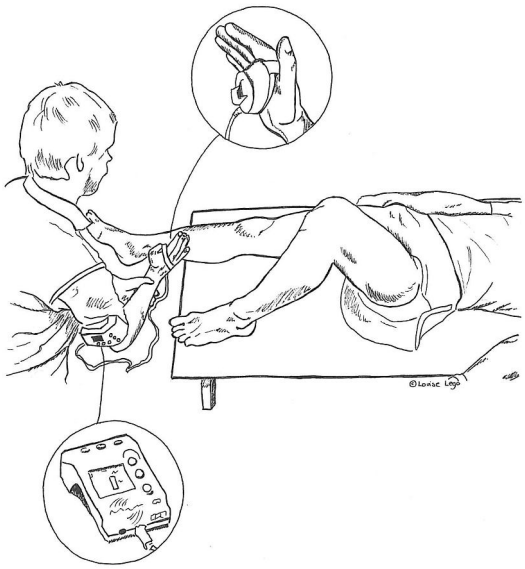
FACULTY OF HEALTH SCIENCES
UNIVERSITY OF COPENHAGEN

Clinical Outcome Measures for Physically Active Individuals with Hip and Groin Pain

Development, evaluation and application



Kristian Thorborg

Department of Orthopaedic Surgery

Amager Hospital

Copenhagen University, Denmark

PhD Thesis 2011

Clinical Outcome Measures for Physically Active Individuals with Hip and Groin Pain

Development, evaluation and application

Kristian Thorborg

From the Department of Orthopaedic Surgery
Amager Hospital, Copenhagen University,
Denmark, 2011

DET SUNDHEDSVIDENSKABELIGE FAKULTET
KØBENHAVNS UNIVERSITET

I shall be telling this with a sigh
Somewhere ages and ages hence:
two roads diverged in a wood, and I –
I took the one less traveled by,
And that has made all the difference.

From the "The Road Not Taken", Robert Frost (1916)

Contact address:

Kristian Thorborg, PT, M. Sportsphysio.
Department of Orthopaedic Surgery,
Amager Hospital,
Copenhagen University, Denmark
Italiensvej 1
DK-2300 Copenhagen S
Denmark

E-mail: kristianthorborg@hotmail.com

Supervisors:

Per Hölmich, Associate professor, MD
Department of Orthopaedic Surgery,
Amager Hospital,
Copenhagen University, Denmark

Ewa Roos, Professor, PT, PhD
Institute of Sports Science
and Clinical Biomechanics,
University of Southern Denmark, Denmark

Peter Magnusson, Professor, PT, DMSc
Institute of Sportsmedicine, Bispebjerg Hospital,
Copenhagen University, Denmark

Evaluation Committee:

Michael Kjær, Professor, MD
Institute of Sports Medicine Copenhagen
Bispebjerg Hospital,
Copenhagen University, Denmark

Peter Brukner, Associate Professor, MD
Melbourne University, Australia

Roland Thomee, Docent, PT
Department of Orthopaedics, Institute of Clinical
Sciences at Sahlgrenska Academy,
Sahlgrenska University Hospital,
University of Gothenburg, Sweden

ACKNOWLEDGEMENTS

First of all I would like to thank my mentor and main supervisor Per Hölmich, for giving me the opportunity to pursue my primary research interest concerning physically active patients with hip and/or groin pain. Ever since our first telephone conversation, you have encouraged me to get involved with research, and I cannot think of a better supervisor within this topic, anywhere in the world.

I would like to thank my two supervisors Professor Peter Magnusson and Professor Ewa Roos. I was very fortunate that you agreed to supervise me in my Ph.D. Peter, your cool, calm and collected approach to research has been inspirational and of great importance to me. Ewa, your ambition of always raising the standards is admirable, and has pushed me in the right direction. Furthermore, I want to thank my other co-authors Else-Marie Bartels, Robin Christensen, Andreas Serner and Thomas Møller Madsen for their important contributions.

A special thanks to my "roommate" at the office at Amager Hospital, Jesper Petersen. I appreciate the discussions we have had regarding research, soccer and other (un)important things in life. Our friendship has been an important part of the journey, and life at the office would not have been the same without you. Also thanks to my dear colleague and good friend Thomas Bandholm for your friendship and moral support through the years.

Further I want to thank the staff at the Department of Orthopaedic Surgery for making me feel welcome from the very first day.

This work was kindly supported by grants from Orthopaedic Research Unit, Department of Orthopaedic Surgery, Amager Hospital; Research Foundation of The Capital Region of Denmark; The Danish Arthritis Foundation; The Association of Danish Physiotherapist; Danish Regions and the Lundbeck Foundation.

I would like to acknowledge American Journal of Sports Medicine (AMJSM Sage publications, Inc), British Journal of Sports Medicine (BMJ Publishing Group Ltd.) and Scandinavian Journal of Medicine & Science in Sports (John Wiley and Sons, Ltd.) for giving permission to reprint original work in this thesis.

Last and most importantly, I would like to thank my wonderful wife, Marianne, for your love and support through the years. Going to Australia on my research exchange, with you and the children, was without a doubt the best experience during the entire Ph.D. Seeing the joy on Mathilde's and Jonathan's faces when meeting you all at the beach in the sunny afternoons in Wollongong will be stuck in my mind forever.

Kristian Thorborg, Copenhagen, Dec 2010

TABLE OF CONTENTS

LIST OF PUBLICATIONS

This thesis is based on the four publications listed below, which are referred to in the following text by their Roman numerals. All studies have been carried out at the Department of Orthopaedic Surgery, Amager Hospital in the period from January 2008 to December 2010.

I

Thorborg K, Roos EM, Bartels EM, Petersen J, Hölmich P

Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: a systematic review

Br J Sports Med. 2010 Aug 10. [Epub ahead of print]

II

Thorborg K, Hölmich P, Christensen R, Petersen J, Roos EM

The Copenhagen Hip and Groin Outcome Score (HAGOS): development and validation according to the COSMIN check list

Submitted

III

Thorborg K, Petersen J, Magnusson P, Hölmich P

Clinical assessment of hip strength using a hand-held dynamometer is reliable

Scand J Med Sci Sports. 2010: 20:493-501

IV

Thorborg K, Serner A, Petersen J, Moller Madsen T, Magnusson P, Hölmich P

Hip adduction and abduction strength profiles in elite soccer players: Implications for clinical evaluation of hip adductor muscle recovery after injury

Am J Sports Med. 2010 Oct 7. [Epub ahead of print]

ABBREVIATIONS

ABD	Abduction
ADD	Adduction
ADL	Activities of daily living
BMI	Body mass index
COSMIN	Consensus-based standards for the selection of health measurement instruments
DOM	Dominant
ER	External rotation
ES	Effect size
EXT	Extension
FLEX	Flexion
GPE	Global perceived effect
HAGOS	Hip and groin outcome score
HOS	Hip outcome score
HOOS	Hip dysfunction and osteoarthritis outcome score
MHHS	Modified Harris hip score
IR	Internal rotation
HHD	Hand-held dynamometer
LSI	Lower limb symmetry index
ICC	Intraclass correlation coefficient
ICF	International classification of functioning
MIC	Minimal detectable change
MID	Minimal important difference
MMT	Manual muscle test
Nm	Newton meter
NDOM	Non-dominant
PRO	Patient-reported outcome
QOL	Quality of life
SF-36	Short-form 36
SEM	Standard error of measurement
SI	International system of units
SDC	Smallest detectable change
SRM	Standardised response mean
WHO	World health organization

DEFINITIONS

Construct validity

The degree to which the scores of a measurement instrument are consistent with a priori hypotheses, based on the assumption that the instrument validly measures the construct to be measured.[1]

Criterion validity

The degree to which scores of a measurement instrument are an adequate reflection of a "gold standard".[1]

Disability

Disability in this thesis encompasses the health dimensions within the methodological framework of The International Classification of Functioning, Disability and Health (ICF) as categorized in one of three levels; impairment (body structure and function), activity limitations (activities), and participation restrictions (participation).[2]

Internal consistency

The degree of interrelatedness among the items e.g. in a questionnaire.[1]

Longstanding hip and/or groin pain

Pain in the hip and groin region of more than 6 weeks' duration is defined as longstanding in nature.[3]

Measurement error (variation)

The systematic and random error (variation) of a patient's score that is not attributed to true changes in the construct to be measured.[1]

Patient-Reported Outcome (PRO)

A PRO is any report coming directly from patients about a health condition and its treatment.[4,5] PRO questionnaires include items, instructions and guidelines for scoring and interpretation and are used to measure these patient reports.[4,5]

Physical activity and inactivity

Physical activity refers to “any force exerted by skeletal muscles that results in energy expenditure above resting level”.[6] Physical inactivity is defined as less than 2.5 hours per week of moderate activity.[7] In this thesis an individual doing any physical activity above resting level, for at least 2.5 hours a week, is referred to as physically active.

Psychometric properties

Psychometrics is the discipline concerned with measurement of variables in tests and questionnaires and has more recently been introduced in health-related fields.[8] Psychometric properties in this thesis are defined as measurement properties of tests concerning validity, reliability, and responsiveness.

Psychometric theory

Classical test theory and item response theory are different expressions of psychometric theory. Classical test theory assumes that an observed score may be decomposed into a “true” score and an “error” score. The term "classical" is seen in contrast to the more recent psychometric theories such as item response theory. Item response theory has also been used to develop and internally validate measures. Item response theory assumes that the test-scale is unidimensional and creates an interval-scaled measure.[8]

Reliability

The extent to which scores for the same patients are unchanged for repeated measurements over time.[1]

Responsiveness

The ability of a an instrument to detect change over time in the construct to be measured.[1]

Smallest Detectable Change

The Smallest Detectable Change (SDC), also referred to as the Minimal Detectable Change (MDC) or Smallest Real Change (SRC), defines which changes in a measurement that fall outside the measurement error.[9]

THESIS AT A GLANCE

Study	Question	Methods	Results	Conclusion
I	Do patient-reported outcome (PRO) questionnaires with adequate measurement qualities for physically active patients with hip and/or groin disability exist?	A systematic review of the reliability, validity and responsiveness of available PRO questionnaires assessing patients with hip and/or groin disability	41 studies, involving 12.779 patients, were included as our final data for reviewing. A total of 13 PRO questionnaires were identified in the included studies. Twelve PRO questionnaires considered the hip region and one questionnaire considered the groin region.	PRO questionnaires for young to middle aged physically active patients with hip and/or groin disability are lacking. A new PRO questionnaire should be developed for young to middle-aged physically active patients with hip and/or groin disability.
II	Can a valid, reliable and responsive PRO questionnaire for young to middle-aged physically active patients with hip and/groin pain be developed?	A new PRO questionnaire was developed including 101 patients with hip and/or groin pain. In a prospective study, validity, reliability and responsiveness of the new questionnaire was assessed.	The new PRO questionnaire Copenhagen Hip and Groin Outcome Score (HAGOS) consists of 6 separate subscales assessing pain, symptoms, function in daily living, function in sport and recreation, participation in physical activities and hip and/or groin-related quality of life. Test-retest reliability was substantial, and a priori set hypotheses concerning construct validity and responsiveness was confirmed	HAGOS has adequate measurement qualities for the assessment of symptoms, activity limitations, and participation-restrictions in physically active patients with longstanding hip and/or groin pain. HAGOS is recommended for use in interventions where the patient's perspective and health-related quality of life are of primary interest.
III	Can a reliable clinical measure of hip muscle strength be developed using a hand-held dynamometer?	The absolute test–retest measurement variation concerning strength assessments of hip ABD, ADD, ER, IR, FLEX and EXT, was investigated in 9 healthy subjects, using a Hand-held dynamometer	The reliability of individual hip strength measurements was between 2-13% (SEM%) in the individual hip strength measurements. Standardised strength assessment procedures of hip ABD, ER, IR and FLEX, with test–retest measurement variation below 5%, hip ADD below 6% and hip EXT below 8%, can be performed	The hand-held dynamometer is easy to administer and produces a small measurement variation, making it possible to determine even small changes in hip strength.
IV	Is isometric hip ADD strength larger in the dominant compared to the non-dominant limb, in soccer players?	Maximal unilateral isometric hip ADD and ABD strength on the dominant and non-dominant side were measured in 100 elite soccer players, with a hand-held dynamometer, using the newly developed and reliable test procedure	In elite soccer players the dominant side was stronger than the non-dominant side for both isometric hip ADD and ABD strength, corresponding to a 3% and 4% difference, respectively. The isometric hip ADD/ABD ratio was not different between the dominant and non-dominant limb.	There is a marginal, but clinically irrelevant, isometric hip ADD and ABD strength difference between the dominant and the non-dominant limb in elite soccer players. Contralateral isometric hip ADD strength can therefore be used as a simple clinical reference-point of full recovery of hip ADD muscle strength.

INTRODUCTION

Hip and groin pain is a common problem in the general population,[10-12] and is often related to physical function and sporting activity.[10,12,13] Pain in the hip and groin region in physically active patients is usually characterised by longstanding symptoms that can be difficult to fully recover from.[12,14]

Different treatment strategies are used concerning physically active patients with hip and groin pain, including different medical, exercise and operative interventions.[3,15-19] Novel treatment methods such as hip arthroscopy, incipient groin hernia repair, ultrasound-guided corticosteroid injections and specific exercise regimens, are advancing rapidly in the management of young to middle-aged physically active patients with hip and groin pain.[3,15-22] However, for the evaluation of treatment outcome in physically active patients with hip and groin disability, reliable, valid and responsive measurement tools are lacking. This means that novel treatment regimes are currently being developed without measurement instruments capable of evaluating their effectiveness.

Prevalence of hip and/or groin pain

The prevalence of hip pain in the general population (defined as hip pain during the last 12 months) is approximately 10%, and increases with age.[23] Pain in the hip and groin region in physically active patients is usually characterised by longstanding symptoms that in many cases do not resolve within 6-12 months.[12,14] Groin pain has especially been reported in sports such as football (soccer) (Figure 1) and ice-hockey,[24] and approximately 10-20% of all injuries in football and ice-hockey are hip and/or groin injuries.[10,25,26]

Figure 1. Football (Soccer)

Football is one of the most popular sports in the world, and it is estimated that more than 500 million people play football world wide.[27] In Denmark, it is estimated that 500.000 are

playing organised football, and that 90% of all males and 20% of all females have tried to play football.[27] The prevalence of hip and/or groin pain in Danish elite football has been documented to be 40-70%.[28,29] Hip and/or groin pain therefore constitutes a large problem and effective treatments for hip and/or groin pain are needed.

Definitions on longstanding hip and/or groin pain

Pain from the hip is difficult to localise and define. According to Birrell et al.,[30] this is due to three main reasons: *"first, the joint is not superficially located, so pain arising from structures in and around the hip joint can be felt across a broader region; second, pain from structures outside the hip—for example, the low back, the groin, and the urinary and genital tracts may also be associated with pain in the hip region (referred pain); third, it is unclear whether there is a specific topographical area that can usefully be distinguished as ''the hip''*.[30]

Birrell et al., previously developed and validated a pre-shaded drawing, covering the ''bathing trunk area,'' to be used for defining the presence of hip pain in studies of patients attending primary care (Figure 2).[30] The use of such a drawing has the advantage of allowing standardisation between different observers for the purposes of multicentre clinical studies. They showed that subjects whose pain satisfies both a pictorial and a verbal definition (where the patient uses the word ''hip'') have the strongest relation to indicators of hip disease.[30] This approach has been recommended when a specific definition is required for ascertaining individuals for study.

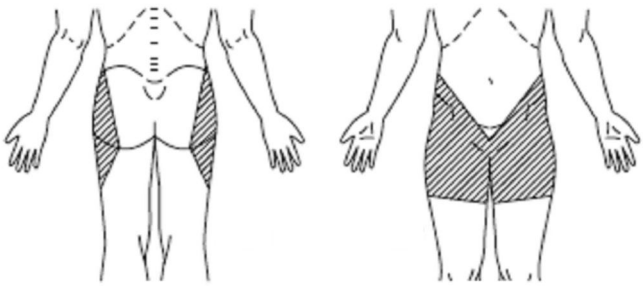


Figure 2. "Bathing trunk area", adapted from Birrell et al.,[30] 2005.

Studies have shown that patients with hip and groin pathology often report symptoms which are not restricted to the hip region.[31-34] The groin seems to be the most symptomatic region when patients report pain related to pathological conditions involving the hip joint.[31-34] These studies seem to confirm experiences from clinical practice, where patients reporting groin symptoms, often do not describe their symptoms as being located to the hip. However, as previously mentioned, the hip and groin regions have never been precisely anatomically defined and therefore merely reflect individual and cultural beliefs.[30] However, since a large majority of health care professionals and patients refer to the medial part of the hip region as the "groin", it is problematic only labelling this region as the "hip".

Hip and/or groin pain does not refer to any specific pathology, and patients with hip and/or groin pain are therefore a large heterogeneous group of people suffering from a variety of different pathological conditions. The hip and groin region is a complex anatomical region, and validated diagnostic tools for differentiation of musculoskeletal diagnoses in this region is lacking.[35-38] Many patients with hip and/or groin pain often seem to have more than one diagnosis or clinical entity,[31,32] however, their symptoms, activity limitations and participation restrictions are often very similar.

The term longstanding groin pain has previously been used in the literature,[39] but no general consensus on this definition exists. In a recent systematic review, on the effects of treatments for longstanding groin pain, longstanding groin pain was defined as groin pain of more than 6 weeks duration.[3]

Outcome measures used in intervention studies involving patients with longstanding hip and/or groin pain

Several systematic reviews evaluating the efficacy of different treatment modalities for patients with hip and/or groin disability, exist.[3,15-19] Numerous types of outcomes are evaluated in the individual studies. Symptoms, pain, muscle strength, return to sporting activity, and patient satisfaction (with treatment) are the most common outcomes measures

evaluated in these studies (Table 1). Only two of these studies consider the use and the quality of these outcome measures.[17,18] Robertson et al.[18] state in their systematic review on hip arthroscopy that: *"In the absence of well-validated outcome instruments to evaluate non-arthritic hip problems, leniency was given regarding outcome measures"*. In the study by Machotka et al.[17] the authors commented that no reliable and validated outcome measures were used in the included studies. The Modified Harris Hip Score (MHHS), a patient-reported outcome evaluating pain, function and activities of daily living (ADL), was used in three of the five included studies in the systematic review by Robertson et al.[18] The lack of focus on the quality of outcome measures is a general tendency in systematic reviews, which are often mainly concerned with obvious methodological qualities, such as randomisation procedures, control groups, blinding, compliance, drop-out, intention to treat etc.[40] Measurement properties have rarely been evaluated in the same methodologically stringent manner.

This means that novel treatment methods, such as hip arthroscopy, incipient groin hernia repair, ultrasound-guided corticosteroid injections and specific exercise regimens, are advancing rapidly in the management of young to middle-aged physically active patients with hip and groin pain,[3,15-22] without reliable and valid outcome measures to evaluate their effectiveness.

Table 1. Outcome measures used in studies included in six systematic reviews investigating the effectiveness of treatments for physically active patients with hip and/or groin pain

Author	Diagnosis	Treatment(s)	Included studies	Follow-up	Outcome measures	Measurement qualities of outcome measures
Robertson et al.,[18] 2006	Labral tears	Hip arthroscopy	5 case-series	Up to 3.5 years	Patient satisfaction rates, MHHS	In the absence of well-validated outcome instruments to evaluate non-arthritic hip problems, leniency was given regarding outcome measures
Swan and Wolcott,[19] 2006	Athletic hernias	Open hernia repair, laparoscopic hernia repair	1 RCT 1 non-randomised trial 13 case series	Up to 5 years	Time to return to play	No information
Caudill et al., [15] 2008	Sports hernias	Open hernia repair, laparoscopic hernia repair	25 case-series	1.5 months – 3.9 years	Return to full activity, return to sport, % excellent results, % satisfactory results, pain, symptom improvement	No information
Choi et al.,[16] 2008	Osteitis pubis, osteomyelitis of the pubic symphysis	Conservative therapy, injection therapy, surgical correction, (osteitis pubis) Antibiotics (ostemyolitis)	13 case series 3 case reports (osteitis pubis) 1 case-series 7 case-reports (osteomylitis)	Up to 14 months	Symptoms, pain (VAS), return to play, full recovery, return to running	No information
Jansen et al., [3] 2008	Longstanding groin pain	Exercise programs, stretching programs, injection therapies, surgical management: tenotomy, open hernia repair, laparoscopic hernia repair, neurectomy, neurolysis	2 RCT's 43 case-series	1 month -12 years	Full recovery, VAS, symptom improvement, recurrence, adductor strength, % excellent results, return to pre injury level, return to training, return to competition/play/sport, EMG	No information
Machotka et al.,[17] 2009	Groin pain	Exercise therapy, manual therapy, massage, stretching, electrotherapy, additional physical activity	1 RCT 2 Case-series 2 Case reports	Not specified	Not specified	No studies with reliable and valid outcome measures included

The International Classification of Functioning, Disability and Health (ICF)

Outcome measures can be related to body functions and structure (impairments), activities (activity limitations) and participation (participation restrictions) according to the ICF model.[2] Environmental factors that interact with all these components are also included. **Body functions** are physiological functions of body systems (including psychological functions) and **Body structures** are anatomical parts of the body (e.g. organs, limbs and their components), **activities** are the execution of tasks or actions by an individual, **participation** is involvement in a "real life" situation, activity limitations are difficulties an individual may have in executing activities, participation restrictions are problems an individual may experience in involvement in "real life" situations, while **environmental factors** make up the physical and social environment in which people live and conduct their lives. **Personal factors** are also included in the model but are not classified (Figure 3).[2]

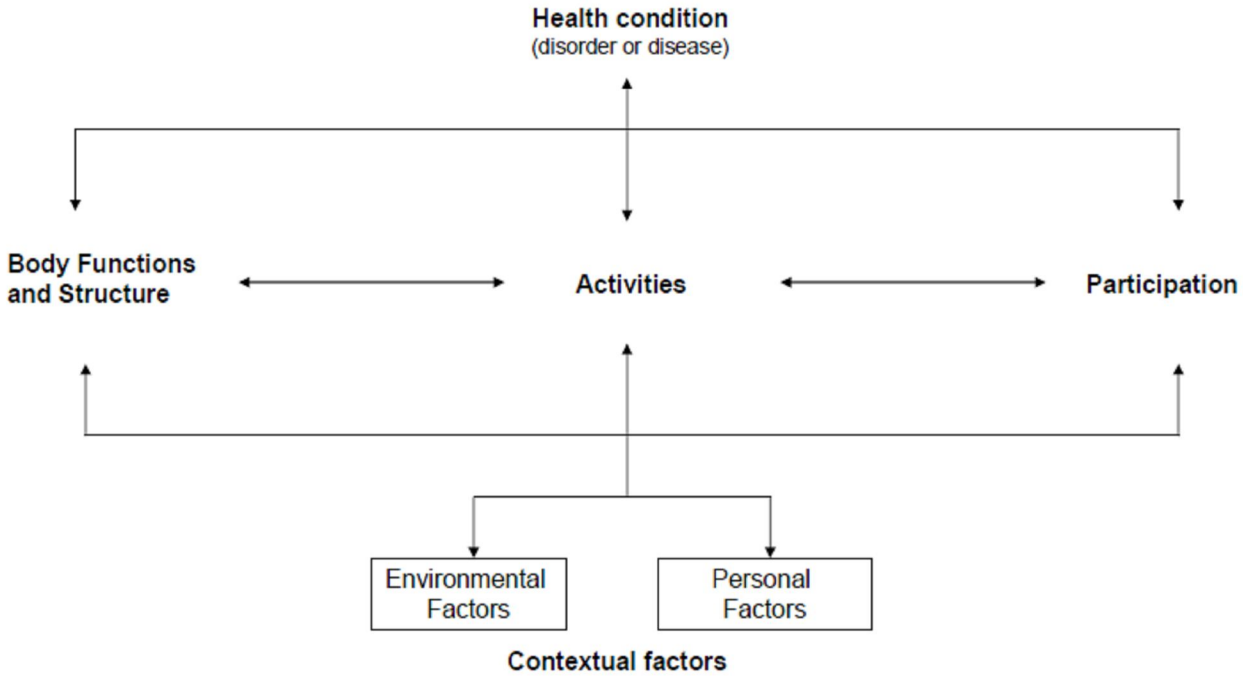


Figure 3. ICF model of disability. Adapted from WHO, 2002.[2]

Muscle strength testing

Muscle strength refers to the amount of force a muscle can produce with a single maximal effort. Size of muscle cells (the contractile component) and the ability of nerves to activate them (the neural component) are related to muscle strength.[41] Concentric muscle action occurs when the muscle shortens and joint movement occurs as tension develops. Eccentric muscle action occurs when external resistance exceeds muscle force and the muscles lengthen while developing tension. Isometric muscle action occurs when a muscle generates force and attempts to shorten but cannot overcome the external resistance.[41]

Hip strength assessment plays an important role in the clinical examination of the hip and groin region, and clinical outcome measures quantifying hip muscle strength are needed.[42] Decreased muscle strength seems to be a consistent finding in patients with hip and groin pathology.[43-45] In a randomised controlled trial including patients with longstanding groin pain, a larger increase in isometric hip adduction (ADD) muscle strength ($p<.001$) was documented in patients, who were treated with an active treatment approach, than patients who were treated with a passive approach. The active treatment approach was an exercise program aimed at improving the coordination and strength of the muscles stabilising the pelvis and hip joints, in particular the adductor muscles, and 79% of the patients treated with the active program returned to sport without groin pain. The passive treatment approach consisted of laser, transverse friction massage, stretching and transcutaneous electrical nerve stimulation of the adductor muscles at the pubic insertion point, and only 14% of the patients treated with this program returned to sport without groin pain.[46] Furthermore, decreased hip ADD strength in football and ice-hockey players, concentrically[47] and eccentrically,[48] seems to increase the risk of sustaining a groin injury.

Traditionally, manual muscle testing (MMT) has been used in the clinical assessment of hip muscle strength[49] (Figure 4).

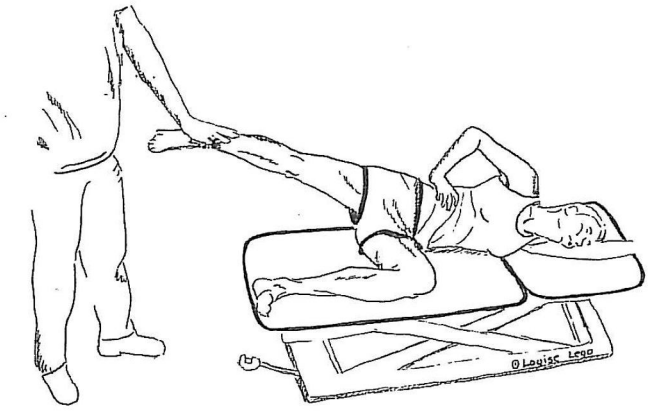


Figure 4. Manual muscle testing (MMT) of the hip

MMT evaluates muscle strength from 0-5, and defines the 6 levels as: 0 (Gone) no contraction felt, 1 (Trace) muscle can be felt to tighten, but cannot produce movement, 2 (Poor) produces movement with gravity eliminated, but cannot function against gravity, 3 (Fair) can raise part against gravity, 4 (Good) can raise part against outside resistance as well as against gravity, 5 (Normal) can overcome a greater amount of resistance than "good" muscle.[50] The advantage of MMT is that no equipment is needed, and the procedure is easy and quick to use. However, MMT has certain limitations when testing patients stronger than 3.[51] Most physically active patients, with no severe disability, score 5 in MMT, despite having muscular deficits. Hence a clear ceiling effect is present when testing these patients. A recent study have showed that the measurement error of MMT between testers is 1, indicating that it is very difficult for different testers to distinguish between the "nearest" levels.[52] Moreover, a classical study from 1956 showed that muscle-strength deficits up to 50%, assessed by quantitative measurement methods (dynamometer), could not be identified by MMT.[53]

In physically active patients with longstanding groin pain, a manual assessment method of hip ADD and hip flexion (FLEX) by Hölmich et al. have been proposed.[42] This method divides muscle strength into one of three levels; weak, intermediate and strong. Kappa values of the intra-observer reliability of this procedure ranged from 0.58-0.72, and the kappa values of the inter-observer reliability ranged from 0-0.22, indicating that the procedure is observer-dependable.[42] As with the 0-5 assessment method, this kind of scale may be able to distinguish between weak and strong patients, but cannot quantify degrees of strength or weakness.

Figure 5. The adduction strength test by Holmich et al.,[42] with permission

The method therefore seems to crude a measure for the detection of small but still relevant hip strength differences. The adduction strength test by Holmich et al.,[42] is performed with the patients lying in the supine position. The examiner stands at the end of the couch with hands and

lower arms between the feet of the subject to hold them apart. The feet of the subject point straight up, and the subject presses them together with maximal force, without lifting the legs or pelvis (Figure 5).[42]

Similar testing procedures, such as the squeeze test, have been introduced.[45] The squeeze test is quantified by using the cuff of a sphygmomanometer, which is placed between the knees, and the athletes are then instructed to squeeze the cuff as hard as they can using both legs. The highest pressure displayed on the sphygmomanometer dial (to the nearest 5 mmHg) during the test is then recorded.[45] Malliaris et al.,[45] showed that athletes with groin pain had reduced hip adduction pressure (force) in the squeeze test of approximately 20%, compared to healthy controls. However, when applying the squeeze test the measured pressure is produced by hip adduction of both legs. By testing each leg individually, it would be possible to achieve a greater depiction of the actual muscle strength in hip ADD in both the injured and uninjured limb, therefore, a unilateral and reliable quantitative strength assessment method seems warranted for physically active patients with hip and groin pain.

The hand-held dynamometer (HHD) is a quantitative measurement method for assessing muscle strength that has been used since the 1940's.[54] It is a portable measurement device, and it has previously been used for assessing hip muscle strength [48,55-57] The procedure is inexpensive and easy to administer, which makes it suitable for the clinical setting (Figure 6). HHD is a valid measure of muscle strength,[58,59] but different factors have been shown to influence the reliability of HHD when assessing muscle strength.

Mechanically, HHD is reliable,[60] and the experienced tester will produce good intra-tester reliability.[61-63] However, if the strength of the person being tested exceeds the tester's strength, reliability is compromised.[64,65] Furthermore, the testing position and the lever arm used for testing also seem to influence the reliability of the procedure.[66]

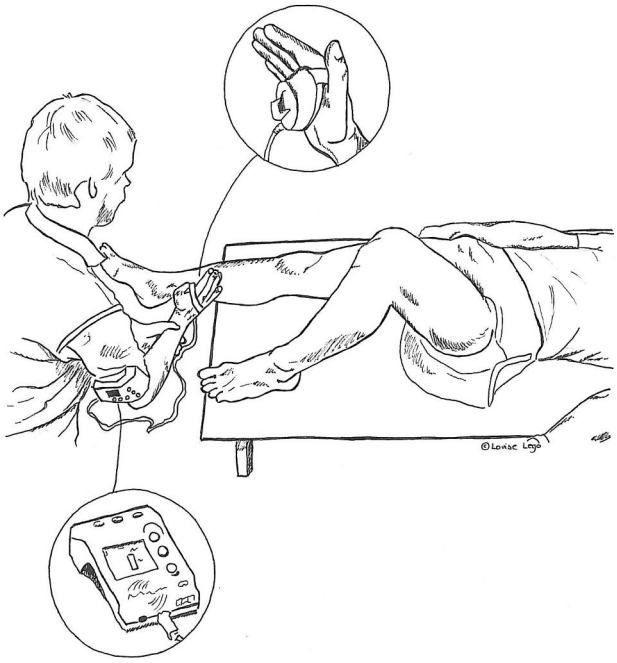


Figure 6. Hip ADD strength testing (make test, isometric contraction) using a hand-held dynamometer,[67] with permission

When using HHD, both isometric testing (make-test, Figure 6) and eccentric testing (break test, Figure 7) can be performed.[68] The reliability between make and break tests have been investigated in different studies with conflicting results. Bohannon[62] did not find a clear difference between the relative reliability of the two procedures while testing elbow flexion strength, while Stratford et al.,[59] showed that the relative reliability of the break test (ICC 2.1= 0.87) was lower than for the make test (ICC 2.1=0.95) when testing elbow flexor strength ($p<.05$).

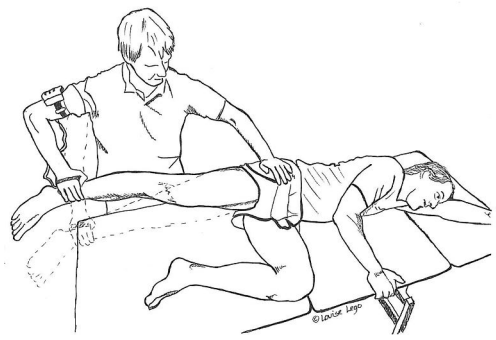


Figure 7. Hip ABD strength testing (break test, eccentric contraction),[70] with permission

A study by Burns et al.[69] showed that in a break test, greater strength values are recorded with faster angular velocities, and differences in angular velocities can therefore affect the reliability of this procedure.

Eccentric strength testing has shown greater strength values than isometric testing,[59,62] but a high correlation exists between the two types of tests (contraction types), and more than 65% of the variance produced by one type of test can be explained by the force produced during the other type of test.[58,59] This means that the make and break test presumable measure the same construct (maximal voluntary strength), just under different conditions.[58] Both tests have clinical advantages and disadvantages that should be considered before use. An advantage of the make test is that isometric loading induces less stress to the musculoskeletal system than eccentric loading,[59,62] thus minimising the risk of injury and delayed-onset muscle soreness.[71,72] In situations in clinical practice where eccentric testing is not feasible due to the pathological state of the patient, isometric testing should be preferred.

At present, different procedures have been reported concerning the positioning of the persons being tested.[66,73] However, consensus on a standardised procedure to determine isometric hip muscle strength using HHD does not exist. A study by Krause et al.,[66] showed that the relative reliability of hip ADD and ABD strength measurements in the side-lying position is better when using a long lever compared with a short lever. Furthermore, the relative reliability was better when using a long lever with a bench for stabilisation in testing hip ADD, compared with a long lever without bench stabilisation[66]. Kelln et al.[74] used an even longer lever when testing hip strength, by applying the hand-held dynamometer in different positions on the foot, when testing hip flexion (FLEX), extension (EXT), adduction (ADD), abduction (ABD), internal rotation (IR) and external rotation (ER) strength to offer testers the greatest possible mechanical advantage over the subjects. Both studies indicate that using long levers for hip strength assessments in healthy subjects produce satisfactory intra- and inter-tester reliability.[66,74]

When measuring muscle strength the SI unit (system of international units) for force is Newton. To express maximal torque, force is multiplied by the moment arm to get the SI unit Newton meter (Nm). The moment arm is the distance from the axis of rotation to the application centre of the load cell of the dynamometer. Nm is then normalised to bodyweight to get the maximal torque pr. Kilo bodyweight (Nm/kg bodyweight), and this is often referred to as muscle strength.[41,75] In this way, the influence of differences in body weight and limb lengths, known to have an isolated effect on muscle strength, has been controlled for.[41,75]

Strength data obtained by HHD can be used clinically in different ways. One possibility is to use normative values. However, normative values do often not exist for different age groups and levels of physical activity, and is therefore typically not an option. Another possibility is to use the unaffected limb as a control. A lower limb symmetry index (LSI) can then be calculated by dividing the strength of the affected limb by the unaffected limb.[76] Generally, it has been suggested that lower-extremity strength deficits of less than 10% on the injured side compared to the uninjured side should be considered the clinical milestone before returning an athlete to sport following an injury.[76,77] More specifically, the achievement of a hip ADD/ABD ratio of more than 90% and an hip ADD strength equal to that of the contralateral side has been recommended before returning to sport after an adductor strain.[78] However, Thorborg et al.[70] showed, that eccentric hip ADD symmetry cannot be assumed in injury-free soccer players. In fact, the dominant side was 14% stronger than the non-dominant side with regards to eccentric hip ADD strength, although hip ABD strength was similar. This finding of asymmetric eccentric hip ADD strength in injury-free soccer players, between the dominant and non-dominant leg,[70] indicates that using contralateral eccentric hip ADD strength as a reference-point for muscle recovery may be questionable.

Strength ratios, such as the hip ADD/ABD strength ratio, have previously been used for research purposes, and seem to be a relevant measure of hip strength,[48] especially in athletes with bilateral groin symptoms, where the contralateral limb cannot be used as a reference point. For clinical evaluation of the individual athlete, the obvious advantage is that this testing method does not require any age, limb-length or weight adjustment, since the player can act as his own control. This makes the testing method ideal for quick assessments in the busy clinical situation. However, the reliability of such ratio's are often not reported. Just as with the LSI, a ratio is the result of two tests, including the individual measurement variation of each test. These measures are therefore often less reliable, compared to one test or movement direction, and measurement variation is increased.

Patient-reported outcome questionnaires

There is a general consensus that Patient-Reported Outcomes (PROs) should serve as the gold standard in the assessment of musculoskeletal conditions, where the patient's perspective and health-related quality of life are of primary interest.[4,79-81] A PRO is any report coming directly

from patients about a health condition and its treatment.[4,5] The need for reliable and valid PRO instruments is emphasised in a study by Marshall et al.,[82] who demonstrated that clinical trials using unpublished measurement instruments were more likely to report positive effects of treatment than clinical trials using published instruments. Therefore, in order to properly evaluate the large spectrum of treatment strategies and regimens for young to middle-aged physically active patients with hip and groin pain, knowledge of validity, reliability and responsiveness of PRO questionnaires are needed. In the literature, validated patient-reported outcome measures exist for physically active patients with musculoskeletal conditions such as shoulder,[83,84] knee[85-88] and ankle pathology,[89-91] including both traumatic and overuse scenarios, and these instruments have played an important role in the evaluation of different treatments,[92-94] where the patient's perspective and health-related quality of life is of primary interest.

Validity, reliability and responsiveness of clinical outcome measures

The concepts of validity, reliability and responsiveness are important to understand when measurement instruments are used in research or clinical practice. In a recent international consensus process, including leading experts in the field of psychology, epidemiology, statistics and clinical medicine from all over the world, consensus on the taxonomy, terminology and definitions of measurement properties for outcome measurements instruments was reached.[95] It was formulated in a COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist.[96] The COSMIN taxonomy comprises three domains (reliability, validity and responsiveness), which contain the measurement properties. Interpretability is also part of the taxonomy and the checklist, although it was not considered a measurement property (Figure 8).

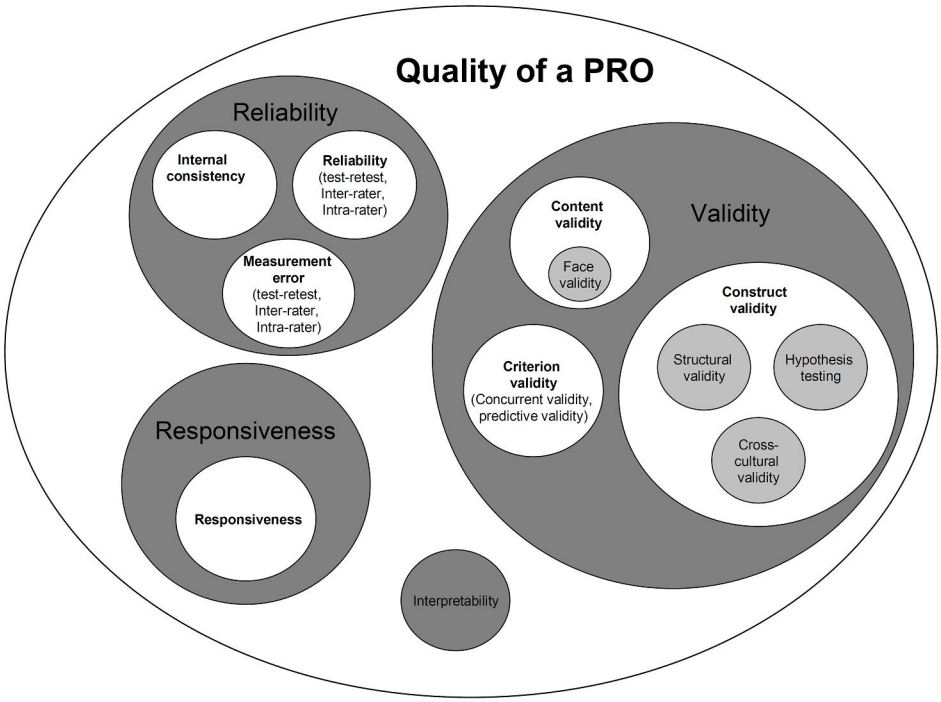


Figure 8. The COSMIN taxonomy. Adapted from www.cosmin.nl.

Reliability encompasses internal consistency, reliability (reproducibility) and measurement error. Validity encompasses content validity, including face validity, criterion validity and construct validity. Internal consistency is the degree of interrelatedness among the items e.g. in a questionnaire.[1]. Reliability is the extent to which scores for the same patients are unchanged for repeated measurements over time.[1]. Measurement error is the systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured.[1] Criterion validity is the degree to which scores of a measurement instrument are an adequate reflection of a "gold standard".[95] Construct validity is the degree to which the scores of a PRO instrument are consistent with a priori hypotheses, based on the assumption that the PRO instrument validly measures the construct to be measured.[1] Responsiveness is defined as the ability of a the instrument to detect change over time in the construct to be measured.[1] Interpretability is the degree to which one can assign qualitative meaning to an instrument's quantitative scores or change in scores.[1]

AIMS AND HYPOTHESES

The overall aim of this thesis was to develop outcome measures for physically active individuals with hip and groin pain. In the development process important measurement aspects such as reliability, validity and responsiveness were analysed. The primary hypothesis of the thesis was:

"It is possible to develop reliable, valid and responsive outcome measures for physically active individuals with hip and groin pain"

The specific aims and hypotheses of the present thesis were:

- To review the reliability, validity and responsiveness of available patient-reported questionnaires (PROs) assessing patients with hip and/or groin disability. The hypothesis was that patient-reported outcome measures with adequate measurement qualities for young physically active patients with hip and/or groin disability would be lacking.

- To develop and evaluate a new PRO questionnaire for physically active patients with hip and/or groin pain. The hypothesis was that a valid, reliable and responsive PRO questionnaire for young to middle aged physically active patients with hip and/groin pain could be developed.

- To examine the absolute test–retest measurement variation concerning various clinical strength assessments of hip abduction (ABD), adduction (ADD), external rotation (ER), internal rotation (IR), flexion (FLEX) and extension (EXT), in physically active subjects. The hypothesis was that an intra-tester, inter-day reliable clinical measure of hip muscle strength could be developed using a hand-held dynamometer.

- To compare isometric hip adduction (ADD) and abduction (ABD) strength on the dominant and non-dominant side in injury-free soccer players. The hypothesis was that hip ADD strength in soccer players would be larger in the dominant (DOM) compared to the non-dominant limb (NDOM).

STUDY I

Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: a systematic review

Material and methods

A systematic review of the literature concerning assessment of hip and/or groin disability was performed. The purpose with the review was to identify PRO questionnaires for patients with hip and/or groin disability, and to evaluate the psychometric properties of these outcome measures.

Search strategy

The following bibliographic databases were searched: MEDLINE via Pubmed (from 1945 to January 2009), EMBASE via OVID (from 1980 to January 2009), CINAHL via Ebesco (from 1982 to January 2009), Cochrane Central Register of Controlled Trials (up to January 2009), PsycInfo via OVID (from 1806/1987 to January 2009), SportsDiscus (up to January 2009) and Web of Science (from 1900 to January 2009). Our search strategy was:

Hip OR groin OR inguinal hernia
AND
Outcome assessment* OR self assessment* OR questionnaire*
AND
Reliability OR validity

The terms were searched as key words, in MEDLINE named MESH terms, in other databases Key words, where possible and also as "free-text" words appearing anywhere in the reference fields. From the retrieved and selected references, reference lists were checked for further relevant studies. Finally, specific searches for identified questionnaires were carried out, and experts in the field were contacted for possible additional references.

Study selection

Two reviewers (KT and EMB) independently carried out the selection of possible studies for inclusion from the retrieved references, based on titles and abstracts. All possible eligible studies were obtained in full and evaluated based on the inclusion criteria. Excluded studies were identified and presented with the reasons for exclusion (Figure 9).[97,98] The exclusion sequence in Figure 8 was chosen to decrease time expenditure, and criteria that were directly assessable from title, abstract or methods were evaluated first, while criteria that needed scrutinising the paper were chosen as second exclusion criteria. We included studies which fulfilled the following criteria:

1. The retrieved study was published in English, German or French, as a full report.

2. Psychometric properties in the study were evaluated with Classical Test Theory.[8]

3. The main purpose of the study was to evaluate one or more psychometric properties of a PRO questionnaire, including patients with hip and/or groin disability.

4. The study included a PRO questionnaire specifically concerning hip or/and groin disability, containing items related to impairment (body functions and structure), disabilities (activities), or participation problems (participation), according to The International Classification of Functioning, Disability and Health.[2]

5. Data on hip and/or groin disability could be separated from disabilities of other anatomical regions.

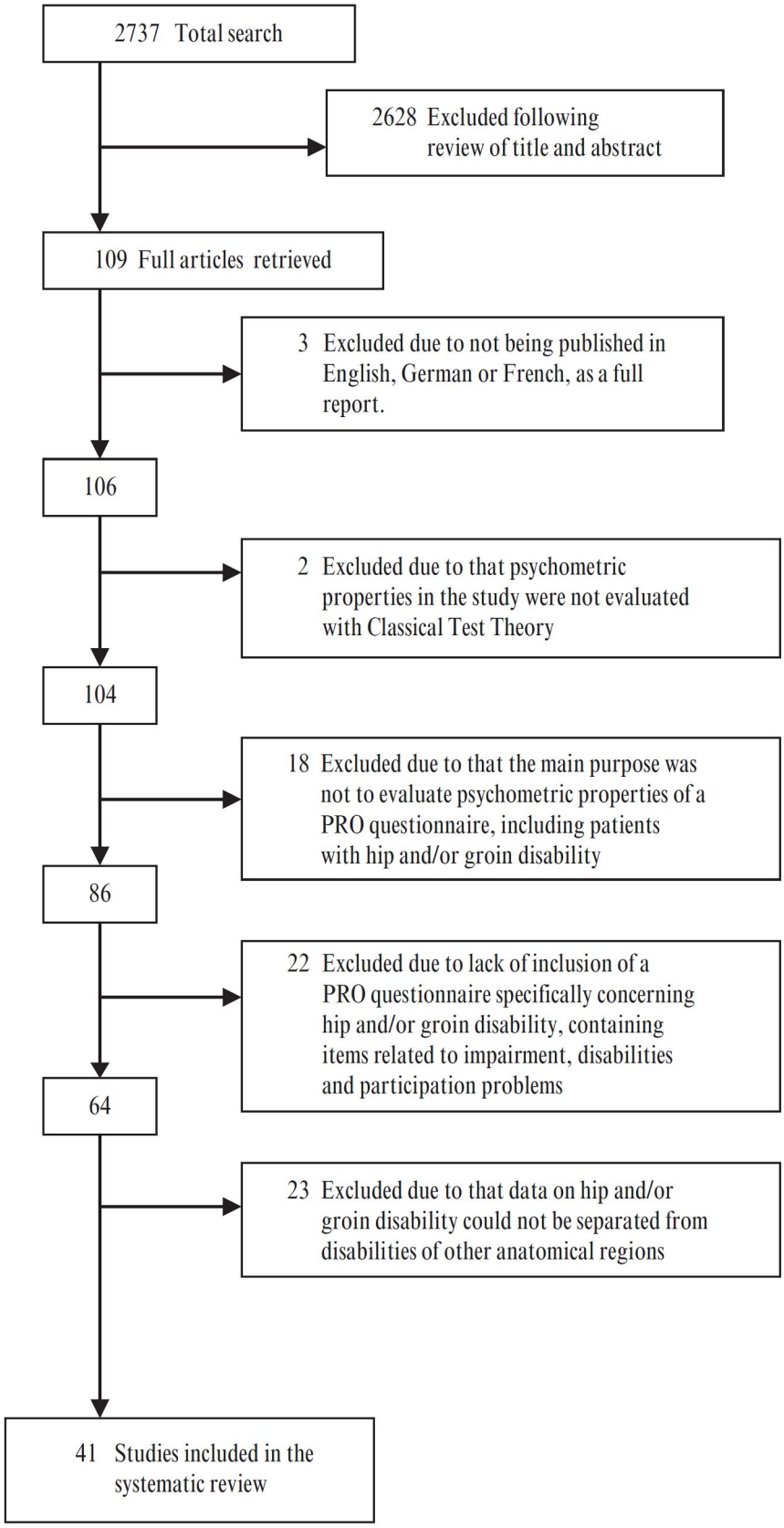


Figure 9. Selection of publications for the systematic review

Characteristics of studies and instruments

The descriptive data in each study had to provide information on psychometric properties evaluated in the study, time of administration, target population (diagnosis/clinical features), study population, and mode of administration. Extracted information from the identified questionnaires included full name of the questionnaire, abbreviation of the name of the questionnaire, assessment dimensions, and number of rating scales.

Data extraction and evaluation of psychometric properties

Based upon the guidelines for systematic reviews,[98] we used a criteria list for evaluative purposes and explicitly described the operationalisation of it (Appendix A). The criteria list in question was recently published by Terwee et al.[99] and is suited to give information on PRO questionnaires and their psychometric properties, where group comparisons are needed. This criteria list has recently been applied in other systematic reviews,[100-102] and we considered it the best available instrument for our purpose. Methodological issues of the criteria list were discussed and refined in the study group, which is in accordance with recommendations in the original article.[99] The original criteria list by Terwee et al. 2007 did not include inter-tester reliability,[99] but we decided to add the evaluation of inter-tester reliability, since some of the included studies in the present review used observer-administration and assessed the inter-tester reliability of this procedure in their study.

The present criteria list evaluated the psychometric properties: content validity, internal consistency, construct validity, floor and ceiling effects, test-retest reliability, inter-tester reliability, agreement, responsiveness and interpretability. Only for PRO questionnaires where observer-administration was introduced, inter-tester reliability was included in the overall quality evaluation. In the present study un-weighted kappa statistics were used to calculate the inter-tester reliability of the initial ratings by the two reviewers, since the ratings are considered nominal.[103]

Results

The total search identified 2737 publications. Following the screening of titles and abstracts, 2628 publications were excluded. Out of the remaining 109 publications, which were read in full, 68 publications were excluded since they did not fulfil our predefined inclusion criteria (Figure 9), leaving 41 studies, involving 12.779 patients, as our final data for reviewing (Table 2, Appendix B).

In three situations we found publications containing information on psychometric properties of PRO questionnaires based upon the evaluation of the same group of patients: (1)[104,105]; (2)[106-108]; and (3).[109-111] These may therefore be considered as one study. We did not exclude any of these part-studies, since each part included different measurement aspects and/or results.

A total of 13 PRO questionnaires were identified in the included studies (Table 3). Twelve PRO questionnaires considered the hip region and one questionnaire considered the groin region.

Table 3. Included PRO questionnaires for patients with hip and/or groin disability

Abbreviation	Full name	Measurement dimension(s)	Number of rating scales
AAOS-HS	American Academy of Orthopaedic Surgeons Hip Score	Symptoms, pain, mobility, ADL	1
HOS	Hip Outcome Score	ADL, sport	2
HOOS	Hip dysfunction and Osteoarthritis Outcome Score	Pain, symptoms, ADL, sport/recreation function, QOL	5
HRQ	Hip Rating Questionnaire	Global impact of arthritis, pain, walking, ADL	1
IPQ	Inguinal Pain Questionnaire	Pain, ADL, QOL	None*
LISH	Lequesne Index of Severity for Osteoarthritis of the Hip	Pain, walking, ADL	1
MHHS	Modified Harris Hip Score	Pain, function, ADL	1
NHS	Nonarthritic Hip Score	Pain, symptoms, ADL, activities	1
OHS	Oxford Hip Score	Pain, ADL	1
PASI	Patient Specific Index	Pain, symptoms, ADL, QOL	1
R-WOMAC-FS	Reduced Western Ontario and McMaster Universities Osteoarthritis index Function Score	ADL/physical function	1
THAOQ	Total Hip Arthroplasty Outcome Questionnaire	Pain, ADL, satisfaction/expectation	None†
WOMAC	Western Ontario and McMaster Universities Osteoarthritis index	Pain, stiffness, ADL/physical function	1–3

*IPQ includes 18 individual items.
†THAOQ includes three forms: a baseline form (15 questions); a history form (26 questions); a postoperative form (13 questions).
ADL, activities of daily living; QOL, quality of life; rom, range of motion.

The PRO questionnaires were assessed in three main target populations: Total hip replacement, hip osteoarthritis, and various forms of hip and groin pain or dysfunction (Table 4, Appendix C). The inter-tester reliability of the independent ratings based upon the criteria list was good (k= 0.79, CI 95% 0.73 - 0.84).[103] Disagreement was mainly caused by reading errors where one of

the reviewers had overlooked specific information on a specific psychometric property. Uncertainty or disagreement only had to be resolved by discussion with the third reviewer on two occasions, regarding internal consistency and agreement. The ratings of the questionnaires in the individual studies can be found in Table 4 (Appendix C). The ratings of the included questionnaires are synthesised in a summary and presented in Table 5.

Table 5. Summary of the quality assessment of the included questionnaire

Name of questionnaire	Content validity	Internal consistency	Construct validity	Floor and ceiling effect	Test–retest reliability	Inter-tester reliability	Agreement	Responsiveness	Interpretability
AAOS-HS						+†			
HOS	−	±‡	+‡	+‡	+‡	NA	+‡	+‡	+‡
HOOS	+	±*,†	+*,†	±*+†	+*,†	NA	±†	+*,†	+*,†
HRQ		±*	±*		+*	NA	±*	±*	+*
IPQ		±§	±§		−§	NA			
LISH		±*,†,¶	+*,¶±†		+*,†	±†		±*	−†
MHHS			+‡			NA			±‡
NHS	+	±‡	±‡		±‡	NA			
OHS	+	±*	+*	+*	+*	NA	±*	±*	+*
PASI	+		+*		+*	+*		±*	+*
R-WOMAC-FS		±*	±*			NA			
THAOQ	−		−*		±*	NA			
WOMAC		±*,†	±*+†	±*	±*,†	−*	−*	−*±†	+*,†

+ (Positive rating); ± (indeterminate rating); −(negative rating); blank (no information available); NA (not assessed).
*Total hip replacement.
†Hip osteoarthritis.
‡Hip arthroscopy.
§Groin-hernia repair.
¶Non-specific hip pain.

STUDY II

The Copenhagen Hip And Groin Outcome Score (HAGOS): Development and validation according to the COSMIN check list

Material and methods

Development of the questionnaire

The methodological framework for developing and evaluating a PRO questionnaire included the following steps: 1) identification of a specific patient population, 2) item generation, 3) item reduction, and 4) determination of the validity, reliability and responsiveness. Steps 1 and 2 involved developing a preliminary version of the questionnaire, which is described in the methods section. Step 3 involved testing the individual items and subscales of the preliminary version by analysing patient responses. Based upon these analyses, a final version of the questionnaire was decided upon. Step 4 involved testing the final version of the questionnaire for validity, reliability and responsiveness. Steps 3 and 4 are described in the results section. A flowchart of the complete study process is shown in Figure 10.

Population identification

The goal of this instrument is to evaluate hip and/or groin disability related to impairment (body functions and structure), activity limitations (activities) and participation restrictions (participation) according to the International Classification of Functioning, disability and health (ICF),[2] in young to middle-aged physically active patients with hip and/or groin pain. Disability in this study encompasses the health dimensions within the methodological framework of ICF as categorised in one of three levels: impairment (body structure and function), activity limitations (activities), and participation restrictions (participation).[2] The objective would be to achieve a quantitative measure of the patient's hip and groin disability according to the different levels of the ICF. The measure should reflect the patient's perception of his/her disability as well as his/her actual disability. Physically active patients refer to any patient who is physically active at least 2.5 hours a week.[2]

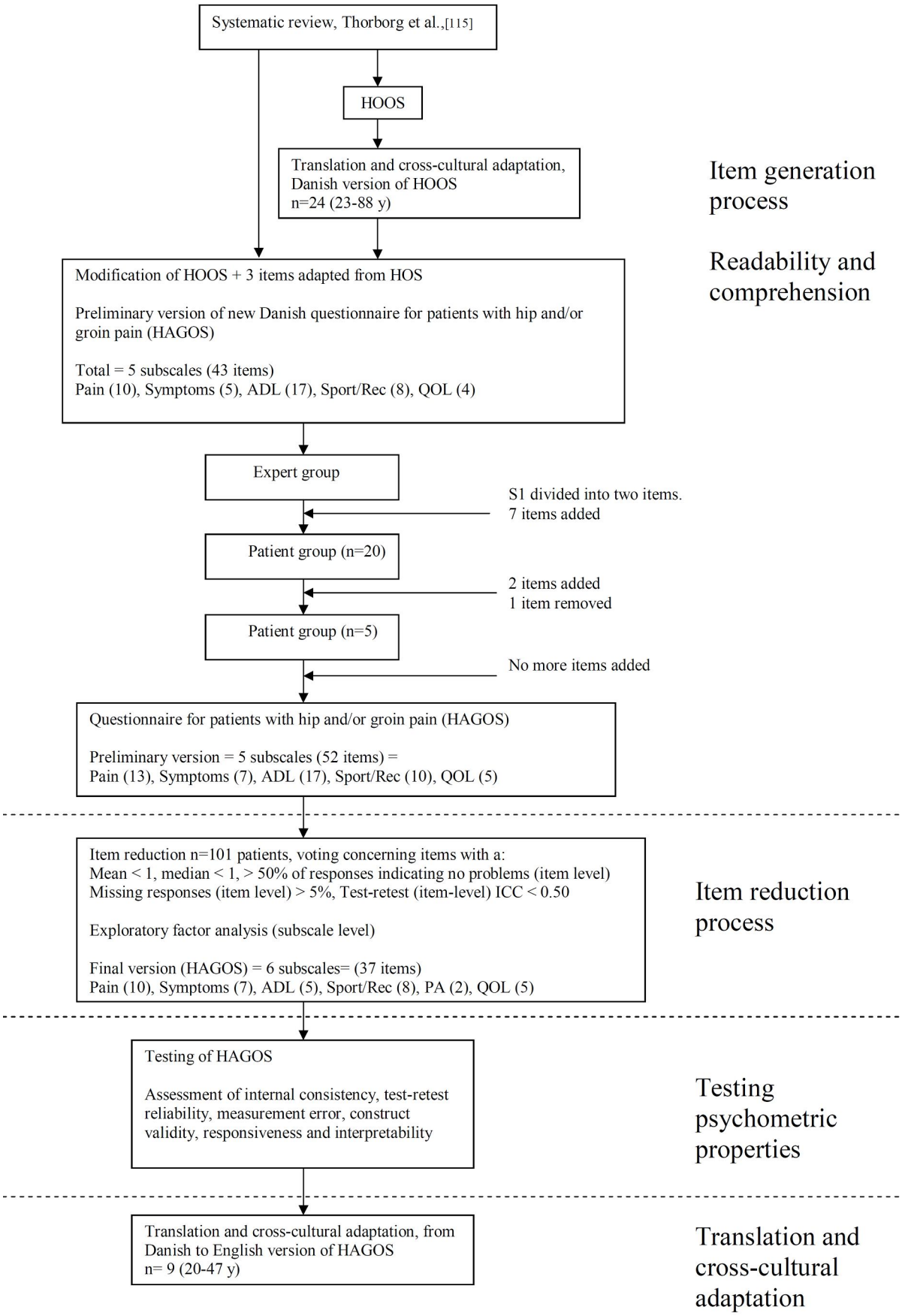


Figure 10. Flowchart of the study process

The groin is anatomically located in the anterior-medial part of the hip region, and the hip and groin region share vascular and neural supply.[112] The pathologies of the hip joint and the groin often present simultaneously and the symptoms can be overlapping.[31-34] This makes the hip and groin a complex anatomical region where validated diagnostic tools for differentiation of musculoskeletal diagnoses are lacking.[35-38] We therefore chose not to restrict our measurement instrument to be evaluated in a patient group with a specific diagnosis, but instead we wanted to focus on the commonalities of hip and/groin pain in physically active patients.

The patient flow is presented in Figure 11. Patients from primary and secondary care, who were at least 18 years of age, were screened by a specialist within the area of musculoskeletal examination of hip and/or groin pain in younger physically active patients. If the specialist suspected that hip and/groin pain was not of musculoskeletal origin, the patient was referred for further investigation and was not invited to participate in the study. All other patients presenting with hip and/or groin pain were considered eligible for the study, and were invited to participate. These patients were informed about the purpose of the research by the people responsible for the study, and written consent was obtained from those who agreed to participate. A self-reported questionnaire was used to screen for inclusion and exclusion of the patients who agreed to participate in the study. Patients seeking medical care presenting with hip and/or groin pain were included if they fulfilled all the following criteria; 1) had received treatment for their hip and/or groin pain 2) were restricted in their activities due to hip and/or groin pain 3) had hip and/or groin pain in the previous 14 days 4) had hip and/or groin pain of more than 6 weeks' duration 5) had hip and/or groin pain located in one of five predefined regions in a pain drawing (region 3,6,7,8 or 9, Figure 12), and 6) were physically active for more than 2.5 hours per week. Patients with self-reported limiting co-morbidities[113] were excluded from the study. The pain drawing (Figure 12) was adapted from methods for determining location of pain used in previous studies,[30,114] and pain of more than 6 weeks' duration has previously been defined as longstanding in nature concerning the population under study.[3]

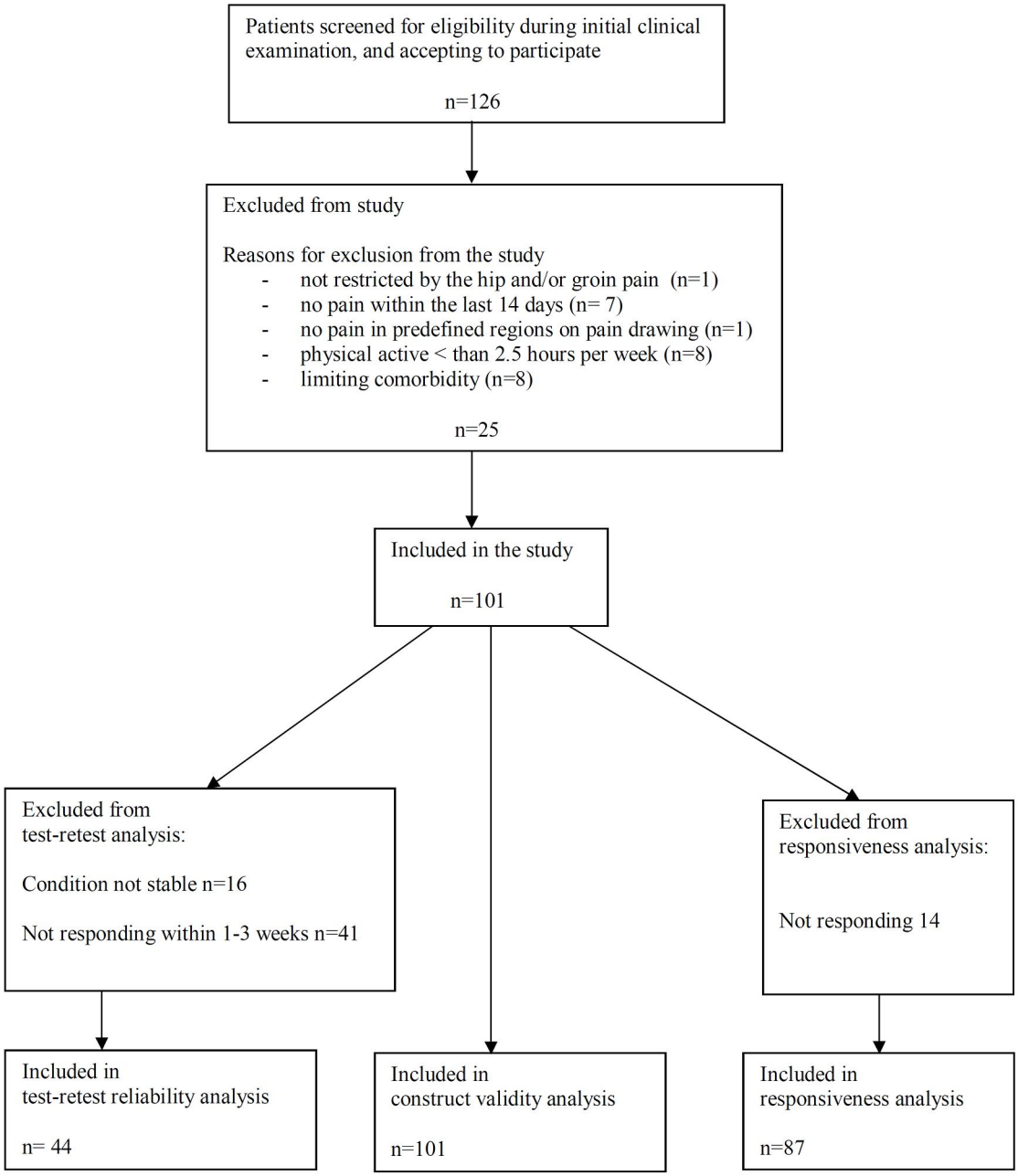


Figure 11. Clinical study profile

Item generation

The item generation phase included the following steps: a systematic review of the literature,[115] a focus group involving experts and individual patient interviews. The systematic review identified existing PROs that showed adequate measurement qualities or promise concerning validity, reliability and responsiveness when assessing patients with hip and/or groin disability.[115] The Hip dysfunction and Osteoarthritis Outcome Score (HOOS) and the Hip Outcome Score (HOS) were

found to be promising tools for patients with hip and/or groin disability, however the HOOS questionnaire had only been validated in patients with hip osteoarthritis or following total hip replacement, and the HOS in patients following hip arthroscopy. Therefore the items were not necessarily addressing our target group of young to middle-aged physically active patients with hip and/or groin pain.[115]

The HOOS was chosen as a template for the development of a new PRO questionnaire, because HOOS consists of items and subscales related to body functions and structure, activities and participation according to the ICF. It shows excellent measurement qualities in patients with hip disability for all dimensions. HOOS consists of five subscales: Pain, Symptoms, Function in daily living (ADL), Sport and Recreation function (Sport/Rec), and hip-related Quality of Life (QOL).[116] Furthermore, HOOS includes a format that is user-friendly, self-explanatory, and is already adopted in hip rehabilitation research worldwide.[115] We therefore decided to translate and cross-culturally adapt the HOOS from the original Swedish version into a Danish version according to existing guidelines [117,118] in a process that included 24 patients with hip disability (http://www.koos.nu/). We then incorporated and adapted relevant items from the HOS (Sports subscale), because this subscale contains some items that were not present in HOOS but might have been relevant.[119-121]. The items from the HOS were named SP7, SP9, and SP10 (Table 6, Appendix D).

Groin problems are common in physically active people and HOOS and HOS address dimensions, such as sport, that are relevant to young to middle-aged physically active people.[115] However, HOOS and HOS do not include groin-related questions, only questions related to the hip. This is problematic since young to middle-aged physically active patients often report groin symptoms[31,32,34] and often do not describe their symptoms as being located in the hip.[115] All questions in the new outcome questionnaire were therefore rephrased so they referred to the term “hip and/or groin”, instead of the term “hip” alone, to improve the face validity of the questionnaire. We found this appropriate based on existing data which have shown that patients with hip and groin pathology, often report symptoms that do not seem to be restricted to one of these anatomical regions,[31,32,34] recognising that these regions have never been precisely defined anatomically, and therefore merely reflect individual and cultural beliefs.[30] By using the

term “hip and/or groin”, we believe that the questionnaire covers a body region that also refers to the frontal and medial part of the hip region (the groin) which patients often refer to as a separate region.[30] The new questionnaire was therefore named the Copenhagen Hip And Groin Outcome Score, abbreviated to HAGOS (Appendix E, F and G).

Expert focus group

The second step involved interviewing experts in the field. Three doctors and four physiotherapists with experience and special expertise in treating physically active patients with hip and/or groin pain were interviewed. The experts underwent a semi-structured interview in which they were asked to fill out the preliminary version, while commenting on issues related to questions they felt were missing, the questionnaire’s readability and its ease of comprehension. The purpose of the interview was to identify relevant items that were missing and to improve the readability and comprehension of the questionnaire.

The experts commented that the introductory information on the questionnaire, where patients were asked to report disability related to the previous week, was problematic. The experts stated that many patients with hip and groin disability have had the problem for a long time and due to their disability, may not have performed these activities at all during the previous week, and therefore would not be able to answer this question in a valid way. It was therefore decided to add the following introductory information: *If an item does not pertain to you or you have not experienced it in the past week please make your “best guess” as to which response would be the most accurate*. This solution has previously been used in the format of the Western Ontario Rotator Cuff Index (WORC) and the Western Ontario Shoulder Instability Index (WOSI).[83,84] Because the current outcome questionnaire is not only a measure of actual disability but also perceived disability we found this solution appropriate. Based upon the focus group involving the experts, item S1 from the original HOOS[116] was divided into S1 and S2, since discomfort and clicking were considered to be different symptomatic aspects. Furthermore, six items, named P12, P13, SP5, SP6, Q4 and Q5, were added after suggestions by the experts (Table 6, Appendix D).

Patient interviews

The final step in the item generation process was to interview patients with hip and/or groin disability individually. Individual patients were specifically chosen for an interview, so that there would be representation of sex, age, type of injury, time from initial injury and severity of symptoms. The preliminary questionnaire was piloted on patients until data saturation was achieved. The patients underwent a semi-structured interview in which they were asked to fill out the preliminary version, while commenting on issues related to questions they felt were missing, the questionnaire readability and its ease of comprehension. This process included 25 patients, 12 male and 13 female (34±11 years). Twenty patients were interviewed individually before data saturation was achieved and two items were added, P2 and SP8 (Table 6, Appendix D). Furthermore, several patients mentioned that they did not understand the meaning of Q3 from the original HOOS: *How much are you troubled with lack of confidence in your hip?*[116] Even though the main purpose of this process was not to omit items we decided that the item had to be removed because too many patients did not understand the meaning of the question. This new preliminary version was piloted on five patients and did not require further modification. The preliminary questionnaire consisted after item generation of 52 items in five subscales (Symptoms (7), Pain (13), ADL (17), Sport/Rec (10), QOL (5).

Methodological testing and evaluation of measurement qualities of the new patient-reported questionnaire using the COSMIN checklist

Internal consistency

Internal consistency is the degree of interrelatedness among the items.[1]. A principal component factor analysis was performed on the individual subscales to assess their structural validity. Failure to load on a single major factor suggests that the items do not all measure the same construct. Chronbachs alfa was calculated per subscale and a score above 0.70 was taken as an indication of sufficient homogeneity of the items in the subscale.[99,122]

Test-retest reliability

Test-retest reliability is the extent to which scores for the same patients are unchanged for repeated measurements over time.[1] Intraclass Correlation Coefficients (ICC) were reported and

test-retest ICC should be ≥ 0.70 for all subscales.[99,122] Test-retest reliability was evaluated after 1 to 3 weeks in 44 stable patients. Patients reported at the retest whether their hip and/or groin pain was “better”, “not changed” or “worse” since the initial test. Patients reporting scores as “unchanged” were considered stable and included in test-retest reliability analysis.[95,96]

Measurement error

Measurement error is the systematic and random error of a patient’s score that is not attributed to true changes in the construct to be measured.[1] The Smallest Detectable Change (SDC), which is the threshold for determining clinical changes beyond measurement error was calculated on the basis of the Standard Error of Measurement (SEM) of the test-retest reliability.[99,123]

Construct validity

Construct validity is the degree to which the scores of a PRO instrument are consistent with a priori hypotheses, based on the assumption that the PRO instrument validly measures the construct to be measured.[1] Construct validity was studied by correlating the subscale scores of the HAGOS with the subscales of the Short Form-36 (SF-36). SF-36 (acute version, 1.0) was used because it is a PRO measure that contains relevant domains for assessing physically active patients with reduced physical function and pain.[124-126] SF-36 is a generic measure of health status which comprises eight subscales: Physical Functioning (PF), Role-Physical (RP), Bodily Pain (BP), General Health (GH), Vitality (VT), Social Functioning (SF), Role-Emotional (RE) and Mental Health (MH). The SF-36 is a valid and reliable instrument also when used in the Danish population.[127-129] Convergent and divergent evidence was examined by assessment of the associations between the HAGOS and SF-36 by the use of Spearman correlation. This construct validity was determined by cross-sectional comparison of the questionnaires when first administered.

A priori hypotheses were formulated.[95,96] We expected the highest correlations when comparing the scales that are supposed to measure similar constructs. Since the HAGOS is designed to measure physical health in patients with hip and/or groin pain rather than mental health, we expected to observe generally higher correlations between the HAGOS subscales and the SF-36 subscales of Physical Function, Physical Role and Bodily Pain (convergent construct

validity) than between the HAGOS subscales and the SF-36 subscales of Mental Health, Vitality, Role-Emotional, Social Functioning, and General Health (divergent construct validity).

Furthermore, we hypothesized that the correlation between the HAGOS subscales ADL and Sport/Rec and the SF-36 subscale Physical Function was at least 0.5, and higher than for the other HAGOS subscales. The correlation between the SF-36 subscale Pain and HAGOS subscales Pain and Symptoms should be at least 0.5 and 0.4 respectively, and higher than for the other HAGOS subscales. At last, for the subscale QOL, which hypothetically relates to both physical and mental health, we expected a correlation of at least 0.4 to the SF-36 subscale Mental Health.

Responsiveness

Responsiveness is defined as the ability of a PRO instrument to detect change over time in the construct to be measured.[1] For evaluating responsiveness, a Global Perceived Effect (GPE) score, where the patients rate their condition in one of seven categories was used. Patients were asked to rate possible change in their condition since the initial administration (baseline) in relation to their hip and/or groin pain. The GPE had the following answer options; much better (3) better (2) somewhat better (1) no change (0) somewhat worse (-1) worse (-2) much worse (-3). A priori hypotheses were formulated for responsiveness.[95,96] We hypothesised that the change in scores of the six subscales of the HAGOS between the initial administration and the 4-month administration would correlate with the GPE score, and that the correlation was at least 0.4 for all subscales. Furthermore, Standardised Response Mean (SRM) and Effect Size (ES) should be higher for patients who reported their condition to be better or much better, than patients reporting no change, only somewhat better or worse on the GPE score. SRM and ES should also be lower for patients reporting worse or much worse than patients reporting no change or only somewhat better or worse on the GPE score.

Interpretability

Interpretability is the degree to which one can assign qualitative meaning to an instrument's quantitative scores or change in scores.[1] Interpretability includes: the distribution of total scores and change scores in the study sample and in relevant subgroups, floor and ceiling effects, estimates of Minimal Important Change (MIC) and/or Minimal Important Difference (MID).[130]

Floor and ceiling effects are present if the questionnaire fails to demonstrate a worse score in the patients demonstrating signs of clinical deterioration and an improved score in patients who show clinical improvement as this can be an indication that a scale is not sufficiently comprehensive. In this study, floor and ceiling effects were defined to be present if more than 15% of the patients were reporting worst (0) or best (100) possible score.[99,131]

Statistical analyses

A sample size ≥100 patients and 7 times the number of items in the scale has been recommended for factor analysis.[99] Unidimensionality of the different subscales was assessed by exploratory factor analysis using Principal Component Analysis with varimax rotation in SPSS.[132] Median values were imputed in situations where missing values existed. Eigenvalues and factor loading patterns were used to identify and extract factors.[8] Items with the lowest factor loading were sequentially deleted until only one Eigenvalue above 1 was produced. The relative test-retest reliability has been calculated based on a linear mixed model (with participants handled as random effects). To estimate the test-retest reliability of the HAGOS subscales, ICCs (3.1, two-way mixed effects model absolute agreement) with 95% confidence intervals were calculated.[8]

Measurement error was expressed as the Standard Error of Measurement (SEM), which was calculated as Standard deviation (SD) x √1-ICC, where SD is the SD of all scores from the participants.[8,133] The SEM was used for calculating the SDC at the individual level, calculated as SEM x1.96 x √2, and at the group level calculated as SEM x 1.96 x √2 / √n.[134,135] Internal consistency, or inter-item correlation, was assessed by calculation of Cronbach's alpha of the baseline values.[8]

Convergent and divergent validity of the HAGOS and the SF-36 were investigated by Spearman's correlation coefficient. Likewise, associations on responsiveness were then measured by correlating the GPE with the change scores of each HAGOS subscale at the 4-month assessment, using Spearman's correlation coefficients. Furthermore, to evaluate the responsiveness of the HAGOS, two distribution-based statistics were evaluated concerning different groups of GPE: (1) the SRM, calculated as the mean change in score divided by the standard deviation of the change; and (2) the ES, equal to the mean change in score divided by the standard deviation of the

baseline score.[8] Both SRM and ES are calculated at the 4-month assessment, compared with baseline.

RESULTS

Prospective clinical study

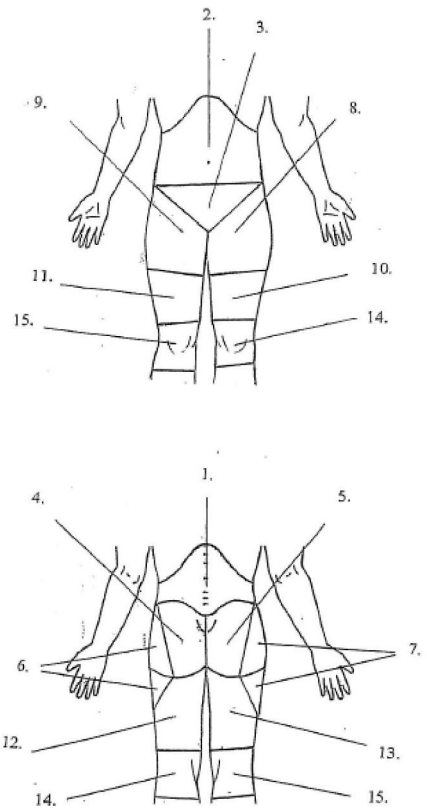
A prospective clinical study was designed to assess validity, reliability and responsiveness. The study was conducted at the Orthopaedic Department, Amager Hospital, Copenhagen. The Danish ethics committee of the Capital Region, and the Danish Data Protection Agency approved the study. Patients were recruited from primary and secondary care. One hundred and twenty-six patients were screened for eligibility during a clinical consultation by a doctor or a physiotherapist. One hundred and one patients were included in the study and they completed the HAGOS and SF-36 questionnaires at the initial consultation. Patients were sent the HAGOS after 1 week and asked to complete the questionnaire a second time and return it by mail as soon as possible. At the 4-month follow-up, the HAGOS and the GPE score were sent by mail, and completed at home. At the 4-month follow-up, patients who did not respond within 3 weeks received one reminder via email or telephone. Eighty-seven patients (87%) responded at the 4-month follow-up (Figure 11). The clinical study included 50 women and 51 men, mean age 36 years, range 18 to 63 years. Patient characteristics including age, height, weight, BMI, physical activity level, pain duration and pain medication use are shown in Table 7.

Table 7. Baseline characteristics

	Total (n=101)	Men (n=51)	Women (n=50)
Age, years mean (SD), Range	36 (11), 18-63	33 (8), 18-53	39 (12), 18-63
Weight, kg mean (SD), Range	74 (13), 32-104	81 (10), 62-104	67 (12), 32-96
Height, cm mean (SD), Range	176 (9) 159-198	182 (7), 166-198	169 (5), 159-180
Pain duration:			
>6 weeks	1 (1%)	1 (2%)	0 (0%)
>12 weeks	11 (11%)	9 (18%)	2 (4%)
>6 months	14 (14%)	8 (16%)	6 (12%)
>12 months	75 (74%)	33 (65%)	42 (84%)
Pain medication use:			
None,	80 (80%)	47 (92%)	33 (66%)
Paracetemol/NSAID,	18 (18%)	4 (2%)	14 (28%)
Opioids	3 (3%)	0 (0%)	3 (6%)
Physical activity:			
≥2.5 h/week	27 (27%)	11 (22%)	16 (32%)
≥ 5 h/week	40 (40%)	22 (43%)	18 (36%)
≥ 10 h/week	34 (34%)	18 (35%)	16 (32%)
BMI, kg/m^2, mean (SD), Range	23.78 (2.97), 17-31.05	24.51 (2.13) 20-31.05	23.4 (3.49), 17.7-31
Primary physical activity form:			
Cycling	26 (26%)	8 (16%)	18 (36%)
Soccer	18 (18%)	18 (35%)	0 (0%)
Running	15 (15%)	10 (20%)	5 (10%)
Strength training/fitness	13 (13%)	8 (16%)	5 (10%)
Other(s)	29 (29%)	8 (16%)	22 (44%)

n, number of patients; m, mean; SD, Standard Deviation; kg, Kilograms; NSAID, Non Steroidal Anti-Inflammatory Drugs; BMI, Body Mass Index; m^2, meter x meter; %, Percentage of patients.

Localisation of pain according to body region was reported by all patients and the results are shown in Figure 12.



1.	Back region	(16%)
2.	Upper abdominal region	(0%)
3.	Lower abdominal region	(11%)
4.	Left buttock	(9%)
5.	Right buttock	(11%)
6.	Left hip/buttock	(27%)
7.	Right hip/buttock	(18%)
8.	Left hip/groin/ thigh	(52%)
9.	Right hip/ groin/ thigh	(57%)
10.	Left anterior thigh	(5%)
11.	Right anterior thigh	(7%)
12.	Left posterior thigh	(4%)
13.	Right posterior thigh	(4%)
14.	Left knee	(11%)
15.	Right knee	(7%)

Figure 12. Pain drawing showing percentages of included patients (n=101) indicating pain in 15 predefined regions at baseline.

Content validity

Item reduction

Based upon the first and second administration of the preliminary HAGOS version (Table 6, Appendix D), item reduction was performed using the following strategy, which incorporated both quantitative and qualitative components: Individual items at the first administration (baseline), that had a median score of < 1, and/or a mean score of <1, and/or where more than 50% of the respondents reported no problems, and/or more than 5% of patients had a missing response to an

item, and/or a test-retest reliability (ICC 3.1, agreement) coefficient of less than 0.50 were considered possibly irrelevant for the population under study. For all 14 items identified as possibly irrelevant, four members of the study group voted about whether these individual items should be removed or not. Each member was told to consider the feasibility of each item based upon content, relevance, patient response and measurement qualities. Each member had one vote and items were removed if at least 3/4 voted for their removal. If 2 were for and 2 were against, consensus was sought by further discussion concerning the relevance of the item. Based upon this, 13 of the 14 items deemed possibly irrelevant were removed. Items P5 and P12 were removed from the Pain subscale. From the ADL subscale items A1, A3, A4, A6, A8, A9, A10, A11, A13, A14, A15, A17 were removed. Q4 also was considered for removal due to an ICC below 0.5, but it was decided to keep this item, since only one person in the study group voted for its removal (Table 6, Appendix D). After this process, the questionnaire consisted of 38 items in five subscales (symptoms (7), Pain (11), ADL (5) Sport/Rec (10) and QOL (5).

Internal consistency

Factor analysis of the five individual subscales showed that the items in the Symptom, ADL and QOL subscales loaded on one factor with Eigenvalues of 3.2 (46% of the variance), 3.3 (66% of the variance), and 2.9 (58% of the variance), respectively. Factor analysis of the Pain subscale showed that two factors with an Eigenvalue greater than 1 were produced. Factor analysis was repeated sequentially omitting item 13 “Do you have any pain when squeezing your legs together?” and the subscale only loaded on one factor, with an Eigenvalue of 5.6 (56% of the variance), and item P13 was therefore removed from the questionnaire. Factor analysis of the Sports subscale showed that two factors with an Eigenvalue greater than 1 were produced. Items 9 and 10 seemed to form a separate subscale and these were omitted from the Sports subscale and further tested as a separate subscale. Items 1 to 8 in the Sports scale loaded on a single factor, with an Eigenvalue of 5.3 (66% of the variance) and items 9 and 10 loaded on a single factor, with an Eigenvalue of 1.8 (89% of the variance) and this new subscale was named participation in Physical Activity (PA). The final version of the HAGOS then held 37 items in six separate subscales: Pain (10 items), Symptoms (7 items), ADL (5 items), Sport/Rec (8 items), PA (2 items) and QOL (5 items) (Appendix E and F). For each of the six HAGOS subscales, Chronbach’s alpha were above 0.79, indicating a sufficient homogeneity of all items in the subscales (Table 8).

Table 8. Descriptive statistics and test-retest reliability of HAGOS

Patients with hip and/or groin pain (n=44)	Test mean (SD)	Retest mean (SD)	Difference test-retest mean (SD)	SEM	SDC(ind)	SDC(group)	ICC (95% CI)	Chronbachs Alpha
Pain	62.3(20.6)	64.8(20.8)	2.6(9.6)	6.8	18.8	2.8	0.89(0.80-0.94)	0.91
Symptoms	56.5(16.7)	58.6(17.9)	2.1(9.0)	6.4	17.7	2.7	0.86(0.76-0.92)	0.79
ADL	68.6(23.5)	68.8(24.7)	0.1(10.1)	7.2	20.0	3.0	0.91(0.85-0.95)	0.87
Sport/Rec	45.0(26.0)	44.9(27.5)	-0.1(11.6)	8.0	22.2	3.3	0.91(0.84-0.95)	0.93
PA¶	25.9(30.7)	26.2(27.7)	0.3(17.8)	12.2	33.8	5.2	0.82(0.69-0.90)	0.87
QOL	33.4(15.8)	37.3(15.9)	3.9(8.4)	6.4	17.7	2.7	0.84(0.68-0.91)	0.81

ICC, intra-class correlation coefficient(3.1, agreement); CI, confidence interval; SEM, standard error of measurement, SDC(ind), smallest detectable change at the individual level; SDC(group), smallest detectable change at group level ; SD, standard deviation; A normalised score (100 indicating no symptoms and 0 indicating extreme symptoms) is calculated for each subscale; **¶,** n=43

Testing the final version of HAGOS

Missing data

HAGOS: Few individual items were missing. At baseline, 9 items from a total of 101 patients x 37 items = 0.2%, were missing. A total score could be calculated for all subjects for all subscales except for PA, where a total score could be calculated for all but one subject. At retest, 1 item of 44 patients x 37 items = 0.1% was missing. Test-retest analyses could be performed for 44 subjects for all subscales except for PA, where test-retest analysis could be calculated for 43 subjects. At the 4-month follow-up, 21 items of 87 patients x 37 items = 0.7%, were missing.

SF-36: Few individual items were missing. At the baseline measurement, 7 items of 101 patients x 36 items = 0.2% were missing. A total score could be calculated for all subjects for all subscales.

Test-retest reliability and measurement error

Table 8 shows ICCs, SEM and SDC of all subscales of the HAGOS. Retest was completed within a mean of 11 days, and a range of 7 to 21 days. For all subscales of the HAGOS, the ICCs were between 0.82 and 0.92 indicating good test-retest reliability. The SDC at the individual level ranged from 17.7 to 33.8 points and at group level from 2.7 to 5.2 points for the different subscales.

Construct validity

Generally higher correlations were found between the HAGOS subscales and the SF-36 subscales of Physical Function, Physical Role and Bodily Pain (convergent construct validity) than between the HAGOS and the SF-36 subscales of Mental Health, Vitality, Role-Emotional, Social Functioning, and General Health (divergent construct validity) (Table 9).

As hypothesized, the correlations between the HAGOS subscales ADL and Sport/Rec and the SF-36 subscale Physical Function were at least 0.5, and higher than for the other HAGOS subscales (Pain, Symptoms, PA and QOL). The correlations between the HAGOS subscales Pain and Symptoms and the SF-36 subscale Bodily Pain were at least 0.5 and 0.4 respectively and as hypothesized, higher than for the HAGOS subscales PA and QOL, but not higher than for the HAGOS subscales Pain and Symptoms. The subscale QOL was moderately correlated to the SF-36 subscale Mental Health, at 0.38 but did not reach the hypothesized threshold of being at least 0.4.

Table 9. Spearman's correlation coefficients (*rho*) determined when comparing the six dimensions in HAGOS to the eight different subscales in SF-36; N=101

	SF-36 Physical Function	SF-36 Physical Role	SF-36 Bodily Pain	SF-36 General Health	SF-36 Vitality	SF-36 Social Functioning	SF-36 Emotionel Role	SF-36 Mental Health
HAGOS Pain	0.67*	0.32*	0.64*	0.34*	0.22*	0.25*	0.08	0.17
HAGOS Symptoms	0.57*	0.22*	0.56*	0.34*	0.18	0.10	0.07	0.17
HAGOS ADL	0.76*	0.42*	0.68*	0.31*	0.19	0.35*	0.18	0.23*
HAGOS Sport/recreation	0.73*	0.32*	0.57*	0.29*	0.27*	0.35*	0.15	0.31*
HAGOS PA	0.37*	0.34*	0.23*	0.23*	0.30*	0.15	0.05	0.31*
HAGOS QOL	0.56*	0.36*	0.45*	0.32*	0.34*	0.32*	0.10	0.38*

*Significant correlation, $p<0.01$.

Responsiveness

As hypothesized, change in the six subscales of the HAGOS correlated with the GPE score, and the correlation was at least 0.4 for all subscales. As hypothesized, ES and SRM were lower for patients reporting worse or much worse than patients reporting somewhat worse, no change or somewhat better on the GPE score, for all subscales. Furthermore, ES and SRM for all subscales were higher for patients who reported their condition to be better or much better, than patients reporting no change or only somewhat better or worse on the GPE score (Table 10).

Table 10. Responsiveness

	GPE Score	"Much worse" and "worse"	"Somewhat worse" and "not changed" and "somewhat better"	"Much better" and "better"
	Total (n=87)	Total (n= 7)	Total (n=46)	Total (n=34)
HAGOS	Spearman Rho	SRM, ES	SRM, ES	SRM, ES
Pain	0.59*	-0.81, -0.63	0.23, 0.19	1.13, 1.12
Symptoms	0.68*	-0.77, -0.60	0.27, 0.16	1.27, 0.90
ADL	0.58*	-1.10, -0.89	0.08, 0.05	0.90, 0.77
Sport/Rec¶	0.61*	-0.96, -0.95	0.16, 0.10	1.01¤, 1.00¤
PA¶	0.56*	-0.88, -1.29	0.01§, 0.01§	1.08, 1.18
QOL	0.69*	-1.51, -0.84	0.21, 0.19	1.46, 1.78

ES, Effect Size; GPE, Global Perceived Effect; Standardised Response Mean, SRM; n, number of patients); ¶, n= 86; §, n=45; ¤, n=33; *Significant Spearman (*Rho*) correlation, p<0.01.

Interpretability

Floor and ceiling effects, predefined as present if more than 15% of the patients were reporting worst (0) or best (100) possible score, were found for the HAGOS subscales PA and ADL at some time points. Much larger floor and ceiling effects (40-80%) were seen for some of the SF-36 subscales. The distributions of total scores and change scores in the study sample and in relevant subgroups are presented in Tables 10 and 11, and floor and ceiling effects of the HAGOS and SF-36 are presented in Table 11.

Table 11. HAGOS score, baseline and 4-months assessment and SF-36 score, baseline assessment.

	Mean	SD	Median	Range	Floor effects	Ceiling effects
HAGOS Baseline (n=101)						
Pain	64.0	19.7	68	10-95	0	0
Symptoms	56.9	18.5	61	11-89	0	0
ADL	68.1	23.2	70	0-100	1(1)	9 (8.9)
Sport/Rec	45.5	25.9	44	0-100	1 (1.0)	2 (2.0)
PA ┼	25.8	29.0	13	0-100	39 (39)	3 (3.0)
QOL	33.5	16.1	35	5-75	0	0
HAGOS 4-months (n=87)						
Pain	73.4	19.4	75	30-100	0	5
Symptoms	67.8	20.2	68	18-100	0	4
ADL	75.8	22.9	80	15-100	0	19 (18.8)
Sport/Rec ¶	56.9	27.2	56	3-100	0	7 (7)
PA¶	36.1	34.2	25	0-100	28 (28)	7 (6.9)
QOL	45.6	23.4	45	5-95	0	0
SF-36 Baseline (n=101)						
SF-36 PF	70.5	19.7	75	20-100	0	3 (3.0)
SF-36 RP	65.6	35.2	75	0-100	13 (12.9)	40 (39.6)
SF-36 BP	54.3	20.0	61	0-84	3 (3)	0
SF-36 GH	74.5	18.3	77	20-100	0	7 (6.9)
SF-36 VT	62.2	19.3	65	5-100	0	1 (1)
SF-36 SF	90.1	18.2	100	12.5-100	0	67 (66.3)
SF-36 RE	86.8	28.3	100	0-100	6 (5.9)	79 (78.2)
SF-36 MH	77.5	15.3	80	28-100	0	3 (3)

Standard Deviation, SD

┼, n=100; ¶, n=86

STUDY III

Clinical assessment of hip strength using a hand-held dynamometer is reliable

Material and methods

Nine healthy participants were included in the study. 5 male, mean ± SD, age = 27 ± 5 years, height = 184 ± 7 cm, weight = 80 ± 8 kg and 4 female, mean ± SD, age = 25 ± 4 years, height =165 ± 8 cm, weight = 57± 7 kg. Only participants with no history of injury to the hip and groin region were included. All participants had to be physically active for at least 2.5 hours a week. The participants did not report any medical conditions compromising their physical function. The participants were instructed to maintain their regular training regimens throughout the experimental period, but exercising on the day prior to the test was not allowed. The participants had no prior HHD test experience.

Testing set-up

The testing was performed in a clinical examination room at the Department of Orthopaedic Surgery, Amager Hospital. The testing set-up included a portable HHD and an examination table. Muscle strength was tested with the Power track II commander (Figure 6). The dynamometer was calibrated on each test-day and all test procedures were standardised.

A physiotherapist (KT) with previous experience using the Hand Held Dynamometer (HHD) did all the testing. All strength tests were isometric strength test, also known as make tests.[68] Test and retest were performed with a one-week interval at the same time of the day. Each subject performed hip flexion, extension, abduction, adduction, external rotation and internal rotation, in two different testing positions for each movement direction.

Testing procedures

The test positions were chosen based upon procedures often applied in clinical settings.[50,66,73] It included 12 isometric tests, which were divided into 6 antagonistic pairs to avoid certain movement directions being repeated in succession (Appendix H). The participants were told to

stabilise themselves by holding on to the sides of the table with their hands. The examiner applied resistance in a fixed position and the person being tested exerted a 5 second isometric maximum voluntary contraction (MVC) against the dynamometer and the examiner.

The testing sequence of the 6 antagonistic pairs was randomised at the initial testing session, and this testing sequence was kept in the same order at the re-test session. After the participants were instructed in the procedures, they were asked to perform one isometric sub-maximal contraction into the investigators hand, to ensure that the correct action by the participant was performed. Then an additional practice trial, in the form of a MVC against the HHD was applied. The individual test was administered four times to reduce a possible learning effect. The highest value of 4 consecutive measurements and the mean of the 3 highest values are presented, since these procedures are commonly applied in MVC testing. The highest value is from hereafter referred to as the "best" value.

There was a 30 seconds rest-period existed between each trial, and after the fourth and the eighth test a five-minute rest-period was introduced. These rest-periods were introduced to avoid a decline in strength across trials due to fatigue.[68] The standardised command by the examiner was "go ahead-push-push-push-push and relax". The whole testing session took approximately one hour (Detailed information on the individual testing procedures can be found in Appendix H).

Calculation of strength ratio

Hip adduction/abduction strength-ratio's in the Supine Position (HADD/HABD-SUP), and in the Side-lying Position (HADD/HABD-SLP) and Hip Internal Rotation/ External Rotation strength-ratio in Prone Position (HIR/HER-PP) and in the Sitting Position (HIR/HER-SIP) were calculated based upon the individual strength measurements of each movement direction.

Statistical analysis

Distributions of variables are presented as mean ± one standard deviation (SD). The average of the test days and mean differences from test day 1 to test day 2 are presented. All the dependent variables demonstrated a normal distribution (Kolmogorov-Smirnov) and parametric tests were applied. Paired t-tests were used to examine if there was a systematic difference between test and

retest. Relative reliability is the degree to which individuals maintain their position in a sample with repeated measurements. To assess relative reliability Intraclass Correlation Coefficient (ICC) 2.1. coefficients (Two-way random model, consistency definition) with the corresponding 95% confidence interval (95% CI) was calculated. Absolute reliability is the degree to which repeated measurements vary for individuals, and was expressed as the SEM, which was calculated as SD× √1-ICC, where SD is the SD of all scores from the participants.[133] SEM is also presented as a SEM % by dividing the SEM with the average of the test and retest values. The SEM was used for calculating the Minimal Detectable Change (MDC) and was calculated as SEM×1.96×√2, to construct a 95% confidence interval.[133] A level of $P < 0.05$ was chosen to indicate statistical significance. Grubb's test was used to detect outliers in the individual test, and these were removed (Grubbs, 1969; www.graphpad.com, 2008).

Results

The reliability of individual hip strength measurements is presented in Table 12. Measurement variation was between 2-13% (SEM%) in the individual hip strength measurements. Hip extension-prone position-short lever (HE-PP-SL) was the only test where measurement variation was above 10%. No systematic differences were present when the best value of 4 measurements was used. A systematic difference was present in hip abduction-supine position (HABD-SUP), when calculating the mean of the three best measurements repetitions.

Table 12.Reliability of hip strength assessment

	Test (*N*) mean (SD)	Retest (*N*) mean (SD)	Difference test–retest (*N*) mean (SD)	Paired *t*-test	ICC (CI 95%)	SEM	SEM (%)	MDC
HABD-SLP								
Best of 4 reps.	128.9 (25.0)	126.4 (18.6)	−2.5 (15.8)	0.647	0.74 (0.21–0.94)	10.9	8.5	30.1
(M) 3 best reps.	125.9 (24.8)	120.3 (19.4)	−5.6 (15.6)	0.312	0.76 (0.24–9.94)	10.7	8.7	29.6
HABD-SUP								
Best of 4 reps.	144.2 (23.2)	143.9 (26.4)	−0.3 (5.8)	0.867	0.97 (0.89–0.99)	4.2	2.9	11.6
(M) 3 best reps.	139.5 (23.1)	135.8 (26.1)	−3.7 (4.7)	0.046*	0.98 (0.92–1.00)	3.4	2.5	9.4
HADD-SLP								
Best of 4 reps.	146.2 (23.0)	152.6 (24.4)	6.3 (16.8)	0.290	0.75 (0.22–0.94)	11.6	7.8	32.1
(M) 3 best reps.	141.9 (22.1)	145.6 (23.7)	3.7 (15.1)	0.488	0.78 (0.30–0.95)	10.5	7.3	28.9
HADD-SUP								
Best of 4 reps.	135.1 (30.0)	139.0 (30.9)	3.9 (11.2)	0.329	0.93 (0.73–0.98)	7.8	5.7	21.6
(M) 3 best reps.	130.1 (28.7)	134.7 (30.6)	4.6 (14.2)	0.359	0.89 (0.57–0.97)	9.6	7.2	26.5
HE-PP-LL								
Best of 4 reps.	214.1 (38.7)	215.4 (51.8)	1.3 (24.3)	0.873	0.86 (0.50–0.97)	16.6	7.7	45.9
(M) 3 best reps.	207.3 (40.0)	209.3 (50.7)	2.0 (22.3)	0.795	0.88 (0.56–0.97)	15.3	7.4	42.4
HE-PP-SL								
Best of 4 reps.	229.4 (56.3)	231.8 (65.6)	2.3 (40.3)	0.866	0.78 (0.30–0.95)	27.8	12.1	76.8
(M) 3 best reps.	218.3 (51.7)	218.4 (65.3)	0.2 (36.5)	0.988	0.81 (0.36–0.95)	24.9	11.4	68.9
HER-PP								
Best of 4 reps.	135.5 (36.4)	131.6 (36.4)	−3.9 (6.1)	0.116	0.99 (0.93–1.00)	3.5	3.0†	9.7
(M) 3 best reps.	130.1 (37.0)	127.6 (34.3)	−2.5 (6.7)	0.329	0.98 (0.91–1.00)	4.9	3.8†	13.5
HER-SIP								
Best of 4 reps.	129.7 (19.7)	131.0 (22.1)	1.3 (8.2)	0.639	0.92 (0.70–0.98)	5.8	4.4	16.0
(M) 3 best reps.	123.7 (18.9)	126.0 (22.1)	2.3 (6.6)	0.317	0.95 (0.79–0.99)	4.5	3.6	12.4
HF-SIP								
Best of 4 reps.	270.0 (49.0)	278.4 (43.0)	8.4 (19.3)	0.225	0.91 (0.66–0.98)	13.5	4.9	37.3
(M) 3 best reps.	258.8 (48.0)	271.2 (41.9)	12.4 (17.5)	0.067	0.92 (0.70–0.98)	12.5	4.7	34.5
HF-SUP								
Best of 4 reps.	212.6 (38.4)	222.3 (43.3)	9.8 (14.0)	0.070	0.94 (0.76–0.99)	9.8	4.5	27.1
(M) 3 best reps.	207.5 (36.0)	213.5 (42.2)	6.0 (14.9)	0.264	0.93 (0.71–0.98)	10.1	4.8	27.9
HIR-PP								
Best of 4 reps.	117.3 (20.6)	121.4 (24.6)	4.1 (10.5)	0.274	0.89 (0.60–0.97)	7.4	6.2	20.5
(M) 3 best reps.	114.1 (20.3)	117.4 (23.6)	3.4 (10.1)	0.349	0.89 (0.60–0.98)	7.1	6.1	19.6
HIR-SIP								
Best of 4 reps.	135.8 (26.5)	138.0 (23.3)	2.2 (15.9)	0.686	0.80 (0.33–0.95)	10.8	7.9	29.9
(M) 3 best reps.	128.0 (24.1)	132.4 (21.0)	4.4 (13.8)	0.368	0.81 (0.37–0.95)	9.6	7.4	26.6

*$P<0.05$.

†$n=8$.

M, mean; N, newton; ICC, intra-class correlation coefficient; CI, confidence interval; SEM, standard error of measurement; MDC, minimal detectable change; SD, standard deviation; reps., repetitions; HABD-SLP, hip abduction-side-lying position; HADD-SUP, hip adduction-supine position; HE-PP-LL, hip extension-prone position-long lever; HE-PP-SL, hip extension-prone position-short lever; HER-PP, hip external rotation-prone position; HER-SIP, hip external rotation-sitting position; HF-SIP, hip flexion-sitting position; HF-SUP, hip flexion-supine position; HIR-PP, hip internal rotation-prone position; HIR-SIP, hip internal rotation-sitting position.

Hip abduction-supine position (HABD-SUP, SEM=2-3%) generally showed less measurement variation than hip abduction-sidelying position (HABD-SLP, SEM=8-9%). Hip extension-prone position-long lever showed less measurement variation (HE-PP-LL, SEM=7-8%) than hip extension-prone position-short lever (HE-PP-SL, SEM=11-12%).

Hip flexion-sitting position (HF-SIP, SEM=4-5%) and hip flexion-supine position (HF-SUP, SEM=4-5%), hip internal rotation-prone position (HIR-PP, SEM=6-7%) and hip internal rotation-sitting position (HIR-SIP, SEM=7-8%), hip external rotation-prone position (HER-PP, SEM=3-4%) and hip

external rotation-sitting position (HER-SIP, SEM=3-5%) showed comparable measurement errors despite the different test-positions used for each movement direction. One outlier was detected for HER-PP when using the maximum value of 4 measurements and when calculating the mean of the three maximum values of 4 measurements, and these data were removed.

Table 13. Reliability of hip antagonist strength ratios

	Test mean (SD)	Retest mean (SD)	Difference test–retest mean (SD)	Paired *t*-test	ICC (95% CI)	SEM	SEM (%)	MDC
HADD/HABD-SLP								
Best of 4 reps.	1.16 (0.25)	1.22 (0.17)	0.05 (0.19)	0.423	0.63 (−0.10–0.90)	0.13	10.9	0.36
(M) 3 best reps.	1.16 (0.26)	1.22 (0.18)	0.06 (0.17)	0.308	0.69 (0.10–0.92)	0.12	10.1	0.33
HAAD/HABD-SUP								
Best of 4 reps.	0.96 (0.16)	1.02 (0.17)	0.06 (0.10)	0.175	0.76 (0.25–0.94)	0.08	7.9	0.22
(M) 3 best reps.	0.94 (0.16)	1.00 (0.17)	0.06 (0.10)	0.109	0.81 (0.35–0.95)	0.07	7.2	0.19
HIR/HER-PP								
Best of 4 reps.	0.91 (0.17)	0.93 (0.12)	0.02 (0.10)	0.519	0.76 (0.25–0.94)	0.08	7.9	0.22
(M) 3 best reps.	0.88 (0.15)	0.92 (0.13)	0.04 (0.07)	0.159	0.72 (0.17–0.93)	0.08	8.6	0.22
HIR/HER-SIP								
Best of 4 reps.	1.05 (0.11)	1.06 (0.15)	0.02 (0.09)	0.632	0.74 (0.20–0.93)	0.06	5.7	0.17
(M) 3 best reps.	1.04 (0.13)	1.06 (0.13)	0.02 (0.09)	0.511	0.76 (0.24–0.93)	0.06	5.7	0.17

*$P<0.05$.

M, mean; N, newton; ICC, intra-class correlation Coefficient; CI, confidence interval; SEM, standard error of measurement; MDC, minimal detectable change; SD, standard deviation; reps., repetitions; HADD/HABD-SLP, adduction/abduction strength ratios in the side-lying position; HADD/HABD-SUP, adduction/abduction strength ratios in the supine position; HIR/HER-PP, hip internal/hip external rotation-prone position, HIR/HER-SIP, hip internal/hip external rotation-siting position.

The reliability of HADD/HABD strength-ratio's, and HIR/HER strength-ratio's are presented in Table 13. Measurement variation was between 5-11% for the different measurements. No systematic differences were present. HADD/HABD-SLP (SEM=10-11%) was the only strength-ratio where measurement variation was above 10%. In HADD/HABD-SUP (SEM=7-8%), HIR/HER-PP (SEM=7-9%) and HIR/HER-SP (SEM= 5-9%) measurement variation was below 10%.

STUDY IV

Hip adduction and abduction strength profiles in elite soccer players: Implications for clinical evaluation of hip adductor muscle recovery after injury

Material and methods

100 male elite soccer players from 6 different clubs were included in the study. All players (age = 24±4 years, (mean ± SD), height = 183±7 cm, weight = 79±7 kg) played at the third best level (2. division, region East) in the national competition in Denmark. The players were semi-professional and trained 3-4 times a week, and played 1-2 games weekly.

The players were tested in February 2009 during the mid-season competition-break leading up to the second half of the season, which started approximately one month later. Injured players were not included. All players were tested when they attended a training session (before training), and only players that participated in the following training session were included in the study. No exercise/training or match was allowed on the day prior to the test. Players were excluded if they were unable to participate in the training as a result of injury at the time of testing, or were taking any pain medication. A self-reported questionnaire was given to the subjects before testing to determine limb dominance and injury history. The preferred lower limb for kicking a ball was defined as the dominant side. Injury history, restricted to hip and/or groin pain within the previous year, was recorded to avoid recall bias related to the collection of retrospective injury data from several body regions.[136] The self-reported questionnaire was administered by a physiotherapist (TMM) who was not involved in the strength testing.

Testing procedures

The testing set-up included a portable hand-held dynamometer (PowerTrack II Commander) and an examination table. The hand-held dynamometer was calibrated before testing and all test procedures were standardised. Maximal voluntary isometric hip ADD and hip ABD strength for the dominant and non-dominant side were tested for all the participants. The testing sequence was

randomised before the testing sessions to avoid systematic bias. Isometric hip ADD and hip ABD strength was measured in a “make” test in the supine position as introduced by Thorborg et al.[67] and calculation of the hip ADD/ABD ratio was based on these measurements. Substantial intra-tester, inter-day reliability of the isometric strength testing and the hip ADD/ABD ratio has previously been reported.[67]

The participants were placed in the supine position and were told to stabilise themselves by holding on to the side of the table with their hands. The examiner, a physiotherapist (AS), applied resistance in a fixed position 5 cm proximal to the proximal edge of the lateral malleolus, and the person being tested exerted a 5 second maximum isometric voluntary contraction against the dynamometer. After the participants were instructed in the procedures, they were asked to perform one practice trial. The highest of the 4 subsequent measurements were used in the analysis. If the last measurement was the highest another measurement was conducted until no further force increase was measured. There was a 30-second rest period between each trial. The rest-period was introduced to avoid a decline in strength across trials due to fatigue.[68] The standardised command by the examiner was “go ahead-push-push-push-push-relax”. The same examiner (AS), who was blinded to side dominance and pain during testing, did all the measurements.

After each strength test trial players were asked to rate any pain during testing on a numerical Visual Analog Scale (VAS) and this was recorded. The players reporting pain during strength testing were asked to specify the localisation of the pain and this was also recorded. Data from players reporting pain during testing was excluded from the analysis of isometric hip strength, since pain has been shown to reduce force production.[137]

Limb length was measured from the most prominent point on the anterior superior iliaca spine in the supine position to 5 cm proximal to the proximal edge of the lateral malleolus. Limb length was used to calculate torque and all force values were weight adjusted and reported as Newton-meters per. kg. body weight (Nm/kg).

Statistical analysis and data reduction

Data are presented as mean ± one standard deviation (SD). All data were statistically examined for normality of distribution (Kolmogorov-Smirnov). The non-dominant hip ABD dataset was the only dataset not being normally distributed, and all data in the analyses including the non-dominant hip ABD dataset was therefore log-transformed. Furthermore, two-sided F test was not significant for any unpaired comparisons thus parametric statistics were applied. Paired and unpaired student's t-tests were used appropriately. A level of $p < 0.05$ was chosen to indicate statistical significance. Grubb's test was used to detect outliers in the individual test, and these were removed (www.graphpad.com, 2008).

Data from 14 players were excluded from the analysis of isometric hip ADD and hip ABD torque for the following reasons; 1 player reported lateral hip pain during testing, 1 player reported pain from a bruise in the area where the load cell was applied during testing, 1 player had no preferred dominant limb, 1 player was not able to train due to groin pain, and finally 10 players reported groin pain during hip ADD testing corresponding to the area of the adductor muscles' proximal tendons, insertion or the pubic symphysis. It was decided to use the data for the 10 players in a post-hoc analysis of hip torque comparing players with groin pain during hip ADD testing to players with a pain-free test to investigate whether groin pain (VAS = 3.9 ± 2.1) during hip ADD testing has any implications for hip ADD/ABD ratio values.

All data were checked for outliers. One outlier in the group of players with a pain-free test was detected in the dominant hip ABD data set and this outlier was removed, because the test value indicated that an error had occurred in the measurement procedure or the data collection during the dominant hip ABD testing concerning this subject.

RESULTS

Of the 86 players included in this study, 69% reported that they had experienced hip and/or groin pain during the previous year. Of the 10 players with groin pain during testing 80% reported that they had experienced hip and/or groin pain during the previous year.

Hip torque values are shown in Figure 13 and 14. The dominant side was stronger than the non-dominant side for both isometric hip adduction (2.45±0.54 vs. 2.37±0.48 Nm/kg, p=0.02) and hip abduction (2.35±0.33 vs. 2.25±0.31 Nm/kg, p<0.001), corresponding to a 3% and 4% difference, respectively. Isometric hip adduction was greater than isometric hip abduction for both dominant (2.44±0.53 vs. 2.35±0.33Nm/kg, p=0.04) and non-dominant side (2.37±0.48 vs. 2.26±0.33 Nm/kg, p=0.03).

Isometric hip adduction/abduction ratio was not different between the dominant (1.04±0.18) and non-dominant side (1.06±0.17) (p=0.40). A post-hoc analysis showed that isometric hip adduction/abduction ratio was significantly lower in players with groin pain during hip adduction testing compared to players with a pain-free test (0.80±0.14) (p<0.001) (Figure 15).

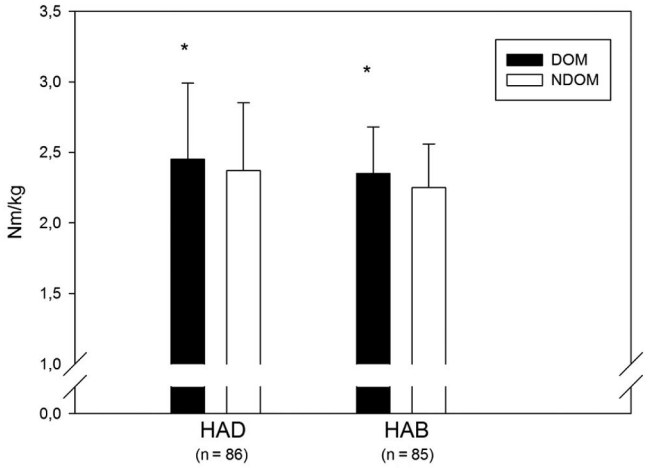


Figure 13. Hip adduction versus hip abduction strength profiles in soccer players. Hip adduction (HAD), hip abduction (HAB), dominant (DOM), non-dominant (NDOM), number of subjects included in the analysis (n). Limb length was used to calculate torque and all force values were weight adjusted and reported as Newton-meters per. kg body weight (Nm/kg). * Denotes statistically significant (p<0.05) within-group difference between HAD and HAB for DOM and NDOM side, respectively.

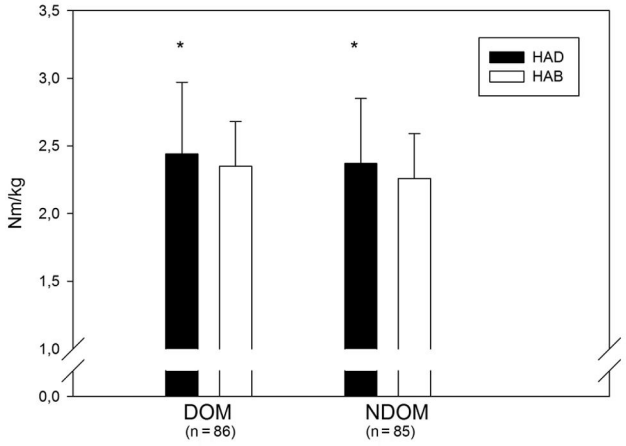


Figure 14. Hip adduction and abduction torque profiles on the dominant and non-dominant side. Hip adduction (HAD), hip abduction (HAB), dominant (DOM), non-dominant (NDOM), number of subjects included in the analysis (n). Limb length was used to calculate torque and all force values were weight adjusted and reported as Newton-meters per. kg body weight (Nm/kg). * Denotes statistically significant ($p<0.05$) within-group difference between DOM and NDOM side for HAD and HAB, respectively.

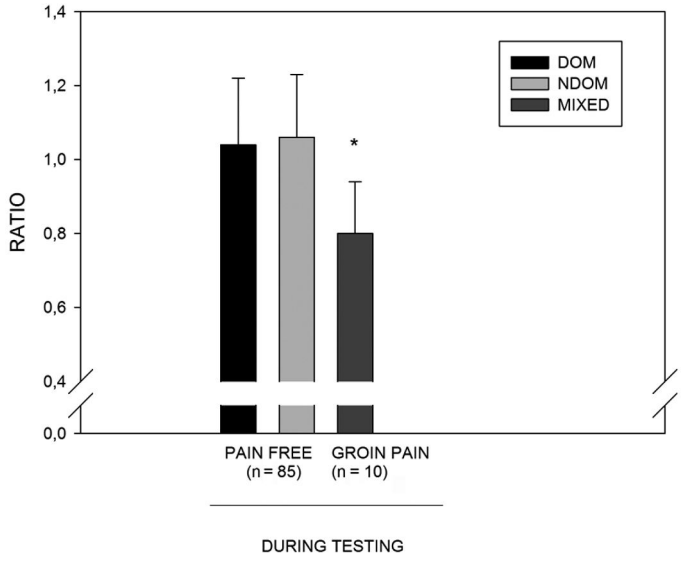


Figure 15. Hip adduction/abduction ratio. Hip adduction (HAD), hip abduction (HAB), dominant (DOM), non-dominant (NDOM), number of subjects included in the analysis (n). The group of 10 players with groin pain during testing represented 4 dominant and 6 non-dominant limbs (MIXED). * Denotes statistical significance ($p<0.001$) between-group difference between MIXED, and DOM and NDOM, respectively.

DISCUSSION

This thesis is based on four studies concerning the development and evaluation of outcome measures for use in the assessment of young to middle-aged physically active individuals with hip and groin pain. The outcome measures developed in this thesis (HAGOS and hip strength measurements) cover the three dimensions body functions and structure (impairment), activities (activity limitations) and participation (participation restrictions) as described in the conceptual framework of the ICF[2] (Figure 16). Where study I and II deals with outcome measures (PROs) related to body functions and structure (BFS), activities and participation, study III and IV deals with the assessment of hip muscle strength related to body functions and structure.

The subscale QOL from HAGOS seems to be the only measurement instrument that can be difficult to link to a specific level in the ICF-model, since it contains more generalised questions concerning the individuals' perception of hip and/or groin disability. Furthermore, "quality of life" as such is not included in the ICF.[138]

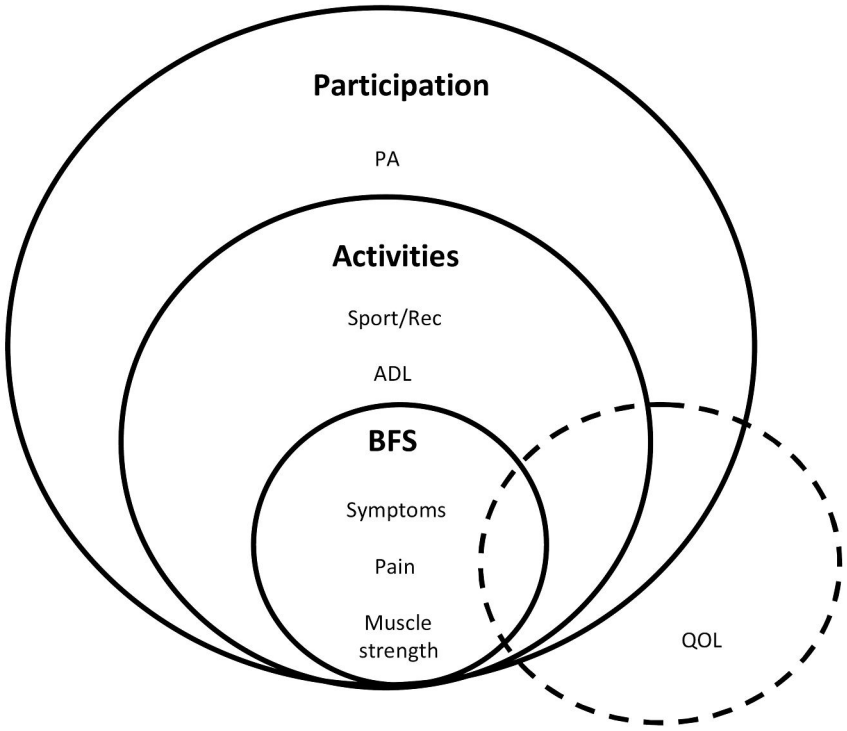


Figure 16. Hip strength measurements and HAGOS subscales relation to the conceptual framework of the ICF.

Patient-reported outcome questionnaires concerning physically active individuals with hip and groin pain

In study I, the literature was systematically reviewed concerning PRO questionnaires assessing hip and groin disability, and their psychometric properties were evaluated. Study I showed that HOS, MHHS and NHS are the only PROs which have been evaluated in a younger group of patients (mean age < 50 years) with hip and/or groin disability.

Evaluation of patient-reported outcome following hip arthroscopy

HOS has adequate psychometric properties when assessing young patients (mean < 50 years) undergoing hip arthroscopy and should be considered for this purpose. HOS involves patient-specific options, such as the use of non-applicable boxes for irrelevant questions.[119] Although patient-specific response options offer the advantage of identifying patient-relevant issues, they are not yet universally accepted by researchers.[139] The lack of standardisation of the items under study means that the scale cannot be considered the same in each patient, and the numeric score may not hold a common meaning.[139] Furthermore, the value of analysing the data statistically and calculating parameters such as means and correlations is questionable, and researchers should consider this issue before implementing this instrument.

Harris Hip Score is the most widely used instrument when assessing hip disability.[140,141] In study I, we only included the Modified Harris Hip score (MHHS), and not the original Harris Hip Score (HHS) since this instrument cannot be considered a true PRO questionnaire, because it is a composite score that combines patient-reported information and physical assessment performed by an observer. The MHHS only contains the patient-reported part of the HHS, and is currently and widely used in the assessment of young and active patients undergoing hip arthroscopy.[18,142-144] Study I showed that the psychometric properties of the MHHS have not been adequately assessed in the young population, and we cannot recommend the use of the MHHS in studies assessing younger and active patients undergoing hip arthroscopy. Furthermore, study I clearly identified that valid, reliable and responsive PRO questionnaires, assessing groin disability, are lacking in general.[115]

Development of a new patient-reported questionnaire (HAGOS)

In study II, HAGOS was therefore developed. HAGOS is to our knowledge, the first PRO questionnaire developed for young to middle-aged physically active patients with longstanding hip and groin pain using a prospective research design.

Study II is one of the first studies following the full COSMIN checklist in the development and testing of a PRO instrument - a checklist based on the recent international consensus process involving leading experts in the development and testing of PRO questionnaires.[95,96] With respect to the COSMIN checklist we found it easy to use and helpful in the design-phase of study II. The purpose of the COSMIN checklist is to evaluate the methodological quality of studies concerning measurement properties of PRO instruments. However, it is important to be aware that the COSMIN checklist is not yet aimed for a specific evaluation of the quality of the PRO instruments themselves.[95,96] In study II, we therefore had to rely on criteria for what constitutes adequate measurement qualities previously proposed by different authors.[99,122] In order to assess the quality of PRO instruments, we agree with the COSMIN panel that future consensus regarding criteria for what constitutes adequate measurement qualities should be included in the COSMIN recommendations,[130] to ensure methodological standardisation of this part of the process as well.

Content validity (HAGOS)

In contrast to the development of many previous PROs concerning hip disability,[115] the HAGOS meets the standards for the development of a PRO instrument, by including patients in the development process.[8,99] A study by Martin et al. 2009,[145] involving patients comparable to the patients in the study II, showed that large discrepancies exist between clinicians and patients, when they to rate the importance of different questions related to the patients hip problems.[145] This study by Martin et al. indicate that these patients perceive questions related to sports and recreation and social-emotional aspects to be of most importance.[145] This seems to be in accordance with the results of study II, in which the lowest baseline scores existed in the subscales Sports and Recreation, participation in Physical Activity and hip and/or groin related Quality of Life.

Internal consistency (HAGOS)

Unidimensionality of a (sub)scale indicates that all the items measure the same aspect.[8] The factor structures of the preliminary HAGOS subscales Pain and Sport/Recreation, developed in study II, were not unidimensional. Therefore, remodelling the factor structure of these subscales to create a new subscale, participation in Physical Activity (PA), seemed warranted. In the process of remodelling the factor structure, we removed one item in the Pain subscale, since this item did not conceptually fit under any of the other factors. This item asks about pain when *"squeezing legs together"* and may be difficult for patients to comprehend, since this is not a frequent activity or movement that all patients perform. This item was originally included by the expert panel in study II, and may represent a more clinical way of thinking, since the adductor squeeze is an important clinical test, performed in this population.[31,32,42,146] The factor analysis revealed that two items formed a separate subscale concerning the ability to participate in physical activity (PA). The PA subscale seems highly relevant for the population that it is intended for, because the inability to fully participate in sports and other physical activities often is one of the most frustrating aspects for these individuals.

Test-retest reliability and measurement error (HAGOS)

In study II, the ICC values were adequate for all subscales indicating adequate test-retest reliability at the group level.[99,122] The SDC for the subscales ranged from 15 to 18 points for the subscales Pain, Symptoms, ADL, Sport/Recreation and QOL. For the PA subscale, the SDC was 34 points. Changes above SDC values can be considered real changes at the individual level. The large SDC values at the individual level ($SDC_{individual}$) in the current study, is a common finding in PRO questionnaires,[147,148] indicating that patient-reported questionnaires can be difficult to use at the individual level, due to their incapacity to detect minimal but still clinically important changes.[123] At the group level, the SDC (SDC_{group}) ranged from 2.7 to 5.2 points for the different subscales, which means that changes above 5 points in group mean scores can be detected with 95% confidence. The fact that the SDC_{group} is much smaller than the corresponding $SDC_{individual}$ implies that the HAGOS is much better at detecting changes at the group level.

Construct validity (HAGOS)

Validation of instruments assessing PRO is a challenge since no gold standard is available for comparisons.[130] Instead construct validity has been assessed by correlating the new measure with already existing well-validated measures measuring similar constructs (convergent construct validity) and dissimilar constructs (divergent construct validity).[130] Because HAGOS is the first PRO for physically active patients with hip and/or groin pain, no ideal instrument for comparison existed. In study II, we therefore chose to use the SF-36 since this is a well-validated measure,[127-129] with adequate measurement qualities, which has been used in similar populations with similar musculoskeletal complaints from other anatomical regions.[124-126]

Responsiveness (HAGOS)

Responsiveness is a very important measurement quality in an outcome score,[122] because it is an indication of the PRO's ability to detect when patients are undergoing relevant clinical changes.[99,122] In the COSMIN process, it was recommended that appropriate measures to evaluate responsiveness are the same as those for hypotheses testing and construct validity with the only difference being that the hypotheses should focus on the change score of the instrument.[130] Therefore this approach for assessing responsiveness was chosen for study II.

The GPE score is only based on one transition question, and has therefore been assumed to be less reliable than a multi-item instrument.[149] However, despite its possible lack of measurement precision, all a priori hypotheses concerning responsiveness of all the HAGOS subscales were confirmed in study II, and showed high correlations between the GPE score and the change scores of the HAGOS subscales ranging from 0.56 to 0.69. Furthermore, study II showed that effect sizes for the different subscales for patients reporting to be "better" or "much better" ranged from 0.9 to 1.2 for Symptoms, Sport/Recreation and PA, whereas it was 0.77 for ADL and 1.78 for QOL. It indicates that more patients are needed for a clinical trial where the ADL subscale is used as the primary outcome. Conversely fewer patients are needed when QOL subscale is the primary outcome that needed when using the subscales Symptoms, Sport/Recreation and PA as primary outcomes, in clinical trials.

Interpretability (HAGOS)

In study II, only a few patients reported a floor or ceiling score for the HAGOS, indicating that both improvement and deterioration over time can be detected. The exception was the subscale PA where 39 subjects reported worst possible score (floor effect) at the initial administration and 28 patients reported worst possible score at the 4-month administration. A floor effect of the PA subscale was, however, not surprising considering the response options in these items. The answer options to the questions concerning the ability to participate in physical activities ranges from "always" to "never". It is not possible to participate to a degree less than "never", and therefore the large number of patients answering "never" to these questions does not seem problematic for the subscale because further deterioration is not possible. Instead we believe that the floor effect for this subscale emphasize the relevance of these items for the population under study. For the ADL subscale, a ceiling effect was present at the 4-month assessment. Again, this is hardly surprising since the items concerning function and Activities of Daily Living (ADL) are usually not the most important for the population under study.[145] However, for patients with severe hip and groin pain, assessing their limitations in daily activities may still be relevant.

Large ceiling effects were seen in the SF-36 for the subscales Physical Role, Social Functioning and Emotional Role, indicating that these subscales may not be very relevant for the population in study II. However, for the subscales Physical Functioning and Bodily Pain, which were primarily used for testing convergent validity in the current study, no floor and ceiling effects existed.

The Minimal Important Change (MIC) or the Minimal Important Difference (MID) has been proposed for establishing cut-points for minimal, but still patient-relevant clinical improvements. The MIC is the smallest change in score (within a patient) in the construct that can be measured that patients still perceive as important.[130] The MID is the smallest difference in the construct that can be measured (between patients) that is considered important.[130] There is an ongoing debate in the literature, about which methods should be used to determine the MIC and/or the MID of a PRO instrument.[130] Within the COSMIN Delphi process, no consensus on standards for assessing MIC or MID could be reached,[130] which is also reflected in the large variation in reporting and interpretation of these concepts in the literature.[149] However, it has been shown that under many circumstances, when patients with a chronic disease are asked to identify

minimal change, the estimates fall very close to half a standard deviation.[150] The MIC of the HAGOS subscales would fall between 10 and 15 points for the six subscales, using this approach (Table 11). We recognise that future research on the interpretability of PRO instruments may provide new evidence that necessitates a different approach. Until then, we agree with Norman et al. that applying the rule of thumb that the estimates of the MIC fall very close to half a standard deviation does not seem inappropriate in the absence of more specific information.[150]

Methodological limitations of studies concerning the development and evaluation of PROs (study I and II)

The criteria list (Appendix A) used in Study I was developed to evaluate psychometric properties of PRO questionnaires based on Classical test theory.[8] Item response theory is a relatively new method to evaluate questionnaires in health care, and has some potential advantages over Classical test theory.[8,151] The Rasch model, a mathematical model applied in item response theory, has been used to develop and internally validate measures, and it uses a logistic function that creates an interval-scaled measure. Our criteria list in study I was only developed to evaluate psychometric properties of questionnaires based upon Classical test theory, and this is a limitation of study I, but a limitation we could not avoid with the present available data. Another limitation of study I is that no gold standard exists to evaluate psychometric properties of PRO questionnaires, and our chosen criteria list may therefore be disputed. There are other criteria lists available,[122,152] but none of these have such detailed criteria for adequate measurement properties as the criteria list published by Terwee et al.[99]

In the future, criteria that evaluate methods and results of studies using item response theory models must be developed,[99] since this method has gained acceptance. Moreover, studies on developing and/or evaluating questionnaires based on item response theory are now more frequent. In study II we used classical test theory to evaluate the HAGOS because the sample size (n=101) of study II was too small for Rasch analysis, since we needed a sample size of at least 200 patients for this.[153] Rasch analysis should certainly be considered for possible improvements of the HAGOS in the future when a larger sample size can be included. Another limitation of study II is that HAGOS was only tested in Denmark (Danish). However, based upon the experiences of HOOS, which was originally developed in Sweden (Swedish),[116] this does not appear to be a

barrier to translation into other languages. Since Danish is not a world language, we decided to translate and cross-culturally adapt the HAGOS into an English version according to existing guidelines.[117,118] This version is given in Appendix F.

Strength testing of physically active individuals with and without groin pain

Reliable hip muscle strength measurements make it possible to objectively determine whether changes in hip strength have occurred over time. Reliable hip muscle strength measurements can also provide a screening tool for the detection of hip muscle weaknesses in healthy individuals, which has been shown to be a risk factor for sustaining a groin injury.[47,48]

Test-retest reliability of hip strength measurements using HHD

The main findings in study III were that standardised strength assessment procedures of hip ABD, ADD, IR, ER, FLEX and EXT can be performed in a clinical setting with small measurement variation. In 11 of the 12 applied tests, strength changes above 10% (SEM%) can be considered to be ‘‘real’’ changes in healthy individuals. No systematic differences between test and retest were found when the best value of four repetitions was used.

Study III is, to our knowledge, the first study investigating the test–retest measurement variation of the hip ADD/ABD strength ratio and hip IR/ER strength ratio. The hip ADD/ABD strength ratio was first introduced by Tyler et al. (2001).[48] The hip IR/ER strength ratio has not been described previously in the literature, and the present study shows that reliable measurements of this procedure can be obtained, both in the prone and in the sitting position. However, the clinical relevance of the hip IR/ER strength ratio and its possible implications need to be investigated in future studies.

Clinical application of muscle strength testing using HHD

Belt stabilisation is often used for clinical and for research purposes when using HHD. We deliberately avoided the use of belts or other stabilisation aids, because we wanted to make a simple testing set-up without extra equipment and with very simple instructions. We wanted the measurement procedures to be easy to learn, administer and implement in the clinical setting. The self-stabilization method, where the patients hold on to the table, worked well during testing

without causing stabilisation problems, and our results suggest that this seems to be an acceptable method.

Study III demonstrated that testing in the supine position apparently produce less measurement variation than in the side-lying position, when testing hip ABD and ADD strength. A possible explanation for this could be that stabilisation issues concerning the person being tested are completely eliminated in the supine position, compared with the side-lying position.
Studies investigating the reproducibility of hip strength measurement using HHD have often only reported the relative reliability.[154,155] However, relative reliability does not provide a cut-off score for delineating a true change from the measurement variation, which is necessary for valid clinical decision making. Another problem with reporting only the relative reliability is that it does not provide an insight into the absolute reliability obtained in different studies and with different testing procedures, making it difficult to choose the most relevant measurement procedure for a certain clinical problem. Therefore, the application of absolute parameters such as the SEM has been advocated.[133] We therefore decided to present both the ICC and the SEM.

Hip adduction and abduction strength profiles in soccer players

Study IV investigated the isometric hip ADD and hip ABD strength in elite soccer players, using the HHD method as described in study III. Study IV showed that isometric hip ADD and ABD strength is only marginally greater on the dominant compared with the NDOM side in soccer players. A between-side difference of 3% to 4% in isometric hip ADD and hip ABD strength, respectively, in soccer players is small, especially when you compare that to eccentric hip ADD strength testing in which the difference between the DOM and the NDOM side can reach up to 14% in soccer players.[70] The reason for the seemingly larger difference between DOM and NDOM hip ADD strength during eccentric testing compared with isometric testing is unknown. However, because eccentric muscle (EMG) activity in the hip adductors is substantial during the early swing phase of a soccer kick,[156,157] it is possible that an improved eccentric hip ADD strength adaption on the dominant side is a result of improved coordination of the motor system from repetitive kicking.

Clinical application of muscle strength testing in soccer rehabilitation

Study III showed that the absolute measurement variation for the testing procedure used in study IV is 6% (SEM%) for isometric hip adduction testing and 3% (SEM%) for isometric hip ABD. This indicates that a 3% hip ADD strength difference between the DOM and NDOM side would not be possible to measure at the individual level. A 3% difference in hip ADD strength between the DOM and NDOM side therefore does not seem to be of any clinical relevance. Hence, isometric hip ADD strength of the DOM and NDOM side should be considered equal for the individual. This has important clinical implications, as the strength of the contralateral side can be used as a relevant reference-point of full recovery of hip ADD strength.

A study from 1968, investigating isometric hip ADD and ABD torque in 80 healthy, non-soccer-playing subjects, of similar age and using a similar test procedure as in study III, showed a 7% greater hip ADD strength compared with hip ABD strength,[158] which is in accordance with the results of the present study. The results of study IV imply that the isometric hip ADD/ABD ratio is approximately 1.05 in soccer players, irrespective of which side is tested. Clinically, this finding has implications because it indicates that the ipsilateral hip ADD/ABD ratio can be used as a guideline during evaluation of recovery of hip adduction strength in soccer players with bilateral groin injury. The advantage of using the ipsilateral side as a guideline is that current or previous injury to the other limb will not interfere with this measurement procedure. This is relevant because groin pain in athletes often exists bilaterally,[32] In these cases, using the contralateral limb as a reference-point is not valid. By testing each limb individually, we believe that we achieve a greater depiction of the actual muscle strength in hip ADD in both the injured and uninjured limb compared with using a squeeze test, in which the produced pressure (force) will be a combination of the hip ADD strength of both legs.[45,159] In subjects with pain elicited during the squeeze test, the maximal force production will be determined mainly by the painful side in subjects with unilateral groin pain, and mainly by the most painful side in subjects with bilateral groin pain. Therefore, we think it is important to consider these aspects clinically when using a bilateral squeeze test or a unilateral strength tests for clinical evaluation of hip adductor muscle strength and recovery.

Groin pain affecting muscle strength in athletes

Groin pain is often aggravated during important soccer skills such as kicking, accelerating, and sudden change of direction.[24] A post-hoc analysis performed for the data of study IV revealed that the isometric hip ADD/ABD torque ratio was 24% lower in soccer players with groin pain (VAS, 3.9 ± 2.1) during hip adduction testing compared to players with a pain-free test. A decreased hip ADD/ABD ratio during maximal exertion might have implications for soccer players and their ability to perform at the maximum level.

It seems that groin pain during hip ADD testing negatively affects the ability to produce maximal hip ADD torque and thereby decreases the hip ADD/ABD ratio to approximately 80%. This finding is in accordance with the study by Malliaras et al,[45] who reported athletes with groin pain to have a reduced hip ADD pressure (force) in the squeeze test of approximately 20% compared with healthy controls. Unfortunately, the authors did not report whether pain was experienced during testing. Not recording possible pain during testing seems to be a general tendency in studies finding hip ADD torque reductions in athletes.[48,159-161] Previously, it has been shown that acutely reducing pain during strength testing in the shoulder region immediately increases the maximal force production.[137] Because pain may be a possible confounder when testing muscle strength, we suggest reporting pain during testing in future studies concerning athletes and hip strength measurements, and in addition that the relationship between reported pain and hip ADD strength is investigated further.

Methodological limitations of studies concerning hip strength measurements (study III and IV)

A limitation of study III is that we only investigated the intra-tester reliability, and we did not examine the inter-tester reliability. Inferior strength of the tester, compared to the subject being tested, is a possible factor affecting inter-tester reliability, when using HHD.[65,74,162,163] Therefore, the present study's results can only be extrapolated to the intra-tester situation. However, we preferred to use a long lever arm in the individual tests whenever possible, so that the tester's strength greatly exceeded the strength of the participant. Hip EXT testing using a short lever was the only test where measurement error was above 10%, which could very well be because tester strength in this measurement did not greatly exceed the isometric hip EXT force of the participant, and in general was difficult to perform. Another limitation of study III is that we

chose to perform isometric testing (make test) instead of eccentric testing (break test),[68] even though eccentric strength testing has shown greater strength values.[62] Isometric loading induces less stress to the musculoskeletal system than eccentric loading, which is relevant when testing individuals presenting with pathology.[62] Because our long-term goal was to develop a test suitable for both healthy individuals and individuals presenting with hip and/or groin pathology, we decided that a less stressful test was better suited for this purpose.

A limitation of study IV was the high proportion of soccer players reporting hip and/or groin pain during the previous year. Soccer players are commonly subjected to hip and/or groin pain, and about 70% of all players included in study IV reported hip and/or groin pain during the previous year, which is in accordance with previous reporting.[161] Hip and/or groin pain during the previous year may influence the force values in the population under study, however, because hip and/or groin pain is highly prevalent in soccer players at this level, we chose to include players if they were participating fully in training, even though they reported hip and/or groin pain during the previous year, because we wanted the test to be performed on the population for which it was intended, namely active elite soccer players. Therefore, excluding soccer players because of previous history of hip and/or groin pain would mean that our data could only be extrapolated to a very small and potentially irrelevant subgroup of players. Whether the data from study IV can be extrapolated to all elite soccer players or other athletes is unknown, but an advantage of using the contralateral or ipsilateral side for comparison instead of generalised normative values is that the wide range of individual muscle strength will not be a problem. However, it is unknown whether players, who are stronger or weaker than those in study IV display different strength symmetry indexes.

CONCLUSION

- Reliable, valid and responsive questionnaires for the evaluation of patient-reported outcome in physically active patients with hip and/or groin pain are lacking.

- A new patient-reported questionnaire (HAGOS) was developed. HAGOS has adequate measurement qualities for the assessment of symptoms, activity limitations and participation-restrictions in physically active patients with longstanding hip and/or groin pain. HAGOS is the first instrument validated for this group of patients and is recommended for use in interventions where the patient's perspective and health-related quality of life are of primary interest.

- The assessment of hip muscle strength using hand-held dynamometry is easy to administer clinically and produces a small measurement variation, making it possible to determine even small changes in hip strength.

- A marginal difference in hip adduction and abduction strength between the dominant and the non-dominant limb in elite soccer players exist, but it is within the measurement variation of the test procedure. Contralateral isometric hip adduction strength can therefore be used as a simple clinical reference-point of full recovery of hip adduction muscle strength in soccer players. Furthermore, it is suggested that the ipsilateral hip adduction/abduction strength ratio is used as a guideline for evaluating hip adduction strength recovery in soccer players with bilateral groin problems.

PERSPECTIVES

Outcome measures for young to middle-aged physically active patients with hip and groin pain have been lacking up to this point. However, this thesis shows, that the patient-reported outcome questionnaire HOS can be used for evaluating patients undergoing hip arthroscopy. Furthermore, a new patient-reported outcome questionnaire (HAGOS) can be recommended in future studies, evaluating young to middle-aged physically active individuals with longstanding hip and/or groin pain, assessing patient-reported symptoms, activity limitations and participation restrictions.

The thesis also shows that assessing hip muscle strength with hand-held dynamometry is a reliable and easy method that can be used for both research and clinical purposes. The supine position offers an advantage in the assessment of isometric hip adduction and abduction strength because it produces a small measurement variation and easily can be applied in individuals, who are either unable to, or who have great difficulties in producing sufficient force in the side-lying position to overcome gravity, due to either muscle weakness or pain. Clinically, this procedure therefore seems to be ideal for a broad range of patients, with varying muscle strength and deficits. The use of isometric hip adduction strength as a clinical reference-point of full recovery of hip adduction muscle strength in soccer players gives the clinician a possibility to assess hip muscle strength and detect specific hip strength deficits. This has relevance when planning and monitoring an individual rehabilitation program, increasing the possibility of bringing back players without muscle strength impairments. In the future this may be an important tool when trying to minimise the high risk of recurrence, which is a major problem for soccer players with hip and/or groin injuries.

I have in this thesis provided reliable and valid measures to be used in the assessment and evaluation of different subgroups of physically active individuals with hip- and groin-related problems, so that comparison of data and outcome from different studies is possible in the future. The possibility of using HAGOS and specific hip strength assessments will hopefully change the primary research approach within the area of young to middle-aged physically active patients with hip and groin pain, from descriptive case-series using non-validated outcome measurements to more experimental designs, including randomised controlled trials, using reliable, valid and responsive outcome measures

SUMMARY

Clinical outcome measures for physically active individuals with hip and groin pain: Development, evaluation and application

Background

Hip and groin pain is a common problem in the general population, and is often related to reduced physical function and activity level. Novel interventions are advancing rapidly in the management of hip and groin disability. However, for the evaluation of treatment outcome in physically active patients with hip and groin disability, reliable, valid and responsive measurement instruments are lacking.

Aim of the thesis

The overall aim of this thesis was to develop outcome measures for physically active patients with hip and groin pain. In the development process important measurement aspects such as reliability, validity and responsiveness were analysed. The purpose of the thesis was to develop a patient-reported outcome (PRO) questionnaire related to impairment, activities and participation in young physically active patients with hip and/or groin pain. Furthermore, a clinical measure of hip strength was developed and tested for test-retest reliability in young healthy subjects. Hip adduction and hip abduction strength was then tested in the dominant and non-dominant limb, to investigate whether strength symmetry between limbs can be assumed in soccer players, and whether the non-affected limb can be used as a reference-point for the evaluation of hip adduction recovery after injury.

Methods

The first study was a systematic review of the reliability, validity and responsiveness of available PRO questionnaire assessing patients with hip and/or groin disability (study I). In the second study a new patient reported outcome measure was developed including 101 patients with hip and/or groin pain in a prospective study (study II). The third study examined the absolute test–retest measurement variation concerning strength assessments of hip abduction (ABD), adduction (ADD), external rotation (ER), internal rotation (IR), flexion (FLEX) and extension (EXT), in 9 healthy

subjects, using a Hand-Held Dynamometer (HHD) (study III). The fourth study included 100 elite soccer players and maximal unilateral isometric hip adduction and abduction strength on the dominant and non-dominant side were measured with a hand-held dynamometer, using the newly developed and reliable test procedure (study IV).

Results

The systematic review showed that reliable, valid and responsive PRO questionnaires for young to middle aged physically active patients with hip and groin pain are lacking. A new PRO questionnaire named the Copenhagen Hip And Groin Outcome Score (HAGOS) was therefore developed. HAGOS consists of 6 separate subscales assessing pain (10 items), symptoms (7 items), function in daily living (5 items), function in sport and recreation (8 items), participation in physical activities (2 items) and hip and/or groin-related quality of life (5 items). Test-retest reliability was substantial, with an Intra-Class Coefficient (ICC) ranging from 0.82-0.91 for the 6 subscales. A priori set hypotheses concerning construct validity and responsiveness were confirmed, with correlation coefficients for construct validity ranging from 0.37-0.76, $p<0.01$, and for responsiveness ranging from 0.56-0.69, $p<0.01$.

Furthermore, this thesis shows that standardised strength assessment procedures of hip ABD, ER, IR and FLEX, with test–retest measurement variation below 5%, hip ADD below 6% and hip EXT below 8%, can be performed. In elite soccer players the dominant side was stronger than the non-dominant side for both isometric hip ADD (2.45±0.54 vs. 2.37±0.48 Nm/kg, $p=0.02$) and hip ABD (2.35±0.33 vs. 2.25±0.31 Nm/kg, $p<0.001$), corresponding to a 3% and 4% difference, respectively. Isometric hip ADD/ABD ratio was not different between the dominant (1.04±0.18) and non-dominant leg (1.06±0.17) ($p=0.40$). A post-hoc analysis showed that isometric hip ADD/ABD ratio was significantly lower in players with groin pain during hip ADD testing compared to players with a pain-free test (0.80±0.14) ($p<0.001$)

Conclusion

Due to the lack of reliable, valid and responsive questionnaires for the evaluation of patient-reported outcome in physically active patients with hip and/or groin pain a new patient-reported questionnaire was developed (HAGOS). HAGOS has adequate measurement qualities for the

assessment of symptoms, activity limitations and participation restrictions in physically active patients with longstanding hip and/or groin pain. HAGOS is the first instrument validated for this group of patients and is recommended for use in interventions where the patient's perspective and health-related quality of life are of primary interest.

Furthermore, the HDD is easy to administer and produces a small measurement variation, making it possible to determine even small changes in hip strength. The marginal strength difference between hip ADD in the dominant and the non-dominant limb is within the measurement variation of the test procedure, and contralateral isometric hip ADD strength can therefore be used as a simple clinical reference-point of full recovery of hip ADD muscle strength in soccer players. Furthermore, it is suggested that the ipsilateral hip ADD/ABD strength ratio is used as a guideline for evaluating hip ADD strength recovery in soccer players with bilateral groin problems.

SUMMARY IN DANISH / SAMMENFATNING PÅ DANSK

Kliniske effektmål for fysisk aktive individer med hofte- og lyskesmerter: Udvikling, evaluering og anvendelse

Baggrund

Hofte- og lyskesmerter er et stort problem i den almene befolkning, og er ofte relateret til reduceret fysisk funktions- og aktivitetsniveau. Nye behandlingsparadigmer udvikles hele tiden i forhold til behandlingen af hofte- og lyskeproblemer. Men valide, reproducerbare og sensitive kliniske måleredskaber til evaluering af behandlingen af fysisk aktive patienter med hofte- og lyskeproblemer mangler indenfor dette område.

Formålet med afhandlingen

Det overordnede formål med denne afhandling er at udvikle kliniske effektmål til fysisk aktive patienter med hofte- og lyskesmerter. I udviklingsprocessen er vigtige måleaspekter såsom reproducerbarhed, validitet og sensitivitet overfor forandringer over tid analyseret. Et patient-rapporteret spørgeskema (PRO) i forhold til funktions-, aktivitets- og deltagelsesniveauet hos fysisk aktive patienter med hofte- og/eller lyskesmerter vil blive udviklet og valideret. Desuden vil en klinisk måling af hoftestyrke blive udviklet og testet for test-retest reproducerbarhed hos yngre raske forsøgspersoner. Hofteabduktions- og hofteadduktionsstyrke blev testet i det dominante og det ikke-dominante ben, for at undersøge om egal hofteadduktions-styrke i benene kan forventes hos fodboldspillere, og om det ikke-afficerede ben kan anvendes som en valid referenceværdi i evalueringen af hofteadduktions-styrken efter en lyskeskade

Metoder

Den første undersøgelse er en systematisk litteraturgennemgang af tilgængelige PRO spørgeskemaer og deres af reproducerbarhed, validitet og sensitivitet overfor forandringer over tid, i forhold til at vurdere patienter med hofte- og/eller lyskeproblemer (studie I). I den anden undersøgelse var målet at udvikle et nyt patient-rapporteret spørgeskema til fysisk aktive patienter (n=101) med hofte- og/eller lyskesmerter i et prospektivt studie (studie II). I den tredje

undersøgelse (studie III) var formålet at undersøge den absolutte test/retest målevariation af hoftemuskelstyrken i de seks forskellige hoftebevægelser, henholdsvis hofte abduktion (ABD), adduktion (ADD), udad rotation (ER), indad rotation (IR), fleksion (FLEX) og ekstension (EXT) hos raske forsøgspersoner (n=9), ved hjælp af et håndholdt dynamometer (HHD). I det fjerde studie blev 101 elite fodboldspillere inkluderet og maksimal isometrisk hofte ADD og ABD styrke på den dominante og ikke-dominante side blev målt med et håndholdt dynamometer, ved at anvende den nyligt udviklede og pålidelige test-metode (studie IV).

Resultater

Den systematiske oversigtsartikel viser at der ikke findes reproducerbare, valide og sensitive patient-rapporterede spørgeskemaer til fysisk aktive patienter med hofte- og/eller lyskesmerter. Et nyt spørgeskema "Copenhagen Hip And Groin Outcome Score (HAGOS)" blev udviklet og består af 6 separate delskalaer der vurderer smerte (10 spørgsmål), symptomer (7 spørgsmål), funktion i dagligdagen (5 spørgsmål), funktion inden for sport og fritid (8 spørgsmål), deltagelse i fysiske aktiviteter (2 spørgsmål) og hofte- og/eller lyske-relateret livskvalitet (5 spørgsmål). Test-retest reproducerbarheden var god, med en ICC fra 0.82-0.91 for de seks subskalaer. A priori hypoteser i forhold til begrebs-validitet og sensitivitet for kliniske forandringer over tid blev bekræftet, med korrelationskoefficienter for validitet fra 0,37 til 0,76, $p<0,01$, og for sensitivitet fra 0,56 til 0,69, $p<0,01$.

Undersøgelsen viser, at test-retest variationen af standardiserede styrkeprocedurer for hofte ABD, ER, IR og FLEX, er under 5%, for hofte ADD under 6% og for hofte EXT under 8%. Hos elitefodboldspillere var det dominante ben stærkere end det ikke-dominante ben for både isometrisk hofte ADD (2.45±0.54 vs. 2.37±0.48 Nm/kg, $p=0.02$) og hofte ABD (2.35±0.33 vs. 2.25±0.31 Nm/kg, $p<0.001$), svarende til henholdsvis 3% og 4% forskel. Isometrisk hofte ADD/ABD ratio var ikke forskellig mellem det dominante (1.04±0.18) og det ikke-dominante ben (1.06±0.17) ($p=0.40$). En post-hoc analyse viste at isometrisk hofte ADD/ABD ratio var signifikant lavere hos spillere med lyskesmerter under hofte ADD testningen, sammenlignet med spillere med en smertefri test (0.80±0.14) ($p<0.001$).

Konklusion

På baggrund af manglende reproducerbare, valide og sensitive spørgeskemaer til at vurdere patient-rapporteret helbredstilstand hos fysisk aktive patienter med hofte- og/eller lyskesmerter udvikles et nyt patient-rapporteret spørgeskema (HAGOS). HAGOS har tilstrækkelige målekvaliteter til at kunne vurdere symptomer, aktivitetsbegrænsninger og deltagelsesrestriktioner hos fysisk aktive patienter med langvarige hofte og/eller lyskesmerter. HAGOS er det første måleinstrument som er valideret til denne gruppe af patienter og anbefales til brug i forbindelse med evaluering af behandlinger, hvor patientens perspektiv og sundhedsrelaterede livskvalitet har det primære fokus.

Det er desuden vist at HDD er nemt at anvende og udviser kun en lille målevariation, hvilket gør metoden ideel at anvende til både forsknings- og klinisk brug, da det muligt at bestemme selv små ændringer i hoftemuskelstyrke. Den marginale forskel mellem hofte ADD styrke på det dominante ben og det ikke-dominante ben hos fodboldspillere er indenfor målevariationen af testproceduren, og kontralateral isometrisk hofte ADD kan derfor anvendes som en simpel klinisk referenceværdi for normalisering af hofte ADD styrken hos fodboldspillere. Desuden, foreslår vi at den samsidige hofte ADD/ABD ratio bruges som udgangspunkt for en evaluering af hofte ADD styrkens normalisering hos fodboldspillere med bilaterale lyskeproblemer.

REFERENCES

1. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, De Vet HCW**:** The COSMIN checklist manual**.** [http://www.cosmin.nl].

2. World Health Organization Towards a Common Language for Functioning, Disability and Health, ICF. 2002. [http://www.who.int/classifications/icf/training/icfbeginnersguide.pdf].

3. Jansen JA, Mens JM, Backx FJ, Kolfschoten N, Stam HJ. Treatment of longstanding groin pain in athletes: a systematic review. *Scand J Med Sci Sports* 2008; **18**: 263-274.

4. 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 in Health* 2007; **10:** S125-S137.

5. U.S.FOOD AND DRUG ADMINISRATION. Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. *Draft guidance*. 2006.

6. Caspersen, CJ, Powell, KE, and Christensen, GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep* 1985; **100**: 2.

7. WHO. The World Health Organization (WHO). Physical Inactivity.The World Health Organization (WHO).[www.who.int/en/]

8. Streiner D.L, Norman G.R. *Health Measurement Scales. A Practical Guide to Their Development and Use*. 2003. New York: Oxford University Press.

9. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health and Quality of Life Outcomes* 2006; **4**.

10. Nielsen AB, Yde J. Epidemiology and Traumatology of Injuries in Soccer. *Am J Sports Med* 1989; **17:** 803-807.

11. Rutkow IM, Robbins AW. Demographic, Classificatory, and Socioeconomic Aspects of Hernia Repair in the United-States. *Surg Clin North Am* 1993; **73:** 413-426.

12. Van der Waal JM, Bot SDM, Terwee CB, van der Windt DAWM, Bouter LM, Dekker J. The course and prognosis of hip complaints in general practice. *Ann Behav Med* 2006; **31:** 297-308.

13. Bay-Nielsen M, Perkins FM, Kehlet H. Pain and functional impairment 1 year after inguinal herniorrhaphy: A nationwide questionnaire study. *Ann Surg* 2001; **233:** 1-7.

14. Fricker PA, Taunton JE, Ammann W. Osteitis Pubis in Athletes - Infection, Inflammation Or Injury. *Sports Med* 1991; **12:** 266-279.

15. Caudill P, Nyland J, Smith C, Yerasimides J, Lach J. Sports hernias: a systematic literature review. *Br J Sports Med* 2008; **42:** 954-964.

16. Choi H, McCartney M, Best TM. Treatment of Osteitis Pubis and Osteomyelitis of the Pubic Symphysis in Athletes: A Systematic Review. *Br J Sports Med* 2008; [Epub ahead of print].

17. Machotka Z, Kumar S, Perraton LG. A systematic review of the literature on the effectiveness of exercise therapy for groin pain in athletes. *Sports Med Arthrosc Rehabil Ther Technol* 2009; **1:** 5.

18. Robertson WJ, Kadrmas WR, Kelly BT. Arthroscopic management of labral tears in the hip: a systematic review of the literature. *Clin Orthop Relat Res* 2007; **455:** 88-92.

19. Swan KG, Wolcott M. The athletic hernia - A systematic review. *Clin Orthop* 2007; 78-87.

20. Muschaweck U, Berger L. Minimal Repair technique of sportsmen's groin: an innovative open-suture repair to treat chronic inguinal pain. *Hernia* 2010; **14:** 27-33.

21. Schilders E, Bismil Q, Robinson P, O'Connor PJ, Gibbon WW, Talbot JC. Adductor-related groin pain in competitive athletes. Role of adductor enthesis, magnetic resonance imaging, and entheseal pubic cleft injections. *J Bone Joint Surg Am* 2007; **89:** 2173-2178.

22. Standaert CJ, Manner PA, Herring SA. Expert opinion and controversies in musculoskeletal and sports medicine: femoroacetabular impingement. *Arch Phys Med Rehabil* 2008; **89:** 890-893.

23. Picavet HSJ, Schouten JSAG. Musculoskeletal pain in the Netherlands: prevalences, consequences and risk groups, the DMC3-study. *Pain* 2003; **102:** 167-178.

24. Lynch SA, Renstrom PAFH. Groin injuries in sport - Treatment strategies. *Sports Med* 1999; **28:** 137-144.

25. Ekstrand J, Hilding J. The incidence and differential diagnosis of acute groin injuries in male soccer players. *Scand J Med Sci Sports* 1999; **9:** 98-103.

26. Pettersson M, Lorentzon R. Ice hockey injuries: a 4-year prospective study of a Swedish elite ice hockey team. *Br J Sports Med* 1993; **27:** 251-254.

27. Beck, R. Fodbold er langt vigtigere end vi tror. *Politiken* 2009.

28. Thorborg, K, Serner, A., Petersen, J, Moller Madsen, T, Magnussonn, P, and Hölmich, P. Hip adduction and adduction strength profiles in elite soccer players; Implications for clinical evaluation of hip adductor muscle recovery after injury. *Am J Sports Med* 2010; [Epub ahead of print].

29. Thorninger, R. Lyskesmerter. *Dansk Sportsmedicin* 2004; **2**: 22-26.

30. Birrell F, Lunt M, Macfarlane GJ, Silman AJ. Defining hip pain for population studies. *Ann Rheum Dis* 2005; **64:** 95-98.

31. Bradshaw CJ, Bundy M, Falvey E. The diagnosis of long-standing groin pain - a prospective clinical cohort study. *Br J Sports Med* 2008; **42**: 851-854.

32. Holmich P. Long-standing groin pain in sportspeople falls into three primary patterns, a "clinical entity" approach: a prospective study of 207 patients. *Br J Sports Med* 2007; **41:** 247-252.

33. Lesher JM, Dreyfuss P, Hager N, Kaplan M, Furman M. Hip joint pain referral patterns: A descriptive study. *Pain Medicine* 2008; **9:** 22-25.

34. Philippon MJ, Maxwell RB, Johnston TL, Schenker M, Briggs KK. Clinical presentation of femoroacetabular impingement. *Knee Surgery Sports Traumatology Arthroscopy* 2007; **15:** 1041-1047.

35. Jansen JA, Mens JM, Backx FJ, Stam HJ. Diagnostics in athletes with long-standing groin pain. *Scand J Med Sci Sports* 2008; **18:** 679-690.

36. Leibold MR, Huijbregts PA, Jensen R. Concurrent criterion-related validity of physical examination tests for hip labral lesions: a systematic review. *J Man Manip Ther* 2008; **16:** E24-E41.

37. Martin RL, Enseki KR, Draovitch P, Trapuzzano T, Philippon MJ. Acetabular labral tears of the hip: examination and diagnostic challenges. *J Orthop Sports Phys Ther* 2006; **36:** 503-515.

38. Martin RL, Irrgang JJ, Sekiya JK. The diagnostic accuracy of a clinical examination in determining intra-articular hip pain for potential hip arthroscopy candidates. *Arthroscopy* 2008; **24:** 1013-1018.

39. Ekberg O, Persson NH, Abrahamsson PA, Westlin NE, Lilja B. Longstanding Groin Pain in Athletes - A Multidisciplinary Approach. *Sports Med* 1988; **6:** 56-61.

40. Olivo SA, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: A systematic review. *Phys Ther* 2008; **88:** 156-175.

41. McArdle WD, Katch FI, Katch VL. *Exercise physiology*. 2007. Baltimore: Lippincot Williams & Wilkins.

42. Holmich P, Holmich LR, Bjerg AM. Clinical examination of athletes with groin pain: an intraobserver and interobserver reliability study. *Br J Sports Med* 2004; **38:** 446-451.

43. Arokoski MH, Arokoski JPA, Haara M, Kankaanpaa M, Vesterinen M, Niemitukia LH, Helminen HJ. Hip muscle strength and muscle cross sectional area in men with and without hip osteoarthritis. *J Rheumatol* 2002; **29:** 2185-2195.

44. Cetin C, Sekir U, Yildiz Y, Aydin T, Ors F, Kalyon TA. Chronic groin pain in an amateur soccer player. *Br J Sports Med* 2004; **38:** 223-224.

45. Malliaras P, Hogan A, Nawrocki A, Crossley K, Schache A. Hip flexibility and strength measures: reliability and association with athletic groin pain. *Br J Sports Med* 2009; **43:** 739-744.

46. Holmich P, Uhrskou P, Ulnits L, Kanstrup IL, Nielsen MB, Bjerg AM, Krogsgaard K. Effectiveness of active physical training as treatment for long-standing adductor-related groin pain in athletes: randomised trial. *Lancet* 1999; **353:** 439-443.

47. O'Connor DM. Groin injuries in professional rugby league players: a prospective study. *J Sports Sci* 2004; **22:** 629-636.

48. Tyler TF, Nicholas SJ, Campbell RJ, Mchugh MP. The association of hip strength and flexibility with the incidence of adductor muscle strains in professional ice hockey players. *Am J Sports Med* 2001; **29:** 124-128.

49. Overgaard S, Fredberg U. [The clinical examination of hip joints in adults]. *Ugeskr Laeger* 2004; **166:** 2544-2547.

50. Kendall FP, McCreary EK, Provance PG. *Muscles testing and function*. 1993. Baltimore, Williams & Wilkins.

51. Wadsworth CT, Krishnan R, Sear M, Harrold J, Nielsen DH. Intrarater reliability of manual muscle testing and hand-held dynametric muscle testing. *Phys Ther* 1987; **67:** 1342-1347.

52. Mahony K, Hunt A, Daley D, Sims S, Adams R. Inter-tester reliability and precision of manual muscle testing and hand-held dynamometry in lower limb muscles of children with spina bifida. *Phys Occup Ther Pediatr* 2009; **29:** 44-59.

53. Beasley WC. Influence of method on estimates of normal knee extensor force among normal and postpolio children. *Phys Ther Rev* 1956; **36:** 21-41.

54. Newman LB. A new device for measuring muscle strength; the myometer. *Arch Phys Med Rehabil* 1949; **30:** 234-237.

55. Cichanowski HR, Schmitt JS, Johnson RJ, Niemuth PE. Hip strength in collegiate female athletes with patellofemoral pain. *Med Sci Sports Exerc* 2007; **39:** 1227-1232.

56. Ireland ML, Willson JD, Ballantyne BT, Davis IM. Hip strength in females with and without patellofemoral pain. *Journal of Orthopaedic & Sports Physical Therapy* 2003; **33:** 671-676.

57. Niemuth PE, Johnson RJ, Myers MJ, Thieman TJ. Hip muscle weakness and overuse injuries in recreational runners. *Clin J Sport Med* 2005; **15:** 14-21.

58. Bohannon RW. Hand-held compared with isokinetic dynamometry for measurement of static knee extension torque (parallel reliability of dynamometers). *Clin Phys Physiol Meas* 1990; **11:** 217-222.

59. Stratford PW, Balsor BE. A comparison of make and break tests using a hand-held dynamometer and the Kin-Com. *JOSPT* 1994; **19:** 28-32.

60. Bohannon RW, Andrews AW. Accuracy of Spring and Strain Gauge Hand-held Dynamometers. *JOSPT* 1989; **10:** 323-325.

61. Bohannon RW. Test-Retest Reliability of Hand-Held Dynamometry During A Single Session of Strength Assessment. *Phys Ther* 1986; **66:** 206-209.

62. Bohannon RW. Make tests and break tests of elbow flexor muscle strength. *Phys Ther* 1988; **68:** 193-194.

63. Magnusson SP. Clinical strength testing. *Rehab Manag* 1993; **6:** 38-45.

64. Wadsworth C, Nielsen DH, Corcoran DS, Phillips CE, Sannes TL. Interrater reliability of hand-held dynamometry: effects of rater gender, body weight, and grip strength. *JOSPT* 1992; **16:** 74-81.

65. Wikholm JB, Bohannon RW. Hand-held Dynamometer Measurements: Tester Strength Makes a Difference. *JOSPT* 1991; **13:** 191-198.

66. Krause DA, Schlagel SJ, Stember BM, Zoetewey JE, Hollman JH. Influence of lever arm and stabilization on measures of hip abduction and adduction torque obtained by hand-held dynamometry. *Arch Phys Med Rehabil* 2007; **88:** 37-42.

67. Thorborg K, Petersen J, Magnusson SP, Holmich P. Clinical assessment of hip strength using a hand-held dynamometer is reliable. *Scand J Med Sci Sports* 2010; **20:** 493-501.

68. Sisto SA, Dyson-Hudson T. Dynamometry testing in spinal cord injury. *J Rehabil Res Dev* 2007; **44:** 123-136.

69. Burns SP, Spanier DE. Break-technique handheld dynamometry: relation between angular velocity and strength measurements. *Arch Phys Med Rehabil* 2005; **86:** 1420-1426.

70. Thorborg K, Couppe C, Petersen J, Magnusson P, Holmich P. Eccentric Hip Adduction and Abduction Strength in Elite Soccer Players and Matched Controls A Cross-Sectional Study. *Br J Sports Med* 2009; [Epub ahead of print].

71. Friden J, Sjostrom M, Ekblom B. Myofibrillar damage following intense eccentric exercise in man. *Int J Sports Med* 1983; **4:** 170-176.

72. Friden J, Sfakianos PN, Hargens AR. Muscle soreness and intramuscular fluid pressure: comparison between eccentric and concentric load. *J Appl Physiol* 1986; **61:** 2175-2179.

73. Pua YH, Wrigley TW, Cowan SM, Bennell KL. Intrarater test-retest reliability of hip range of motion and hip muscle strength measurements in persons with hip osteoarthritis. *Arch Phys Med Rehabil* 2008; **89:** 1146-1154.

74. Kelln BM, Mckeon PO, Gontkof LM, Hertel J. Hand-held dynamometry: Reliability of lower extremity muscle testing in healthy, physically active, young adults. *Journal of Sport Rehabilitation* 2008; **17:** 160-170.

75. Kreighbaum E, Barthels KM. *Biomechanics - A Qualitative Approach for Studying Human Movement.* 1996. Needham Heights: Allyn & Bacon.

76. Augustsson J, Thomee R, Karlsson S. Ability of a new hop test to determine functional deficits after anterior cruciate ligament reconstruction. *Knee Surgery Sports Traumatology Arthroscopy* 2004; **12:** 350-356.

77. Orchard J, Best TM, Verrall GM. Return to play following muscle strains. *Clin J Sport Med* 2005; **15:** 436-441.

78. Nicholas SJ, Tyler TF. Adductor muscle strains in sport. *Sports Med* 2002; **32:** 339-344.

79. Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. *BMJ* 2010; **340:** c186.

80. Speight J, Barendse SM. FDA guidance on patient reported outcomes. *BMJ* 2010; **340:** c2921.

81. Timmins N. NHS goes to the PROMS. *Br Med J* 2008; **336:** 1464-1465.

82. Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. *Br J Psychiatry* 2000; **176:** 249-252.

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

84. Kirkley A, Alvarez C, Griffin S. The development,and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: The Western Ontario Rotator Cuff Index. *Clin J Sport Med* 2003; **13:** 84-92.

85. Irrgang JJ, Snyder-Mackler L, Wainner RS, Fu FH, Harner CD. Development of a patient-reported measure of function of the knee. *J Bone Joint Surg Am* 1998; **80:** 1132-1145.

86. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee injury and osteoarthritis outcome score (KOOS) - Development of a self-administered outcome measure. *JOSPT* 1998; **28:** 88-96.

87. Roos EM, Roos HP, Ekdahl C, Lohmander LS. Knee injury and osteoarthritis outcome score (KOOS) - validation of a Swedish version. *Scand J Med Sci Sports* 1998; **8:** 439-448.

88. Visentini PJ, Khan KM, Cook JL, Kiss ZS, Harcourt PR, Wark JD. The VISA score: an index of severity of symptoms in patients with jumper's knee (patellar tendinosis). Victorian Institute of Sport Tendon Study Group. *J Sci Med Sport* 1998; **1:** 22-28.

89. Martin RL, Irrgang JJ, Burdett RG, Conti SF, Van Swearingen JM. Evidence of validity for the Foot and Ankle Ability Measure (FAAM). *Foot Ankle Int* 2005; **26:** 968-983.

90. Robinson JM, Cook JL, Purdam C, Visentini PJ, Ross J, Maffulli N, Taunton JE, Khan KM. The VISA-A questionnaire: a valid and reliable index of the clinical severity of Achilles tendinopathy. *Br J Sports Med* 2001; **35:** 335-341.

91. Roos EM, Brandsson S, Karlsson J. Validation of the foot and ankle outcome score for ankle ligament reconstruction. *Foot Ankle Int* 2001; **22:** 788-794.

92. 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:** 331-342.

93. Kongsgaard M, Kovanen V, Aagaard P, Doessing S, Hansen P, Laursen AH, Kaldau NC, Kjaer M, Magnusson SP. Corticosteroid injections, eccentric decline squat training and heavy slow resistance training in patellar tendinopathy. *Scand J Med Sci Sports* 2009; **19:** 790-802.

94. Roos EM, Engstrom M, Lagerquist A, Soderberg B. Clinical improvement after 6 weeks of eccentric exercise in patients with mid-portion Achilles tendinopathy -- a randomized trial with 1-year follow-up. *Scand J Med Sci Sports* 2004; **14:** 286-295.

95. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010; **63:** 737-745.

96. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HC. 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.

97. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 1999; **354:** 1896-1900.

98. Van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. *Spine* 2003; **28:** 1290-1299.

99. Terwee CB, Bot SDM, De Boer MR, van der Windt DAWM, Knol DL, Dekker J, Bouter LA, De Vet HCW. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; **60:** 34-42.

100. Bot SDM, Terwee CB, van der Windt DAWM, Bouter LM, Dekker J, De Vet HCW. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Ann Rheum Dis* 2004; **63:** 335-341.

101. De Boer MR, Moll AC, De Vet HCW, Terwee CB, Volker-Dieben HJM, Van Rens GHMB. Psychometric properties of vision-related quality of life questionnaires: a systematic review. *Ophthalmic Physiol Opt* 2004; **24:** 257-273.

102. Eechaute C, Vaes P, Van AL, Asman S, Duquet W. The clinimetric qualities of patient-assessed instruments for measuring chronic ankle instability: a systematic review. *BMC Musculoskelet Disord* 2007; **8:** 6.

103. Altman DG. *Practical statistics for medical research.* 1991 London: Chapman and Hall.

104. Wright JG, Young NL. The patient-specific index: asking patients what they want. *J Bone Joint Surg Am* 1997; **79:** 974-983.

105. Wright JG, Young NL. A comparison of different indices of responsiveness. *J Clin Epidemiol* 1997; **50:** 239-246.

106. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br* 1996; **78:** 185-190.

107. Dawson J, Fitzpatrick R, Murray D, Carr A. Comparison of measures to assess outcomes in total hip replacement surgery. *Qual Health Care* 1996; **5:** 81-88.

108. Fitzpatrick R, Dawson J. Health-related quality of life and the assessment of outcomes of total hip replacement surgery. *Psychology & Health* 1997; **12:** 793-803.

109. Stucki G, Meier D, Stucki S, Michel BA, Tyndall AG, Elke R, Theiler R. [Evaluation of a German questionnaire version of the Lequesne cox- and gonarthrosis indices]. *Z Rheumatol* 1996; **55:** 50-57.

110. Stucki G, Meier D, Stucki S, Michel BA, Tyndall AG, Dick W, Theiler R. [Evaluation of a German version of WOMAC (Western Ontario and McMaster Universities) Arthrosis Index]. *Z Rheumatol* 1996; **55:** 40-49.

111. Stucki G, Sangha O, Stucki S, Michel BA, Tyndall A, Dick W, Theiler R. Comparison of the WOMAC (Western Ontario and McMaster Universities) osteoarthritis index and a self-report format of the self-administered Lequesne-Algofunctional index in patients with knee and hip osteoarthritis. *Osteoarthritis Cartilage* 1998; **6:** 79-86.

112. Falvey EC, Franklyn-Miller A, McCrory P. The groin triangle: a patho-anatomic approach to the diagnosis of chronic groin pain in athletes. *Br J Sports Med* 2008.

113. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. *Arthritis Rheum* 2003; **49:** 156-163.

114. Benjamin S, Morris S, McBeth J, Macfarlane GJ, Silman AJ. The association between chronic widespread pain and mental disorder: a population-based study. *Arthritis Rheum* 2000; **43:** 561-567.

115. Thorborg K, Roos E, Bartels EM, Petersen J, Holmich P. Validity, Reliability And Responsiveness Of Patient-Reported Outcome Questionnaires When Assessing Hip And Groin Disability: A Systematic Review. *Br J Sports Med* 2009; [Epub ahead of print].

116. Nilsdotter AK, Lohmander LS, Klassbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS) - validity and responsiveness in total hip replacement. *Bmc Musculoskeletal Disorders* 2003; **4**.

117. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000; **25:** 3186-3191.

118. Guillemin F, Bombardier C, Beaton D. Cross-Cultural Adaptation of Health-Related Quality-Of-Life Measures - Literature-Review and Proposed Guidelines. *J Clin Epidemiol* 1993; **46:** 1417-1432.

119. Martin RL, Kelly BT, Philippon MJ. Evidence of validity for the hip outcome score. *Arthroscopy* 2006; **22:** 1304-1311.

120. Martin RL, Philippon MJ. Evidence of validity for the hip outcome score in hip arthroscopy. *Arthroscopy* 2007; **23:** 822-826.

121. Martin RL, Philippon MJ. Evidence of reliability and responsiveness for the hip outcome score. *Arthroscopy* 2008; **24:** 676-682.

122. Lohr KN, Aaronson NK, Alonso J, Burnam MA, Patrick DL, Perrin EB, Roberts JS. Evaluating quality-of-life and health status instruments: Development of scientific review criteria. *Clin Ther* 1996; **18:** 979-992.

123. Terwee CB, Roorda LD, Knol DL, De Boer MR, De Vet HCW. Linking measurement error to minimal important change of patient-reported outcomes. *J Clin Epidemiol* 2009; **62:** 1062-1067.

124. Ashby E, Grocott MP, Haddad FS. Outcome measures for orthopaedic interventions on the hip. *J Bone Joint Surg Br* 2008; **90:** 545-549.

125. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *J Clin Epidemiol* 1997; **50:** 79-93.

126. Patel AA, Donegan D, Albert T. The 36-item short form. *J Am Acad Orthop Surg* 2007; **15:** 126-134.

127. Bjorner JB, Damsgaard MT, Watt T, Groenvold M. Tests of data quality, scaling assumptions, and reliability of the Danish SF-36. *J Clin Epidemiol* 1998; **51:** 1001-1011.

128. Bjorner JB, Thunedborg K, Kristensen TS, Modvig J, Bech P. The Danish SF-36 Health Survey: translation and preliminary validity studies. *J Clin Epidemiol* 1998; **51:** 991-999.

129. 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; **51:** 1189-1202.

130. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, De Vet HCW. 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:** 539-549.

131. McHorney CA, Tarlov AR. Individual-Patient Monitoring in Clinical-Practice - Are Available Health-Status Surveys Adequate. *Qual Life Res* 1995; **4:** 293-307.

132. De Vet HCW, Ader HJ, Terwee CB, Pouwer F. Are factor analytical techniques used appropriately in the validation of health status questionnaires? A systematic review on the quality of factor analysis of the SF-36 - Response. *Qual Life Res* 2005; **14:** 1223-1224.

133. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research* 2005; **19:** 231-240.

134. Busija L, Osborne RH, Nilsdotter A, Buchbinder R, Roos EM. Magnitude and meaningfulness of change in SF-36 scores in four types of orthopedic surgery. *Health Qual Life Outcomes* 2008; **6:** 55.

135. de Vet HC, Bouter LM, Bezemer PD, Beurskens AJ. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. *Int J Technol Assess Health Care* 2001; **17:** 479-487.

136. Gabbe BJ, Finch CF, Bennell KL, Wajswelner H. How valid is a self reported 12 month sports injury history? *Br J Sports Med* 2003; **37:** 545-547.

137. Ben-Yishay A, Zuckerman JD, Gallagher M, Cuomo F. Pain inhibition of shoulder strength in patients with impingement syndrome. *Orthopedics* 1994; **17:** 685-688.

138. Wade DT, Halligan P. New wine in old bottles: the WHO ICF as an explanatory model of human behaviour. *Clin Rehabil* 2003; **17:** 349-354.

139. Jolles BM, Buchbinder R, Beaton DE. A study compared nine patient-specific indices for musculoskeletal disorders. *J Clin Epidemiol* 2005; **58:** 791-801.

140. Riddle DL, Stratford PW, Bowman DH. Findings of extensive variation in the types of outcome measures used in hip and knee replacement clinical trials: A systematic review. *Arthritis & Rheumatism-Arthritis Care & Research* 2008; **59:** 876-883.

141. Stavem K, Arnesen O. Use of hip and knee clinical scoring systems in prosthesis surgery in Norwegian hospitals. *Int Orthop* 2005; **29:** 301-304.

142. Bardakos NV, Vasconcelos JC, Villar RN. Early outcome of hip arthroscopy for femoroacetabular impingement: the role of femoral osteoplasty in symptomatic improvement. *J Bone Joint Surg Br* 2008; **90:** 1570-1575.

143. Byrd JW, Jones KS. Arthroscopic femoroplasty in the management of cam-type femoroacetabular impingement. *Clin Orthop Relat Res* 2009; **467:** 739-746.

144. Philippon MJ, Briggs KK, Yen YM, Kuppersmith DA. Outcomes following hip arthroscopy for femoroacetabular impingement with associated chondrolabral dysfunction: minimum two-year follow-up. *J Bone Joint Surg Br* 2009; **91:** 16-23.

145. Martin RL, Mohtadi NG, Safran MR, Leunig M, Martin HD, McCarthy J, Guanche CA, Kelly BT, Byrd JW, Clohisy JC, Philippon MJ, Sekiya JK. Differences in physician and patient ratings of items used to assess hip disorders. *Am J Sports Med* 2009; **37:** 1508-1512.

146. Verrall GM, Slavotinek JP, Barnes PG, Fon GT. Description of pain provocation tests used for the diagnosis of sports-related chronic groin pain: relationship of tests to defined clinical (pain and tenderness) and MRI (pubic bone marrow oedema) criteria. *Scand J Med Sci Sports* 2005; **15:** 36-42.

147. De Boer MR, De Vet HCW, Terwee CB, Moll AC, Volker-Dieben HJM, Van Rens GHMB. Changes to the subscales of two vision-related quality of life questionnaires are proposed. *J Clin Epidemiol* 2005; **58:** 1260-1268.

148. Quintana JM, Escobar A, Bilbao A, Arostegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after hip joint replacement. *Osteoarthritis Cartilage* 2005; **13:** 1076-1083.

149. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, Croft P, De Vet HCW. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol* 2010; **63:** 524-534.

150. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003; **41:** 582-592.

151. Davis AM, Perruccio AV, Canizares M, Tennant A, Hawker GA, Conaghan PG, Roos EM, Jordan JM, Maillefert JF, Dougados M, Lohmander LS. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. *Osteoarthritis Cartilage* 2008; **16:** 551-559.

152. Bombardier C, Tugwell P. Methodological Considerations in Functional Assessment. *J Rheumatol* 1987; **14:** 6-10.

153. 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:** 336-345.

154. Click FP, Bellew JW, Pitts TA, Kay RE. Reliability of stabilised commercial dynamometers for measuring hip abduction strength: a pilot study. *Br J Sports Med* 2003; **37:** 331-334.

155. Scott DA, Bond EQ, Sisto SA, Nadler SF. The intra- and interrater reliability of hip muscle strength assessments using a handheld versus a portable dynamometer anchoring station. *Arch Phys Med Rehabil* 2004; **85:** 598-603.

156. Brophy RH, Backus SI, Pansy BS, Lyman S, Williams RJ. Lower extremity muscle activation and alignment during the soccer instep and side-foot kicks. *JOSPT* 2007; **37:** 260-268.

157. Charnock BL, Lewis CL, Garrett WE, Queen RM. Adductor longus mechanics during the maximal effort soccer kick. *Sports Biomechanics* 2009; **8:** 223-234.

158. Murray MP, Sepic SB. Maximum isometric torque of hip abductor and adductor muscles. *Phys Ther* 1968; **48:** 1327-1335.

159. Mens J, Inklaar H, Koes BW, Stam HJ. A new view on adduction-related groin pain. *Clin J Sport Med* 2006; **16:** 15-19.

160. Crow JF, Pearce AJ, Veale JP, Vanderwesthuizen D, Coburn PT, Pizzari T. Hip adductor muscle strength is reduced preceding and during the onset of groin pain in elite junior Australian football players. *J Sci Med Sport* 2010; **13**: 202-204.

161. Hanna CM, Fulcher ML, Elley CR, Moyes SA. Normative values of hip strength in adult male association football players assessed by handheld dynamometry. *J Sci Med Sport* 2010; **13** : 299-303.

162. Agre JC, Magness JL, Hull SZ, Wright KC, Baxter TL, Patterson R, Stradel L. Strength testing with a portable dynamometer: reliability for upper and lower extremities. *Arch Phys Med Rehabil* 1987; **68:** 454-458.

163. Lu TW, Hsu HC, Chang LY, Chen HL. Enhancing the examiner's resisting force improves the reliability of manual muscle strength measurements: comparison of a new device with hand-held dynamometry. *J Rehabil Med* 2007; **39:** 679-684.

164. Kirmit L, Karatosun V, Unver B, Bakirhan S, Sen A, Gocen Z. The reliability of hip scoring systems for total hip arthroplasty candidates: assessment by physical therapists. *Clin Rehabil* 2005; **19:** 659-661.

165. Klassbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score - An extension of the Western Ontario and McMaster Universities Osteoarthritis Index. *Scand J Rheumatol* 2003; **32:** 46-51.

166. de Groot I, Reijman M, Terwee CB, Bierma-Zeinstra SM, Favejee M, Roos EM, Verhaar JA. Validation of the Dutch version of the Hip disability and Osteoarthritis Outcome Score. *Osteoarthritis Cartilage* 2007; **15:** 104-109.

167. Johanson NA, Charlson ME, Szatrowski TP, Ranawat CS. A self-administered hip-rating questionnaire for the assessment of outcome after total hip replacement. *J Bone Joint Surg Am* 1992; **74:** 587-597.

168. Lequesne MG, Mery C, Samson M, Gerard P. Indexes of severity for osteoarthritis of the hip and knee. Validation--value in comparison with other assessment tests. *Scand J Rheumatol Suppl* 1987; **65:** 85-89.

169. Ludwig FJ, Melzer C, Grimmig H, Daalmann HH. [Cross cultural adaptation of the lequesne algofunctional indices for german speaking patients with osteoarthritis of the hip and the knee]. *Rehabilitation (Stuttg)* 2002; **41:** 249-257.

170. Dawson J, Linsell L, Doll H, Zondervan K, Rose P, Carr A, Randall T, Fitzpatrick R. Assessment of the Lequesne index of severity for osteoarthritis of the hip in an elderly population. *Osteoarthritis Cartilage* 2005; **13:** 854-860.

171. Potter BK, Freedman BA, Andersen RC, Bojescul JA, Kuklo TR, Murphy KP. Correlation of short form-36 and disability status with outcomes of arthroscopic acetabular labral debridement. *Am J Sports Med* 2005; **33:** 864-870.

172. Christensen CP, Althausen PL, Mittleman MA, Lee JA, McCarthy JC. The nonarthritic hip score: reliable and validated. *Clin Orthop Relat Res* 2003; 75-83.

173. Fitzpatrick R, Morris R, Hajat S, Reeves B, Murray DW, Hannen D, Rigge M, Williams O, Gregg P. The value of short and simple measures to assess outcomes for patients of total hip replacement surgery. *Qual Health Care* 2000; **9:** 146-150.

174. Dawson J, Fitzpatrick R, Frost S, Gundle R, Lardy-Smith P, Murray D. Evidence for the validity of a patient-based instrument for assessment of outcome after revision hip replacement. *J Bone Joint Surg Br* 2001; **83:** 1125-1129.

175. Ostendorf M, van Stel HF, Buskens E, Schrijvers AJP, Marting LN, Verbout AJ, Dhert WJA. Patient-reported outcome in total hip replacement - A comparison of five instruments of health status. *J Bone and Joint Surg Br* 2004; **86B:** 801-808.

176. Gosens T, Hoefnagels NHM, De Vet RCW, Dhert WJA, van Langelaan EJ, Bulstra SK, Geesink RGT. The "Oxford Heup Score" - The translation and validation of a questionnaire into Dutch to evaluate the results of total hip arthroplasty. *Acta Orthopaedica* 2005; **76:** 204-211.

177. Garbuz DS, Xu M, Sayre EC. Patients' outcome after total hip arthroplasty: a comparison between the Western Ontario and McMaster Universities index and the Oxford 12-item hip score. *J Arthroplasty* 2006; **21:** 998-1004.

178. Naal FD, Sieverding M, Impellizzeri FM, von KF, Mannion AF, Leunig M. Reliability and Validity of the Cross-Culturally Adapted German Oxford Hip Score. *Clin Orthop Relat Res* 2008.

179. Wright JG, Rudicel S, Feinstein AR. Ask Patients What They Want - Evaluation of Individual Complaints Before Total Hip-Replacement. *J Bone Joint Surg Br* 1994; **76B:** 229-234.

180. Wright JG, Young NL, Waddell JP. The reliability and validity of the self-reported patient-specific index for total hip arthroplasty. *J Bone Joint Surg Am* 2000; **82:** 829-837.

181. Whitehouse SL, Lingard EA, Katz JN, Learmonth ID. Development and testing of a reduced WOMAC function scale. *J Bone Joint Surg Br* 2003; **85:** 706-711.

182. Liang MH, Katz JN, Phillips C, Sledge C, Catsbaril W. The Total Hip-Arthroplasty Outcome Evaluation Form of the American-Academy-Of-Orthopedic-Surgeons - Results of A Nominal Group-Process. *J Bone Joint Surg Am* 1991; **73A:** 639-646.

183. Katz JN, Phillips CB, Poss R, Harrast JJ, Fossel AH, Liang MH, Sledge CB. The validity and reliability of a Total Hip Arthroplasty Outcome Evaluation Questionnaire. *J Bone Joint Surg Am* 1995; **77:** 1528-1534.

184. Soderman P, Malchau H. Validity and reliability of Swedish WOMAC osteoarthritis index: a self-administered disease-specific questionnaire (WOMAC) versus generic instruments (SF-36 and NHP). *Acta Orthop Scand* 2000; **71:** 39-46.

185. Angst F, Aeschlimann A, Steiner W, Stucki G. Responsiveness of the WOMAC osteoarthritis index as compared with the SF-36 in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention. *Ann Rheum Dis* 2001; **60:** 834-840.

186. Nilsdotter AK, Roos EM, Westerlund JP, Roos HP, Lohmander LS. Comparative responsiveness of measures of pain and function after total hip replacement. *Arthritis & Rheumatism-Arthritis Care & Research* 2001; **45:** 258-262.

187. Escobar A, Quintana JM, Bilbao A, Azkarate J, Guenaga JI. Validation of the Spanish version of the WOMAC questionnaire for patients with hip or knee osteoarthritis. Western Ontario and McMaster Universities Osteoarthritis Index. *Clin Rheumatol* 2002; **21:** 466-471.

188. Roorda LD, Jones CA, Waltz M, Lankhorst GJ, Bouter LM, van der Eijken JW, Willems WJ, Heyligers IC, Voaklander DC, Kelly KD, Suarez-Almazor ME. Satisfactory cross cultural equivalence of the Dutch WOMAC in patients with hip osteoarthritis waiting for arthroplasty. *Ann Rheum Dis* 2004; **63:** 36-42.

189. Soohoo NF, Vyas RM, Samimi DB, Molina R, Lieberman JR. Comparison of the responsiveness of the SF-36 and WOMAC in patients undergoing total hip arthroplasty. *J Arthroplasty* 2007; **22:** 1168-1173.

190. Soininen JV, Paavolainen PO, Gronblad MA, Kaapa EH. Validation study of a Finnish version of the Western Ontario and McMasters University osteoarthritis index. *Hip Int* 2008; **18:** 108-111.

191. Franneby U, Gunnarsson U, Andersson M, Heuman R, Nordin P, Nyren O, Sandblom G. Validation of an Inguinal Pain Questionnaire for assessment of chronic pain after groin hernia repair. *Br J Surg* 2008; **95**: 488-93.