measuring objects with the ordinal edge of the ruler using the CTT paradigm obviously leads to mixed results, in that, the measurements will differ depending on which part of the ruler is applied.

Figure 6. (Illustration from John Brodersen)[153]



Ordinal

Interval

Ordinal versus interval scale

4.7.2 Validating scales with Item-Response Theory

As CTT methods are not equipped to handle categorical/ordinal data, a solution is to use item-response theory (IRT).[103, 104, 152, 154] Whereas, CTT emphasizes only item parameters (e.g., the summary score of items), the focus in IRT models ”is on item and person parameters, which are non-linear transformations of raw scores, and on variances of these estimates”(Andrich, 1988).[130, 155] Thus, in IRT models, both the item and the person scores are manifested as a single value on the same latent variable. Intuitively, this makes sense, also in terms of the fact that it is the person’s ability that is the object of measurement, and not just the difficulty of the test.[155]

4.7.3 Fundamental measurement and Questionnaires

In the physical sciences, fundamental measurement entails a so-called “2-way frame of reference” (Andrich p. 17).[130] For example, to measure the weight of an object, a force must be applied. The measured mass of the object (the output) is an expression of the interaction between the object and the force (an agent) acting on it. Likewise, for measurement to take place using latent variables, a similar 2-way frame of reference must be established.[130, 152] The target population must be clearly identified, that is, patients with a specific pathological condition to be measured (e.g., ACL deficiency). The patients correspond to the objects to be measured. Then, the relevant scale(s) used to measure the objects must be constructed (e.g., Symptoms, Range of Motion, or Activity). This is analogous to the force when measuring a mass.

4.7.4 The Guttman Scale and Unidimensionality

People can be characterized by multiple properties; however, in order to construct a latent variable for measurement, the variable must be identified and mapped on a single real number line. This is known as a *unidimensional* construct (Andrich p. 9).[130] Unidimensionality implies that comparisons in differences of degree (and not just kind) can be made. Unidimensionality is the basis of IRT and can be explained in terms of Guttman scaling. A Guttman scale is considered the ideal in measurement theory.[104, 156] If a scale exhibits a Guttman pattern, then the scale will be unidimensional (NOTE: This is an ideal that does not occur in reality; however, the closer, the better.). A Guttman pattern is illustrated in Figure 7. The table shows ten hypothetical persons with ACL deficiency who might be asked to complete a questionnaire on the occurrence of pain when participating in different activities. For simplicity, the response options are scored dichotomously: "0" for no pain, and "1" for pain during the activity. The persons are represented by letters (rows), and the items addressing different activities are represented by numbers (columns). The pattern observed shows an example of a perfect Guttman structure in a 2-way frame of reference. Person A experienced pain in only one item and is the least affected person. Item J is the least pain-provoking item, with only one person affirming the item. This pattern is rather convenient, for the persons and items are all lined up across from each other. However, which patients were most (and least) affected was unknown before applying the questionnaire, and which activity was most and least pain provoking for the patients was unknown. Therefore, in reality, the data would not line up as neatly as is the case in Figure 7. In Figure 8, the scores do not follow any pattern – or do they? As mentioned, *a priori*, it was not known which patient would affirm most items or which item would be affirmed by most patients. However, if the persons and items are reordered from highest to lowest according to the number of items affirmed, and the number of persons who affirmed the items, the pattern in Figure 8 is transformed to the pattern in Figure 9. In Figure 9, it becomes apparent that person I (blue) has affirmed all items and was most affected, person H affirmed only one item and thus was least affected.

Figure 7.

Items

Persons	1	2	3	4	5	6	7	8	9	10
A	1	0	0	0	0	0	0	0	0	0
B	1	1	0	0	0	0	0	0	0	0
C	1	1	1	0	0	0	0	0	0	0
D	1	1	1	1	0	0	0	0	0	0
E	1	1	1	1	1	0	0	0	0	0
F	1	1	1	1	1	1	0	0	0	0
G	1	1	1	1	1	1	1	0	0	0
H	1	1	1	1	1	1	1	1	0	0
I	1	1	1	1	1	1	1	1	1	0
J	1	1	1	1	1	1	1	1	1	1

A Guttman pattern (Brodersen 2006)[153]

Figure 8.

Items

Persons	1	2	3	4	5	6	7	8	9	10
A	1	1	1	1	1	1	1	0	0	1
B	1	1	1	1	1	1	1	1	1	0
C	0	1	0	1	0	0	0	0	0	0
D	1	1	1	1	0	1	0	0	0	0
E	1	1	0	0	1	0	0	0	0	0
F	0	1	1	1	1	1	1	0	0	0
G	1	1	1	1	1	1	0	0	0	1
H	0	1	0	0	0	0	0	0	0	0
I	1	1	1	1	1	1	1	1	1	1
J	0	1	0	0	1	1	0	0	0	0

A more realistic pattern (Brodersen 2006)[153]

Thus, item two in Figure 9 (red), which was affirmed by all persons, is the most pain-provoking item. The yellow responses are those, which deviate from the expected response, as compared to the expected Guttman scale as in Figure 7. It is apparent that some patients have responded in an unexpected manner to certain items when considering the level of difficulty of the item relative to

Figure 9.

Persons	Items:	2	5	4	6	1	3	10	9	7	8
	H	1	0	0	0	0	0	0	0	0	0
	C	1	0	1	0	0	0	0	0	0	0
	J	1	1	0	1	0	0	0	0	0	0
	E	1	1	0	1	1	0	0	0	0	0
	D	1	0	1	1	1	1	0	0	0	0
	F	1	1	1	1	0	1	0	0	1	0
	G	1	1	1	1	1	1	1	0	0	0
	A	1	1	1	1	1	1	1	0	1	0
	B	1	1	1	1	1	1	0	1	1	1
	I	1	1	1	1	1	1	1	1	1	1

A probabilistic Guttman pattern (Brodersen 2006)[153]

all the other items, and considering the person's ability relative to all the other persons completing the questionnaire. For example, person D did not affirm item five, despite the fact that item five provoked pain in seven other persons and was the second most pain-provoking item. Conversely, person D was provoked by item three, which is clearly less pain provoking than item five. This pattern is reflective of what actually happens in real life. Sometimes observed responses are not as expected. There is a discrepancy between the observed and the expected response. IRT models allow this type of probabilistic imperfect response pattern as seen in the yellow-green zone of Figure 9. Another way to visualize this probabilistic relationship can be seen in Figure 10. The likelihood of affirming the item, (score of 1), decreases as the likelihood of not affirming the item increases (score of 0). For the pain scale, this means that as the zero increases, the probability of being affected (having pain) gets smaller.[153]

Figure 10.

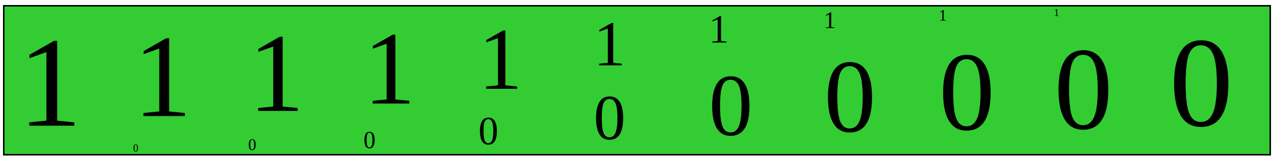


Figure 10 shows the probability of one response being replaced by another. (Brodersen 2006)[153]

4.7.5 The Rasch model

The Rasch model is a probabilistic item-response function,[152] where the dichotomous model is simplest to explain in practical terms.

The dichotomous Rasch model is:

$$p(\beta) = \frac{e^{(\beta - \delta)}}{1 + e^{(\beta - \delta)}} \qquad (1)$$

As can be seen in formula (1), the Rasch model is a logistic function, which states simply that the probability "p" of a person with the ability "β" to affirm an item with a level of difficulty "δ", is a function of the natural log of the difference between these two parameters.[130, 152, 157]

This means that when the parameter estimate of the item difficulty (δ) is equal to the parameter estimate of the person's ability (β), there will be a 50/50 chance of affirming the item, because $\beta - \delta$ will then be zero, and e^0 is equal to one. Thus, by substituting $e^{(\beta - \delta)}$ with e^0 in formula (1), the probability becomes ½. When the item parameter (δ) is greater than the person ability parameter (β), the likelihood of affirming goes from 0.5 towards zero (the probability decreases). Conversely, if β is greater than δ, the probability increases from 0.5 towards one. As the Rasch model calculates the probability of a given response, it does not make distributional assumptions. The importance of this becomes apparent when considering the purpose of questionnaires. Items in a questionnaire are meant to measure the ability of the person responding to the questions, such that the set of items constitutes the scale (agent) used to quantify the person's (object) ability in a 2-way frame of reference.

Possibly the best analogy to explain the concept of the Rasch model is to consider high jumpers of different abilities attempting to jump over a bar set at different heights.[158] The probability that a jumper will clear the bar depends on the ability of the jumper and the height of the bar. The height at which the jumper clears the bar on every other attempt, fifty-percent likelihood, is that particular jumper's level of ability. The height of the bar is of course the level of difficulty of the task (item). Note that in this context, the bar height is assumed to be a unidimensional construct; the higher the bar, the more difficult the objective.

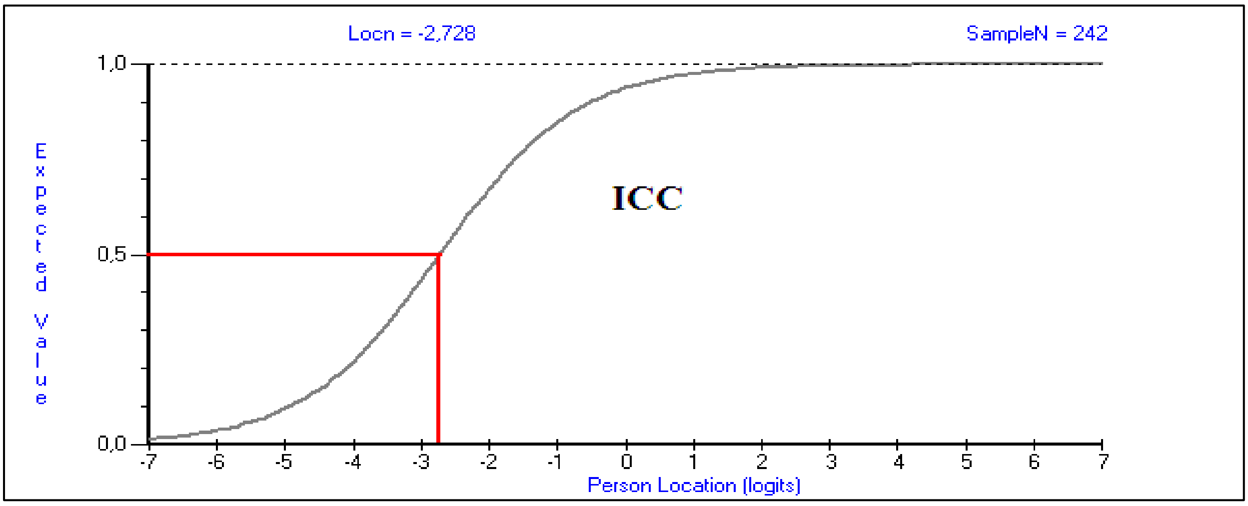
4.7.6 Item characteristic curve (ICC)

This probability of "clearing the bar" can be rendered as an *item characteristic curve* (ICC) [130, 159] (not to be confused with intra-class correlation coefficients), which is a graphical expression of the Rasch model in equation (1). Thus, the ICC predicts that as a person's ability increases on the latent variable (x-axis), so will the probability that the person will affirm the item (y-axis). Figure 11 exhibits an example of an ICC. The point of inflection of the curve as it is projected onto the latent scale (x-axis) is known as the *item threshold.* This is the point at which a person has a fifty percent likelihood of affirming or not affirming the item (red line). The item threshold is considered the item's level of difficulty, as each item will have a different location along the latent variable. The Rasch model calculates the ICC as the expected value of an item based on the way all persons have responded to all items and the way in which all items have been responded to by all persons. Note that the ICC is the ideal curve, if the data were to fit a Guttman structure.[130, 159] The item in Figure 11 has a threshold of -2.728 on the latent scale. Hence, persons with this ability will have a fifty percent likelihood of affirming or not affirming the item. The result can go either way. Persons with a higher ability than -2.728 will have an increased likelihood to affirm the item, and persons under -2.728 will have a decreased likelihood to affirm the item.

4.7.7 Model misfit

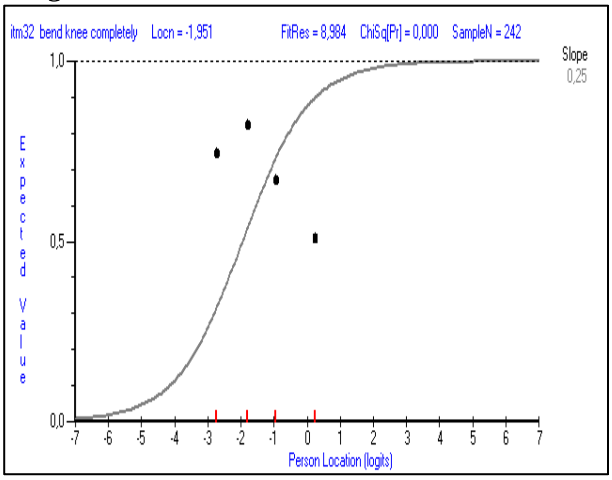
The values of the observed data are plotted and compared with the expected values as generated by the model. Various approaches are used to assess the degree of fit to the Rasch model, although they all employ a likelihood ratio test, which produces a Chi-square statistic.[158, 160-162]

Figure 11.



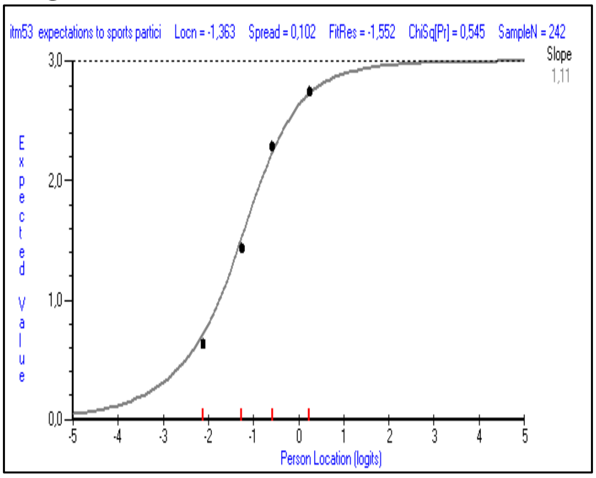
The Item Characteristic Curve: The item threshold (red line) is the point of inflection.

Figure 12A



Misfit to model

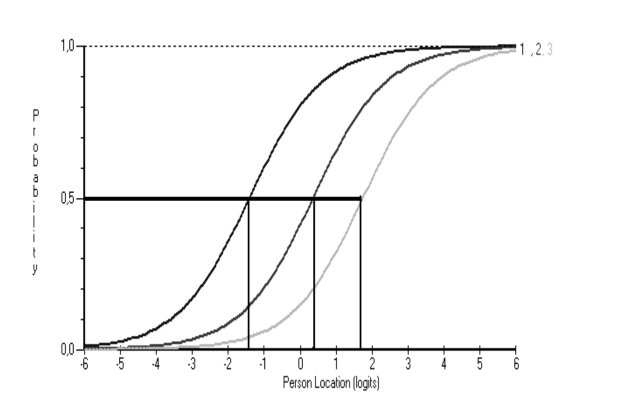
Figure 12B



Good fit to the model

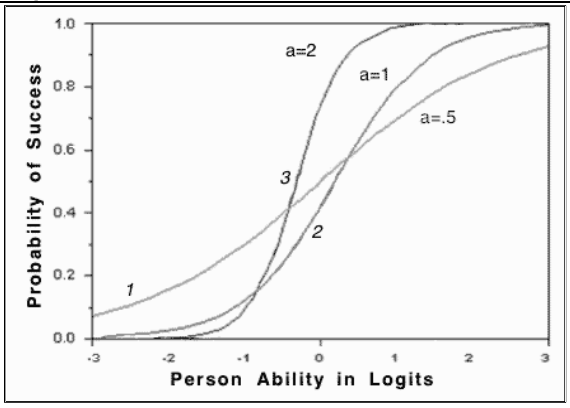
Figure 12A shows an example of observed data with significant lack of fit to the expected model. Conversely, Figure 12B shows a "well-fitting" item where the observed data for all persons nicely converge on the expected ICC. The items in Figures 12A and 12B are, in fact, examples taken from the analyses carried out in paper three of this study. Interestingly, 12A is one of the items that did not function as a polytomous partial credit item in the analysis of KOOS,[2] as shown in Figure 4. In the present study, this same item was applied to patients pre- and post-ACL reconstruction using a dichotomous response structure: "Can you bend your knee completely?": "Yes"/"No". As can be seen in Figure 12A, the item showed severe lack of fit (fit residual = 8.984 [normative range ± 2.5][163] with a chi-square probability of $p<0.000005$), and as can be visualized clearly in Figure 12A, the observed data points diverge substantially from the theoretically derived ICC curve. This means that the item does not work for patients with ACL deficiency in either polytomous or dichotomous forms. Conversely, the item in Figure 12B exhibits adequate fit (fit residual = -1.552, chi-square probability = 0.545) and the observed data points can be seen to visually converge on the ICC. Another aspect worth mentioning, in terms of the ICC, is that in an item set within a Rasch model, all item curves rendered by a Rasch model have the same level of discrimination (and thus slope).[130, 133, 164] If not, then the curves will interact, which could indicate multidimensionality or other anomalies, such as *non-uniform Differential Item Functioning* (see next section).[133] Figure 13A shows three hypothetical items rendered by a Rasch model. Other IRT models purposely model this interaction, such as the 2-parameter logistic model (2-PLM) [165, 166] (see Figure 13B). The only concern with this model is that it allows for a reversal of item difficulty levels for different person ability levels. Item one is easiest at minus one logit with a probability of .30 of affirming the item. As the level goes up to plus two logits, the item is now the most difficult item with a probability of just .70. Items 2 and particularly three are completely reversed relative to item one. It is difficult to imagine when this would actually be the case in real life in the context of ACL deficiency.

Figure 13A.



Rasch IRT model (3 items – same slopes)

Figure 13B.



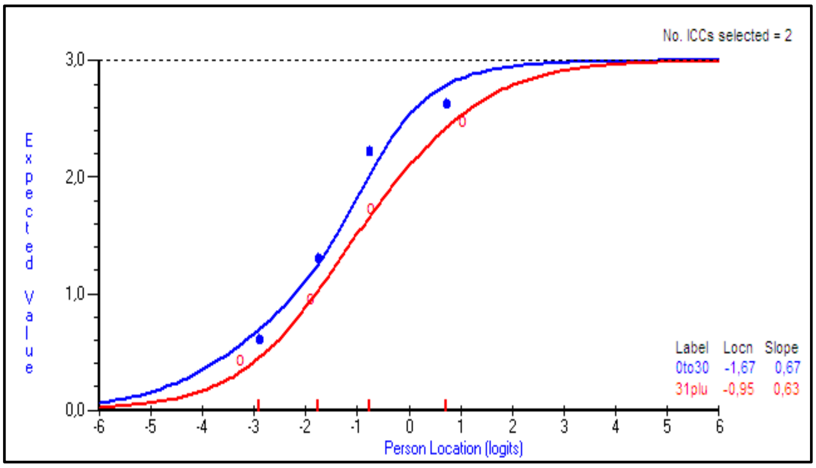
2-PLM IRT model (3 items – different slopes)

4.7.8 Sources of misfit

Fit of the observed data to the theoretically expected ICCs is paramount to establish unidimensionality.[104, 133, 167] If even a single item shows significant “misfit” as in Figure 12A, this can undermine the dimensionality of the entire scale.[168] Sources of misfit can include *multidimensionality,* interactions between subgroups of persons, also called *Differential Item Functioning* (DIF), and interactions among the responses to the items themselves, known as *Local Response Dependency* (LD).[142, 169, 170]

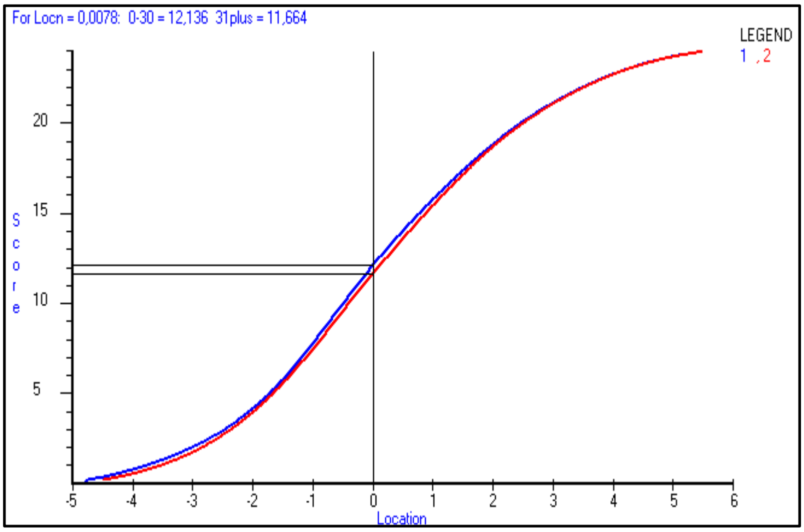
Multidimensionality indicates that a subset of items measures a different latent construct than the other items in the scale. Interactions between item-responses and person factors, such as gender or age, create a type of item bias known as differential item functioning (DIF), which leads to spurious results if not addressed.[168] There are two types of DIF, *non-uniform* and *uniform* DIF[142]. Uniform DIF is present when the ICCs for a covariate, such as age, deviate in a uniform manner across the spectrum of the latent variable (Figure 14.). In other words, the responses are systematically different along the entire range of the latent variable. For example, younger patients may respond differently to an item than older patients, as is the case in Figure 14, which shows an item concerning difficulty when crawling on hands and knees. The blue curve shows that persons thirty years old and under are not as affected (they have less difficulty on the latent scale) as those over thirty. This type of bias can cause misfit, but moreover, it implies that a correction of values should be implemented in order to compare test scores, for example from before and after a treatment intervention.[168] Uniform DIF can lead to misinterpretation of score results, but can, and must be identified, and resolved.[162, 168, 171, 172] When uniform DIF is found, a solution is to ascertain the degree to which the DIF affects the test-score for the group of items and adjust for this difference. This method of compensating for uniform DIF is referred to as DIF *test-equating.*[168] Figure 15 shows how the DIF in Figure 14 is reflected in the test score for the whole dimension (the latent variable). It shows that for comparison of these patients, it is necessary to heighten the test scores of the over-thirty year-old subjects by maximally 0.37 points in the bottom and middle region of the scale, that is, the range corresponding to ± one logit on the latent ADL variable, and corresponding to between eight and fourteen on the raw-score scale. The actual DIF-equating scores, used to generate the curves in Figure 15, are seen to right of the graph.

Figure 14.



Uniform DIF by age-group: 0-30 scores are higher than 31plus.

Figure 15.



DIF-equating. The 31plus patients (red) are slightly lower on the test score.

DIF equating scores:

Score	AGEGRP 31+	≤30
1	1.00	1.16
2	2.00	2.19
3	3.00	3.19
4	4.00	4.19
5	5.00	5.21
6	6.00	6.23
7	7.00	7.27
8	8.00	8.30
9	9.00	9.33
10	10.00	10.35
11	11.00	11.36
12	**12.00**	**12.37 Max**
13	13.00	13.36
14	14.00	14.34
15	15.00	15.32
16	16.00	16.29
17	17.00	17.26
18	18.00	18.22
19	19.00	19.17
20	20.00	20.13
21	21.00	21.09
22	22.00	22.05
23	23.00	23.00

Non-uniform DIF is characterized by interacting ICCs between subgroups, in fact, in the same way as in the example of the 2-PLM in Figure 13B. Truly excessive non-uniform DIF that cannot be resolved may require removal of the item, as DIF in particular effects the overall model fit.[168, 171]

Interestingly, the search for DIF can also induce *artificial DIF*.[173] Artificial DIF is an artifact, which can surface due to simultaneous estimations of DIF in the presence of substantial DIF in one item. This underscores the fact that misfit of single items influences the entire item set. In addition, artificial DIF can surface in the presence of moderate DIF for a single item if the number of items is small. Different methods are suggested to ascertain the difference between real and artificial DIF, which of course is crucial to ensure appropriate removal or retention of items.[173]

Local Response Dependency can also affect model fit and is a phenomenon that until recently had not really been thoroughly addressed. LD indicates that a response in one item depends on the response to another item. In practical terms, it involves some degree of redundancy between items. If you can walk a hundred yards, then you can also walk ten yards. This does not mean that the lesser item needs to be removed. It can be included in a *composite item*,[174] in that retention of items is desirable in order to retain as much relevant item content as possible.[162, 171, 174]

4.7.9 Different Rasch models

As mentioned above, the dichotomous Rasch model is the simplest way to explain the mechanisms of the model. However, various Rasch models exist, and in particular, special cases of the Rasch model have been formulated to allow for polytomous response data (for example the Partial Credit Model (PCM),[103] the Rating Scale Model (RSM),[104] and Graphical Loglinier Rasch Models (GLLRM)).[162] The PCM is a modification of the dichotomous model and allows for the probability of multiple response options to be mapped out along the latent variable. The item in Figure 4 is an example of a partial credit structure. The RSM is a special case of the PCM and assumes equidistance between item thresholds. For example, the distance between the vertical black lines in Figure 13A (and for all items in an instrument) is assumed to be equal.[104] GLLRM excels at identifying DIF and LD and is more sensitive to a broader spectrum of violations to the Rasch model.[162, 171, 175, 176]

4.7.10 Reliability, Responsiveness, and Known-Groups Validity

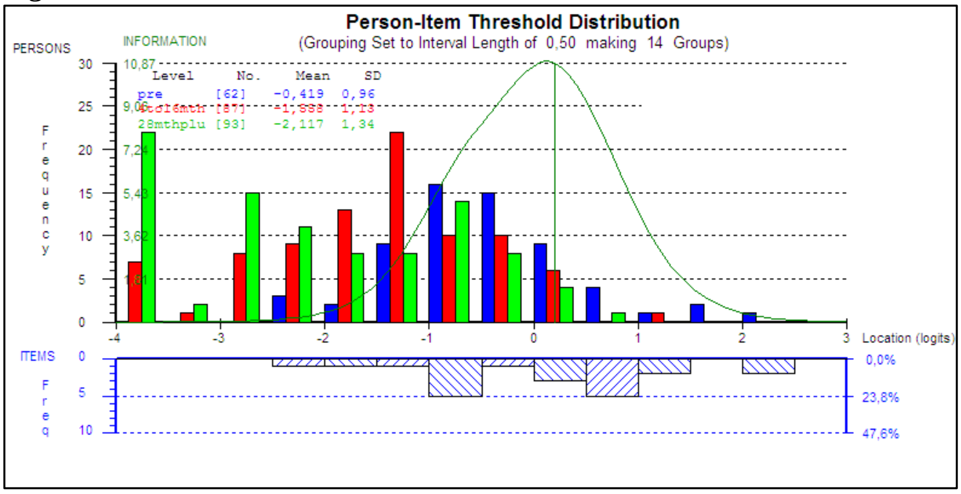
Test reliability is also an essential part of the validation process and can be problematic, because any statement concerning the stability, reproducibility, and internal consistency of a test relates equally as much to the population being tested as it does to the test itself.[177, 178] A summated score must accurately reflect the underlying trait being measured. If the trait is stationary, the score must convey this, and is thus reliable.[179] Conversely, the score must also be *responsive* to change in the underlying trait, if change for example due to treatment truly occurs. As Borsboom states, variations in the attribute being measured must have a causal effect on the values of the outcome measure.[131]

Reliability in CTT is defined by the “true score”, the observed score, and the error score.[178] Many methods are used to calculate reliability; however, baseline techniques used as a reference for others are the Kuder-Richardson Formula 20 (KR-20), for dichotomously scored data, and Chronbach’s alpha (gives a lower bound) for polytomous items.[155, 179] The drawback to these tests, within the framework of CTT, is that they are sample dependent. The reliability indices can be increased by increasing the number of items.[178] A practical example of this was seen in Table 3 in which the two WOMAC domains of Pain and Symptoms, which consist of few items, exhibited low alpha coefficients compared with the 17-item ADL domain. This was despite the fact that the ADL domain exhibited greatest misfit to the Rasch model. In Rasch models, reliability is assessed using the *Person Separation Index* (PSI). The PSI is calculated in virtually the same manner as alpha, except that it is generated from the data after the Rasch factor has been extracted, that is, on unidimensional data. Table 3 shows how the PSI and alpha are virtually the same when the data fit the Rasch model, as can be seen with the two non-WOMAC domains. This supports the implication that unidimensional scales possess internal consistency and are thus reliable.[155, 179]

Responsiveness can be assessed using different methods such as *Receiver Operating Characteristic* (ROC) curves to determine the performance of diagnostic classification tests. ROC curves are applicable when there is a single variable (e.g., ACL deficiency) with two categories by which persons are classified, such as ACL deficient: Yes/No.[180] The ROC Curve procedure is essentially the same idea as described in section 4.1.4 regarding diagnostic rates. Yet another variation of this theme is *Known-groups validity* assessment, which involves comparison of group means in either a within-subject repeated-measures design or between-subject comparison of different groups. The instrument should be able to measure differences between the groups in order to validate the intervention, but also to validate the measure itself. For example, it is assumed that patients will benefit from reconstruction of the ACL. Therefore, there should be a difference in the mean values of the summary scores between preoperative and postoperative patients. Figure 16 shows an example of this concept using results from the Rasch analyses in the present study. The graph shows the frequency of person scores relative to the item thresholds on a latent Symptoms trait (Sensation of Slackness – see papers two and three). The severity of the scale increases from left to right. In other words, the blue columns represent the patients who are most affected in terms of the latent trait (this is the preoperative group) and the green columns are the least affected persons (these are the patients, who are at least twenty-eight months post-op). The pattern indicates that the severity of the problems decreases across the groups from preoperative to postoperative decreases, which are the expected results in terms of the known-groups scenario. The fact that groups two and three have a statistically significant reduction in the level of symptoms relative to group one warrants the comparison of these groups using ROC curve analysis. For example, Group 3 can be used as a reference relative to Group 1.

However, true proof of benefit requires an external base of evidence to substantiate that the patients actually have improved after ACL reconstruction, and such a *gold standard* does not in fact exist. In order to truly confirm causal relationships and demonstrate external criterion validity, some substantive reference base, which should mirror the latent trait, and yet is independent of it, is needed. As previously mentioned in the section on validity, this might be biomechanical or biomedical indices known to reflect the condition of ACL deficiency.

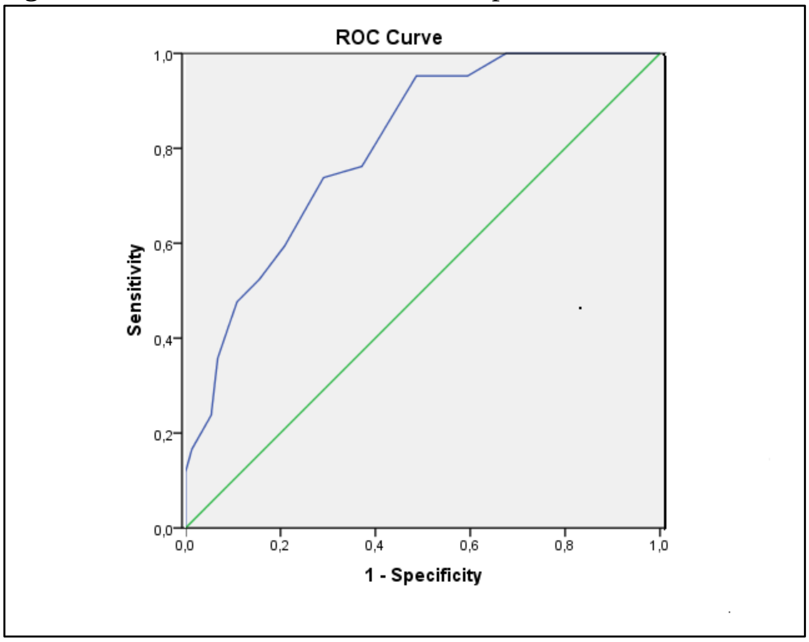
Figure 16.



Example of known-groups difference. The severity of the trait diminishes from right to left. Thus the green group is least affected (≥ 28 months post operative).

A ROC curve generated from the same Slackness scale as seen in Figure 17, comparing Group 1 with Group 3, shows respectable results, in that the area under the ROC curve is 0.805. The drawback to this analysis is that Group 3 is not a gold standard, as there are certainly patients who are not "completely" ACL sufficient.

Figure 17 shows a ROC curve of Groups 1 and 3.



Area Under the Curve Test Result Variable (s): Slackness				
			Asymptotic 95% Confidence Interval	
Area	Std. Error[a]	Asymptotic Sig.[b]	Lower Bound	Upper Bound
0.805	0.034	0.000	0.737	0.872
ROC analysis performed with SPSS, Version 18.				

A final note on the Rasch model in terms of reliability and construct validity is the fact that when an item set fits a Rasch model, the scale will possess what is known as *statistical sufficiency*.[181, 182] This means that the summary score will yield all relevant information about the person, and it is impossible to extract more intrinsic information than is contained in the total score. The dichotomous and partial-credit Rasch models are the only IRT models known to satisfy the requirements of sufficiency.[177, 181, 182]

4.8 Rasch analysis in Physical Therapy

Rasch IRT has been used to evaluate the psychometric properties of some PROs used in the field of Physical Therapy. A Medline search combining "Rasch analysis" and "Physical Therapy" shows an increase in studies involving Rasch analysis over the past several years. The Abilhand Measure of Manual Ability is a "Rasch-built" PRO for rheumatoid arthritis, which dates back to 1998.[183] The same authors have more recently created other instruments using similar methods.[184-189] Campbell et al.[190] published a seminal paper in 1995 on construct validation of a test for infant motor performance using Rasch Rasch analysis, and the study was proclaimed paper of the year in *Physical Therapy*.[191] In terms of knee-specific PRO instruments, the IKDC – Subjective Knee Form has been subjected to a G-IRT Graded Response Model (used for Likert Scales),[192] as well as Rasch analysis using the partial credit model.[193] The Lysholm Scale has recently undergone a modification and validation using Rasch modeling,[194] and the Oxford Knee Score,[195] the KQoL-26, [196] the OAKHQOL,[197] and particularly the WOMAC scores [106, 198-200] have been assessed using the Rasch model. Notwithstanding, no condition-specific PRO for ACL deficiency has yet been constructed using the Rasch model.

4.9 Development and Validation of PROs Using Rasch Analysis – A Summary

The advantage of using Rasch IRT to construct and validate PROs is that it provides construct validation. The construct is generated from theoretical considerations about the latent variable based on empirical and clinical experiences with the condition of interest. The instrument is assembled (items generated) using the information and qualitative characteristics of the construct(s) derived from the target patient groups and experts. The specific target group to be measured is confronted with the measurement scale and the output scores are acquired. An attempt is made to fit an appropriate Rasch model to the data. If there is adequate fit, the scale will exhibit unidimensional invariant measurement for the selected group of individuals. The power of Rasch analysis lies in its confirmatory nature. It simply allows us to ascertain the validity of our measure based on the manner in which people have responded and not by qualitative assumptions of how we believe they ought to have responded. These qualitative assumptions, which optimally are based on solid empirical observations, are the nucleus of constructing a scale to be tested. In formal terms, if the observed response converges on the expected response, as specified by the Rasch model, then the output score is the sufficient statistic. Thus, it yields all necessary information about that person, and the score can be used for comparison across treatment and over time.[160] Only sum scores derived with the Rasch model can be used in this way.

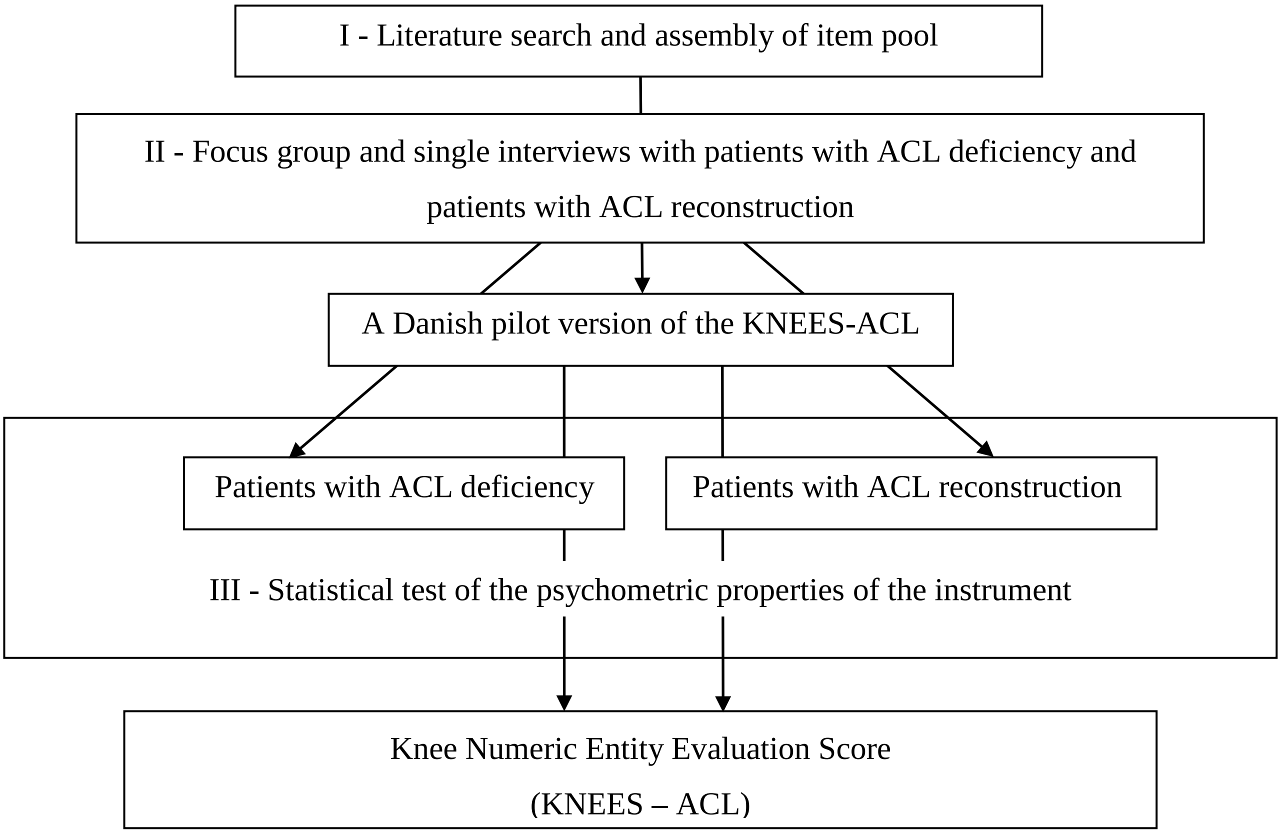
4.10 The Objective of this PhD Study

The overall purpose of this study was to create a PRO questionnaire, which could be used for pretest-posttest studies of treatment for ACL deficiency. This required identifying item content for patients with knee disease, confirming the content relevance and coverage of the items in groups of patients with ACL deficiency and ACL reconstruction, and testing the psychometric properties of this content on a cohort of patients with ACL deficiency and ACL reconstruction. The aims of the three papers were:

1) To review studies where PRO instruments were used to assess outcome in patients with knee disease in order to identify possibly relevant items and constructs for patients with ACL deficiency and ACL reconstruction.
2) To assess the content validity of the identified items from the review and construct a pilot questionnaire for psychometric testing and validation
3) To statistically assess the psychometric properties of the constructed PRO on patients with ACL deficiency and patients with ACL reconstruction.

4.11 The steps of the PhD study

The Roman numerals in the flow chart correspond to the steps of the study described in each of the papers.



5. Article 1 (accepted for publication)

Ensuring Face Validity in Patient-Related Outcome Scores – a matter of content

Jonathan D. Comins [a,b,c,*]

Michael R. Krogsgaard [a]

John Brodersen [b]

[a] Section for Sports Traumatology M51, Department of Orthopedic Surgery, Copenhagen University Hospital – Bispebjerg, Bispebjerg Bakke 23, 2400 Copenhagen NV, Denmark

[b] Department of General Practice, University of Copenhagen, Øster Farimagsgade 5, entrance Q, ground and 1st floor, P.O.B. 2099, 1014 Copenhagen K, Denmark

[c] Sahva A/S, Borgervænget 5, 2100 Copenhagen, Denmark Ø

Key words: Face validity, Content validity, PRO, Knee, ICF

* Address for corresponding author:

Section for Sports Traumatology M51, Department of Orthopedic Surgery, Copenhagen University Hospital – Bispebjerg, Bispebjerg Bakke 23, 2400 Copenhagen NV

Phone: +45 6155 2418

Fax: +45 3532 7946

E-mail: jocom@sund.ku.dk

Abstract

Patient reported outcome (PRO) questionnaires are increasingly used to measure treatment effect in patients with knee pathology. PRO's commonly used to assess outcome in patients with knee conditions can be generic, knee-specific, or condition-specific. Most PRO´s have been created on the basis of clinician-based consensus and are not patient-centered. Items (questions plus their response options) in PROs can be generated by clinicians or through patient interviews. Items created by clinicians possess face validity. The objective of this study was to find all existing PRO items with potentially relevant content for patients with knee pathology. An exhaustive literature search was conducted for PRO questionnaires in English, German, and Scandinavian languages used to assess outcome in patients with knee pathology. The items from the collected PROs were assessed for content redundancy and item reduction was carried out to isolate items of unique content. These items were grouped into one of the components of the ICF classification system. Thirty-one PRO's used for assessment of patients with knee problems were identified, yielding a total of 539 items. Approximately 70 percent of these items consisted of redundant content matter and were reduced to a pool of 157 items of unique content. The identified items can be used to build condition-specific PRO questionnaires for patients with different types of knee pathology.

Introduction

Knee disease imparts substantial economic, physical, and psychosocial consequences on society. The socioeconomic impact of knee osteoarthritis (OA) alone is just under a third of a percent of the gross national product in Western countries [1]. This corresponds annually to nearly half a billion Euros in Denmark, a country of 5.4 million inhabitants. In Western countries, OA affects 10 percent of the general population and roughly thirty percent of persons aged seventy [1, 2]. Rupture of the anterior cruciate ligament significantly increases the risk of knee OA [3, 4] and is present in nearly a quarter of patients with knee OA [5].

Recommendations for treatment of knee disease are increasingly based on trials where patient reported outcome (PRO) questionnaires are used as primary outcome. Frobell and colleagues recently used a well known PRO questionnaire to conclude that patients with ACL rupture are subjected to surgical over treatment [6], despite evidence that the PRO measure employed in the study was found to be insufficient for application on these patients [7, 8]. Such conclusions can obviously influence recommendations for clinical treatment guidelines and thus clinical decision-making. Therefore, it is essential to confirm that the outcome instrument used for specific disease conditions adequately measures the effect of treatment in that group, for meaningful interpretation of results. This is not always the case when PROs are used, as treatment outcome for specific conditions are often measured using instruments designed for broader or completely other purposes [7-11].

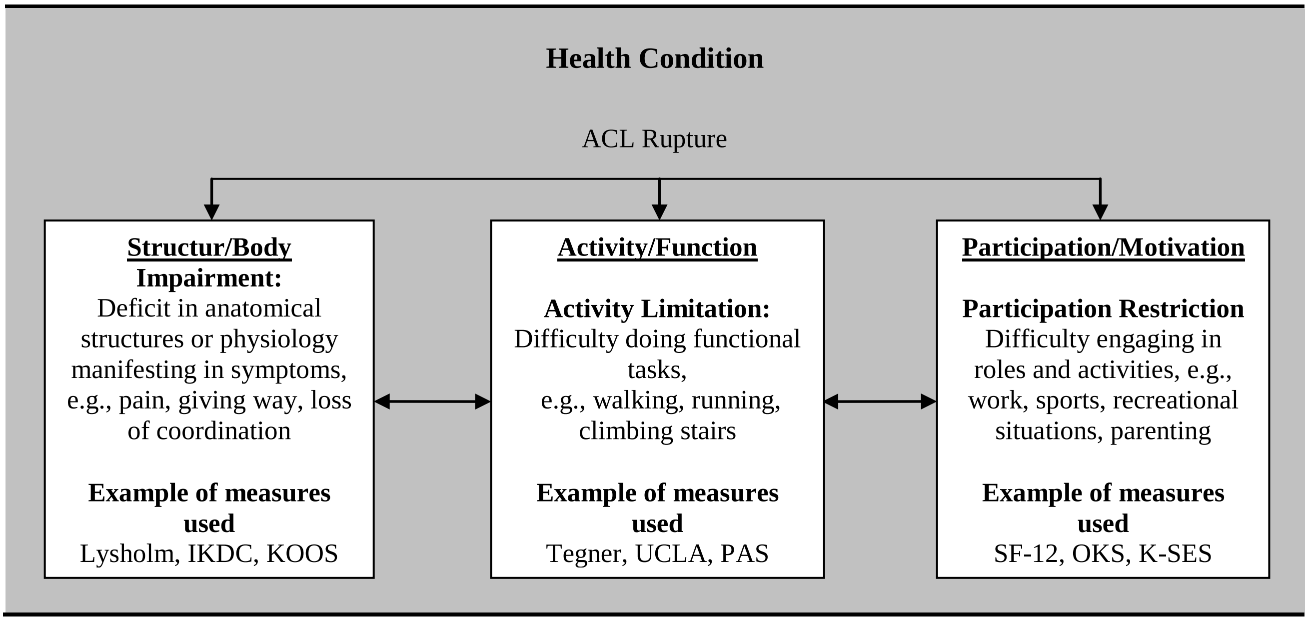
What are PROs?

A PRO is “any report coming directly from subjects without interpretation of the physician or others about how they function overall or feel in relation to a condition and its therapy” [12]. PRO data are collected via standardised questionnaires designed to measure underlying constructs not directly measurable, such as pain (or other symptoms), or the ability to carry out functional tasks. Individual items (questions plus their response categories) are grouped into one or more domains, depending on the concept they represent. A domain can be used as a measurement construct or scale only if it can be shown that the addition of the raw scores from each item is mathematically justifiable [7, 10, 13-15]. The importance of PROs in health care assessment is increasingly being recognised, as PROs address the patient's own perception of function and treatment effect. Conversely, clinical and para-clinical outcomes seek to quantify pathology from the perspective of the clinician [16, 17]. The American Food and Drug Administration (FDA) strongly recommends the use of PROs stating "physical examination and performance testing is not sufficient to measure what may be most important from the patients perspective" [18].

Development of PROs

The first step in developing PRO questionnaires is to choose a theoretical framework to ensure the integrity and content relevance of the constructs to be measured [17, 19, 20]. Conventional classification concepts in health related PROs, are Impairment (symptoms), Activity limitations (functioning), Participation restrictions, and Health related quality of life (HRQoL). The concepts of impairment, activity limitations, and participation have been explicitly defined in the International Classification of Functioning, Disability, and Health (ICF) (WHO, 2001) and are broadly used to classify patients in terms of health status. The ICF can be instrumental in establishing a common conceptual reference [21-24]. Impairment is defined as the loss or abnormality of psychological, physiological, or anatomical structure or function, and largely equates to symptoms and signs [17, 22-26]. Impairment relates to dysfunction at the bodily level manifesting in such problems as restricted mobility and pain, but also in depression and anxiety. Impairment is conceived as some deviation from the norm in an individual's biomedical status. Activity limitations are any restriction or lack of ability to perform activities within some normative range [17, 21]. This might include problems with dressing, walking, or personal care. Activity limitations are often referred to as disability or decreased functional status [17, 21, 26, 27]. Instruments that assess the impact or treatment of disease at this level typically target the construct of activity. This is essential when the objective of a study is to assess activities of daily living, for example. Participation is defined as a person's involvement in life situations. The construct of participation is closely related to activity, but differs in that it emphasizes the degree to which a person is able to and actually does take part in areas of life, regardless of their level of impairment or the degree to which their ability to perform activities is limited [17, 21, 26, 27]. A bilateral amputee sprinter would be considered highly impaired, yet at an exceptionally high participation level. Impairment, activity, and participation are well-suited constructs for determining and planning interventions, and particularly important in studies of interventions intended to increase patients' participation in life [17, 21, 26-28]. Constructs targeting lifestyle, psychosocial, and general satisfaction are somewhat problematic in the ICF model. Alternative models are likely more appropriate to deal with existential aspects of HRQoL [29]. Figure 1 shows the structural model of the ICF system as applied to instruments for outcome assessment of patients with ACL deficiency. The arrows indicate a dynamic interaction between the components of the model. Items and constructs can be placed within one or another component depending on the perspective of the author of the instrument. This can seem arbitrary. However, it is important to keep in mind that the ICF is intended as a classification tool which can be useful for conceptualizing and constructing PROs. The ICF is not a measure in itself. It is analogous to a tool box, where the ruler in this context would be the PRO questionnaire itself.

Figure 1. The ICF model as applied to ACL rupture



Specific as opposed to generic instruments

PROs can be generic, anatomically specific, or condition specific. Generic PROs are developed for overall assessment of Health Status regardless of underlying pathology and diagnostic criteria. Instruments such as the Medical Outcome Study - Short Form (SF-36) and the Sickness Impact Profile are examples of generic PROs. Anatomically specific PROs are constructed for patients with pathology associated with a specific organ or anatomic region, such as the knee or hip, without regards to type of pathology. Knee-specific instruments such as the International Knee

Documentation Committee Knee score (IKDC), the Lysholm Score, and the Knee Injury and Osteoarthritis Outcome Score (KOOS) are examples of anatomically specific PROs. Instruments addressing specific pathologies in specific organs are said to be condition-specific, for example the Western Ontario Meniscal Evaluation Tool (WOMET) for meniscus injuries. The need for condition-specific PROs is increasingly emphasized [30-32]. The main weakness with generic instruments is that they are not designed to target areas of concern in particular patient populations and thus most likely enquire about issues that are irrelevant for specific patient groups [33, 34]. Asking patients to answer non-relevant questions can alienate respondents and increase the potential for missing data [29, 33]. Therefore, condition-specific instruments are preferred as they provide information that is more detailed and are more sensitive to disease-specific and treatment-specific effects [32, 35].

Face and content validity

Item content should be drawn from the most appropriate source in order to confirm that items in PRO instruments measure concepts that are relevant for the targeted patient group [15, 17, 21, 36]. Items are typically derived from two sources; pre-existing PROs and individual clinical experts or consensus groups; focus group and individual interviews with patients [14, 17, 28]. While items measuring symptoms are best generated from the perspective of the patient, with possible input from the clinician, items used to measure participation, psychosocial attributes, and HRQoL can really only be derived from the patient [17]. Items obtained from pre-existing instruments through literature reviews will possess face validity [37]. Mosier states that the instrument ”is considered to be valid, if the sample of items appears to the subject matter experts to represent adequately the total universe of appropriate test questions“ [37, 38]. Item content relevancy and coverage (content validity) is confirmed by confronting the particular patient group with the items in question. Tanner and colleagues selected eleven knee-specific questionnaires to assess which items in the instruments were most relevant for different knee pathologies based on consensus interviews with orthopedic specialists [31]. The authors constructed a composite 111-item pilot questionnaire to assess condition-specific item-content relevancy through semi-structured focus group interviews with three separate diagnostic groups of knee patients: ACL-deficient, meniscus tears, and knee OA [31]. Others have used similar methods for the purpose of item generation [39-42]. Thus, for the purpose of creating new PROs, review of existing instruments can be particularly relevant in the initial stages of item generation.

Aim

The objective of the present study was to find all PRO questionnaires with knee-specific content, to identify items with unique content contained within them, to classify these items according to the ICF model.

Methods

A literature search was conducted in the Medline, CINHAL, EMBASE, and PEDRO databases to find existing PROs used to assess knee pathology. The search was designed to capture articles which identified questionnaires used to assess outcome for patients with knee ligament and/or knee cartilage injuries, and/or knee OA. Questionnaires written in English, German, and Scandinavian languages were included. Hierarchically, the search term “questionnaire” included all categories of patient related outcomes (e.g., PRO, POEM, etc.). Thus, the terms: ”knee”, “knee joint”, “articular ligaments”, “articular cartilage”, and “knee injury” were combined with the dependent response term “questionnaire”. The original search was conducted in April 2009. The search was progressively updated in Medline and the reference lists of the included papers were further scrutinized to search for knee-specific PROs.

Titles generated from the initial search were screened to identify abstracts in which questionnaires were used to assess treatment effects relating to knee pathology. Unique instruments were noted. Identified abstracts were scrutinized to determine the specific application of the PRO for knee patients. Identified measures were excluded if they were:

: Not a PRO; the instrument was not a questionnaire, but contained only objective, clinician-based measures of outcome

: Condition non-specific; the instrument was designed for other or more general pathologies such as inflammatory disease (e.g., rheumatory arthritis), extra-articular injuries (e.g., tibiofibular joint), bony fractures, tumours, acute trauma, emergency room conditions, or for evaluation of acute-care post-operative status

: Anatomically non-specific; the instrument was designed for other anatomic regions or was solely generic

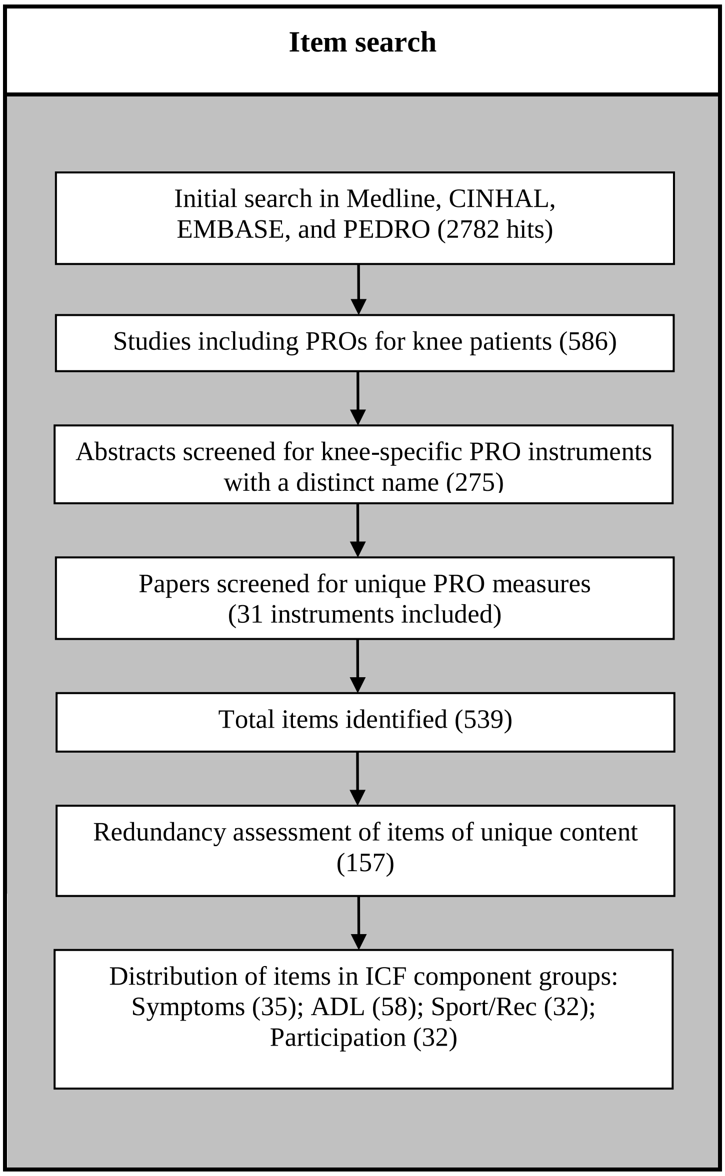
Assessment of item content redundancy and item reduction

After knee-specific PRO instruments were identified, all individual constructs, such as pain, function, activity level, as well as the items comprising these domains, were distinguished. Qualitative assessment of the meaningful content of each item was conducted, as the main purpose of this study was to extract only items of unique content from PROs used to assess knee patients. Item-content redundancy was assessed in the following way: All the collected items from the different PROs were extracted from the host instrument and placed into a single grouping and the main topic or theme of each item was then ascertained. These items were stripped of their grammatical structure to extract the meaningful content of the item. For example, an item involving walking ability:"Are you able to walk on rough ground, inclines, or negotiate curves?" became "walking ability in uneven terrain." The result was a single group of truncated items consisting only of item content matter. The wording and structure of the response scales for the items was not considered, as only the content of the item attribute was of concern. Finally, the truncated items were classified according to the most appropriate ICF component (Table 2).

Results

The search and screening pathways are shown schematically in figure 2. The original search yielded 2782 articles. From these, 586 studies were identified where PRO questionnaires were mentioned in conjunction with knee pathology. Screening of article titles resulted in 275 abstracts, which in turn yielded 81 measures with unique names. Thirty-one of these instruments were judged to have met the inclusion criteria of being a knee-specific PRO.

Figure 2. The search pathway for PROs used to assess knee pathology.



The final 31 instruments included 87 separate sub-domains, such as pain or other symptoms, where the majority of these domains across the instruments were redundant. Table 1 shows the 31 PRO instruments in alphabetical order. The 31 instruments contained 539 items, which after qualitative assessment of item topic redundancy was reduced to 157 items of discrete item content. The final 157 list of non-redundant items as related to the ICF classification system are shown in Table 2.

As can be seen in Table 2, thirty five items were classified into the Impairment (Symptoms) component of the ICF, 58 items in the Activity of Daily Living (Daily Function) component, 32 items in Sport and Recreational Activity (Sport Function), and 32 items belonged to the Participation (Psychosocial) ICF component. Seventeen of the 35 items, which were classified in the Symptoms domain, consisted of item content where pain with or without movement of the knee was the main topic. The other item topics were symptoms relating mostly to mechanical sensations, such as grinding or clicking during movement. In addition, items concerning variations of impinged motion and the sensation of mechanical instability and numbness were represented. In the Function in Daily Activities domain, the 58 item topics were quite evenly distributed throughout topics of walking at different levels of difficulty, daily mobility, such as getting in and out of cars and bed, work abilities, and a broad range of other functional topics during daily life. The Sports and Recreational Activity items consisted of 32 items dealing with high-level activity demands in different functional contexts, such as type of sports, type of movement, or intensity level of competition. The Participation domain consisted of 31 items with an array of psychosocial topics, which could influence the level and ability to participate in normal life activities, from emotional distress at work to social isolation.

Table 1. PRO instruments used to assess knee outcome

	PRO instrument	Ref.
1	**Activity Rating Score (ARS)**	**41**
2	**British Orthopedic Association Score (BOAS)**	**43**
3	**Cincinnati Knee Rating System(CKRS)**	**44**
4	**Crosby/Insall**	**45**
5	**Hospital for Special Surgery Score (HSS)**	**46**
6	**Hughston Clinic knee self-assessment questionnaire**	**47**
7	**International Knee Documentation Committee (IKDC)**	**48**
8	**Knee injury and Osteoarthritis Outcome Score (KOOS)**	**49**
9	**Knee self-efficacy scale (K-SES)**	**50**
10	**Knee Outcome Survey (KOS)**	**51**
11	**Knee Society Score (KSS)**	**52**
12	**KOOS – PS**	**53**
13	**Kujala**	**54**
14	**Lequesne index**	**55**
15	**Lower Extremity Functional Scale (LEFS)**	**56**
16	**Lysholm**	**57**
17	**McGill Pain Questionnaire**	**58**
18	**Mohtadi (ACL-QOL)**	**59**
19	**Musculoskeletal Function Assessment Questionnaire (XSMFA)**	**60**
20	**Orthopadische Arbeitsgruppe Knie (OAK)**	**61**
21	**Oxford Knee Score (OKS)**	**62**
22	**PFPS Pain Severity Scale (PSS)**	**63**
23	**Physical Activity Scale (PAS)**	**64**
24	**SF-12**	**65**
25	**Shelbourne**	**66**
26	**Stanford Health Assessment Questionnaire (HAQ)**	**67**
27	**Tegner Activity Scale**	**68**
28	**Total Knee Function Questionnaire (TKFQ)**	**69**
29	**UCLA Activity Scale**	**70**
30	**Western Ontario and McMaster Universities Osteoarthritis Index function scale (WOMAC)**	**71**
31	**WOMET**	**30**

Table 2. Items with unique content as related to the ICF

Impairment (Symptoms)	Activity (Function in ADL)	Activity (Function in Sport/Rec)	Participation (Psycho-social)
1. Pain when turning/twisting knee	36. Walking - independent outdoors	94. Jogging 3-4 X pr week	126. Feeling relaxed and tranquil
2. Pain when walking on flat surface	37. Walking – flat surface	95. Straight running in sport	127. Feel depressed or sad
3. Pain when walking in uneven surface	38. Walking - uneven surface (e.g. woods, hills)	96. Run with change of direction	128. Being reminded of knee injury
4. Pain when walking up stairs	39. Limited walking distance	97. Run with sudden stopping	129. Anxiety over contact sport
5. Pain when walking down stairs	40. Walking with assistive device	98. Running with turning/pivoting	130. General safety worries
6. Pain when jumping/hopping	41. Walking down stairs	99. Running with turning/pivoting on injured knee	131. Worries about lifestyle and family activities
7. Pain when running	42. Walking up stairs	100. Running in woods	132. Self confidence due to knee injury
8. Pain when getting up from sitting, lying, or crawling.	43. Walking down stairs with hand rail	101. Full competition running	133. Insecurity over re-injury
9. Pain after running	44. Walking up stairs with hand rail	102. Jumping/hop	134. Emotional distress in ADL or at work
10. Limping	45. Standing still	103. Full competition jumping and landing	135. Energy level
11. Pain when kneeling	46. Sitting in squatting position	104. Hop on injured knee	136. Fear of injury worsening due to sport/activity
12. Pain when crouching	47. Lying in bed	105. Side to side hop from one leg to the other leg	137. Fear of knee giving way during sport
13. Pain when straightening knee	48. Sitting with bended knee	106. Competitive sports	138. Fear of external factors (slippery floors etc.)
14. Pain when bending knee	49. Putting on stockings	107. Recreational soccer	139. Apprehension during sport due to knee injury
15. Pain at night	50. Taking off stockings	108. Amateur level soccer	140. Social isolation
16. Pain when sitting	51. Washing hair	109. Division III soccer or lower	141. Ability to go all out during sport
17. Pain when lying	52. Standing up from sitting	110. Elite division I - II soccer	142. Level of participation in sport of first choice
18. Pain when standing	53. Getting into car	111. Elite contact sports (handball, basketball, hockey, athletics, tennis and squash	143. Level of participation in sport of second choice
19. Clicking in knee	54. Getting out of car	112. Alpine skiing	144. Level of participation in fitness and physical training
20. Grinding in knee	55. Getting into bed	113. Nordic skiing	145. Level of pleasure in life
21. Crunching in knee	56. Getting out of bed	114. Swimming	146. Shyness because of injury
22. Other sounds from knee	57. Assistance getting in/out of bed	115. Non-contact sports	147. Acceptance of limitations due to injury
23. Limited range of motion in knee	58. Getting in/out of shower/bathtub	116. Recreational tennis/squash	148. Changes in expectations to sport
24. Stiffness in knee in the morning	59. Getting on/off toilet	117. Horseback riding	149. Fulfillment of competition needs
25. Stiffness in knee later in the day	60. Transfer in bed	118. Recreational cycling	150. Changes in lifestyle
26. Locking of knee	61. Moving into crouched position	119. Elite cycling	151. Being reminded of injury
27. Catching of knee	62. Walking or crawling on knees	120. Long distance cycling	152. Magnitude of problem
28. Something slipping in knee	63. Bending down to floor	121. Hard physical training	153. Focus on injury
29. Strength reduction of knee	64. Pivoting on knee	122. Stretching	154. Satisfaction with knee
30. Loss of knee control	65. Twisting on knee	123. Strength exercises for lower extremities	155. Trusting the knee
31. Giving way	66. Bicycling	124. Slightly limited deep knee bends	156. Comparison of activity level before/after injury
32. Thigh is thinner than the other	67. Going out dancing	125. Unlimited deep knee bends	157. Participation in sport with/without symptoms
33. Knee swelling	68. Running after small children		
34. Impinged knee movement	69. Running to catch bus/train		
35. Numbness in/around knee	70. Jumping from a pier to a boat		
	71. Bending forwards/backwards		
	72. Turning or pivoting		
	73. Light activities		
	74. Carrying heavy objects		
	75. Unable to work – sick leave		
	76. Light work		
	77. Moderate work		
	78. Part time work		
	79. Full time work		
	80. Change of occupation		
	81. Heavy industrial labor		
	82. Hobbies or recreational activities		
	83. Light home activities		
	84. Heavy home activities		
	85. Dusting and watering flowers/plants		
	86. Vacuuming and mowing grass		
	87. Heavy housework		
	88. Light housework		
	89. Gardening		
	90. Moving a table – distance		
	91. Shopping		
	92. Carrying groceries		
	93. Moving around in a rocking boat		

Discussion

We assembled an item pool stemming from all questionnaires in English, German, and Scandinavian languages that target patients with knee ligament, meniscus, and cartilage injury. These items were classified according to the ICF classification system.

Tanner et al. [31] collected 111 items selected from questionnaires recommended by only five surgeons, and Tanners items were restricted to those found in the English language. We performed an analysis of item content redundancy by isolating the meaningful content. It is unclear how Tanner dealt with content redundancy. Thus, their coverage of item content was not comprehensive and may have included redundant items.

Our items can be used in the development PRO measures for patients with knee pathology. The items are knee-specific, which means, in principle, they can be used to construct knee scores for different knee pathologies. However, in this case, it is essential to confront each diagnostic group with the content of each item through focus group - and individual cognitive interviews to confirm comprehensiveness and content validity, and to explore whether other items or constructs can emerge. Only four previously developed knee-specific PROs have included the process of patient confrontation, and just two of these were condition-specific [32]. It is also possible that certain items could be used within a “core” instrument with content covering multiple diagnostic groups of knee pathology, for example, patients with meniscus, cartilage, or ligament injuries, or combinations of these. Other items would be condition-specific for individual diagnostic groups.

The separation of items into a core and condition-specific groups has not previously been proposed for knee PROs. The items in SF-36 were extracted from the original long-form MOS 116 item set, a core instrument for assessment of general health and well-being [72]. However, the items and domains of the MOS instruments (i.e., SF-20, SF-36, and SF-12) were never intended to be condition-specific. Regardless of the path used to construct the questionnaire, once the content has been derived and the PRO measure assembled, appropriate methods to confirm the dimensionality and other measurement properties of the instrument must be employed. This lack of confirmation is a fundamental flaw in earlier knee PRO development, although this has been done for other conditions [4].

We used a novel approach to isolate the meaningful content of items by stripping the items of their grammatical structure and discarding the response options. In this way, we could eliminate cultural, linguistic, and grammatical contexts, since our search included instruments from different cultures and languages. This made it possible to easily identify items that coincided in content regardless of the original language. The truncated items were grouped into one of the ICF components once the item content was isolated, allowing a standardized conceptual reference platform from which the items can be combined to explore the development of new measures. This approach is supported by the findings of Wang and colleagues [32]. They concluded, in a recent critical review of knee PROs that condition-specific instruments are preferred, stating, "Because no standardized knee instrument exists, clinicians and researchers must assess an instruments utility based on their specific disease and patient population of interest." The fact that we found at least 157 items with distinct content defining different attributes of knee dysfunction supports the importance of confirming the content relevance of each item in the specific patient group. There are multiple reasons why items should be condition-specific for distinct diagnostic groups. The activity or task may be too easy or difficult for the patient group being tested, which results in a ceiling or a floor effect. This happens, for example, when young cruciate ligament injured athletes are asked to complete questions designed to measure function in patients with OA [7]. If response scores to particular items lie at the top or the bottom of the spectrum of response options, then the item cannot yield valid information about the particular condition. This means the item cannot differentiate between patients or measure change over time [7, 14]. Also, an item might perform differently for patients with similar, but not the same pathologies, because functional difficulties can vary considerably depending on which anatomical structures (ligament, muscle/tendon, neural, cartilage, bone, etc.) are affected. Moreover, there may be cultural, gender, or age-dependent differences in the way certain items are perceived by patients. Condition-specificity has not previously been addressed adequately, except in one case [32].

We may not have found all PROs relevantly used to assess patients with knee pathology. The WOMET was missed in the original search because the Medical Subject Headings (MeSH) term "knee" does not encompass the hyphenated term "knee-specific". Tanner (2007) mentioned the WOMET, but also used the term "knee-specific" [31]. However, with the inclusion of 31 questionnaires in five languages containing 539 items, we maintain that the probability of item content saturation is high. The fact that approximately 70 percent of the original item pool was redundant supports this argument.

Another problem when searching for items is that the concepts of the measurement domains from which they originate is so heterogeneously interpreted and understood. Even PROs purporting to measure outcome in the same patient groups are made up of a multitude of different domains, sub-domains, constructs, scores, and scales, and there is no consensus as to which domains items belong. This is reflected in the way items are grouped within instruments and in the way in which measurement constructs are operationalised. For example, Marx and colleagues, the creators of the Activity Rating Scale (ARS) [41], discuss the "arbitrariness" of ranking different types of sports activities in terms of level of difficulty. Is tennis more or less difficult than football? Can they be compared? This issue is particularly relevant when scores from items that represent different constructs such as activities or symptoms are added together to yield a sum score. Marx and associates mention the Tegner Activity Score as having Gutman scaling properties. Gutman structure is a basic requirement a scale must fulfil to yield valid measurement as an interval scale. It assumes that an item-response score quantitatively subsumes itself and any item below it on the scale and no items above it. However, this must be mathematically confirmed through testing using appropriate methods such as Item Response Theory (IRT) models (e.g., Rasch analysis). Neither the Tegner Score nor the ARS have been subjected to this type of construct validation. In fact, very few PROs have been assessed using IRT. Most PRO instruments are validated using Classical Test Theory (CTT), such as correlation or factor analyses. It is increasingly recognized that IRT, particularly Rasch analysis, is the most appropriate method to validate PROs [7]. The only knee-specific PROs assessed with Rasch analysis are the OKS [73], the Lysholm Score [74], and the KOOS [7]. The KOOS was found to be insufficient for use on patients who had received an ACL reconstruction [7].

As a result of this study, we present a collection of items that possess face validity for patients with knee pathology. The items can be used to construct content relevant condition-specific PROs.

Conflict of interest statement

No conflicts of interest to declare. All authors contributed substantially to the study and writing of the manuscript. Each of the authors has read and concurs with the content of the final manuscript.

Acknowledgements

The study was made possible by funding from Sahva A/S and the Danish Agency for Science, Technology, and Innovation. David Stodolsky, PhD, is thanked for grammatical corrections to the manuscript.

References

[1] Woo J, Lau E, Lau CS, Lee P, Zhang J, Kwok T, Chan C, Chiu P, Chan KM, Chan A, Lam D. Socioeconomic impact of osteoarthritis in Hong Kong: utilization of health and social services, and direct and indirect costs. Arthritis Rheum 2003;49(4):526-34.

[2] Lohmander LS, Roos EM. Clinical update: treating osteoarthritis. Lancet 2007;370(9605):2082-4.

[3] Lohmander LS, Englund PM, Dahl LL, Roos EM. The long-term consequence of anterior cruciate ligament and meniscus injuries: osteoarthritis. Am J Sports Med 2007;35(10):1756-69.

[4] Gelber AC, Hochberg MC, Mead LA, Wang NY, Wigley FM, Klag MJ. Joint injury in young adults and risk for subsequent knee and hip osteoarthritis. Ann Intern Med 2000;133(5):321-8.

[5] Hill CL, Seo GS, Gale D, Totterman S, Gale ME, Felson DT. Cruciate ligament integrity in osteoarthritis of the knee. Arthritis Rheum 2005;52(3):794-9.

[6] Frobell RB, Roos EM, Roos HP, Ranstam J, Lohmander LS. A randomized trial of treatment for acute anterior cruciate ligament tears. N Engl J Med 2010;363(4):331-42.

[7] Comins J, Brodersen J, Krogsgaard M, Beyer N. Rasch analysis of the Knee injury and Osteoarthritis Outcome Score (KOOS): a statistical re-evaluation. Scand J Med Sci Sports 2008;18(3):336-45.

[8] Comins J, Brodersen J, Krogsgaard M. Treatment for acute anterior cruciate ligament tear. N Engl J Med 2010;363(19):1871-3.

[9] Zarins B. Are validated questionnaires valid? J Bone Joint Surg Am 2005;87(8):1671-2.

[10] Hambly K, Griva K. IKDC or KOOS? Which measures symptoms and disabilities most important to postoperative articular cartilage repair patients? Am J Sports Med 2008;36(9):1695-704.

[11] Brodersen J, Thorsen H, Cockburn J. The adequacy of measurement of short and long-term consequences of false-positive screening mammography. J Med Screen 2004;11(1):39-44.

[12] Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, O'Neill R, Kennedy DL. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Health 2007;10 Suppl 2:S125-S137.

[13] Cano SJ, Hobart JC. The problem with health measurement. Patient Prefer Adherence 2011;5:279-90.

[14] Brodersen J, McKenna SP, Doward LC, Thorsen H. Measuring the psychosocial consequences of screening. Health Qual Life Outcomes 2007;5:3.

[15] Andrich D. Controversy and the Rasch model: a characteristic of incompatible paradigms? Med Care 2004;42(1 Suppl):I7-16.

[16] McKenna SP, Doward LC, Kohlmann T, Mercier C, Niero M, Paes M, Patrick D, Ramirez N, Thorsen H, Whalley D. International development of the Quality of Life in Depression Scale (QLDS). J Affect Disord 2001;63(1-3):189-99.

[17] Doward LC, McKenna SP. Defining patient-reported outcomes. Value Health 2004;7 Suppl 1:S4-S8.

[18] Sloan JA, Halyard MY, Frost MH, Dueck AC, Teschendorf B, Rothman ML. The Mayo Clinic manuscript series relative to the discussion, dissemination, and operationalization of the Food and Drug Administration guidance on patient-reported outcomes. Value Health 2007;10 Suppl 2:S59-S63.

[19] Keszei AP, Novak M, Streiner DL. Introduction to health measurement scales. J Psychosom Res 2010;68(4):319-23.

[20] Doward LC, Meads DM, Thorsen H. Requirements for quality of life instruments in clinical research. Value Health 2004;7 Suppl 1:S13-S16.

[21] Doward LC, Gnanasakthy A, Baker MG. Patient reported outcomes: looking beyond the

label claim. Health Qual Life Outcomes 2010;8:89.

[22] Stucki G. Developing human functioning and rehabilitation research. Part I: Academic training programs. J Rehabil Med 2007;39(4):323-33.

[23] Stucki G, Cieza A, Melvin J. The International Classification of Functioning, Disability and Health (ICF): a unifying model for the conceptual description of the rehabilitation strategy. J Rehabil Med 2007;39(4):279-85.

[24] Stucki G, Celio M. Developing human functioning and rehabilitation research. Part II: Interdisciplinary university centers and national and regional collaboration networks. J Rehabil Med 2007;39(4):334-42.

[25] Stucki G, Reinhardt JD, Grimby G. Organizing human functioning and rehabilitation research into distinct scientific fields. Part II: Conceptual descriptions and domains for research. J Rehabil Med 2007;39(4):299-307.

[26] Stucki G, Melvin J. The International Classification of Functioning, Disability and Health: a unifying model for the conceptual description of physical and rehabilitation medicine. J Rehabil Med 2007;39(4):286-92.

[27] Stucki G, Grimby G. Organizing human functioning and rehabilitation research into distinct scientific fields. Part I: Developing a comprehensive structure from the cell to society. J Rehabil Med 2007;39(4):293-8.

[28] Swaine-Verdier A, Doward LC, Hagell P, Thorsen H, McKenna SP. Adapting quality of life instruments. Value Health 2004;7 Suppl 1:S27-S30.

[29] McKenna SP, Doward LC, Niero M, Erdman R. Development of needs-based quality of life instruments. Value Health 2004;7 Suppl 1:S17-S21.

[30] Kirkley A, Griffin S, Whelan D. The development and validation of a quality of life-measurement tool for patients with meniscal pathology: the Western Ontario Meniscal Evaluation Tool (WOMET). Clin J Sport Med 2007;17(5):349-56.

[31] Tanner SM, Dainty KN, Marx RG, Kirkley A. Knee-specific quality-of-life instruments: which ones measure symptoms and disabilities most important to patients? Am J Sports Med 2007;35(9):1450-8.

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

[33] McKenna SP, Doward LC. Integrating patient-reported outcomes. Value Health 2004;7 Suppl 1:S9-12.

[34] Streiner DL, Norman G.R. Health Measurement Scales - a practical guide to their development and use. Oxford: Oxford University Press, 2008.

[35] Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. J Clin Epidemiol 2003;56(1):52-60.

[36] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010;19(4):539-49.

[37] MOSIER CI. A critical examination of the concepts of face validity. Educ Psychol Meas 1947;7(2):191-205.

[38] Sireci SG. The construct of content validity. Social Indicators Research 1998;45:83-117.

[39] Baro E, Carulla J, Cassinello J, Colomer R, Mata JG, Gascon P, Gasquet JA, Herdman M, Rodriguez CA, Sanchez J, Valentin V. Development of a new questionnaire to assess patient perceptions of cancer-related fatigue: item generation and item reduction. Value Health 2009;12(1):130-8.

[40] Brodersen J, Thorsen H, McKenna S, Doward L. Assessing psychosocial/quality of life outcomes in screening: how do we do it better? J Epidemiol Community Health 2005;59(7):609.

[41] Marx RG, Stump TJ, Jones EC, Wickiewicz TL, Warren RF. Development and evaluation of an activity rating scale for disorders of the knee. American Journal of Sports Medicine 2001;29(2):213-8.

[42] Webster KE, Feller JA, Lambros C. Development and preliminary validation of a scale to measure the psychological impact of returning to sport following anterior cruciate ligament reconstruction surgery. Phys Ther Sport 2008;9(1):9-15.

[43] Waters TS, Bentley G. Patellart resurfacing in total knee arthroplasty - A prospective,

randomized study. Journal of Bone and Joint Surgery-American Volume 2003;85A(2):212-7.

[44] Barber-Westin SD, Noyes FR, McCloskey JW. Rigorous statistical reliability, validity, and responsiveness testing of the Cincinnati knee rating system in 350 subjects with uninjured, injured, or anterior cruciate ligament-reconstructed knees. American Journal of Sports Medicine 1999;27(4):402-16.

[45] Crosby EB, Insall J. Recurrent Dislocation of Patella - Relation of Treatment to Osteoarthritis. Journal of Bone and Joint Surgery-American Volume 1976;58(1):9-13.

[46] Insall JN, Ranawat CS, Aglietti P, Shine J. Comparison of 4 Models of Total Knee-Replacement Prostheses. Journal of Bone and Joint Surgery-American Volume 1976;58(6):754-65.

[47] Hooper DM, Morrissey MC, Drechsler WI, McDermott M, McAuliffe TB. Validation of the Hughston Clinic subjective knee questionnaire using gait analysis. Medicine and Science in Sports and Exercise 2001;33(9):1456-62.

[48] Irrgang JJ, Anderson AF. Development and validation of health-related quality of life measures for the knee. Clinical Orthopaedics and Related Research 2002;(402):95-109.

[49] Roos EM, Roos HP, Ekdahl C, Lohmander LS. Knee injury and osteoarthritis outcome score (KOOS) - validation of a Swedish version. Scandinavian Journal of Medicine & Science in Sports 1998;8(6):439-48.

[50] Thomee P, Wahrborg P, Borjesson M, Thomee R, Eriksson BI, Karlsson J. A new instrument for measuring self-efficacy in patients with an anterior cruciate ligament injury. Scandinavian Journal of Medicine & Science in Sports 2006;16(3):181-7.

[51] Irrgang JJ, Snyder-Mackler L, Wainner RS, Fu FH, Harner CD. Development of a patient-reported measure of function of the knee. Journal of Bone and Joint Surgery-American Volume 1998;80A(8):1132-45.

[52] Insall JN, Dorr LD, Scott RD, Scott WN. Rationale of the Knee-Society Clinical Rating System. Clinical Orthopaedics and Related Research 1989;(248):13-4.

[53] Perruccio AV, Lohmander LS, Canizares M, Tennant A, Hawker GA, Conaghan PG, Roos EM, Jordan JM, Maillefert JF, Dougados M, Davis AM. The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS) - an

OARSI/OMERACT initiative. Osteoarthritis and Cartilage 2008;16(5):542-50.

[54] Kujala UM, Jaakkola LH, Koskinen SK, Taimela S, Hurme M, Nelimarkka O. Scoring of Patellofemoral Disorders. Arthroscopy 1993;9(2):159-63.

[55] Lequesne MG, Mery C, Samson M, Gerard P. Indexes of Severity for Osteo-Arthritis of the Hip and Knee - Validation Value in Comparison with Other Assessment Tests. Scandinavian Journal of Rheumatology 1987;85-9.

[56] Binkley JM, Stratford PW, Lott SA, Riddle DL. The lower extremity functional scale (LEFS): Scale development, measurement properties, and clinical application. Physical Therapy 1999;79(4):371-83.

[57] Lysholm J, Gillquist J. Evaluation of Knee Ligament Surgery Results with Special Emphasis on Use of A Scoring Scale. American Journal of Sports Medicine 1982;10(3):150-4.

[58] Allen RA, Weinmann RL. The Mcgill-Melzack Pain Questionnaire in the Diagnosis of Headache. Headache 1982;22(1):20-9.

[59] Mohtadi N. Development and validation of the quality of life outcome measure (questionnaire) for chronic anterior cruciate ligament deficiency. American Journal of Sports Medicine 1998;26(3):350-9.

[60] Wollmerstedt N, Kirschner S, Bohm D, Faller H, Konig A. Design and evaluation of the extra short musculoskeletal function assessment questionnaire XSMA-D. Zeitschrift fur Orthopadie und Ihre Grenzgebiete 2003;141(6):718-24.

[61] Muller W, Biedert R, Hefti F, Jakob RP, Munzinger U, Staubli HU. Oak Knee Evaluation - A New Way to Assess Knee Ligament Injuries. Clinical Orthopaedics and Related Research 1988;(232):37-50.

[62] Dunbar MJ, Robertsson O, Ryd L, Lidgren L. Appropriate questionnaires for knee arthroplasty - Results of a survey of 3600 patients from the Swedish Knee Arthroplasty Registry. Journal of Bone and Joint Surgery-British Volume 2001;83B(3):339-44.

[63] Laprade JA, Culham EG. A self-administered pain severity scale for patellofemoral pain syndrome. Clin Rehabil 2002;16(7):780-8.

[64] Thomee P, Wahrborg P, Borjesson M, Thomee R, Eriksson BI, Karlsson J. Self-efficacy,

(10) Hambly K, Griva K. IKDC or KOOS? Which measures symptoms and disabilities most important to postoperative articular cartilage repair patients? Am J Sports Med 2008 Sep;36(9):1695-704.

(11) Comins J, Brodersen J, Krogsgaard M. Treatment for acute anterior cruciate ligament tear. N Engl J Med 2010 Nov 4;363(19):1871-3.

(12) Brodersen J, Thorsen H, Cockburn J. The adequacy of measurement of short and long-term consequences of false-positive screening mammography. J Med Screen 2004;11(1):39-44.

(13) Snyder-Mackler L, Risberg MA. Who needs ACL surgery? An open question. J Orthop Sports Phys Ther 2011 Oct;41(10):706-7.

(14) Comins J, Brodersen J, Krogsgaard M, Beyer N. Rasch analysis of the Knee injury and Osteoarthritis Outcome Score (KOOS): a statistical re-evaluation. Scand J Med Sci Sports 2008 Jun;18(3):336-45.

(15) Krogsgaard MR, Brodersen J, Comins J. A scientific approach to optimal treatment of cruciate ligament injuries. Acta Orthop 2011 Jun;82(3):389-90.

(16) Streiner DL, Norman G.R. Health Measurement Scales - a practical guide to their development and use. Oxford: Oxford University Press; 2008.

(17) Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Health 2007 Nov;10 Suppl 2:S125-S137.

(18) McKenna SP, Doward LC, Kohlmann T, Mercier C, Niero M, Paes M, et al. International development of the Quality of Life in Depression Scale (QLDS). J Affect Disord 2001 Mar;63(1-3):189-99.

(19) Doward LC, McKenna SP. Defining patient-reported outcomes. Value Health 2004 Sep;7 Suppl 1:S4-S8.

(20) Brodersen J, Thorsen H, McKenna S, Doward L. Assessing psychosocial/quality of life outcomes in screening: how do we do it better? J Epidemiol Community Health 2005 Jul;59(7):609.

(21) Doward LC, McKenna SP. Defining patient-reported outcomes. Value Health 2004 Sep;7 Suppl 1:S4-S8.

(22) Doward LC, Gnanasakthy A, Baker MG. Patient reported outcomes: looking beyond the label claim. Health Qual Life Outcomes 2010;8:89.

(23) Lasch KE, Marquis P, Vigneux M, Abetz L, Arnould B, Bayliss M, et al. PRO development: rigorous qualitative research as the crucial foundation. Qual Life Res 2010 Oct;19(8):1087-96.

(24) Mosier CI. A critical examination of the concepts of face validity. Educational and Psychological Measurement 1947;7:191-205.

(25) Sireci SG. The construct of content validity. Social Indicators Research 1998;45:83-117.

(26) McKenna SP, Doward LC, Niero M, Erdman R. Development of needs-based quality of life instruments. Value Health 2004 Sep;7 Suppl 1:S17-S21.

(27) Doward LC, Meads DM, Thorsen H. Requirements for quality of life instruments in clinical research. Value Health 2004 Sep;7 Suppl 1:S13-S16.

(28) Brodersen J, McKenna SP, Doward LC, Thorsen H. Measuring the psychosocial consequences of screening. Health Qual Life Outcomes 2007;5:3.

(29) Swaine-Verdier A, Doward LC, Hagell P, Thorsen H, McKenna SP. Adapting quality of life instruments. Value Health 2004 Sep;7 Suppl 1:S27-S30.

(30) Brodersen J, McKenna SP, Doward LC, Thorsen H. Measuring the psychosocial consequences of screening. Health Qual Life Outcomes 2007;5:3.

(31) Comins J, Krogsgaard M, Brodersen J. Ensuring Face Validity in Patient-Related Outcome Scores - a matter of content. 2011.

(32) Brodersen J, Doward L, Thorsen H, McKenna S. Writing Health-Related Items for Rasch Models - Patient Reported Outcome Scales for Health Sciences: From Medical Paternalism to Patient Autonomy. In: Kreiner S, Christensen KB, Mounir M., editors. Rasch Related Models and Methods for Health Sciences. Wiley. In press.; 2012.

(33) Stucki G, Melvin J. The International Classification of Functioning, Disability and Health: a unifying model for the conceptual description of physical and rehabilitation medicine. J Rehabil Med 2007 May;39(4):286-92.

(34) Stucki G, Cieza A, Melvin J. The International Classification of Functioning, Disability and Health (ICF): a unifying model for the conceptual description of the rehabilitation strategy. J Rehabil Med 2007 May;39(4):279-85.

(35) Brodersen J, Thorsen H, Kreiner S. Consequences of screening in lung cancer: development and dimensionality of a questionnaire. Value Health 2010 Aug;13(5):601-12.

(36) Brodersen J, Thorsen H. Consequences of Screening in Breast Cancer (COS-BC): development of a questionnaire. Scand J Prim Health Care 2008;26(4):251-6.

(37) Collins D. Pretesting survey instruments: an overview of cognitive methods. Qual Life Res 2003 May;12(3):229-38.

(38) Willis GB, Reeve BB, Barofsky I. The use of cognitive interviewing techniques in quality of life and patient-reported outcomes assessment. In: Lipscomb J, Gotay CC, Snyder C, editors. Outcome Assessment in Cancer: Measures, Methods, and Applications.New York: Cambridge University Press; 2005. p. 610-23.

(39) Willis GB, Gonzalez A. Methodological issues in the use of survey questionnaires to assess the health effects of torture. J Nerv Ment Dis 1998 May;186(5):283-9.

(40) Fortune-Greeley AK, Flynn KE, Jeffery DD, Williams MS, Keefe FJ, Reeve BB, et al.

Using cognitive interviews to evaluate items for measuring sexual functioning across cancer populations: improvements and remaining challenges. Qual Life Res 2009 Oct;18(8):1085-93.

(41) McKenna S, Wilbum J, Thorsen H, Brodersen J. Production of new language versions of PRO scales. In: Kreiner S, Christensen KB, Mounir M., editors. Rasch Related Models and Methods for Health Science. Wiley. In press.; 2012.

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

(43) Chwalisz K, Stark-Wroblewski K. The subjective experiences of spouse caregivers of persons with brain injuries: a qualitative analysis. Appl Neuropsychol 1996 Feb;3(1):28-40.

(44) Barzansky B, Berner E, Beckman CR. Evaluation of a clinical program. Applying the concept of trustworthiness. Eval Health Prof 1985 Jun;8(2):193-208.

(45) Schou L, Hostrup H, Lyngso EE, Larsen S, Poulsen I. Validation of a new assessment tool for qualitative research articles. J Adv Nurs 2011 Dec 13.

(46) Brodersen J, Thorsen H, Cockburn J. Validity of short-term consequences of cancer prevention and screening activities? J Clin Oncol 2005 Jan 1;23(1):244-5.

(47) Brodersen J, Thorsen H, Kreiner S. Validation of a condition-specific measure for women having an abnormal screening mammography. Value Health 2007 Jul;10(4):294-304.

(48) Stucki G, Cieza A, Melvin J. The international classification of functioning, disability and health: A unifying model for the conceptual description of the rehabilitation strategy. Journal of Rehabilitation Medicine 2007 May;39(4):279-85.

(49) Stucki G, Melvin J. The international classification of functioning, disability and health: A unifying model for the conceptual description of physical and rehabilitation medicine. Journal of Rehabilitation Medicine 2007 May;39(4):286-92.

(50) Stucki G, Reinhardt JD, Grimby G. Organizing human functioning and rehabilitation research into distinct scientific fields. Part II: Conceptual descriptions and domains for research. Journal of Rehabilitation Medicine 2007 May;39(4):299-307.

(51) Stucki G. Developing human functioning and rehabilitation research - Part I: Academic training programs. Journal of Rehabilitation Medicine 2007 May;39(4):323-33.

(52) Stucki G, Celio M. Developing human functioning and rehabilitation research - Part II: Interdisciplinary university centers and national and regional collaboration networks. Journal of Rehabilitation Medicine 2007 May;39(4):334-42.

(53) Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 2007 Dec 15;57(8):1358-62.

(54) Irrgang JJ, Anderson AF. Development and validation of health-related quality of life

measures for the knee. Clinical Orthopaedics and Related Research 2002 Sep;(402):95-109.

(55) Conaghan PG, Emerton M, Tennant A. Internal construct validity of the Oxford Knee Scale: evidence from Rasch measurement. Arthritis Rheum 2007 Dec 15;57(8):1363-7.

(56) Brodersen J, Meads DM, Kreiner S, Thorsen H, Doward L, McKenna SP. Methodological Aspects of Differential Item Functioning in the Rasch Model. Journal of Medical Economics 2007;10(3):309-24.

8. Discussion of article 2

8.1 Major Findings

Of the 157 items of unique content from the literature search of non condition-specific PROs, fifty-two items from the initial item pool were endorsed by the first focus group. The most important issues confirmed in the patient interviews were symptoms, functional limitations (particularly in ADL and recreational tasks), and existential aspects such as job security, social isolation, and worries about lifestyle and family activities. Other existential themes such as romantic or sexual dysfunction due to knee problems were found irrelevant. Several new item topics emerged in the second focus group interview. These were endorsed by subsequent respondents. “Slackness” and “Looseness” were verbatim terms that were closely related to other symptoms and activity items, which were grouped around the verbatim items to form separate constructs. In all, fifty-five items spanning six suggested-constructs were found to cover the spectrum of outcome for the target patient group. These items and constructs were used to assemble a pilot questionnaire for field-testing and psychometric validation.

8.2 Assessment of Methods

Through an extensive literature search, an item pool consisting of items which possessed face validity was generated. Focus group interviews in the target patient groups were used to assess which items from the pool could be endorsed. Endorsement was based on the mention of the topic in open discussion, as well as verbal probing, and then direct confirmation of all items. Certain items such as “giving way” were endorsed, but further probing in subsequent interviews disqualified these items as being too obscure and ambiguous. Items themes that were new were based on verbatim descriptions of the particular topic. These items and themes were subsequently probed, also in single debriefing interviews until the item groupings were stable. By the end of the third focus group interview, no new topics emerged, and no topics were deemed irrelevant by the respondents. Fifty-four items were distributed across six conceptual domains, which the respondents also endorsed as relevant. Single debriefing interview confirmed the content relevance of the instrument, and one more item was generated. The method of using item content from instruments previously created for patients with knee problems is a viable means of generating possibly relevant item content. Cognitive interviewing techniques on an item-by-item basis to address content relevance, content coverage, and comprehension are highly effective in conjunction with verbal probing. Item content generated based on verbatim expressions seems to be an effective tool.

8.3 Justification of Conclusion

The resulting 55-item PRO questionnaire consists of item and domain content endorsed by clinicians and by patients with ACL deficiency. This combination of content sources is constructive and necessary as a basis for the creation of PROs that measure constructs most important to the target group. The major strength of this study is that face validity and content validity of a condition-specific instrument are ensured by combining qualitative techniques. The interviews transitioned from open-ended to semi-structured, although they were never completely open-ended because items from existing instruments and the ICF model were used as a type of interview framework. This allowed for the emergence of new themes beyond those obtained through the literature search, thus satisfying the aspect of content validity known as content coverage, or comprehensiveness. In general, patient confrontation is sparsely used in the development of knee-specific and condition-specific PROs, particularly concerning ACL-specific instruments.[205] Also, the fact that the items and constructs are anchored within the framework of the ICF model places emphasis specifically on the constructs of physical impairment and functional deficit, and allows the components of participation, motivation, and emotional aspects to be included.

8.4 Contribution to Current Knowledge

Of the original 157 items, 43 items were endorsed by the respondents and twelve new items emerged. Thus, 55 items confirmed to be relevant and comprehensive for patients with ACL deficiency and patients with ACL reconstructions are presented. The twelve new items consist of content based on verbatim expressions in patient interviews. Six domains addressing constructs of ADL, Psychosocial consequences, Looseness, Slackness, Symptoms, and Sport/recreational activities have been assembled. The constructs possess face and content validity for Danish patients with ACL deficiency and ACL reconstructed knees. These domains require psychometric assessment through pilot testing to confirm construct validity.

10. Discussion of article 3

10.1 Major Findings

Seven dimensions addressing ADL (8 items), Psychosocial Consequences (5 items), Looseness (4 items), Slackness (7 items), Symptoms (7), Sports Behavior (6 items), and Sports Physical Activity (4 items) exhibited fit to both the PCM Rasch model in RUMM 2030 and the GLLRM in DIGRAM. Thus, all 41 items within these dimensions fit the Rasch model, except item 31 *Pain sitting in chair with knee bent*, which was retained because of overall fit and marginal face- and content relevance. Two items in ADL were removed due to misfit and lack of face validity (items 5 and 9), six items related to stiffness in the symptoms dimension exhibited misfit (items 32 to 37) and were removed; the six Sport avoidance items (40 to 45) showed misfit and must be qualitatively reconsidered. Therefore, of the fourteen items removed from the KNEES-ACL scales, five items could be included in another form in the instrument, most probably in an independent scale.

10.2 Assessment of Methods

Two Rasch approaches were used to assess the scale validity of the item sets. The GLLRM model confirmed the results of the Rasch PCM model and shed extra light on the nature of the items response interdependency (LD) and item-covariate interactions (DIF).

10.3 Justification of Conclusion

Construct validation of the dimensions of the KNEES-ACL is substantiated using the polytomous partial-credit Rasch model and confirmed using the graphical loglinear Rasch model. The methods used to assess the psychometric properties are the most stringent methods available to confirm unidimensionality.

10.4 Contribution to Current Knowledge

Two new scales based on verbatim qualitative interviewing techniques emerged in Danish, which address what patients describe as Looseness and Slackness sensations while participating in functional activities. These scales fit Rasch models in their entirety with no removal of item content. All scales exhibit good internal consistency, targeting, and show clear differences in mean scores between preoperative and postoperative groups. Thus, all scales can be used for invariant assessment of treatment effect between groups, and within-groups with repeated measures. Methods to adjust for LD and DIF are presented.

11. Conclusions of this PhD study

The purpose of this study was to create a condition-specific PRO questionnaire for use in pretest-posttest studies of non-surgical and surgical treatment effect of ACL deficiency. Item content was identified through an extensive literature search of knee-ligament- and cartilage- PROs, whereupon face validity was established. The content relevance of these items was tested in focus groups and individual cognitive interviews of patients pre- and post-ACL reconstruction. Content coverage of the items was found to be insufficient, in that new item content emerged from the interviews. The content relevance of these new items was confirmed in subsequent groups of patients. A pilot PRO was constructed, and finally, the psychometric properties were assessed on a cohort of 242 patients with ACL deficiency and ACL reconstruction. The scales of the KNEES-ACL exhibit face-, content-, and construct-validity. This is the first condition-specific PRO questionnaire that consists of scales, which with certainty can be used for invariant comparison of treatment effect pre- and post- ACL reconstruction. Thus, the KNEES-ACL will potentially allow clinicians to differentiate more reliably between patients who require surgery and those who do not.

12. Implications for research

The results of this study imply that the methods employed for PRO scale construction optimally should follow the sequence used and described in this PhD-project: PROs should be developed starting with a literature search, followed by patient confrontation, and finally psychometric validation using item-response theory. The questionnaire will then possess adequate psychometric properties. Further, the requirements of invariant measurement are met if the Rasch model is used. Without invariant comparison, construct validity cannot be assumed. Earlier methods using CTT can produce arbitrary results, which may just as well reflect the properties of the items rather than the condition of the patient. In summary, utilization of carefully structured qualitative and then quantitative procedures will yield superior results.

Another research implication is the potential to investigate the costs and benefits of surgery, as more accurate methods for determining treatment strategy would have substantial utility. Patients who do not need ACL reconstruction would be more likely to avoid unnecessary surgery, and patients who need surgery would more likely receive timely treatment. Avoidance of unnecessary surgery benefits both the patient and society. Patients avoid pain and loss of income that could result from needless surgical procedures. Moreover, those patients receiving needed surgery promptly can avoid long-term impairments. Society benefits from the cost avoidance of late complications from untreated pathology.

13. Implications for practice

The KNEES-ACL consists of items with high content validity and confirmed psychometric properties. The instrument can be used as an evaluative tool for the establishment and verification of clinical guidelines for efficacious treatment of ACL deficiency. This PRO should be used in preference to existing questionnaires due to its substantiated unidimensional properties.

14. Perspectives

Ideally, we would like to show a correspondence between the subjective responses of the patient and objectively assessed biomechanical deficiency in the knee. The score on each dimension is the response of the patient, and the biomechanical deficiency is captured by kinematic and kinetic analyses. This can be achieved through video-based motion capture, or stereo X-ray techniques. Correlation of the subjective and the objective measurements can yield external criterion validation. The correlation analyses will be carried out most likely using graphical models or structural equation models.

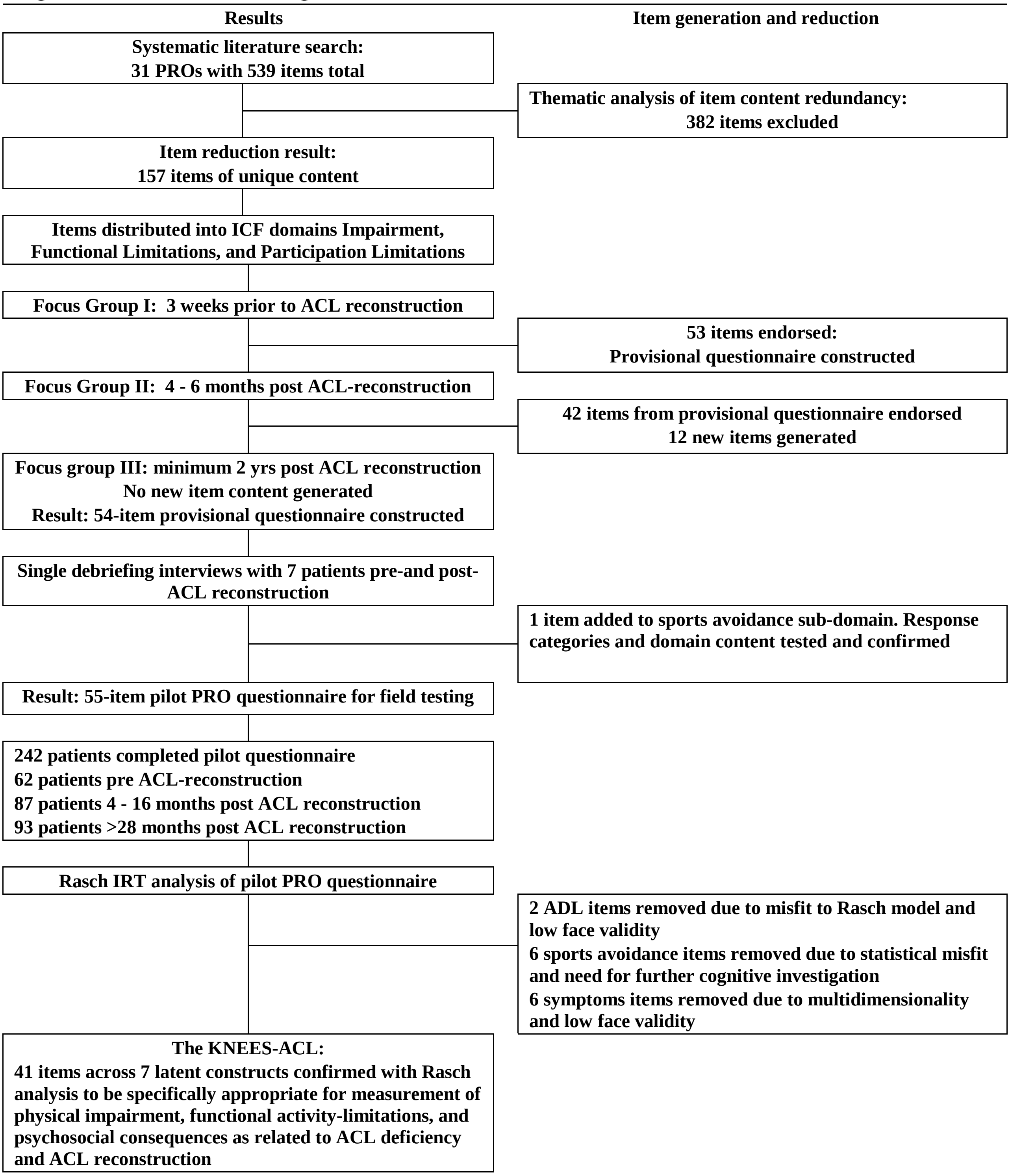
As mentioned in Article 3, the sub-domain of sports avoidance warrants exploration using the above-described qualitative methods on a sub-group of athletes. The sports avoidance items may target this elite group of patients more appropriately. The new scale would not be a part of the KNEES-ACL, but it could be a relevant tool for this patient group. In fact, the initial item pool of 157 items can be used to generate other condition-specific instruments for patients with impairments due to cartilage injury and anterior knee pain.

The pilot versions cannot be used for clinical purposes. Appendices II and III show the English and the Danish versions of the KNEES-ACL. The final KNEES-ACL is not included in this thesis, because the optimal formats still need to be determined; that is, how the appearance and layout will be presented, what the most appropriate clinician/patient interface platforms will be (e.g., digital versus physical formats and design), as well as other practical aspects in terms of instrument application.

Therefore, the English version cannot be considered valid in any way, as it is simply a coarse translation of the Danish version, solely for the purpose of illustration. Because it has not been properly translated to English, it cannot yet be validated on a group of native English speakers. For formal validation, 2-panel translation, qualitative item content analysis, assessment of unidimensionality and invariance must first be carried out in English-speaking patient groups.

15. Trajectory of the study

Figure 1. Flowchart of item generation, reduction, and statistical validation



16. Abbreviations and professional terminology

ACL	Anterior Cruciate Ligament
ACL deficient	A mechanical rupture or lesion of the ACL which causes functional impairments
ADL	Activities of Daily Living
Allograft (ACL)	A transplantation graft procured from the body of a donor
Anterior drawer test	A test to assess the mechanical laxity of the knee in the sagittal plane. The excursion of the tibia relative to the femur.
Autograft	A transplantation graft procured from another region of the body of the person receiving the graft
Condition-specific questionnaire	A condition-specific questionnaire measures a specific condition in a specific population in, as opposed to a generic questionnaire, which measures overall concepts (e.g., health status, anxiety, depression, psychiatric morbidity etc.)
Coper	A person diagnosed with ACL deficiency, who progresses to a level of pre-injury function without signs of impairment
CTT	Classical Test Theory
Differential item functioning, DIF	Differential item functioning is where an item functions differently across subpopulations, such as gender, age, or other exogenous group factors.
False negative test result	An impaired person with a normal test result
False positive test result	A healthy person with an abnormal test result
Focus group interviews	Group interviews where both the issue(s) discussed and the participants in the group are focused.
Generic questionnaire	A generic questionnaire measures overall concepts (e.g., health status, anxiety, depression, psychiatric morbidity etc.) in contrast to condition-specific measures developed to measure a specific condition in a specific population.
GLLRM	Graphical Loglinear Rasch Model
GRF	Ground Reaction Force
IRT	Item Response Theory.
Item	A question with its corresponding response options.
Item truncation	Shorten. Disambiguate. Extract meaningful theme of an item.
KNEES	Knee Numeric Entity Evaluation Score
Latent variable, latent trait	A variable, which is unobservable but is supposed to enter into the structure of a system being studied, such as level of pain.
Non-coper	A person diagnosed with ACL deficiency, who cannot progress to a pre-injury level of function without signs of

	functional impairment
Pilot study	A study, usually on a small scale, carried out prior to the main study, primarily to gain information to improve the efficiency of the main study.
Positive predictive value	The positive predictive value of a test expresses how many persons with a positive result actually have the disease.
Prevalence	The prevalence of e.g. disease is the number of existing persons with the disease in a population at a designated time.
PRO	Patient Related Outcome score. A health-related questionnaire.
Psychometric properties	Psychometrics relates to the measurement of mental abilities and attributes. Psychometric properties are in this context an overall term for the validity and reliability of a psychometric measure.
Sensitivity	The ability of a diagnostic test to identify pathology in patients who present with symptoms that might indicate the presence of pathology
Specificity	The ability of a diagnostic test to rule out pathology in patients who do not present with symptoms
Total score (summary score)	When the numeric responses to items are added together to yield a composite score. Represents the amount of the latent trait the person possesses.
Validity, concurrent	Consists of convergent and divergent validity, for example when a new measure is correlated to an existing measure, the correlation between the measures can converge or diverge.
Validity, construct	The overarching concept of validity. Almost any kind of information about a test can contribute to an understanding of its construct validity. Fundamentally, all validation is construct validation, in the sense that all validity evidence contributes to (or undermines) the empirical foundation or trustworthiness of the score interpretation.
Validity, content	Content validity encompasses content relevance and content coverage. Only questions, response categories, and items that are confirmed to be relevant for the target population will possess content relevance. If all relevant items and response categories addressing an area of interest are included in an instrument, content coverage will be achieved.
Validity, face	If an item appears to be relevant from the perspective of the person applying it, the item has face validity
Validity, known group	Also called extreme groups validity. A test where the measure is given to two groups; one of which has the trait or behaviour, and the other which does not. The former group should score significantly higher (or lower) on the instrument.

Validity, predictive	An example of predictive validity could be a test’s ability to predict whether students tested before they were admitted to university would graduate three years later as bachelors. Concurrent validity and predictive validity are unified under the term criterion validity. For further explanation, see those kinds of validity.

17. Appendix I – Medline search strategy

(knee injury OR “knee injury” OR “injuries” OR osteoarthritis, knee OR “osteoarthritis, knee” OR knee joint OR “knee joint” OR joints OR ligaments, articular OR “ligaments, articular” OR ligament OR cartilage OR “cartilage, articular” OR pathology OR physiology OR physiopathology OR sports OR “sports injuries” OR athletic OR “athletic injuries” OR “pain” OR pain AND functional outcome AND function AND “function” AND outcome AND assessment AND questionnaire AND “questionnaire” AND “questionnaires” AND questionnaires AND “Outcome Assessment (Health Care)” AND “self-efficacy” AND self-efficacy AND scores AND “scores” AND score AND “score” AND “rating scale” AND rating scale AND scales AND patient-related outcome AND PRO AND “PRO” AND POEM AND “POEM”AND self-rated)

18. Appendix II - KNEES-ACL Pilot version (English)

How is your knee?

A questionnaire for individuals with cruciate ligament injury

Developed by --------------: Department of Sports Traumatology, Bispebjerg Hospital, and

Research Unit and Department of General Practice, Faculty of Medicine, University of

Thank you in advance for your help!

The purpose of this questionnaire is to find out how people who have had a cruciate ligament rupture or have had cruciate ligament surgery are doing.

Our goal is to improve the treatment of ligament injuries. You can help us by answering this questionnaire.

Your responses will help us to find the best treatment strategies.

It is important that you answer all the questions.

Copenhagen, Denmark

Your Name:	**Personal No.**	**Date:**

Part 1.

Difficulty in daily activities

Have you - **in the past week** – experienced the following due to your knee injury?

	Not at all	A little	A bit	A lot
1. I have had difficulty walking on level ground.	☐	☐	☐	☐
2. I have had difficulty walking on uneven ground (e.g., in the woods).	☐	☐	☐	☐
3. I have had difficulty walking down stairs.	☐	☐	☐	☐
4. I have had difficulty bending down on knee to pick something up off the floor.	☐	☐	☐	☐
5. I have had difficulty sitting in a chair with knee bent.	☐	☐	☐	☐

Have you - **in the past week** – experienced the following due to your knee injury?

	Not at all	A little	A bit	A lot	N/A
6. I have had difficulty bicycling.	☐	☐	☐	☐	☐
7. I have been unable to crawl on all fours.	☐	☐	☐	☐	☐
8. I have had difficulty squatting.	☐	☐	☐	☐	☐
9. I have been unable to dance.	☐	☐	☐	☐	☐
10. I have had difficulty running.	☐	☐	☐	☐	☐

Psychosocial strain

Have you – **in the past week** – experienced the following due to your knee injury?

	Not at all	A little	A bit	A lot	N/A
11. It has been a mental strain not knowing when my knee would be okay again.	☐	☐	☐	☐	☐
12. It has been a mental strain to make ends meet in daily life because of my knee problems.	☐	☐	☐	☐	☐
13. It has been a mental strain to make ends meet at work because of my knee problems.	☐	☐	☐	☐	☐

Have you – **in the past week** – experienced the following due to your knee injury??

	Not at all	A little	A bit	A lot	N/A
14. It has been a mental strain to make ends meet at home because of my knee problems.	☐	☐	☐	☐	☐
15. It has been a mental strain not being able to participate in hobbies because of knee problems.	☐	☐	☐	☐	☐

Looseness

Have you – **in the past week** – experienced the following due to your knee injury?

	Not at all	A little	A bit	A lot
16. I have felt that my injured knee was un-stabile because of muscle weakness.	☐	☐	☐	☐
17. I have felt that my injured knee was loose when moving around.	☐	☐	☐	☐
18. I have felt that I should monitor my injured knee when moving around.	☐	☐	☐	☐

Have you – **in the past week** – experienced the following due to your knee injury?

	Not at all	A little	A bit	A lot
19. I have felt that I lacked control over my injured knee when moving around.	☐	☐	☐	☐
20. I have felt that I couldn’t count on my injured knee when moving around.	☐	☐	☐	☐

Slackness

Have you – **in the past week** – experienced the following due to your knee injury?

	Not at all	A little	A bit	A lot
21. I have had a slack feeling in my injured knee when moving around.	☐	☐	☐	☐
22. I have spared my injured knee.	☐	☐	☐	☐
23. I have overloaded my ”healthy” knee.	☐	☐	☐	☐

Have you – **in the past week** – experienced the following due to your knee injury?

	Not at all	A little	A bit	A lot	N/A
24. I have experienced shakiness in the injured knee during knee exercises.	☐	☐	☐	☐	☐
25. I have fatigued more quickly in the injured knee compared with uninjured knee during knee exercises.	☐	☐	☐	☐	☐
26. I have had difficulty balancing on my injured knee during knee exercises.	☐	☐	☐	☐	☐

Symptomer

Have you – **in the past week** – experienced the following due to your knee injury?

		Not at all	A little	A bit	A lot
27.	I have had knee pain when twisting/pivoting on my injured knee.	☐	☐	☐	☐
28.	I have had pain in my knee when walking up stairs.	☐	☐	☐	☐
29.	I have had pain in my knee when walking in uneven terrain (e.g., in woods).	☐	☐	☐	☐
30.	I have had knee pain after a long walk.	☐	☐	☐	☐
31.	I have had pain in my knee when sitting in a chair with bended knee.	☐	☐	☐	☐

Have you – **in the past week** – experienced the following?

		Yes	No
32.	I have been able to bend my injured knee completely.	☐	☐
33.	I have been able to extend my injured knee completely.	☐	☐

Have you – **in the past week –** experienced the following due to your knee injury?

	Not at all	A little	A bit	A lot
34. I have had stiffness in my knee in the evening.	☐	☐	☐	☐
35. I have had stiffness in my knee in the morning.	☐	☐	☐	☐
36. I have had swelling of my knee.	☐	☐	☐	☐

Have you – **in the past week –** experienced the following due to your knee injury?

	Not at all	A little	A bit	A lot	N/A
37. I have had a feeling of lost knee control when in motion.	☐	☐	☐	☐	☐
38. I have had knee pain when jumping.	☐	☐	☐	☐	☐
39. I have had knee pain after knee exercises.	☐	☐	☐	☐	☐

Sport and Recreational Activities

40. **In the past week** I have: Tick just one box here

… avoided **all** sport	☐
… avoided **some** forms of sport	☐
… participated in my normal sport	☐

Have you – **in the past week** – experienced the following due to your knee injury?

		Yes	No
41.	I have avoided playing sports because I was told not to by my doctor or therapist.	☐	☐
42.	I have avoided playing sports due to knee pain.	☐	☐
43.	I have avoided playing sports due to knee swelling.	☐	☐
44.	I have avoided playing sports because I've been worried of getting a new injury.	☐	☐
45.	I have avoided playing sports due to worries of injury getting worse.	☐	☐

Have you – **in the past week** – experienced the following due to your knee injury?

		Not at all	A little	A bit	A lot	N/A
46.	I have been more cautious than usual when playing sports.	☐	☐	☐	☐	☐
47.	I have been limited in my capacity to play sports.	☐	☐	☐	☐	☐

Have you – **in the past week** – experienced the following due to your knee injury?

		Not at all	A little	A bit	A lot	N/A
48.	I have had difficulty "going all out" when playing sports.	☐	☐	☐	☐	☐
49.	I have had difficulty changing direction when running.	☐	☐	☐	☐	☐
50.	I have had difficulty stopping suddenly when running.	☐	☐	☐	☐	☐
51.	I have had difficulty jumping.	☐	☐	☐	☐	☐
52.	I have had difficulty landing when jumping.	☐	☐	☐	☐	☐

Have you – **in general** – experienced the following due to your knee injury?

		Not at all	A little	A bit	A lot	N/A
53.	I have had to reduce my expectations to sports.	☐	☐	☐	☐	☐
54.	I feel isolated from the people I used to do sports with before my injury.	☐	☐	☐	☐	☐
55.	I feel that my competitive needs are no longer met.	☐	☐	☐	☐	☐

Part II:

1. You are:
   - ☐ Employed
   - ☐ Student
   - ☐ Looking for work
   - ☐ Receiving disability or similar

---

2. What primary school education do you have? (Tick one box)

- ☐ Elementary school
- ☐ 10th grade
- ☐ High school, technical school
- ☐ Other education
- ☐ N/A

---

3. What vocational education do you have? (Tick one box)

- ☐ No træning
- ☐ Special vocational education
- ☐ Apprentice, apprenticeship, or other training (eg. social and health care worker, nursing assistant, technical assistant)
- ☐ Vocation requiring junior college - associates degree
- ☐ Vocation requiring undergraduate degree (school teacher, Nurse, Journalist)
- ☐ Vocation requiring postgraduate degree (Engineer, Biologist, Architect, MD)
- ☐ Other
- ☐ N/A

4. **What is your current or most recent professional position?**
(Precise details, e.g., Medical secretary not just secretary; smith apprentice, not just blacksmith; agriculturalist, not just farmer; primary school teacher, not just teacher; nursery manager not only an educator, Office Manager of the Treasury, not just Office Manager)

____

5. Do you, or did you have responsibility for employees in your job?
   (How many? Write 0, if none)

|__|__|__|__|

6. Du you live alone?
   (Tick just one box)

   ☐ No ☐ Yes

Thank you for your help!

19. Appendix III - KNEES-ACL Pilot version (Danish)

Hvordan går det med dit knæ?

Spørgeskema til personer med korsbåndsskade

Udarbejdet af xxxxxxx: Idrætskirurgisk Enhed, Bispebjerg Hospital, Afdeling for Almen Medicin, Københavns Universitet og Sahva A/S; xxxxxxxxxx: Forskningsenheden for Almen Medicin, Københavns Universitet; samt xxxxxxx: Idrætskirurgisk Enhed, Bispebjerg Hospital, Københavns Universitet (januar 2011).

På forhånd tak for hjælpen!

Gennem dette spørgeskema håber vi at få at vide, hvordan personer med korsbåndsskade har det før og eventuelt efter en korsbåndsoperation.

Vi vil gerne forbedre behandlingen af korsbåndsskader, og det kan du hjælpe os med ved at besvare skemaet.

Svarene fra dig og andre skal være med til at vise, hvilken behandling virker bedst.

Det er vigtigt, at du svarer på alle spørgsmålene.

Dit Navn:	**Cpr.**	**Dato:**

Del I.

Besvær i dagligdagen

Har du - **i den sidste uges tid** – oplevet følgende på grund af din knæskade?

	Nej, slet ikke	Ja, lidt	Ja, noget	Ja, meget
1. Jeg har haft besvær med at gå på et jævnt underlag.	☐	☐	☐	☐
2. Jeg har haft besvær med at gå på et ujævnt underlag, f.eks. i skoven.	☐	☐	☐	☐
3. Jeg har haft besvær med at gå ned ad trapper.	☐	☐	☐	☐
4. Jeg har haft besvær med at gå ned i knæ, hvis jeg f.eks. skulle samle noget op fra gulvet.	☐	☐	☐	☐
5. Jeg har haft besvær med at sidde på en stol med knæet bøjet.	☐	☐	☐	☐

Har du - **i den sidste uges tid** – oplevet følgende på grund af din knæskade?

	Nej, slet ikke	Ja, lidt	Ja, noget	Ja, meget	Ved Ikke
6. Jeg har haft besvær med at cykle.	☐	☐	☐	☐	☐
7. Jeg har haft besvær med at kravle på alle fire.	☐	☐	☐	☐	☐
8. Jeg har haft besvær med at sidde på hug.	☐	☐	☐	☐	☐
9. Jeg har haft besvær med at danse.	☐	☐	☐	☐	☐
10. Jeg har haft besvær med at løbe.	☐	☐	☐	☐	☐

Psykisk belastning

Har du – **i den sidste uges tid** – oplevet følgende på grund af din knæskade?

	Nej, slet ikke	Ja, lidt	Ja, noget	Ja, meget	Ved ikke
11. Det har været en psykisk belastning, at jeg ikke vidste hvornår knæet kom i orden.	☐	☐	☐	☐	☐
12. Det har været en psykisk belastning, at skulle få hverdagen til at hænge sammen pga. mine knæproblemer.	☐	☐	☐	☐	☐
13. Det har været en psykisk belastning, at skulle få mit arbejdsliv til at fungere pga. mine knæproblemer.	☐	☐	☐	☐	☐

Har du – **i den sidste uges tid** – oplevet følgende på grund af din knæskade?

	Nej, slet ikke	Ja, lidt	Ja, noget	Ja, meget	Ved ikke
14. Det har været en psykisk belastning, at skulle få familielivet til at hænge sammen pga. mine knæproblemer.	☐	☐	☐	☐	☐
15. Det har været en psykisk belastning, ikke at kunne deltage i mine fritidsaktiviteter.	☐	☐	☐	☐	☐

Løshed

Har du – **i den sidste uges tid** – oplevet følgende på grund af din knæskade?

	Nej, slet ikke	Ja, lidt	Ja, noget	Ja, meget
16. Jeg har oplevet, at knæet har været ustabilt pga. manglende kræfter i musklerne.	☐	☐	☐	☐
17. Jeg har haft en følelse af, at knæet var løst, når jeg har bevæget mig.	☐	☐	☐	☐
18. Jeg har følt, at jeg skulle kontrollere knæet, når jeg har bevæget mig.	☐	☐	☐	☐

Har du – **i den sidste uges tid** – oplevet følgende på grund af din knæskade?

	Nej, slet ikke	Ja, lidt	Ja, noget	Ja, meget
19. Jeg har følt, at jeg manglede kontrol over knæet, når jeg har bevæget mig.	☐	☐	☐	☐
20. Jeg har følt, at jeg ikke kunne stole på mit skadede knæ, når jeg har bevæget mig.	☐	☐	☐	☐

Slaphed

Har du – **i den sidste uges tid** – oplevet følgende på grund af din knæskade?

	Nej, slet ikke	Ja, lidt	Ja, noget	Ja, meget
21. Jeg har haft en følelse af slaphed omkring mit skadede knæ, når jeg har bevæget mig.	☐	☐	☐	☐
22. Jeg har skånet det skadede knæ.	☐	☐	☐	☐
23. Jeg har overbelastet det ”raske” knæ.	☐	☐	☐	☐

Har du – **i den sidste uges tid** – oplevet følgende på grund af din knæskade?

	Nej, slet ikke	Ja, lidt	Ja, noget	Ja, meget	Ved Ikke
24. I forbindelse med knæøvelserne har benet rystet.	☐	☐	☐	☐	☐
25. I forbindelse med knæøvelserne er jeg blevet hurtigere træt i det skadede ben i forhold til det andet ben.	☐	☐	☐	☐	☐
26. I forbindelse med knæøvelserne har jeg haft svært ved at holde balancen.	☐	☐	☐	☐	☐

Symptomer

Har du – **i den sidste uges tid** – oplevet følgende på grund af din knæskade?

		Nej, slet ikke	Ja, lidt	Ja, noget	Ja, meget
27.	Jeg har haft smerter, når jeg har vredet knæleddet.	☐	☐	☐	☐
28.	Jeg har haft smerter, når jeg har gået op ad trapper.	☐	☐	☐	☐
29.	Jeg har haft smerter, når jeg har gået på ujævnt underlag, f.eks. i skoven.	☐	☐	☐	☐
30.	Jeg har haft smerter i knæet efter en lang gåtur.	☐	☐	☐	☐
31.	Jeg har haft smerter i knæet, når jeg har siddet på en stol med bøjet knæ.	☐	☐	☐	☐

Har du – **i den sidste uges tid** – oplevet følgende?

		Ja	Nej
32.	Jeg har kunnet bøje mit skadede knæ helt.	☐	☐
33.	Jeg har kunnet strække mit skadede knæ helt.	☐	☐

Har du – **i den sidste uges tid** – oplevet følgende på grund af din knæskade?

	Nej, slet ikke	Ja, lidt	Ja, noget	Ja, meget
34. Jeg har haft stivhed af knæet om aftenen.	☐	☐	☐	☐
35. Jeg har haft stivhed af knæet om morgenen.	☐	☐	☐	☐
36. Jeg har haft hævelse af knæet.	☐	☐	☐	☐

Har du **– i den sidste uges tid –** oplevet følgende på grund af din knæskade?

	Nej, slet ikke	Ja, lidt	Ja, noget	Ja, meget	Ved ikke
37. Jeg har haft en følelse af, at jeg mister styring med knæet, når jeg har bevæget mig.	☐	☐	☐	☐	☐
38. Jeg har haft smerter i knæet, når jeg har hoppet.	☐	☐	☐	☐	☐
39. Efter knæøvelserne har jeg fået ondt i knæet.	☐	☐	☐	☐	☐

Sports- og fritidsaktiviteter

40.	I den sidste uges tid har jeg:	Sæt kun ét kryds her	
		… undgået at dyrke al sport	☐
		… undgået at dyrke nogle former for sport	☐
		… dyrket den sport jeg plejer	☐

Har du – i den sidste uges tid – oplevet følgende på grund af din knæskade?

		Ja	Nej
41.	Jeg har undgået at dyrke sport, fordi jeg har fået at vide af en fysioterapeut eller læge, jeg ikke måtte.	☐	☐
42.	Jeg har undgået at dyrke sport, fordi jeg har haft ondt i knæet.	☐	☐
43.	Jeg har undgået at dyrke sport, fordi knæet har været hævet.	☐	☐
44.	Jeg har undgået at dyrke sport, fordi jeg har været bekymret for at få en ny knæskade.	☐	☐
45.	Jeg har undgået at dyrke sport, fordi jeg har været bekymret for at knæskaden skulle forværres.	☐	☐

Har du – i den sidste uges tid – oplevet følgende på grund af din knæskade?

		Nej, slet ikke	Ja, lidt	Ja, noget	Ja, meget	Ved ikke
46.	Jeg har været mere forsigtig end jeg plejer, når jeg har dyrket sport.	☐	☐	☐	☐	☐
47.	Jeg har været begrænset i at dyrke sport.	☐	☐	☐	☐	☐

Har du – i den sidste uges tid – oplevet følgende på grund af din knæskade?

		Nej, slet ikke	Ja, lidt	Ja, noget	Ja, meget	Ved Ikke
48.	Det har været vanskeligt at give sig fuldt ud med benene, når jeg har dyrket sport.	☐	☐	☐	☐	☐
49.	Jeg har haft besvær med at skifte retning når jeg løb.	☐	☐	☐	☐	☐
50.	Jeg har haft besvær med at bremse op hurtigt.	☐	☐	☐	☐	☐
51.	Jeg har haft besvær med at hoppe.	☐	☐	☐	☐	☐
52.	Jeg har haft besvær med at lande når jeg hoppede.	☐	☐	☐	☐	☐

Har du – overordnet set – oplevet følgende på grund af din knæskade?

		Nej, slet ikke	Ja, lidt	Ja, noget	Ja, meget	Ved ikke
53.	Jeg har måttet nedsætte mine forventninger i forhold til hvor megen sport jeg har kunnet dyrke.	☐	☐	☐	☐	☐
54.	Jeg føler mig isoleret fra de mennesker, jeg plejede at dyrke sport sammen med inden jeg kom til skade.	☐	☐	☐	☐	☐
55.	Jeg føler, at mit konkurrencebehov ikke længere bliver opfyldt.	☐	☐	☐	☐	☐

Del II:

1. Du er:
   - ☐ I arbejde
   - ☐ Under uddannelse
   - ☐ Arbejdssøgende
   - ☐ På førtidspension eller lignende

---

2. Hvilken skoleuddannelse har du?
   (sæt ét kryds ved den seneste)

- ☐ 7.- 9. klasse
- ☐ 10. klasse (folkeskoles udvidede afgangsprøve eller realeksamen)
- ☐ Studentereksamen, hf, højere handelseksamen (hh), højere teknisk eksamen (htx)
- ☐ Anden skoleuddannelse
- ☐ Ved ikke

---

3. Hvilken erhvervsuddannelse har du?
   (sæt ét kryds ved den seneste)

- ☐ Ingen erhvervsuddannelse
- ☐ Specialarbejderuddannelse
- ☐ Efg, lærling, elevuddannelse eller anden faglig uddannelse
  (f.eks. social- og sundhedshjælper, sygehjælper, teknisk assistent)
- ☐ Kort videregående uddannelse under 3 år
  (f.eks. markedsøkonom, maskintekniker, økonoma)
- ☐ Mellemlang videregående uddannelse, 3 – 4 år
  (f.eks. folkeskolelærer, sygeplejerske, journalist, bacheloruddannelse)
- ☐ Lang videregående uddannelse, over 4 år (f.eks. ingeniør, biolog, arkitekt)
- ☐ Anden uddannelse
- ☐ Ved ikke

---

4. Hvad er din nuværende eller seneste erhvervsmæssige stilling?
(nøjagtig angivelse, f.eks.: Lægesekretær, ikke blot sekretær; smedesvend, ikke blot smed; gårdejer, ikke blot landmand; folkeskolelærer, ikke blot lærer; børnehaveleder, ikke blot pædagog; kontorchef i skattevæsenet, ikke blot kontorchef)

____

5. Har du eller havde du nogle underordnede eller ansatte i den stilling?
(skriv antal, hvis ingen skriv 0)

|__|__|__|__|

6. Bor du alene?
(sæt kun ét kryds)

☐ Nej ☐ Ja

Mange tak for hjælpen!

20. Appendix IV – Item pool in Danish

Symptoms

Bevægelse	1.	smerter ved at dreje/vride knæet
	2.	smerter ved at gå på jævnt underlag
	3.	smerter ved at gå på Ujævnt underlag
	4.	smerter ved at gå op ad trapper
	5.	smerter ved at gå ned ad trapper
	6.	smerter i knæet når jeg hopper
	7.	smerte ved løb
	8.	smerter når du rejser dig (siddende/liggende stilling eller på knæ)
	9.	smerter efter jeg har løbet
	10.	jeg halter
	11.	smerter når du ligger på knæ
	12.	smerter i hugsiddende stilling
	13.	smerter ved at strække knæet
	14.	smerter ved at bøje knæet
U. bevægelse	15.	nattesmerter
	16.	smerte ved at sidde
	17.	smerte ved at ligge
	18.	smerte ved at stå
Lyde	19.	murren fra knæet
	20.	knasen fra knæet
	21.	klikkende lyd fra knæet
	22.	andre lyde fra knæet
Bevægelighed	23.	nedsat bevægelighed af knæet
	24.	stivhed af knæet om morgenen
	25.	stivhed af knæet på andre tidspunkter
	26.	låsning af knæet
Styring/kraft	27.	nedsat kraft af knæet
	28.	jeg mister styring
	29.	knæet svigter/give efter
	30.	mit ene lår er tyndere end det andet
Andre	31.	hævelse af knæet
	32.	noget der kommer i klemme i knæet
	33.	noget der smutter i knæet
	34.	knæet hager sig fast
Womet item	35.	følelsesløshed i/omkring knæet*

Function in Activities of Daily Living

Gang	36.	gå ud alene
	37.	besvær med at gå på jævnt underlag
	38.	besvær med at gå på ujævnt underlag – (e.g., skoven, bakker)
	39.	ubegrænset/begrænset distance
	40.	nødt til at gå med stok
Trapper	41.	besvær med at gå ned ad trapper
	42.	besvær med at gå op ad trappe
	43.	besvær med at gå ned ad trapper med gelænder
	44.	besvær med at gå op ad trappe med gelænder
Stille	45.	besvær ved at stå stille
	46.	besvær med at sidde i hug
	47.	besvær med at ligge i seng
	48.	besvær med at sidde med bøjet knæ
Påklædning/hygiejne	49.	besvær med at tage strømper på
	50.	besvær med at tage strømper af
	51.	besvær med at vaske hår
Transfer	52.	besvær med at rejse dig fra siddende
	53.	besvær med at komme ind i en bil
	54.	besvær med at komme ud af i en bil
	55.	besvær med at komme i sengen
	56.	besvær med at komme ud af sengen
	57.	behov for hjælp til at komme ind/ud af seng
	58.	besvær med at komme ind/ud badekar, brusebad
	59.	besvær med at rejse/sætte sig fra toilet
	60.	besvær med at vende i seng
Bevægelse	61.	besvær med at gå ned i knæ /ned på hug
	62.	besvær med at ligge på knæ/ gå på knæ
	63.	besvær med at bøje ned til gulv
	64.	usikkerhed ved at dreje på knæ
	65.	usikkerhed ved at vride på knæ
	66.	besvær med at cykle
	67.	besvær med at tage ud at danse
	68.	besvær med at løbe efter små børn
	69.	besvær med at løbe efter bussen/toget
	70.	besvær med at hoppe fra en båd til kajen
	71.	bøje forover/bagover
	72.	vende eller dreje
	73.	begrænset i lettere aktiviteter
	74.	besvær med at bære tunge objekter
Arbejde	75.	sygemeldt
	76.	let arbejde
	77.	middeltungt arbejde (fx sygehjælper)

	78. deltid
	79. fuldtid
	80. skift af arbejde
	81. tungt arbejde (fx jord/beton)
Hjemmet	82. problemer med vanlige hobbies eller fritidsaktiviteter
	83. lette aktiviteter i hjemmet
	84. tunge aktiviteter i hjemmet
	85. støv af og vand blomster
	86. støvsugning og græsslåning
	87. besvær med tungt husarbejde
	88. besvær med let husarbejde
	89. besvær med havearbejde
	90. flytte bord, distance
	91. besvær med at gå på indkøb
	92. besvær med at bære indkøbsvarer
	93. bevæge dig rundt i en lille gyngende båd

Function in Sport and Recreation

Løb	94. motionsløb 3-4 x ugl.
	95. almindelig løb ligefrem ved sport
	96. løb med retningsskift
	97. løb med opbremsninger
	98. løb med drejende, vridende bevægelser
	99. løb med dreje eller vride på det skadede knæ
	100. orienteringsløb
	101. fuld konkurrence løb - lige ud
Hop	102. hop
	103. fuld konkurrence hop og landing
	104. hop på det skadede ben
	105. sidelæns hop fra det ene ben til det andet
Sportsgrene	106. deltage i konkurrence sport
	107. fodbold (fritidsniveau)
	108. fodbold (seriehold)
	109. fodbold (3. div. og nedad)
	110. elitefodbold (1. og 2. Division)
	111. elite: kontaktidræt (håndbold, basketball, ishockey…) atletik (hop), tennis, squash
	112. alpint skiløb
	113. Langrend skiløb
	114. svømning
	115. øvrig kontaktidræt, tennis (ikke elite)
	116. motions tennis/squash
	117. ridning (heste)
	118. motionscykling
	119. elite cykling
	120. cykling lang distance
Fysisk træning	121. hård fysisk træning kort tid efter skaden eller operation
	122. udspændingsøvelser
	123. styrkeøvelser til benene
	124. lidt begrænset dybe knæbøjninger
	125. normale ubegrænset dybe knæbøjninger

Psycho-social

Emotionelt	126. føler dig roligt og afslappet (tidsrum f.eks. sidste 4 uger)
	127. følt dig trist til mode
	128. frustration over at skulle huske knæet ved sportsaktivitet
	129. angst for kontaktsport
	130. generelle sikkerheds bekymringer
	131. bekymringer omkring din livsstil og aktiviteter med din familie
	132. manglende selvtillid pga din skade
	133. angst for genskade
	134. Problemer med arbejde eller ADL pga følelsesmæssige problemer
	135. fuld af energi
	136. bekymring for at skaden forværres af at dyrke sport/aktivitet
	137. angst for knæsvigt/giver efter ved sport
	138. bekymring for eksterne faktorer (i.e., vådt gulv, etc.)
	139. forsigtighed ved sport
Socialt	140. vanskeligt at se andre mennesker
Begrænsninger	141. vanskelighed ved at give dig helt under sport
	142. begrænsning i at dyrke din 1. prioriterede sport
	143. begrænsning i at dyrke din 2. prioriterede sport
	144. begrænset i at dyrke fitness og fysisk træning
	145. begrænsninger i livsnydelse
	146. tilbageholdenhed pga din skade
	147. accept af dine begrænsninger pga skaden
Forandringer	148. Ændret forventninger til din sport
	149. overvejelser om at dine konkurrencebehov ikke længere bliver opfyldte
	150. omlægning af din livsstil pga knæet
Bevidsthed af skaden	151. hvor ofte mindes problemet
	152. problemstørrelsen
	153. opmærksom på dine knæ problemer
	154. hvor tilfreds er du med dit knæ
	155. stole på knæet
	156. sammenligning af aktivitetsniveau før/efter skaden
	157. deltage i sport med/uden symptomer

21. Reference list for the introduction section of this thesis

Reference List

(1) Guidelines for the PhD Programme - Faculty of Health Science, University of Copenhagen. 2012.

(2) Comins J, Brodersen J, Krogsgaard M, Beyer N. Rasch analysis of the Knee injury and Osteoarthritis Outcome Score (KOOS): a statistical re-evaluation. *Scand J Med Sci Sports* 2008 June;18(3):336-345.

(3) Wittenberg RH, Oxfort HU, Plafki C. A comparison of conservative and delayed surgical treatment of anterior cruciate ligament ruptures. A matched pair analysis. *Int Orthop* 1998;22(3):145-148.

(4) Moksnes H, Snyder-Mackler L, Risberg MA. Individuals with an anterior cruciate ligament-deficient knee classified as noncopers may be candidates for nonsurgical rehabilitation. *J Orthop Sports Phys Ther* 2008 October;38(10):586-595.

(5) Krogsgaard MR. [The anterior cruciate ligament]. *Ugeskr Laeger* 2002 February 25;164(9):1208-1214.

(6) Krogsgaard MR, Brodersen J, Comins J. A scientific approach to optimal treatment of cruciate ligament injuries. *Acta Orthop* 2011 June;82(3):389-390.

(7) Krogsgaard MR. [Sports injuries and arthroscopy--from science to clinical treatment]. *Ugeskr Laeger* 2012 March 12;174(11):719-722.

(8) Frobell RB, Roos EM, Roos HP, Ranstam J, Lohmander LS. A randomized trial of treatment for acute anterior cruciate ligament tears. *N Engl J Med* 2010 July 22;363(4):331-342.

(9) Comins J, Brodersen J, Krogsgaard M. Treatment for acute anterior cruciate ligament tear. *N Engl J Med* 2010 November 4;363(19):1871-1873.

(10) Van de Velde SK, Gill TJ. Treatment for acute anterior cruciate ligament tear. *N Engl J Med* 2010 November 4;363(19):1871-1872.

(11) Alentorn-Geli E, Myer GD, Silvers HJ et al. Prevention of non-contact anterior cruciate ligament injuries in soccer players. Part 2: a review of prevention programs aimed to modify risk factors and to reduce injury rates. *Knee Surg Sports Traumatol Arthrosc* 2009 August;17(8):859-879.

(12) Myer GD, Ford KR, Khoury J, Succop P, Hewett TE. Biomechanics laboratory-based prediction algorithm to identify female athletes with high knee loads that increase risk of

ACL injury. *Br J Sports Med* 2011 April;45(4):245-252.

(13) Myer GD, Ford KR, Hewett TE. New method to identify athletes at high risk of ACL injury using clinic-based measurements and freeware computer analysis. *Br J Sports Med* 2011 April;45(4):238-244.

(14) Hewett TE, Ford KR, Hoogenboom BJ, Myer GD. Understanding and preventing acl injuries: current biomechanical and epidemiologic considerations - update 2010. *N Am J Sports Phys Ther* 2010 December;5(4):234-251.

(15) Hanson AM, Padua DA, Troy BJ, Prentice WE, Hirth CJ. Muscle activation during side-step cutting maneuvers in male and female soccer athletes. *J Athl Train* 2008 April;43(2):133-143.

(16) DiStefano LJ, Padua DA, DiStefano MJ, Marshall SW. Influence of age, sex, technique, and exercise program on movement patterns after an anterior cruciate ligament injury prevention program in youth soccer players. *Am J Sports Med* 2009 March;37(3):495-505.

(17) Padua DA, Marshall SW, Boling MC, Thigpen CA, Garrett WE, Jr., Beutler AI. The Landing Error Scoring System (LESS) Is a valid and reliable clinical assessment tool of jump-landing biomechanics: The JUMP-ACL study. *Am J Sports Med* 2009 October;37(10):1996-2002.

(18) Kaplan Y. Identifying individuals with an anterior cruciate ligament-deficient knee as copers and noncopers: a narrative literature review. *J Orthop Sports Phys Ther* 2011 October;41(10):758-766.

(19) Solomonow M, Krogsgaard M. Sensorimotor control of knee stability. A review. *Scand J Med Sci Sports* 2001 April;11(2):64-80.

(20) Krosshaug T, Slauterbeck JR, Engebretsen L, Bahr R. Biomechanical analysis of anterior cruciate ligament injury mechanisms: three-dimensional motion reconstruction from video sequences. *Scand J Med Sci Sports* 2007 October;17(5):508-519.

(21) Alentorn-Geli E, Myer GD, Silvers HJ et al. Prevention of non-contact anterior cruciate ligament injuries in soccer players. Part 1: Mechanisms of injury and underlying risk factors. *Knee Surg Sports Traumatol Arthrosc* 2009 July;17(7):705-729.

(22) Shultz SJ, Schmitz RJ, Nguyen AD et al. ACL Research Retreat V: an update on ACL injury risk and prevention, March 25-27, 2010, Greensboro, NC. *J Athl Train* 2010 September;45(5):499-508.

(23) Hewett TE, Ford KR, Myer GD. Anterior cruciate ligament injuries in female athletes: Part 2, a meta-analysis of neuromuscular interventions aimed at injury prevention. *Am J Sports Med* 2006 March;34(3):490-498.

(24) Renstrom P, Ljungqvist A, Arendt E et al. Non-contact ACL injuries in female athletes: an International Olympic Committee current concepts statement. *Br J Sports Med* 2008 June;42(6):394-412.

(25) Natri A, Beynnon BD, Ettlinger CF, Johnson RJ, Shealy JE. Alpine ski bindings and

injuries. Current findings. *Sports Med* 1999 July;28(1):35-48.

(26) Krogsgaard M. The cruciate ligaments -- a challenge for the scientific mind. *Scand J Med Sci Sports* 2004 October;14(5):273-274.

(27) Krogsgaard M, Solomonow M. The sensory function of ligaments. *J Electromyogr Kinesiol* 2002 June;12(3):165.

(28) Krogsgaard MR, Dyhre-Poulsen P, Fischer-Rasmussen T. Cruciate ligament reflexes. *J Electromyogr Kinesiol* 2002 June;12(3):177-182.

(29) Krogsgaard M. Rotational instability--the major reason for symptoms after knee ligament injury. *Scand J Med Sci Sports* 2007 April;17(2):97-98.

(30) Johnson RJ. The anterior cruciate ligament problem. *Clin Orthop Relat Res* 1983 January;(172):14-18.

(31) Kowalk DL, Duncan JA, McCue FC, III, Vaughan CL. Anterior cruciate ligament reconstruction and joint dynamics during stair climbing. *Med Sci Sports Exerc* 1997 November;29(11):1406-1413.

(32) Gianotti SM, Marshall SW, Hume PA, Bunt L. Incidence of anterior cruciate ligament injury and other knee ligament injuries: a national population-based study. *J Sci Med Sport* 2009 November;12(6):622-627.

(33) Comins J. Korsbånds-rehabilitering a.m. Shelbourne (ACL-rehabilitation a.m. Shelbourne). *Dansk Sportsmedicin (Danish Sports Medicine)* 1999;3(3):22-25.

(34) Prodromos CC, Han Y, Rogowski J, Joyce B, Shi K. A meta-analysis of the incidence of anterior cruciate ligament tears as a function of gender, sport, and a knee injury-reduction regimen. *Arthroscopy* 2007 December;23(12):1320-1325.

(35) Dragoo JL, Braun HJ, Durham JL, Chen MR, Harris AH. Incidence and risk factors for injuries to the anterior cruciate ligament in national collegiate athletic association football: data from the 2004-2005 through 2008-2009 national collegiate athletic association injury surveillance system. *Am J Sports Med* 2012 May;40(5):990-995.

(36) Krogsgaard MR. [Should badly documented techniques be introduced?]. *Ugeskr Laeger* 2005 October 3;167(40):3761.

(37) Dansk Korsbånds Rekonstruktions Register - Årsrapport 2010. 2010.

(38) Gelber AC, Hochberg MC, Mead LA, Wang NY, Wigley FM, Klag MJ. Joint injury in young adults and risk for subsequent knee and hip osteoarthritis. *Ann Intern Med* 2000 September 5;133(5):321-328.

(39) Hill CL, Seo GS, Gale D, Totterman S, Gale ME, Felson DT. Cruciate ligament integrity in osteoarthritis of the knee. *Arthritis Rheum* 2005 March;52(3):794-799.

(40) Lohmander LS, Englund PM, Dahl LL, Roos EM. The long-term consequence of anterior cruciate ligament and meniscus injuries: osteoarthritis. *Am J Sports Med* 2007

October;35(10):1756-1769.

(41) Felson DT, Naimark A, Anderson J, Kazis L, Castelli W, Meenan RF. The prevalence of knee osteoarthritis in the elderly. The Framingham Osteoarthritis Study. *Arthritis Rheum* 1987 August;30(8):914-918.

(42) Woo J, Lau E, Lau CS et al. Socioeconomic impact of osteoarthritis in Hong Kong: utilization of health and social services, and direct and indirect costs. *Arthritis Rheum* 2003 August 15;49(4):526-534.

(43) Benjaminse A, Gokeler A, van der Schans CP. Clinical diagnosis of an anterior cruciate ligament rupture: a meta-analysis. *J Orthop Sports Phys Ther* 2006 May;36(5):267-288.

(44) Van DP, Gielen J, D'Anvers J et al. MR diagnosis of meniscal tears of the knee: analysis of error patterns. *Arch Orthop Trauma Surg* 2007 November;127(9):849-854.

(45) Van DP, Gielen JL, Vanhoenacker FM, Wouters K, Dossche L, Parizel PM. Stable or unstable tear of the anterior cruciate ligament of the knee: an MR diagnosis? *Skeletal Radiol* 2012 March;41(3):273-280.

(46) Fagan TJ. Nomogram for Bayes Theorem. *N Engl J Med* 1975;(293):257.

(47) Alkjaer T, Simonsen EB, Peter Magnusson SP, Aagaard H, Dyhre-Poulsen P. Differences in the movement pattern of a forward lunge in two types of anterior cruciate ligament deficient patients: copers and non-copers. *Clin Biomech (Bristol , Avon)* 2002 October;17(8):586-593.

(48) Herrington L, Fowler E. A systematic literature review to investigate if we identify those patients who can cope with anterior cruciate ligament deficiency. *Knee* 2006 August;13(4):260-265.

(49) Alkjaer T, Simonsen EB, Jorgensen U, Dyhre-Poulsen P. Evaluation of the walking pattern in two types of patients with anterior cruciate ligament deficiency: copers and non-copers. *Eur J Appl Physiol* 2003 May;89(3-4):301-308.

(50) Eastlack ME, Axe MJ, Snyder-Mackler L. Laxity, instability, and functional outcome after ACL injury: copers versus noncopers. *Med Sci Sports Exerc* 1999 February;31(2):210-215.

(51) Rudolph KS, Eastlack ME, Axe MJ, Snyder-Mackler L. 1998 Basmajian Student Award Paper: Movement patterns after anterior cruciate ligament injury: a comparison of patients who compensate well for the injury and those who require operative stabilization. *J Electromyogr Kinesiol* 1998 December;8(6):349-362.

(52) Hurd WJ, Snyder-Mackler L. Knee instability after acute ACL rupture affects movement patterns during the mid-stance phase of gait. *J Orthop Res* 2007 October;25(10):1369-1377.

(53) Rudolph KS, Axe MJ, Snyder-Mackler L. Dynamic stability after ACL injury: who can hop? *Knee Surg Sports Traumatol Arthrosc* 2000;8(5):262-269.

(54) Shelbourne KD, Trumper RV. Preventing anterior knee pain after anterior cruciate

ligament reconstruction. *Am J Sports Med* 1997 January;25(1):41-47.

(55) Kokmeyer D, Wahoff M, Mymern M. Suggestions from the field for return-to-sport rehabilitation following anterior cruciate ligament reconstruction: alpine skiing. *J Orthop Sports Phys Ther* 2012 April;42(4):313-325.

(56) Fitzgerald GK, Axe MJ, Snyder-Mackler L. Proposed practice guidelines for nonoperative anterior cruciate ligament rehabilitation of physically active individuals. *J Orthop Sports Phys Ther* 2000 April;30(4):194-203.

(57) Thomee R, Kaplan Y, Kvist J et al. Muscle strength and hop performance criteria prior to return to sports after ACL reconstruction. *Knee Surg Sports Traumatol Arthrosc* 2011 November;19(11):1798-1805.

(58) Davids JR. Quantitative gait analysis in the treatment of children with cerebral palsy. *J Pediatr Orthop* 2006 July;26(4):557-559.

(59) McClelland JA, Webster KE, Feller JA. Gait analysis of patients following total knee replacement: a systematic review. *Knee* 2007 August;14(4):253-263.

(60) Minns RJ. The role of gait analysis in the management of the knee. *Knee* 2005 June;12(3):157-162.

(61) Pedersen EN, Alkjaer T, Soballe K, Simonsen EB. Walking pattern in 9 women with hip dysplasia 18 months after periacetabular osteotomy. *Acta Orthop* 2006 April;77(2):203-208.

(62) Winter DA. Biomechanics of human movement with applications to the study of human locomotion. *Crit Rev Biomed Eng* 1984;9(4):287-314.

(63) Hewett TE, Myer GD, Ford KR et al. Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: a prospective study. *Am J Sports Med* 2005 April;33(4):492-501.

(64) Buczek FL, Cavanagh PR. Stance phase knee and ankle kinematics and kinetics during level and downhill running. *Med Sci Sports Exerc* 1990 October;22(5):669-677.

(65) Laursen B, Sogaard K, Sjogaard G. Biomechanical model predicting electromyographic activity in three shoulder muscles from 3D kinematics and external forces during cleaning work. *Clin Biomech (Bristol , Avon)* 2003 May;18(4):287-295.

(66) Lewek MD, Chmielewski TL, Risberg MA, Snyder-Mackler L. Dynamic knee stability after anterior cruciate ligament rupture. *Exerc Sport Sci Rev* 2003 October;31(4):195-200.

(67) Benoit DL, Ramsey DK, Lamontagne M, Xu L, Wretenberg P, Renstrom P. Effect of skin movement artifact on knee kinematics during gait and cutting motions measured in vivo. *Gait Posture* 2006 October;24(2):152-164.

(68) Benoit DL, Ramsey DK, Lamontagne M, Xu L, Wretenberg P, Renstrom P. In vivo knee kinematics during gait reveals new rotation profiles and smaller translations. *Clin Orthop Relat Res* 2007 January;454:81-88.

(69) Andersen MS, Benoit DL, Damsgaard M, Ramsey DK, Rasmussen J. Do kinematic models reduce the effects of soft tissue artefacts in skin marker-based motion analysis? An in vivo study of knee kinematics. *J Biomech* 2010 January 19;43(2):268-273.

(70) Andersen MS, Damsgaard M, Rasmussen J, Ramsey DK, Benoit DL. A linear soft tissue artefact model for human movement analysis: proof of concept using in vivo data. *Gait Posture* 2012 April;35(4):606-611.

(71) Spindler KP, Warren TA, Callison JC, Jr., Secic M, Fleisch SB, Wright RW. Clinical outcome at a minimum of five years after reconstruction of the anterior cruciate ligament. *J Bone Joint Surg Am* 2005 August;87(8):1673-1679.

(72) Johnson DS, Smith RB. Outcome measurement in the ACL deficient knee - what's the score? *Knee* 2001 March;8(1):51-57.

(73) Zarins B. Are validated questionnaires valid? *J Bone Joint Surg Am* 2005 August;87(8):1671-1672.

(74) Garratt AM, Brealey S, Gillespie WJ. Patient-assessed health instruments for the knee: a structured review. *Rheumatology (Oxford)* 2004 November;43(11):1414-1423.

(75) Anderson AF, Federspiel CF, Snyder RB. Evaluation of knee ligament rating systems. *Am J Knee Surg* 1993;6:68-73.

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

(77) Hrubesch R, Rangger C, Reichkendler M, Sailer RF, Gloetzer W, Eibl G. Comparison of score evaluations and instrumented measurement after anterior cruciate ligament reconstruction. *Am J Sports Med* 2000 November;28(6):850-856.

(78) Marx RG, Jones EC, Allen AA et al. Reliability, validity, and responsiveness of four knee outcome scales for athletic patients. *J Bone Joint Surg Am* 2001 October;83-A(10):1459-1469.

(79) Risberg MA, Holm I, Steen H, Beynnon BD. Sensitivity to changes over time for the IKDC form, the Lysholm score, and the Cincinnati knee score. A prospective study of 120 ACL reconstructed patients with a 2-year follow-up. *Knee Surg Sports Traumatol Arthrosc* 1999;7(3):152-159.

(80) Sgaglione NA, Del PW, Fox JM, Friedman MJ. Critical analysis of knee ligament rating systems. *Am J Sports Med* 1995 November;23(6):660-667.

(81) Patrick DL, Burke LB, Powers JH et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value Health* 2007 November;10 Suppl 2:S125-S137.

(82) McKenna SP, Doward LC, Kohlmann T et al. International development of the Quality of Life in Depression Scale (QLDS). *J Affect Disord* 2001 March;63(1-3):189-199.

(83) Doward LC, Meads DM, Thorsen H. Requirements for quality of life instruments in

clinical research. *Value Health* 2004 September;7 Suppl 1:S13-S16.

(84) Doward LC, McKenna SP. Defining patient-reported outcomes. *Value Health* 2004 September;7 Suppl 1:S4-S8.

(85) Sloan JA, Halyard MY, Frost MH, Dueck AC, Teschendorf B, Rothman ML. The Mayo Clinic manuscript series relative to the discussion, dissemination, and operationalization of the Food and Drug Administration guidance on patient-reported outcomes. *Value Health* 2007 November;10 Suppl 2:S59-S63.

(86) Godlee F. Outcomes that matter to patients. *BMJ* 2012 November 1.

(87) Ageberg E, Thomee R, Neeter C, Silbernagel KG, Roos EM. Muscle strength and functional performance in patients with anterior cruciate ligament injury treated with training and surgical reconstruction or training only: a two to five-year followup. *Arthritis Rheum* 2008 December 15;59(12):1773-1779.

(88) Sernert N, Kartus J, Kohler K et al. Analysis of subjective, objective and functional examination tests after anterior cruciate ligament reconstruction. A follow-up of 527 patients. *Knee Surg Sports Traumatol Arthrosc* 1999;7(3):160-165.

(89) Sernert N, Kartus J, Kohler K, Ejerhed L, Brandsson S, Karlsson J. Comparison of functional outcome after anterior cruciate ligament reconstruction resulting in low, normal and increased laxity. *Scand J Med Sci Sports* 2002 February;12(1):47-53.

(90) Reinke EK, Spindler KP, Lorring D et al. Hop tests correlate with IKDC and KOOS at minimum of 2 years after primary ACL reconstruction. *Knee Surg Sports Traumatol Arthrosc* 2011 November;19(11):1806-1816.

(91) Fuchs S, Skwara A, Rosenbaum D. Preliminary results after total knee arthroplasty without femoral trochlea: evaluation of clinical results, quality of life and gait function. *Knee Surg Sports Traumatol Arthrosc* 2005 November;13(8):664-669.

(92) Grindem H, Logerstedt D, Eitzen I et al. Single-legged hop tests as predictors of self-reported knee function in nonoperatively treated individuals with anterior cruciate ligament injury. *Am J Sports Med* 2011 November;39(11):2347-2354.

(93) Mizner RL, Petterson SC, Clements KE, Zeni JA, Jr., Irrgang JJ, Snyder-Mackler L. Measuring functional improvement after total knee arthroplasty requires both performance-based and patient-report assessments: a longitudinal analysis of outcomes. *J Arthroplasty* 2011 August;26(5):728-737.

(94) Noyes FR, Barber-Westin SD. Reconstruction of the anterior and posterior cruciate ligaments after knee dislocation. Use of early protected postoperative motion to decrease arthrofibrosis. *Am J Sports Med* 1997 November;25(6):769-778.

(95) Barber-Westin SD, Noyes FR, Galloway M. Jump-land characteristics and muscle strength development in young athletes: a gender comparison of 1140 athletes 9 to 17 years of age. *Am J Sports Med* 2006 March;34(3):375-384.

(96) Barber-Westin SD, Noyes FR. Objective criteria for return to athletics after anterior

cruciate ligament reconstruction and subsequent reinjury rates: a systematic review. *Phys Sportsmed* 2011 September;39(3):100-110.

(97) Barber-Westin SD, Noyes FR. Factors used to determine return to unrestricted sports activities after anterior cruciate ligament reconstruction. *Arthroscopy* 2011 December;27(12):1697-1705.

(98) Lind M, Menhert F, Pedersen AB. The first results from the Danish ACL reconstruction registry: epidemiologic and 2 year follow-up results from 5,818 knee ligament reconstructions. *Knee Surg Sports Traumatol Arthrosc* 2009 February;17(2):117-124.

(99) Brodersen J, McKenna SP, Doward LC, Thorsen H. Measuring the psychosocial consequences of screening. *Health Qual Life Outcomes* 2007;5:3.

(100) Brodersen J, Doward L, Thorsen H, McKenna S. Writing Health-Related Items for Rasch Models - Patient Reported Outcome Scales for Health Sciences: From Medical Paternalism to Patient Autonomy. In: Kreiner S, Christensen KB, Mounir M., eds. *Rasch Related Models and Methods for Health Sciences*. Wiley. In press.; 2012.

(101) McKenna SP, Doward LC. Integrating patient-reported outcomes. *Value Health* 2004 September;7 Suppl 1:S9-12.

(102) Streiner DL, Norman G.R. *Health Measurement Scales - a practical guide to their development and use*. Oxford: Oxford University Press; 2008.

(103) Andrich D. A rating formulation for ordered response categories. *Psychometrika* 1978;43(4):561-574.

(104) Andrich D. Rating scales and Rasch measurement. *Expert Rev Pharmacoecon Outcomes Res* 2011 October;11(5):571-585.

(105) Roos EM, Roos HP, Ekdahl C, Lohmander LS. Knee injury and osteoarthritis outcome score (KOOS) - validation of a Swedish version. *Scandinavian Journal of Medicine & Science in Sports* 1998 December;8(6):439-448.

(106) Kersten P, White PJ, Tennant A. The visual analogue WOMAC 3.0 scale--internal validity and responsiveness of the VAS version. *BMC Musculoskelet Disord* 2010;11:80.

(107) Stratford PW, Binkley JM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. *Phys Ther* 1996 October;76(10):1109-1123.

(108) Thorborg K, Holmich P, Christensen R, Petersen J, Roos EM. The Copenhagen Hip and Groin Outcome Score (HAGOS): development and validation according to the COSMIN checklist. *Br J Sports Med* 2011 May;45(6):478-491.

(109) McKenna SP, Brodersen J. The chimera of population norms. *Value Health* 2011 June;14(4):618-620.

(110) Twiss J, Meads DM, Preston EP, Crawford SR, McKenna SP. Can we rely on the Dermatology Life Quality Index as a measure of the impact of psoriasis or atopic dermatitis? *J Invest Dermatol* 2012 January;132(1):76-84.

(111) Baron R, Elashaal A, Germon T, Hobart J. Measuring outcomes in cervical spine surgery: think twice before using the SF-36. *Spine (Phila Pa 1976)* 2006 October 15;31(22):2575-2584.

(112) Hobart JC, Williams LS, Moran K, Thompson AJ. Quality of life measurement after stroke: uses and abuses of the SF-36. *Stroke* 2002 May;33(5):1348-1356.

(113) Hagell P, Reimer J, Nyberg P. Whose quality of life? Ethical implications in patient-reported health outcome measurement. *Value Health* 2009 June;12(4):613-617.

(114) Comins J, Krogsgaard M, Brodersen J. Symptoms, function, and psychosocial consequences in patients with anterior cruciate ligament deficiency: Development of a condition-specific questionnaire. 2012 Jan 23; 2012.

(115) Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992 June;30(6):473-483.

(116) Rabin R, de CF. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 2001 July;33(5):337-343.

(117) Brodersen J, Thorsen H, Kreiner S. Validation of a condition-specific measure for women having an abnormal screening mammography. *Value Health* 2007 July;10(4):294-304.

(118) Doward LC, Gnanasakthy A, Baker MG. Patient reported outcomes: looking beyond the label claim. *Health Qual Life Outcomes* 2010;8:89.

(119) Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. *J Clin Epidemiol* 2003 January;56(1):52-60.

(120) MOSIER CI. A critical examination of the concepts of face validity. *Educ Psychol Meas* 1947;7(2):191-205.

(121) Sireci SG. The construct of content validity. *Social Indicators Research* 1998;45:83-117.

(122) Baro E, Carulla J, Cassinello J et al. Development of a new questionnaire to assess patient perceptions of cancer-related fatigue: item generation and item reduction. *Value Health* 2009 January;12(1):130-138.

(123) Brodersen J, Thorsen H, McKenna S, Doward L. Assessing psychosocial/quality of life outcomes in screening: how do we do it better? *J Epidemiol Community Health* 2005 July;59(7):609.

(124) Brodersen J, Thorsen H. Consequences of Screening in Breast Cancer (COS-BC): development of a questionnaire. *Scand J Prim Health Care* 2008;26(4):251-256.

(125) Tanner SM, Dainty KN, Marx RG, Kirkley A. Knee-specific quality-of-life instruments: which ones measure symptoms and disabilities most important to patients? *Am J Sports Med* 2007 September;35(9):1450-1458.

(126) Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing

patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. *Value Health* 2009 November;12(8):1075-1083.

(127) Mckenna S, Wilbum J, Thorsen H, Brodersen J. Production of new language versions of PRO scales. *Rasch Related Models and Methods for Health Science*. Wiley; 2012.

(128) Brodersen J, Thorsen H, Kreiner S. Consequences of screening in lung cancer: development and dimensionality of a questionnaire. *Value Health* 2010 August;13(5):601-612.

(129) McKenna SP, Doward LC, Niero M, Erdman R. Development of needs-based quality of life instruments. *Value Health* 2004 September;7 Suppl 1:S17-S21.

(130) Andrich D. *Rasch Models for Measurement*. Sage Publications, Inc.; 1988.

(131) Borsboom D. *Measuring the mind*. Cambridge University Press; 2005.

(132) Messick S. Meaning and Values in Test Validation: The Science and Ethics of Assessment. *Educational Researcher* 1989;18(2):5-11.

(133) Andrich D. Controversy and the Rasch model: a characteristic of incompatible paradigms? *Med Care* 2004 January;42(1 Suppl):I7-16.

(134) Turk DC, Dworkin RH, Burke LB et al. Developing patient-reported outcome measures for pain clinical trials: IMMPACT recommendations. *Pain* 2006 December 5;125(3):208-215.

(135) Terwee CB, Bot SD, De Boer MR et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007 January;60(1):34-42.

(136) Tennant A, McKenna SP, Hagell P. Application of Rasch analysis in the development and application of quality of life instruments. *Value Health* 2004 September;7 Suppl 1:S22-S26.

(137) Waugh RF, Chapman ES. An analysis of dimensionality using factor analysis (true-score theory) and Rasch measurement: what is the difference? Which method is better? *J Appl Meas* 2005;6(1):80-99.

(138) Andrich D. Controversy and the Rasch model: a characteristic of incompatible paradigms? *Med Care* 2004 January;42(1 Suppl):I7-16.

(139) Andrich D. Understanding resistance to the data-model relationship in Rasch's paradigm: a reflection for the next generation. *J Appl Meas* 2002;3(3):325-359.

(140) Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil* 1989 November;70(12):857-860.

(141) Wright BD, Mok M. Rasch models overview. *J Appl Meas* 2000;1(1):83-106.

(142) Andrich D, Hagquist C. Detection of Differential Item Functioning using Analysis of

Variance. Perth 2004.

(143) Kreiner S. Statistical analysis of complex health and social data. *Soc Sci Med* 1989;29(2):253-258.

(144) Christensen KB. From Rasch scores to regression. *J Appl Meas* 2006;7(2):184-191.

(145) Ford JK, MacCallum RC, Tait M. The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology* 1986;39(2):291-314.

(146) Kaiser HF. The application of electronic computers to factor analysis. *Education and Psychological Measurement* 1960;20:141-151.

(147) Scoboria A, Ford J, Lin HJ, Frisman L. Exploratory and confirmatory factor analyses of the structured interview for disorders of extreme stress. *Assessment* 2008 December;15(4):404-425.

(148) Prieto L, Alonso J, Lamarca R, Wright BD. Rasch measurement for reducing the items of the Nottingham Health Profile. *J Outcome Meas* 1998;2(4):285-301.

(149) Prieto L, Alonso J, Lamarca R. Classical Test Theory versus Rasch analysis for quality of life questionnaire reduction. *Health Qual Life Outcomes* 2003;1:27.

(150) Andrich D. On the distribution of measurements in units that are not arbitrary. *Social Science Information Sur les Sciences Sociales* 2003 December;42(4):557-589.

(151) Andrich D. Distinctions between assumptions and requirements in measurement in the social sciences.Amsterdam: North-Holland; 1989.

(152) Rasch G. *Probabilistic models for some intelligence and attainment tests.* Chicago: The University of Chicago Press; 1960.

(153) Brodersen J. Measuring psychosocial consequences of false-positive screening results - breast cancer as an example. Copenhagen: Faculty of Health Sciences, University of Copenhagen; 2006.

(154) Andrich D. Rating scales and Rasch measurement. *Expert Review of Pharmacoeconomics & Outcomes Research* 2011 October;11(5):571-585.

(155) Andrich D. An Index of Person Separation in Latent Trait Theory, the Traditional KR-20 Index, and the Guttman Scale Response Pattern. *Educational Research and Perspectives* 1982;9:95-104.

(156) Guttman L. The Basis for Scalogram Analysis. In: Stouffer SA, ed. *Measurement and Prediction*.New York: Wiley; 1950.

(157) Andrich D. Educational Measurement: Rasch Models. In: Eva Baker, Barry McGraw, Penellope Peterson, eds. *International Encyclopedia of Education*. 3 ed. Elsevier; 2009:1-15.

(158) Wright BD, Mok MMC. An overview of the family of Rasch measurement models. In:

Smith EV, Smith RM, eds. *Introduction to Rasch measurement*.Maple Grove, MN: JAM Press; 2004:1-25.

(159) Andrich D. An Elaboration of Guttman Scaling with Rasch Models for Measurement. *Sociological Methodology* 1985;33-80.

(160) Andersen EB. Goodness of Fit Test for Rasch Model. *Psychometrika* 1973;38:123-140.

(161) Christensen KB. Fitting polytomous Rasch models in SAS. *J Appl Meas* 2006;7(4):407-417.

(162) Kreiner S, Christensen KB. Validity and Objectivity in Health-Related Scales: Analysis by Graphical Loglinear Rasch Models. *Multivariate and Mixture Distribution Rasch Models - Statistics for Social and Behavioral Sciences*. Springer; 2007:329-346.

(163) Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol* 2007 March;46(Pt 1):1-18.

(164) Ware JE, Jr., Bjorner JB, Kosinski M. Practical implications of item response theory and computerized adaptive testing: a brief summary of ongoing studies of widely used headache impact scales. *Med Care* 2000 September;38(9 Suppl):II73-II82.

(165) Birnbaum A. Some latent trait models and their use in inferring an examinee's ability. *Statistical Theories of Mental Test Scores*.Reading, Mass.: Addison-Wesley; 1968.

(166) Hambleton RK. Emergence of item response modeling in instrument development and data analysis. *Med Care* 2000 September;38(9 Suppl):II60-II65.

(167) Andrich D. Understanding resistance to the data-model relationship in Rasch's paradigm: a reflection for the next generation. *J Appl Meas* 2002;3(3):325-359.

(168) Brodersen J, Meads DM, Kreiner S, Thorsen H, Doward L, McKenna SP. Methodological Aspects of Differential Item Functioning in the Rasch Model. *Journal of Medical Economics* 2007;10(3):309-324.

(169) Andrich D, Kreiner S. Quantifying response dependence between two dichotomous items using the Rasch model. *Applied Psychological Measurement* 2010;34(3):181-192.

(170) Kreiner S, Christensen KB. Analysis of Local Dependence and Multidimensionality in Graphical Loglinier Rasch Models. *Communications in Statistics* 2004;33(1239):1276.

(171) Kreiner S, Christensen KB. Item Screening in Graphical Loglinier Rasch Models. *Psychometrika* 2011;76(2):228-256.

(172) Bjorner JB, Kreiner S, Ware JE, Damsgaard MT, Bech P. Differential item functioning in the Danish translation of the SF-36. *J Clin Epidemiol* 1998 November;51(11):1189-1202.

(173) Andrich D, Hagquist C. Real and Artificial Differential Item Functioning. *Journal of Educational and Behavioral Statistics* 2011;0(0):1-30.

(174) Kreiner S. A Note on Item-rescore Association in Rasch Models. *Applied Psychological Measurement* 2011;35:557-561.

(175) Kelderman H. Logliniear Rasch model tests. *Psychometrika* 1984;49:223-245.

(176) Kreiner S. Validity and objectivity: Reflections on the role and nature of Rasch models. *Nordic Psychology* 2007 October;59(3):268-298.

(177) Kreiner S., Nielsen T. Item Analysis in Digram - Notes on the use of DIGRAM for item analysis by graphical loglinear Rasch models. Department of Biostatistics - University of Copenhagen; 2007.

(178) Traub R.E. Understanding Reliability. Springer; 1991.

(179) Andrich D. Stability of Response, Reliability, and Accuracy of Measurement. *Educational and Psychological Measurement* 1981;41(2):253-262.

(180) Greco NJ, Anderson AF, Mann BJ et al. Responsiveness of the International Knee Documentation Committee Subjective Knee Form in comparison to the Western Ontario and McMaster Universities Osteoarthritis Index, modified Cincinnati Knee Rating System, and Short Form 36 in patients with focal articular cartilage defects. *Am J Sports Med* 2010 May;38(5):891-902.

(181) Andersen EB. Sufficient statistics and latent trait models. *Psychometrika* 1977;42:69-81.

(182) Nielsen T., Kreiner S. Modifying or replacing items: A suggestion for a strategy. 2005.

(183) Penta M, Thonnard JL, Tesio L. ABILHAND: a Rasch-built measure of manual ability. *Arch Phys Med Rehabil* 1998 September;79(9):1038-1042.

(184) Decruynaere C, Thonnard JL, Plaghki L. Measure of experimental pain using Rasch analysis. *Eur J Pain* 2007 May;11(4):469-474.

(185) Vandervelde L, Van den Bergh PY, Goemans N, Thonnard JL. ACTIVLIM: a Rasch-built measure of activity limitations in children and adults with neuromuscular disorders. *Neuromuscul Disord* 2007 June;17(6):459-469.

(186) Caty GD, Arnould C, Stoquart GG, Thonnard JL, Lejeune TM. ABILOCO: a Rasch-built 13-item questionnaire to assess locomotion ability in stroke patients. *Arch Phys Med Rehabil* 2008 February;89(2):284-290.

(187) Caty GD, Arnould C, Thonnard JL, Lejeune TM. ABILOCO-Kids: a Rasch-built 10-item questionnaire for assessing locomotion ability in children with cerebral palsy. *J Rehabil Med* 2008 November;40(10):823-830.

(188) Vandervelde L, Van den Bergh PY, Penta M, Thonnard JL. Validation of the ABILHAND questionnaire to measure manual ability in children and adults with neuromuscular disorders. *J Neurol Neurosurg Psychiatry* 2010 May;81(5):506-512.

(189) Batcho CS, Tennant A, Thonnard JL. ACTIVLIM-Stroke: a crosscultural Rasch-built scale of activity limitations in patients with stroke. *Stroke* 2012 March;43(3):815-823.

(190) Campbell SK, Kolobe TH, Osten ET, Lenke M, Girolami GL. Construct validity of the test of infant motor performance. *Phys Ther* 1995 July;75(7):585-596.

(191) Bond TG, Fox CM. *Applying the Rasch Model - Fundamental Measurement in the Human Sciences*. Mahwah, New Jersey: Lawrence Erlbaum Associates,Inc.; 2001.

(192) Higgins LD, Taylor MK, Park D et al. Reliability and Validity of the International Knee Documentation Comittee (IKDC) Subjective Knee Form. *Joint Bone Spine* 2007;74(6):594-599.

(193) Irrgang JJ, Anderson AF, Boland AL et al. Development and validation of the international knee documentation committee subjective knee form. *Am J Sports Med* 2001 September;29(5):600-613.

(194) Smith HJ, Richardson JB, Tennant A. Modification and validation of the Lysholm Knee Scale to assess articular cartilage damage. *Osteoarthritis Cartilage* 2009 January;17(1):53-58.

(195) Ko Y, Lo NN, Yeo SJ et al. Rasch analysis of the Oxford Knee Score. *Osteoarthritis Cartilage* 2009 September;17(9):1163-1169.

(196) Garratt AM, Brealey S, Robling M et al. Development of the knee quality of life (KQoL-26) 26-item questionnaire: data quality, reliability, validity and responsiveness. *Health Qual Life Outcomes* 2008;6:48.

(197) Goetz C, Ecosse E, Rat AC, Pouchot J, Coste J, Guillemin F. Measurement properties of the osteoarthritis of knee and hip quality of life OAKHQOL questionnaire: an item response theory analysis. *Rheumatology (Oxford)* 2011 March;50(3):500-505.

(198) Davis AM, Badley EM, Beaton DE et al. Rasch analysis of the Western Ontario McMaster (WOMAC) Osteoarthritis Index: results from community and arthroplasty samples. *J Clin Epidemiol* 2003 November;56(11):1076-1083.

(199) Ryser L, Wright BD, Aeschlimann A, Mariacher-Gehler S, Stucki G. A new look at the Western Ontario and McMaster Universities Osteoarthritis Index using Rasch analysis. *Arthritis Care Res* 1999 October;12(5):331-335.

(200) Wolfe F, Kong SX. Rasch analysis of the Western Ontario MacMaster questionnaire (WOMAC) in 2205 patients with osteoarthritis, rheumatoid arthritis, and fibromyalgia. *Ann Rheum Dis* 1999 September;58(9):563-568.

(201) Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009 August 18;151(4):264-9, W64.

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

(203) Wollmerstedt N, Glatzel M, Kirschner S, Schneider J, Faller H, Konig A. [Comparative analysis of patient-centered outcome of total hip and knee arthroplasty]. *Z Orthop Ihre*

Grenzgeb 2006 September;144(5):464-471.

(204) Muller W, Biedert R, Hefti F, Jakob RP, Munzinger U, Staubli HU. Oak Knee Evaluation - A New Way to Assess Knee Ligament Injuries. *Clinical Orthopaedics and Related Research* 1988 July;(232):37-50.

(205) Wang D, Jones MH, Khair MM, Miniaci A. Patient-reported outcome measures for the knee. *J Knee Surg* 2010 September;23(3):137-151.