From THE DEPARTMENT OF NEUROBIOLOGY, CARE
SCIENCES AND SOCIETY, DIVISION OF PHYSIOTHERAPY
Karolinska Institutet, Stockholm, Sweden

THE SINGLE LEG SQUAT IN CLINICAL TESTING.
ASPECTS OF RELIABILITY, VALIDITY, AND ASSOCIATED FACTORS

John Ressman

Stockholm 2024

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2023

ISBN 978-91-8017-227-1
Cover illustration: by formligen.se

THE SINLGE LEG SQUAT IN CLINICAL TESTING. ASPECTS OF RELIABILITY, VALIDITY, AND ASSOCIATED FACTORS

Thesis for Doctoral Degree (Ph.D.)

By

John Ressman

The thesis will be defended in public at Karolinska Institutet, lecture hall H2, Alfred Nobels allé 23, Huddinge.

Friday, January 26, 2024, at 09:00.

Principal Supervisor:
Associate professor, Eva Rasmussen-Barr
Karolinska Institutet
Department of Neurobiology,
Care Sciences and Society
Division of Physiotherapy

Co-supervisor(s):
Associate professor, Wim Grooten
Karolinska Institutet
Department of Neurobiology,
Care Sciences and Society
Division of Physiotherapy

Opponent:
Professor Emeritus, Roland Thomée
Sahlgrenska akademin, Göteborgs Universitet
Department of Health and Rehabilitation
Unit of Physiotherapy

Examination Board:
Professor, Martin Hägglund
Linköpings Universitet
Department of Health, Medicine, and Caring Sciences
Division of Prevention, Rehabilitation, and Community Medicine

Docent, Sofia Ryman Augustsson
Linné Universitet
Department of Sport Science
Faculty of Social Sciences

Professor, Charlotte Häger
Umeå Universitet
Department of community Medicine and Rehabilitation
Division of Physiotherapy

Tillägnad en fantastisk person och vän
En bok till din bokhylla, ett fönster till vår omvärld
Det skulle varit din tes, men den skall vi dryfta vid ett senare tillfälle
Vila i frid

Henrik Bergman, 8 april 1972 – 20 februari 2023

Abstract

Background: The Single Leg Squat (SLS) test is a functional test widely used in clinical settings to examine and evaluate rehabilitation goals. Research indicates that the SLS is reliable when the knee relative to the foot is dichotomously assessed. However, the assessment of functional movements often comprises more complex analyses of the whole kinetic chain with several body segments which highlights the need to develop and test a standardised multi-segmental SLS. Movement quality is an important aspect when using prevention programs in the clinical context, and the SLS can be used for this purpose. As knee injuries are common among athletes and especially among female soccer players, further investigation of the SLS in this population is warranted. Overall, the development of both quantitative and qualitative measurements needs to be studied to improve clinical testing. Clinically, portable marker-less motion capture (MMC) systems are suggested to be an adequate substitute for a three-dimensional analysis system, and one such novel MMC system is the Qinematic™. Before a test such as the SLS, or any other new measurement instrument, can be used in clinical settings, it is important to explore its measurement properties.

Aim: The overall aim of this thesis was to develop and assess aspects of reliability and validity of the SLS among physically active people, and from a biopsychosocial perspective investigate factors associated with the SLS in a sample of female soccer players.

Methods: Study I was a systematic review and meta-analysis that investigated the current literature regarding the intra- and inter-rater reliability of visually assessed SLS, including the Forward Step-Down (FSD) and Lateral Step-Down (LSD) tests. **Study II** was a laboratory-based test-retest reliability and validity study of a three-dimensional MMC system, the Qinematic™. **Study III** was an intra- and inter-rater reliability study of a standardised multi-segmental SLS developed from the findings in Study I. **Study IV** was a cross-sectional observational study using linear regression models to explore if demographic and biopsychosocial factors associated with the outcome of the SLS, assessed as a total score for all segments and as a separate knee segment in a sample of elite and sub-elite female soccer players. **An additional analysis** investigated the possibility of the SLS to discriminate injured soccer players from non-injured players.

Results: In **Study I**, the pooled results of ICC/kappa showed a “moderate” agreement for inter-rater reliability and a “substantial” agreement for intra-rater reliability of the SLS, including the FSD and LSD. In **Study II**, the Qinematic™ showed “substantial relative reliability” but “poor absolute reliability”. Regarding validity, a “moderate” agreement between the visual assessment and Qinematic™ data for various knee angles was shown and the best discriminative ability of the SLS was found at a knee angle of 6°. In **Study III** the proposed multi-segmental SLS showed a “moderate” inter-rater reliability and an

“almost perfect” intra-rater reliability. In **Study IV**, the outcome of the SLS was associated with previous injuries and various demographic-, biomechanical- and psychosocial factors depending on the tested leg. The total score associated with hip strength for both the dominant and the non-dominant leg, and the knee segment associated with division inherency for both the dominant and non-dominant leg.
The additional analysis showed that the SLS was not able to discriminate between players with and without previous or present injuries.

Conclusion: The SLS seems to be a reliable and clinically useful multi-segmental test of movement quality in contrast to the Qinematic™ system. The SLS was, in a sample of female elite and sub-elite soccer players, associated with a variety of biopsychosocial factors when assessed as a total score or as a separate knee segment. The results imply that several factors need to be considered when assessing the SLS among female soccer players such as leg dominance, division inherency, hip strength, and psychosocial factors.

Abstrakt

Bakgrund: Enbensknäböj testet, vilket i den engelsktalande litteraturen benämns som the Single Leg Squat (SLS) test, är ett vanligt förekommande funktionellt rörelsetest vilket bland annat används till att undersöka, träna och utvärdera rörelsekvalité. Forskning visar på att SLS kan bedömas på ett reliabelt sätt när knäts position i förhållande till foten bedöms som klarar/klarar inte. Bedömningen av funktionella rörelser omfattar dock ofta mer komplexa bedömningar av flera olika rörelsesegment i hela kinetiska kedjan vilket belyser behovet av att utveckla ett standardiserat multisegmentellt SLS. Bedömning och träning av rörelsekvalitet är en viktig aspekt i olika skadeförebyggande rörelseprogram och för detta ändamål passar SLS väl. Eftersom knäskador är vanliga bland idrottare och särskilt bland kvinnliga fotbollsspelare, är ytterligare forskning av SLS motiverat i denna population. För att förbättra den kliniska bedömningen av rörelsekvalité finns ett behov av att undersöka och utveckla både kvantitativa och kvalitativa mätmetoder. Som ett substitut till opraktiska och kostsamma tredimensionella rörelseanalyssystem har bärbara markörlösa rörelsesystem föreslagits, och ett sådant system är Qinematic™. Innan ett test som SLS, eller något annat nytt mätinstrument, kan användas i den kliniska vardagen är det viktigt att utforska dess mätegenskaper.

Syfte: Det övergripande syftet med denna avhandling var att utveckla och bedöma olika aspekter av reliabilitet och validitet för SLS hos fysiskt aktiva människor, samt ur ett biopsykosocialt perspektiv undersöka vilka faktorer som associerar med SLS ett urval av kvinnliga fotbollsspelare.

Metoder: Studie I var en systematisk översikt och meta-analys vilken undersökte litteraturen vad gäller intra- och interbedömmar reliabiliteten för ett visuellt bedömt SLS, inklusive Forward Step-Down (FSD) testet och Lateral Step-Down (LSD) testet. **Studie II** var en reliabilitets- och validitetsstudie av ett markörlöst tredimensionellt rörelseanalys system, Qinematic™. **Studie III** var en intra- och interbedömmar reliabilitetsstudie av ett standardiserat multisegmentellt SLS vilket utvecklades från resultaten i studie I. **Studie IV** var en tvärsnittsstudie vilken använde linjära regressionsmodeller för att undersöka om demografiska och biopsykosociala faktorer associerar med resultatet av ett SLS, bedömt som en totalpoäng för alla segment och som ett separat knäsegment i ett urval av kvinnliga fotbollsspelare på elit- och subelit nivå. **Ytterligare analyser, eller tilläggsanalyser,** genomfördes i syfte att undersöka om SLS kan diskriminera skadade fotbollsspelare från icke-skadade spelare.

Resultat: I **Studie I** visade det sammanslagna resultatet av meta-analysen för ICC/kappa en "moderat" interbedömmar reliabilitet och en "substantiell" intrabedömmar reliabilitet av SLS, inklusive FSD och LSD. I **Studie II** visade Qinematic™ en "substantiell" relativ reliabilitet och en "dålig" absolut reliabilitet. Vad gäller validitet uppvisade Qinematic™ en

"moderat" reliabilitet mellan den visuella bedömningen och dess data för olika knävinklar, den bästa diskriminativa validiteten för SLS visade sig vara vid en knävinkel på 6°.
I **Studie III** visade det föreslagna multisegmentella SLS en "moderat" interbedömmar reliabilitet och en "nästan perfekt" intrabedömmar reliabilitet. I **Studie IV** associerade utfallet av SLS med tidigare skador och olika demografiska, biomekaniska och psykosociala faktorer beroende på vilket ben som testades. Bedömningen som en total poäng associerad med höftstyrka för både det dominanta och det icke-dominanta benet, och den enskilda bedömningen av knäsegmentet associerade med spelarnivå (division) för både det dominanta och icke-dominanta benet. **Tilläggsanalysen** visade att SLS inte kunde diskriminera mellan spelare med och utan tidigare eller nuvarande skador

Konklusion: SLS verkar vara ett reliabelt och kliniskt användbart multisegmentellt test för bedömning av rörelsekvalitet i motsats till Qinematic™ systemet. Bedömt som en totalpoäng eller som ett enskilt knä segment associerade SLS med en mängd olika biopsykosociala faktorer i ett urval av kvinnliga fotbollsspelare på elit- och subelit nivå. Resultaten antyder att flera olika faktorer måste beaktas när man bedömer SLS bland kvinnliga fotbollsspelare, såsom bendominans, spelarnivå, höftstyrka och psykosociala faktorer.

List of scientific papers

This thesis is based on the following original articles and manuscript, which are referred to in the text by their Roman numerals.

I. **Ressman J**, Grooten WJA, Rasmussen Barr E. Visual assessment of movement quality in the single leg squat test: a review and meta analysis of inter-rater and intrarater reliability. BMJ Open Sport Exerc Med, 2019. **5**(1): p. e000541. https://doi.org/10.1136/bmjsem-2019-000541

II. **Ressman J**, Rasmussen-Barr E, Grooten WJA. Reliability and validity of a novel Kinect-based software program for measuring a single leg squat. BMC Sports Sci Med Rehabil, 2020. **12**: p. 31. https://doi.org/10.1186/s13102-020-00179-8

III. **Ressman J**, Grooten WJA, Rasmussen-Barr E. Visual assessment of movement quality: a study on intra- and interrater reliability of a multi segmental single leg squat test. BMC Sports Sci Med Rehabil, 2021. **13**(1): p. 66. https://doi.org/10.1186/s13102-021-00289-x

IV. **Ressman J**, Von Rosen P, Grooten WJA, Rasmussen-Barr E. Factors associated with the single leg squat in female soccer players: a cross-sectional study. Submitted manuscript.

This thesis also contains additional analysis.

Contents

List of abbreviations

2D	Two Dimensional
3D	Three Dimensional
95% CI	95% Confidence Interval
AFAQ	Athletic Fear Avoidance Questionnaire
AUC	Area Under the Curve
CLAM	Clamshell
Div	Division
FSD	Forward Step-Down
GAD-7	Generalized Anxiety Disorder-7 items
GK	Ground to Knee
HW	Heel to Wall
ICC	Intraclass Correlation Coefficient
LSD	Lateral Step-Down
MMC	Marker-less Motion Capture
NTA	Net Trajectory Angle
N	Newton
Nm	Peak torque values
NDL	Non-Dominant Leg
NPV	Negative Predictive Value
OR	Odds Ratio
PSS-14	Perceived Stress Scale-14 items
PSQI	Pittsburgh Sleep Quality Index
PPV	Positive Predictive Value
ROC	Receiver Operation Characteristic
SEM	Standard Error of Measurement
SLS	Single Leg Squat

Definition of central concepts

Functional movement tests: Functional movement tests are seen as functionally multi-joint tasks, based on work-specific activities and/or sport-specific skills [1, 2].

Movement Quality: Movement quality is defined as the maintenance of a correct posture, a good vertical alignment of joints and segments, and in addition, a good balance while performing a selected movement [3, 4].

The Single Leg Squat: There is today no universal and well-defined Single Leg Squat (SLS) test existing and in the present thesis "the SLS" will be used as a synonym or collective name for all different Single Leg Squat tests.

The FSD and LSD: The Forward Step-Down (FSD) and Lateral Step-Down (LSD) tests are two tests that are performed and assessed in a similar way as the SLS but differ as they are performed on a 15-20 cm high box [5, 6].

The Movement Continuum Theory of Physical Therapy: The conceptualisation of movement on a continuum that incorporates physical and pathological aspects of movement with social and psychological considerations [7].

Reliability: Is defined as *"The degree to which the measurement is free from measurement error"* [8], and measurement error is defined as *"The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured"* [8].

Validity: Validity is by COSMIN defined as *"the degree to which a health-related patient-reported outcome (HR-PRO) instrument measures the construct(s) it purports to measure"* [8] and is divided into content validity, criterion validity, and construct validity.

Preface

Since childhood, I always loved sports but also watching and observing movements in general. It doesn't matter if a movement is simple or complex, performed at the Olympic games or the playground, they always fill me with joy and interest. So, when I see people struggling with their movement, limping, or moving funnily I cannot stop wondering where the dysfunction is located and how I could correct it.

I first got in contact with the concept of movement screening and movement analyses in 2003 when I was working with a football team where we used preseason movement screening tests. This helped me to understand the complexity of our human movements and the concept of movement quality, and I think it was at this point my research interest started to grow. This thesis explores only one simple movement test, the Single Leg Squat test, which can be seen as one piece of the puzzle in clinical testing. Perhaps this could highlight the complexity of human movements and movement quality.

As an industrial, and part-time, doctoral student I have had one foot in the clinic and one foot in the academic world for the last six years. This has suited me well as I love my craftmanship and enjoy being a clinician. However, my supervisors have had concerns about me being too much of a clinician in my way of transforming into a researcher. They told me to put my doctoral hat on and look at my research from a helicopter perspective and not with the eyes of a clinician. With this thesis, I hope that I from now on, proudly can transform myself into a researcher or clinician whenever I need.

So, wherever I lay my hat that's my home..........

1 Introduction

Movement is a complex phenomenon and an essential part of human life. Often, we do not think about our movements as they are automatically executed without further reflection. However, injuries are sometimes a consequence of activities of daily life and sports and suffering from such an injury can be troublesome in the short or long term. Knee injuries are especially prevalent in some multidirectional sports such as soccer, and females seem to be more prone to severe knee injuries compared with men. Pre-screening programs and specific exercises are commonly used to prevent injuries, and one way of testing athletes for an injury preventive purpose is to use functional movement analyses and tests.

Movement analyses and tests are one of the cornerstones in the physiotherapy profession, used in clinical assessment, for rehabilitation purposes, and in the prevention of injuries. One of several functional movement tests that are commonly used to visually assess movement quality is the Single Leg Squat (SLS) test. In addition to qualitative movement tests, there is also a need to use quantitative measures that more accurately capture human movements. The ongoing digitization has made this possible with the development of portable and marker-less 3-dimensional motion capture (MMC) systems with a consumer-friendly price point.

Before a test such as the SLS, or any other new measurement instrument, can be used in the clinical setting, it is important to explore its measurement properties. Furthermore, to improve clinical testing and test-retest evaluation it is also important to understand which factors that associate with the SLS, and which factors can explain the outcome of the test.

2 Background

2.1 Theoretical framework

This thesis should be interpreted in the context of the Movement Continuum Theory (MCT) of Physical Therapy which was presented by Cott et al. [7] in 1995. The key concept of this theory is movement, but as other professions also are concerned with movement, Cott et al. wished to place the physiotherapy profession into a theoretical context that distinguished the physiotherapy profession from other professions concerned with movement. In their opinion *"Physical Therapists conceptualize movement on a continuum that incorporates physical and pathological aspects of movement with social and psychological considerations"* [7]. The MCT is based on the work of Hislop [9] and the model of pathokinesiology, but it is also clearly influenced by the biopsychosocial model presented by Engle in 1977 [10]. In an overall aspect, the MCT can be seen as the basis for movement assessment and interventions that are affected by biopsychosocial factors such as joint stiffness, muscle weakness, movement quality, work, education, motivation, coping mechanisms, and other personal and psychological factors.

The multidimensional movement continuum reaches from a micro level (molecular) to a macro level (person in society) and contains everything in between such as cells, organs, body parts, the whole body, and the person in a context [7]. The theory recognises that interference (both internal and external) with movement at any level has the potential to affect movement at the macro end of the continuum, and conversely, factors affecting the person at the macro level could affect movement at the micro end [7]. In other words, movement at each level of this continuum is affected by movements of the levels preceding and following and their relationship is dynamic and responsive to changes over time. This means that the levels can be affected in both directions and that the disturbance of internal and external factors can affect both the quantity and quality of the movement.

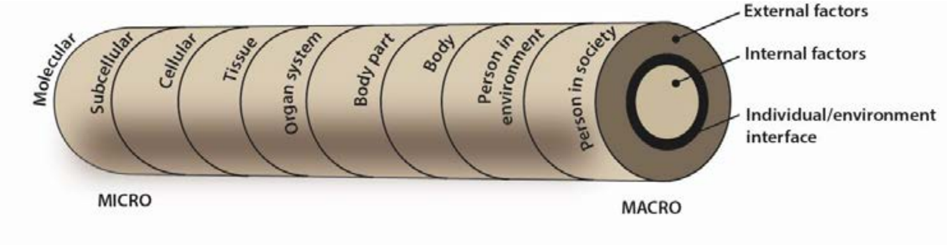


Figure 1. Pictures the MCT of physical therapy as a cylindric multidimensional movement continuum, modified after Cott et al. [7].

2.2 Sports injuries

Acute and overuse injuries are common in sports and many athletes suffer from pain and functional limitation due to their injuries [11, 12]. When looking at sex differences in injury rates among team sports athletes, it is obvious that women have a different injury risk profile than men [13, 14] with a higher risk of concussion [15, 16] and knee injuries [17], compared to men. On the other hand, men have higher injury rates for hip/groin, thigh, and foot injuries compared to women [13, 14].

A recently published meta-analysis on injury incidence rates (IIR) in senior women's soccer [18] reported an overall (match and training) IIR of 5.63/1000 hours for elite players. Amateur players were not so frequently studied, but a higher overall IIR was reported compared to the elite players. For elite players the most common injury type was "joint and ligament" and" muscle and tendon" with an overall IIR of 2.62-2.70/1000 hours when stratified on injury type, the injuries were predominantly in the lower limb [18].

Overuse injuries defined as an "*injury with insidious onset and no known trauma"* [19] have been reported to range from 16-30% of all injuries among female soccer players. This injury type is more commonly seen in the preseason and mostly affects the knee [20, 21]. A common overuse injury in the knee is patellofemoral pain (PFP) which has a prevalence of 9-15% in an active population and is more often reported in females [17, 22-25]. The annual prevalence of an anterior cruciate ligament injury in female soccer players is reported to be between 0.5-6.0%. Women also have a 2-3 times higher risk of suffering from an anterior cruciate ligament injury compared to male players [26]. Furthermore, female soccer players with an anterior cruciate ligament reconstruction have nearly a 5-fold higher rate of a new anterior cruciate ligament injury and a 2- to 4-folder-high rate of another new knee injury than knee-healthy controls [27].

2.3 Injury prevention

In 1992, one of the pioneers in injury prevention, van Mechelen et al. [11], described an injury prevention model and over the years this model has been further developed by Meeuwisse et al. [28, 29] and Bahr and Krosshaug [30]. Models of injury prevention today have advocated complex systems of pattern recognition rather than a focus on the linear causation of separate units of risk factors [31]. The idea is that the nature of injuries is multifactorial and complex, and does not arise from a linear combination of isolated factors, but rather from the interaction among separate factors, which are linked to each other in a non-linear manner [31].

2.3.1 Risk factors for injury

Identification of risk factors for injury has been pointed out as one of the cornerstones in the prevention of sports injuries [11, 28, 30]. Risk factors for injury can be divided into external and internal risk factors. External risk factors are described as factors that the athlete is exposed to, such as weather, equipment, coaches, rules, or turf type. Internal risk factors are those that are specific to an athlete, such as age, sex, coordination, weight, and biomechanical- and psychosocial factors [30]. A biomechanical analysis contains both kinematics and kinetics, where kinematics involves the description of spatial and temporal characteristics of a motion, thus describing position, velocity, and acceleration without any concern for the forces causing the motion [32] and kinetics is the study of the forces that causes the movement [32].

2.3.1.1 The dynamic knee valgus

One biomechanical risk factor that has been extensively researched in both male and female athletes is the so-called dynamic knee valgus which is suggested as important for non-contact anterior cruciate ligament injury [33-36] and for patellofemoral pain [36-38]. In male and female soccer players kinematic analysis of the dynamic knee valgus often estimates the intersegmental relationship and joint angles according to frontal and sagittal plane alignment of different segments [34, 39]. The intersegmental relationship during a dynamic knee valgus is characterised as a postural malalignment of the lower extremity and trunk, such as excessive foot pronation, tibial inward rotation, knee valgus, femoral adduction and inward rotation, contralateral pelvic drop and ipsilateral trunk tilt and contralateral rotation [39-41]. Different parts of this postural malalignment have in addition to patellofemoral pain and the anterior cruciate ligament injury been related to overuse injuries such as low back pain [42, 43], tibial stress fractures [44], iliotibial band syndromes [45], and femuro-acetabular impingement [46].

In cross-sectional studies, an increased dynamic knee valgus measured by kinematics during different weight-bearing single-leg squat tasks has been associated with decreased trunk lateral flexion strength [47-49], decreased hip strength, and altered hip muscle activation [50-52] even though conflicting results exist [50, 51]. However, it seems that the association between hip strength and knee valgus might be conditional to task demand [52, 53] and that the level of influence of hip strength and electromyography activity varies across single-leg squatting, landing tasks, and gender [52]. Furthermore, there seems to be a sex difference regarding the dynamic knee valgus as women without knee injuries or patellofemoral pain show an increased dynamic knee valgus compared to men when performing a functional weight-bearing activity [38]. Finally, it is also suggested that a decreased ankle dorsiflexion contributes to an increased dynamic knee valgus [50, 51, 54].

2.3.1.2 Psychosocial risk factors

Briefly, psychosocial risk factors refer to the influence of social factors on an individual's thoughts and behaviour, and the interaction between these [55]. Studies on risk factors for sports injuries often use the notion of psychosocial risk factors as an inclusive term for different personal and psychological traits which affect, and are affected by, an individual´s social context or environment [56, 57]. Psychosocial risk factors such as fear of avoidance, stress, anxiety, sleep quality, or coping mechanisms are proposed to be associated with an increased risk of both illness and musculoskeletal injuries among athletes in different sports [56-59]. A meta-analysis from 2017 pointed out that life-event stress and strong stress responsivity had the strongest relationship with injury rates in different athletes [56]. Furthermore, a risk factor such as fear of reinjury have been related to a higher risk of suffering a second anterior cruciate ligament injury and to poor performance on different tests in rehabilitation [59, 60]. Hence, there is a need to consider psychosocial risk factors in injury prevention and to improve the instruments used in the clinic and research.

2.3.2 Preventive strategies

Preventive strategies that are used by sports teams and individual athletes are preseason screening [61, 62], assessment through questionnaires [62, 63], and exercise-based programs that are performed during the season [13].

2.3.2.1 Preseason screening

A preseason screening can differ depending on sport, demand, and purpose. Concerning the prevention of musculoskeletal injuries, preseason screening procedures have over the years moved away from isolated assessments of joints and muscles to a more functional approach where functional movements are used to assess the whole kinetic chain [1]. Either with a single functional movement test or a complete screening battery containing several different tests. In this context, functional movement tests are seen as functionally multi-joint tasks, based on work-specific activities and/or sport-specific skills [1, 2].

The predictive validity for screening batteries is poor meaning that they have a poor ability to predict a new injury [64-67]. The same applies to a variety of separate functional movement tests, especially when screening for future anterior cruciate ligament injuries with the vertical drop jump [68, 69], the Single Leg Squat (SLS) test [70], and the Star Excursions Balance Test [68]. However, some contradicting evidence for the predictive validity exists for single-leg landings [71], the Star Excursions Balance Test [72] and the SLS [73]. In addition, the use of screening tests for injury-preventive purposes in sports is debated [4, 74-76]. In a review article, a former major proponent of injury screening invalidated its use in sports [74]. Other authors recognize the limitation of functional screening tests but still claim that screening may be an important strategy to protect an athlete from recurrent injuries [75, 76]. In this sense, there is a need to further improve the measurement properties of the assessment methods applied in the clinical context and to get a better understanding of the complexity underlying an injury situation [75, 76].

2.3.2.2 Assessment through questionnaires

Another way of screening an athlete for injury risk is by the use of questionnaires. Studies from male premier league soccer teams report that the use of questionnaires is the second and third most used test to identify injury risk in athletes [62, 63, 77]. Physiotherapists and team doctors report that the questionnaires include measures of pain, fatigue, quality of sleep/rest, lifestyle habits, and psychological state [63].

2.3.2.3 Exercised-based injury prevention programs

Exercised-based programs that are performed during the season have proven to reduce anterior cruciate ligament injuries for all athletes and all sports [78], especially in females where a 67% reduction of non-contact anterior cruciate ligament injuries has been reported [78]. A meta-analysis focusing only on female soccer players reported that multicomponent exercise-based programmes reduced injuries (all kinds of injuries) by 27% (incidence rate ratio (IRR) 0.73, 95% CI 0.59 to 0.91) and anterior cruciate ligament injuries with 45% (IRR 0.55, 95% CI 0.32 to 0.92) [13].

There are many exercise-based injury preventive programs, for example, FIFA 11+ [79, 80], Knäkontroll [81], and Prevent Injury and Enhanced Performance Programme (PEP) [82]. In general, these programmes focus on improving core strength, balance, and neuromuscular activation of the lower extremities, and the vertical alignment of the lower extremity and trunk during activities similar to anterior cruciate ligament injury mechanisms (e.g., lounging, squatting, cutting, jumping, and landing) [13, 81-83]. In other words, there seems to exist support in favour of “correct” movement patterns and movement quality during the performance of these exercises [83]. Furthermore, several studies report that frequent training of the FIFA 11+ [80, 84], specific core training [85], and neuromuscular training with a focus on the lower extremity and hip [86], alter and improve the vertical alignment of the lower extremity. That will say, improve the athlete’s movement quality.

2.4 Movement quality

One of the pioneers in functional movement screening is Gray Cook who in 2006 presented the Functional Movement Screen (FMS) which was described as an evaluation tool to assess an individual's fundamental movement pattern [1, 87]. The FMS focuses on the recognition of poor movement patterns in the kinetic chain recognised by "weak links" and "compensatory movements" due to poor stability, mobility, joint proprioception, and/or reduced neuromuscular control [1, 88].

The concept of movement quality is often used in conjunction with the assessment of functional movement tests and is closely related to the concept of good/poor movement patterns. Even though authors have claimed that there is no consensus statement in the literature on what defines movement quality [3] the two concepts of movement quality and good/poor movement patterns are sometimes used interchangeably [4]. Movement quality has been described as the maintenance of a correct posture, a vertical alignment of joints and segments, and in addition, a good balance while performing a selected movement [3, 4]. Furthermore, it is also described in terms of the efficiency of a functional movement [67, 89]. In other words, movement quality might be an individual's ability to perform a functional movement in an optimal way [1, 90]. On the contrary, poor movement quality is often seen as a disruption of the "normal" function of synergist muscle function, neuromuscular control, proprioception, joint mobility/stability, and/or muscle flexibility during a functional movement [1, 52, 91-94]. One specific trait of what is defined as poor movement quality, which is commonly assessed during a functional movement test such as the SLS is the presence of uncontrolled oscillatory movement of the knee in a mediolateral direction, clinically observed as repeated knee valgus/varus motion or knee wobbling [95, 96]. Biomechanical studies support that this uncontrolled oscillation can be seen as a movement deviation [97, 98], and two studies have shown that those oscillatory movements are more common in patients with an anterior cruciate ligament injury compared to healthy controls [96, 99]. One reason suggested for those oscillatory movements is a lack of neuromuscular control [96, 99].

Movement quality has further been identified as an independent attribute, as it, unlike other quantitative measures such as power or strength, aims to capture other important aspects of a movement such as movement efficiency and the maintenance of a correct posture, balance, and vertical alignment [3, 89, 100]. This highlights the importance of integrating the concept of movement quality when examining a patient, testing an athlete for return to play, or screening an athlete for an injury preventive purpose [3, 101]. With reliable and valid functional movement tests, the assessment of movement quality may provide a systematic and reliable way to observe and rate (quantify) movements [91]. This makes the assessment of movement quality suitable for clinical examination, test re-test evaluation, and prescribing exercises. Furthermore, a specific focus on movement quality in training (e.g., good knee control when cutting) can also be used as a tool to improve an athlete's performance, for example in a change of direction and cutting manoeuvres in male youth soccer players [102], or as a way to increase power and strength [86, 90].

2.5 Measurements of movement quality

Clinicians use visual assessment to assess different aspects of movement quality, such as posture, balance, vertical alignment, and fluency of a movement, but clinicians cannot use their eyes to "measure", angles, velocity, muscle activity, and forces. For this, quantitative systems using different complexity have been developed. In the clinical setting, the gold standard, three-dimensional (3D) motion analysis systems, are seen as impractical, time-consuming, and expensive [103] and these systems are only used in laboratory settings. However, there are also other promising measures available for quantifying these aspects of movement quality in a clinical setting, such as 2D-motion analysis systems and marker-less motion capture (MMC) systems. One such system is Qinematic™.

2.5.1 Visual assessment

Visual assessment of movement is commonly used in a clinical situation, where the need for quantification of movement variables is low. Visual assessment of movement quality has been seen as an important skill required of clinicians, athletic trainers, and coaches, and an important component in the clinical decision-making process [101, 104, 105]. However, a test that is totally based on visual assessment could have low measurement properties because of the subjective nature of it [106-108]. Therefore, a clinician must be aware of the wide range of available functional tests, their within-subject kinematic variation, and their reliability and validity [105].

The reliability of a test based solely on visual assessment is in general affected by the complexity of the rating scale (dichotomised or multiple classification categories), the number of segments assessed, the definition of the rating criteria (anatomical references or critical features), the velocity of the test, and in addition the training and experience of the assessors [105, 109]. Moreover, the reliability and validity of a functional test might differ with the movement being rated and which population (sex and age) is tested [110].

2.5.2 Three- and two-dimensional motion analysis systems

In the absence of 3D-motion analysis systems, 2D-motion analysis systems have been seen as an acceptable assessment method to quantify frontal plane kinematics in the lower extremity and trunk [51, 103, 111]. Low-cost 2D-analysis systems have been available on the market for a decade, and some of them have comparable precision to leading analysis systems, but at a significantly lower cost [112, 113]. One systematic review and meta-analysis [114] and one systematic review [115] report that 2D-analysis systems are reliable but dependent on the task and the type of reliability that is evaluated and that the agreement with 3D-movement analysis (criterion validity) has shown conflicting or poor results in measuring frontal plane kinematics [114, 115].

Even though 3D- and 2D-motion analysis systems are seen as quantitative and objective measures when analysing movement quality, one should keep in mind that a chosen cut-off for good and poor movement quality may originate from a subjective decision. As pointed out by Roald Bahr [74], a risk factor for future knee injury such as the dynamic knee valgus is continuous and there might be a substantial overlap in movement quality between the injured and uninjured groups compared to a dichotomous risk factor for breast cancer such as the presence or absence of a tumour.

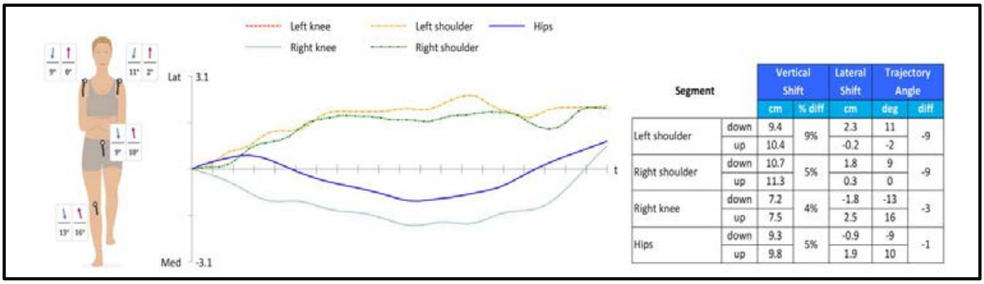
2.5.3 Marker-less motion capture systems

Accurately capturing the patient's kinematics in the clinic requires a portable and marker-less motion capture (MMC) system, since the use of 3D-motion analysis systems in the clinical setting is limited [103, 116, 117]. Microsoft created the Kinect camera for Xbox 360 as a game controller in 2021, and by using a built-in RGB (red, green, and blue) sensor and a skeleton tracking algorithm, the Kinetic system could process 3D kinematic data for clinical purposes [118, 119]. The Kinect sensor has been evaluated on its test-retest reliability, accuracy, and construct validity of gait, balance, posture, functional tasks, and movements in the trunk and upper and lower extremities in a variety of different diagnoses [116, 120]. In general, research points out that MMC systems offer a great potential to expand the 3D-movement analysis to a clinical setting [120, 121], although it is recognised that the utility of such a system might be limited, especially where detailed 3D-motion analysis is required [120, 121]. When looking at specific functional tasks such as squatting and landing, the Kinect system has demonstrated moderate to good validity and reliability [117, 122-126]. However, it has also been reported that the reliability of the Kinect V2 system varies depending on joint angles and positions and that the reliability decreases with increased task complexity, such as double leg squats versus lunges [127].

2.5.3.1 The QinematicTM

Qinematic™ is a portable MMC system that uses the Kinect sensor together with a refined software program (Quickposture™), that has improved the camera's stability and accuracy through a unique tracking algorithm [128]. The system was initially designed to be used by health and wellness providers who aimed to record and visualise basic human movements. It works like a semi-automated service where the users perform seven different functional tasks in the following order: #1. Standing balance; #2. Side bending; #3. Squat (double leg squat) with arms crossed over the chest; #4. Balance on the right leg with arms crossed over the chest; #5. Balance on the left leg with arms crossed over the chest; #6. Squat on the right leg (SLS) with arms crossed over the chest; and #7. Squat on the left leg (SLS) with arms crossed over the chest.

Qinematic™ registers segmental displacement during the whole movement and calculates a Net Trajectory Angle (NTA), which contrasts with other studies that calculate the peak joint angle during a movement [122, 124-127]. For example, during a SLS, the medial and lateral displacement of the knee is measured 30 times per second from the start of the movement to the bottom (down), and back (up). In this way, the NTA represents the angle between the estimated "line of best fit" through the changes in knee position and the vertical axes for each direction, **see Figure 2**.



Segment		Vertical Shift		Lateral Shift	Trajectory Angle	
		cm	% diff	cm	deg	diff
Left shoulder	down	9.4	9%	2.3	11	-9
	up	10.4		-0.2	-2	
Right shoulder	down	10.7	5%	1.8	9	-9
	up	11.3		0.3	0	
Right knee	down	7.2	4%	-1.8	-13	-3
	up	7.5		2.5	16	
Hips	down	9.3	5%	-0.9	-9	-1
	up	9.8		1.9	10	

Figure 2: The biomechanical report of Qinematic™ for the SLS. The net trajectory angle (NTA) estimates the "line of best fit" for the pathway of different key body parts, the table shows 13° of medial displacement for the right knee on the way down, and 16° of lateral displacement on the way up, but only 1.8 cm and 2.5 cm of medial/lateral shift, respectively, blue dotted line in the chart. (**Nota Bene:** This figure was previously published in Study II, no changes have been made, the article is licensed under a Creative Commons Attributions 4.0 International License, https://creativecommons.org/licenses/by/4.0/**).**

2.6 The SLS

The SLS is a functional movement test that is used to assess how a patient performs a knee squatting movement on one leg. It is often used in the clinical context; during rehab as an exercise, to evaluate improvements in a test-retest test situation, and in different screening batteries [88] and exercise-based injury preventive programs [81]. As there exists a variety of SLS, no uniform SLS exists to date. The various SLS presented in the literature are not always, but often referred to as "the SLS". **Table 1** depicts the general differences for some of these tests.

Table 1. The differences of a Single Leg Squat regarding name, performance, and rating criteria

Study	Method/Test	Assessment Criteria
Ageberg et al [129]	**Single-limb mini squat:** -Stance leg; 50° knee flexion -Non-stance leg; slight hip flexion and 80° knee flexion. -Arms; Fingertip support.	-Body segments; knee/foot -Rating scale; Medial or over/lateral to 2nd toe at ≥ 3 out of 5 squats.
Crossley et al [130]	**Single-Leg Squat:** **(N.B. this test is named as an SLS but executed as an FSD).** -Stance leg; on a 20-cm high box -Non-stance leg; pointing forward (as for step down). -Arms; folded across their chest	-Body segments; trunk, pelvis, hip joint and knee joint. -Rating scale; a three-point scale. -Overall impression across 5 trials.
Edmondston et al [131]	**Single-Leg Squat:** -Stance leg; knee flexion 30°. -Non-stance leg; knee flexion 30° -Arms; arms along the side.	-Body segments; trunk -Rating scale; dichotomous, direction of the trunk against/from the non-stance leg.
Frohm et al [88]	**One-Legged Squat:** -Stance leg; squat is performed as deep as possible with the upper body vertically. -Non-stance leg; Hip neutral, knee flexion 90° -Arms; Hands on hip	-Body segments; foot, knee, hip, pelvis and trunk -Rating scale; a four-point scale -Three squats were assessed.
Piva et al [5]	**Lateral Step-Down LSD Test:** -Stance leg; Standing on a 20-cm high box. Bending until the non-stance leg gently touches the floor. -Non-stance leg; positioned over the floor adjacent to the step, maintained with the knee in extension. -Arms; hands on waist.	-Body segments; arm strategy, trunk, pelvic, knee and stance. -Rating scale; dichotomous for each segment and given 0-1 points. Except for the knee where 0-2 points were given. -All five LSD was assessed to a composite score of 0 to ≥4 points.
Stensrud et al [132]	**Single Leg Squat:** -Stance leg; 90° knee flexion. -Non-stance leg; not allowed in front or at side. -Arms; hands-on waist	-Body segments; pelvic, knee and medial/lateral side-to-side movement of the knee. -Rating scale; a three-point scoring scale -The subjects were scored by their poorest performance.

Nota Bene: Parts of this Table were previously published in Study I. Compared to the original Table, some columns are removed but no changes have been made regarding the published text. The article is licensed under a Creative Commons Attributions 4.0 International License, https://creativecommons.org/licenses/by/4.0/).

As can be seen from **Table 1,** the performance of an SLS differs regarding the depth of the squat, the placement of the arms, and the position of the non-weight bearing leg. Some authors use a uni-segmental approach as they only assess one segment at a time (e.g., the knee in relation to the foot), while others have a multi-segmental approach assessing the whole kinetic chain from the foot to the trunk, and rating criteria differ from a dichotomous rating to a 4-point rating scale. Some authors present a specific score for each separate segment while others present a composite score for all segments. In addition, there is also the Forward Step Down (FSD) test and the Lateral Step Down (LSD) test, which are tests that are performed and assessed in a similar way as the SLS but differ as they are performed on a 15-20 cm high box.

Biomechanical studies have shown kinematic and kinetic differences in the various performances of the SLS, e.g. when the non-weight bearing leg is positioned forward (SLS-forward), in the middle (SLS-middle) or backward (SLS-back) in relation to the trunk [133, 134]. In a study by Khuu et al. [133], the SLS-back demonstrated the greatest differences at peak knee flexion compared to SLS-front and SLS-middle with greater kinematics in the trunk, pelvic, hip and knee, and greater kinetics in hip external rotators and knee extensors. There is also evidence for kinematic differences when comparing the SLS with the Forward Step-Down test and Lateral Step-Down test [135, 136]. Martonick et al. [136] used statistical parametric mapping to study the kinematic waveforms during the whole motion and demonstrated that the SLS-back generated greater amounts of trunk, pelvic, and hip flexion for the entire movement compared to the Forward Step-Down test and Lateral Step-Down test. However, the Forward Step-Down test provoked greater amounts of knee abduction than the SLS and Lateral Step-Down test at 26-66% of the movement [136] which also has been confirmed in another study [135]. Subjects with anterior cruciate ligament injuries [96, 137-139] and patellofemoral pain [37, 140] show different kinematics and poorer performance of the SLS compared to non-injured subjects or the non-injured leg. Finally, there are also gender differences in the performance of a SLS where females show greater kinematic hip internal rotation, hip adduction and knee valgus angles compared to males [141-145].

2.6.1 The common denominator in all SLS

Despite the kinematic and kinetic differences of the various SLS described in **Table 1**, one could argue that the various SLS, the Forward Step-Down test and Lateral Step-Down test assess the same construct if considering the definition of a construct as a theoretical concept, theme, or idea based on empirical observations [146]. Firstly, the general movement pattern during all those tests is the same with flexion at the knee, hip and trunk, pelvic tilt, hip adduction, and knee internal rotation and abduction [133, 135], secondly, as they all assess movement quality which has been identified as an independent attribute [3, 89, 100]. The common denominator is that the aim of the SLS is to visually assess the coordination of different body segments in relation to each other (the vertical alignment), knee and trunk stability, balance, and overall motor control (movement efficiency and neuromuscular function).

2.6.2 Measurement properties in general

Before a test instrument can be used in research and clinical practice the measurement properties of the test or instrument need to be evaluated [147]. The Consensus based Standards for the selection of health Measurement Instruments (COSMIN) is an initiative to improve the selection of health measurement instruments based on a Delphi study [8]. The aim was to reach a consensus on which measurement properties are relevant for evaluating health-related patient-reported outcomes (HR-PROs), which terminology and definitions to use, and which design requirements and statistical methods are most proper. Importantly, an HR-PRO is a reported health condition status that comes directly from the patient without any interpretation by a clinician or anyone else [148], in contrast to the evaluation of a functional movement test that is visually assessed by a therapist. However, high methodological quality is important for HR-PROs and in addition for all tests used in research and clinical settings. COSMIN´s taxonomy of measurement properties is divided into three domains: reliability, validity, and responsiveness [8].

The domain reliability contains internal consistency, reliability, and measurement error. Reliability is by COSMIN defined as *"The degree to which the measurement is free from measurement error"* [8] and measurement error is defined as *"The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured" [8]*. Furthermore, reliability is a measure of whether the scores of a test person change over time when the measurements are repeated without any intervention between the test occasions (test-retest reliability), between different raters (inter-rater reliability) and within a rater (intra-rater reliability) [8]. The measurement property measurement error is also divided into test-retest, inter-rater and intra-rater reliability [8].

Validity is by COSMIN defined as *"the degree to which a health-related patient-reported outcome (HR-PRO) instrument measures the construct(s) it purports to measure"* [8] and is divided into content validity, criterion validity and construct validity. Content validity refers to *"The degree to which the content of an HR-PRO instrument is an adequate reflection of the construct to be measured"* [8]. A first aspect of content validity is face validity, which is defined by COSMIN as "*the degree to which (the items of) an HR-PRO instrument indeed looks as though they are an adequate reflexion of the construct to be measured"* [8]. Face validity is a subjective, overall, first impression without any further standards on how to assess it. However, a lack of face validity is a quite strong argument for not using an instrument [149]. Stronger evidence for validity is criterion validity which refers to *"the degree to which the scores of an HR-PRO* *instrument are an adequate reflection of a "gold standard" [8],* the definition implies that criterion validity only can be assessed when a gold standard is available. Criterion validity can further be divided into concurrent and predictive validity, where concurrent validity usually is used for diagnostic purposes and predictive validity for predictive applications [149]. On the other hand, when a gold standard is not available, construct validity should be used. Construct validity is by COSMIN defined as" *the degree to which the scores of an HR-PRO instrument are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the HR-PRO instrument validly measures the construct to be measured"* [8].

The domain responsiveness is by COSMIN defined as *"The ability of an HR-PRO instrument to detect change over time in the construct to be measured"*[8], responsiveness is however a separate domain but closely related to validity [149]. The only difference between validity and responsiveness is that validity refers to the validity of a single score, meanwhile, responsiveness refers to the validity estimated from two different measurements, for example when a test is used longitudinally to measure change over time [149].

2.6.3 Measurement properties for the SLS

When performing a functional movement test such as the SLS, errors can emanate from several sources such as from the equipment and the raters. One main source of error is that humans vary their movements over time and have a great ability to adapt to specific circumstances when performing the test [150]. This variation can be described as the within-subject kinematic variation and is important if the test is to be used for clinical decision-making and repeated measurements [105, 150]. The within-subject kinematic is often measured with the Standard Error of Measurement (SEM) which is a measure of absolute reliability, unlike the Intraclass Correlation Coefficient (ICC) which is a relative reliability measure that provides less clinical meaning [105, 150].

With regard to the differences between all SLS, the Forward Step-Down and Lateral Step-Down the reported within-subject kinematic variation seems to be relatively low with an SEM of less than five degrees and mostly substantial/almost perfect ICC values over 0.60 [49, 103, 105, 132, 151-155]. A systematic review [105] and a systematic review and meta-analysis [156] indicated that the evidence for the intra- and inter-rater reliability of the SLS, when using a dichotomous rating of the knee relative to the foot was adequate. It was also reported that agreement with more complex ratings, such as ≥3-point rating scales and multi-segmental assessments, was acceptable in some of the studies [105].

In general, the SLS seems to have strong face validity in the sense that it has biomechanical and neuromuscular similarities to athletic movements as it simulates common athletic positions such as cutting and landing [144, 151, 152, 157, 158]. Previous studies on healthy subjects have shown that visual assessment of the SLS corresponds well with 3D- and 2D-motion analysis systems in discriminating good and poor SLS performers when assessing the foot, knee, pelvic, hip and trunk [105, 129, 132, 155, 159-163]. However, not all segments assessed in a SLS are equally studied (the knee segment most studied) and not all SLS are represented. To date, only a few studies have investigated the visually assessed SLS predictive and discriminative validity, and they report poor and conflicting results [70, 73, 164, 165]. In an alternative study, Crossley et al. [130] used an expert panel to set consensus about the movement quality on the performance of 34 separate SLS which were assessed as good, fair, or poor. The consensus rating was then compared with three physiotherapists who made their assessments of the performance. The concurrency with the consensus panel was substantial to almost perfect (kappa 0.60-0.80). Finally, one literature review from 2015 aimed to study responsiveness in relation to the SLS, but no such studies were found [166].

2.6.4 Biomechanical factors associated with the outcome of the SLS

To improve clinical testing and test-retest evaluation it is not only important that the test is reliable and valid, but also that there is an understanding of which factors that associate with the test and which factors that can explain the outcome of the test. Kinematic studies indicate that there exists an association between a poor performance of the SLS and a decreased hip muscle strength and an altered hip muscle activation [50-52] and in addition a decreased ankle dorsiflexion [50, 51, 54, 167]. Regarding the visually assessed SLS, four studies reported a positive association between poor SLS performance with decreased hip muscle strength and an altered hip muscle activation [130, 168], while two studies reported no such associations [169, 170]. For ankle dorsiflexion, four studies reported positive associations with reduced ankle dorsiflexion and poor Lateral Step-Down test performers [169-172] meanwhile two studies reported contradicting results for the Forward Step-Down test and the SLS [168, 173].

Other associated factors are fatigue [143] and balance [174], where a fatigue-inducing program (muscular, cardiovascular and respiratory) has shown kinematic changes in the trunk and pelvis during the performance of a SLS. However, no differences were found in postural sway in individuals with and without a visually assessed dynamic knee valgus when performing a SLS [174]. On the other hand, one study on visual assessment of a SLS reported a large significant relationship between good balance and good performance of the SLS [165].

2.6.5 Psychosocial factors associated with the SLS

To measure psychosocial factors Patient-Reported Outcome Measures (PROMs) can be used. A PROM is a self-completed questionnaire that for example could measure fear of avoidance, stress, anxiety, and sleep quality [148, 175]. Psychological- and psychosocial factors are suggested to affect the outcome of rehabilitation after a sports injury [176, 177] and it has been reported that athletes with a higher fear of re-injury do not return to their pre-injury level to the same extent as those having lower levels of fear of re-injury [178, 179]. Fear of re-injury is one of the most common personal reasons for not returning to sports apart from knee issues such as swelling, instability, or weakness [180]. However, studies of physiotherapists in clinical practice imply that PROMs are used to a low extent [181, 182].

Studies on the importance of psychosocial factors in relation to physical function and performance during rehabilitation are sparse [59], and to date, knowledge is limited about how and if various psychosocial factors affect or are associated with the outcome of a test such as the SLS. Nevertheless, studies have reported that the physical performance of single leg hop, and muscle strength might be affected by fear of reinjury during the rehabilitation of anterior cruciate ligament injuries [59, 60]. Furthermore, anxiety has been associated with performance problems in sports and other fields [183, 184], and is also reported as one psychological factor that negatively affects the return to play after an anterior cruciate ligament injury [185]. The link between anxiety disorders and competitive performance is, however, not well understood and it is unclear whether interventions that decrease anxiety are associated with better performance [184]. Hence, there is a need for a more profound understanding of how psychosocial factors affect rehabilitation, but also how these factors affect and are associated with tests and exercises in the screening and rehabilitation situation.

2.7 Rationale

The wide use of the visually assessed SLS in clinical testing and for preventive purposes highlights the need for a further investigation of the test. Unfortunately, there is today no uniform SLS existing as the performance and assessment criteria vary. Research indicates that the SLS when using a dichotomous rating of the knee relative to the foot has adequate reliability investigated across a range of ages. However, the assessment of functional movements often comprises more complex analyses of the whole kinetic chain including several body segments at the same time. This emphasises the need for a reliable multi-segmental SLS based on a proper methodological setup and refined assessment criteria. Prevention programs used in sports emphasise movement quality and knee control [83], and as knee injuries are common among both men and women athletes, those injuries are of great concern in sports and rehabilitation.

Furthermore, sex differences regarding injury rates and profile [13, 14], together with the increased dynamic knee valgus in females compared to men when performing a functional weight-bearing activity [38] justify further investigation of the SLS in female athletes. A better understanding of the outcome of the SLS and its separate segments viewed from a biopsychosocial perspective might aid the clinician to make evidence-based decisions about rehabilitation and return to sport, and to guide safe and effective exercise prescriptions. In relation to that, this thesis focuses on one test, the SLS, it is important to highlight that more than one test or more than one outcome of a PROM is needed to evaluate and assess functional movements in relation to injury or return to play. Therefore, the outcome of the SLS should, in this context be seen as one piece of a puzzle to better understand how the outcome of a functional movement relates to injury or return to play.

To improve clinical testing in the assessment and test-retest evaluation, the development of both quantitative and qualitative measurements needs to be further studied. As an alternative for 3D-analysis systems in the clinical setting a portable MMC system such as the Qinematic™ is of importance for the ongoing healthcare digitisation. In other words, there is also a need for simpler but objective and quantitative methods to capture functional movements in the clinical setting.

3 Overall aim

The overall aim of this thesis was to develop and assess aspects of reliability and validity of the SLS among physically active people, and from a biopsychosocial perspective investigate factors associated with the SLS in a sample of female soccer players.

3.1 Specific aims

Study I

To systematically review and meta-analyse the current literature on intra- and inter-rater reliability of visually assessed SLS, including the Forward Step-Down and Lateral Step-Down tests.

Study II

To establish the reliability and validity of QinematicTM for assessing the SLS. A further aim was to identify different angles or cut-off points of medial knee displacement, during a SLS measured by QinematicTM, that in the best way would match the results of a visually assessed knee-over-foot or knee-medial-to-foot position.

Study III

To investigate the intra- and inter-rater reliability of a standardised multi-segmental SLS with refined assessment criteria.

Study IV

To investigate whether demographics, previous injuries, and biomechanical and psychosocial factors are associated with the outcome of the SLS, assessed as a total score for all segments and as a separate knee segment in a sample of elite and sub-elite female soccer players.

4 Materials and methods

4.1 An overview of study design and subjects

The present thesis contains four studies and additional statistical analysis where the manuscript for **Study IV** is under submission. The set-up for the data collection was similar in **Study II** and **III**, and the additional statistical analyses were calculated on the same subjects as **Study IV**. An overview of the study design, subjects, and statistics are presented in **Table 2.**

Eligibility criteria for inclusion in **Study I** were methodological studies that assessed the inter-rater and/or intra-rater reliability of a visually assessed SLS, Forward Step-Down test, and Lateral Step-Down test. All subjects that showed interest were included. Studies with kinematic, kinetic, and other quantitative measures were excluded. Inclusion criteria for **Study II** and **II** were men and women, aged 18 to 65. Exclusion criteria were an ongoing musculoskeletal injury in the lower extremity, a history of severe knee disorder (ligament- or meniscal rupture and knee replacement), a neurological disease, or a visual deficiency that could not be corrected with eyeglasses. Inclusion criteria for **Study IV** were contracted players who were 16 years or older, and who understood written and spoken Swedish. Exclusion criteria were two-footed players and an ongoing injury that made it impossible to perform the physical tests without pain, or those for whom participation meant an additional risk for injury.

Table 2. Overview of study design, subjects, and statistics used in the included studies.

Studies	Design	Subjects	Statistics
Study I	A systematic review and meta-analyses	31 included studies (n=1136) ♀=454, ♂=360, unknown gender=322 Age: 9-89 years. 65% of the study sample were healthy active subjects. Age: 18-37 years	**Meta-analyses:** Pooled agreement of Intra- and inter-rater reliability
Study II	A laboratory-based test-retest reliability and validity study of a marker-less 3D motion analysis system, Qinematic™	37 healthy active subjects ♂=10, ♀=27 Age: 18-65 years old, mean 34 (±12)	**Systematic differences between test occasions:** Wilcoxon sign ranked test and Bland-Altman plots. **Relative and absolute reliability:** Spearman correlation coefficient, ICC, SEM and SDC **Construct validity:** Agreement-Kappa Diagnostic accuracy-sensitivity, specificity, PPV, NPV and AUC
Study III	An intra- and inter-rater reliability study		**Systematic differences between test occasions:** Wilcoxon sign ranked test. **Relative reliability:** Kappa statistics
Study IV	A cross-sectional observational study	254 female soccer players distributed on: Div. 1: 89 players Div. 2: 51 players Div. 3: 114 players Median age: 22 (min. 16-max. 39)	**Statistical differences:** McNemar´s test, Wilcoxon rank sum test and Wilcoxon signed rank test **Associations to the SLS:** A backward logistic regression analysis
Additional analysis	A cross-sectional observational study		**Predictive/discriminative validity:** Sensitivity, specificity, PPV and NPV.
Abbreviations: ICC=Intraclass correlation coefficient; SEM=Standard error of measurement; SDC=Smallest detectable change; PPV=Positive predictive value; NPV=Negative predictive value; AUC=Area under the curve; 3D= Three dimensional; n=total number of includes subjects; Div=Division; min=minimum; max=maximum.			

4.2 Ethical considerations

Study I-IV were conducted in accordance with the Declaration of Helsinki [186], aiming to protect individuals' rights, integrity and autonomy in research procedures. Written informed consent to participate was obtained for all subjects in **Study II-IV** following written and oral information, and the studies were approved by the Regional Ethical Review Board in Stockholm. **Study II and III** received Ethical approval Dnr: 2016/595-31 with amendment Dnr 2017/318-32, and **Study IV** with Ethical approval Dnr: 2021-03067 with amendment Dnr 2021-05398-02. The data collection for **Study IV** was performed in January – February 2022, and the registration at the United States National Library of Medicine, Clinical Trials Gov was performed 2022-03-01 with the clinical trials identifier: NCT05289284A [187]. The registration in Clin Trials Gov includes a larger prospective project to investigate the SLS predictive validity for future injuries.

As **Study I** was a systematic review and meta-analysis, no ethical approval was needed. However, when conducting such a study it is of equal importance to consider the ethics as there might be a chance to include studies that have overlooked the ethical principles such as autonomy, beneficence, and justice. Therefore, were all included studies checked for ethical approval and informed consent to participate, and all articles met this requirement.

Autonomy of all included subjects in **Study II-IV** was ensured by written and verbal informed consent to participate. **Study IV** required some further ethical considerations compared to **Study I-III** as these studies were based on interventions with biomechanical tests and questionnaires on some psychosocial factors. Possible risks with the biomechanical tests were minimised with proper warming up and precise instructions on how to perform the tests, and when to abort a test. These tests are also used in regular clinical assessment why they are not new or specific for this research. All tests and instructions were given by an experienced physiotherapist (JR). Given that some questions in the psychosocial questionnaire might have been perceived as upsetting by some participants and that the sample comprised of younger female soccer players (> 16 years) actions were taken to be able to support the participants if they wished to talk about the questions after filling in the questionnaires. The purpose of these questionnaires was explicitly explained during the initial information procedure before all testing and the participants were able to ask questions at the information occasion or afterward, if they wished to do so alone. All in all, it is therefore considered that the benefits of **study II-IV** outweigh the risks that could be predicted, which were judged to be small. As no control group was used in **Study I-IV** justice was not a great matter as all participants performed the same tests and answered the same questionnaires.

4.3 Study I

The systematic review and meta-analysis in Study I were conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines [188, 189], and pre-registered in the International Prospective Register of Systematic Reviews (PROSPERO) with registration number CRD42018077822.

4.3.1 Literature search and study selection

The systematic search was conducted in CINAHL, Cochrane Library, Embase, Medline (OVID) and Web of Science databases from inception up until 29 November 2018. The following search concepts were used: SLS, reproducibility of results and observer variation. The searches were performed with the help of two librarians at the Karolinska Institutet University Library.

4.3.2 Data extraction

Data was extracted by two independent researchers (JR, ERB) and any disagreement was solved by a consensus discussion with a third researcher (WG). Extracted information was summarised in tables including study name, number of participants, age/gender, activity level, musculoskeletal disorders, number of examiners and their level of experience, method/test, assessment criteria, and outcome/statistics.

4.3.3 Methodological quality

The Quality Appraisal of Reliability Studies Checklist (QAREL) [190] was used to assess the methodological quality of the included studies. The assessment was conducted in the same manner as for data extraction with two independent researchers (JR, ERB) and a third researcher (WG) if there was any disagreement. QAREL is a reliable instrument specially designed to assess the methodological quality of studies of diagnostic reliability [190, 191] and it consists of 11 items that cover seven principles. For the interpretation of the methodological quality, each item should be considered individually [190].

4.4 Study II

4.4.1 Study design

In Study II, the marker-less 3D motion analysis system Qinematic™ was evaluated regarding its capability to assess the SLS reliably and validly. For the test-retest reliability, subjects performed one session of Qinematic™ on two different occasions six to seven days apart. Afterwards, the construct validity was studied by comparing the quantitative data (degrees) from Qinematic™ with video-recorded SLS that were visually assessed as knee-over-foot or knee-medial-to-foot position (pass/fail), where the visually assessed SLS was used as a reference standard. In total, 37 healthy and active persons (27 women, 10 men) were recruited via verbal announcements and information posters at the Karolinska Institutet in Stockholm.

4.4.2 Data collection

The setup for the data collection in Study II and III were in common. All tests were performed at the movement laboratory at Karolinska Institutet from 21 March to 11 May 2017 and handled by two researchers (JR, WG). Before the tests all subjects filled in a questionnaire regarding demographics and background data. The Qinematic™ system works as a semi-automated service that gives the subjects oral and visual instructions while standing in front of the kinetic camera and a computer touch screen. The test in front of the Qinematic™ system took approximately 10 minutes. Parallel to the Qinematic™ system, two digital video cameras were orthogonally placed in relation to the subject and recorded all trials in the sagittal and frontal planes. The subjects wore tight shorts/tights, a sports top/singlet, or a T-shirt.

4.4.3 Procedures

The standard Qinematic™ movement screening procedure includes seven different functional tasks, but only the Squats (SLS) on the left and right leg were of interest for the purpose of Study II. Our research group studied the reliability and validity of posture, balance and side-bending using Qinematic™ in an earlier study [192]. The visually assessed SLS that was used as a reference standard against the Qinematic™ (index test) was collected in parallel to the Qinematic™ procedure by the two orthogonally placed video cameras. The assessment was dichotomised as having a knee-over-foot position (pass) or a knee-medial-to-foot position (fail) [129], the assessment was done by two researchers (JR, ERB). Before any assessment, eleven randomly chosen video recordings were used to reach consensus a on how to assess the test. Finally, all video recordings were individually assessed by the two raters (JR, ERB), and consensus was reached in all cases without consulting a third party (WG).

4.5 Study III

4.5.1 Study design

Based on previous scientific findings on the reliability of the SLS in Study I, a less complex and well-defined multi-segmental SLS was evolved and evaluated for its reliability in Study III. To improve the methodological standardisation of the study, the study followed the checklist of QAREL [190]. The recruitment of study subjects was the same for Study II and III.

4.5.2 Data collection

The setup for the data collection was the same for Study II and III, see item 4.4.2. The SLS was recorded by two orthogonally placed video cameras. Before the subjects performed the SLS they were first instructed by one researcher (JR) on how to perform the test, and they were at the same time allowed to practice the test. For the actual test, the subjects followed a pre-recorded video clip with precise instructions, the test was performed three times on each leg and started always with the left leg.

4.5.3 Rating procedure

Two experienced physiotherapists with more than 20 years of work experience who used specific movement quality tests daily were invited as raters together with two novice physiotherapists who had been working for about four years and had no greater previous experience in assessing specific movement quality tests. To learn how to assess the test, ten video recordings were sent to each one of the raters together with the rating instructions. One week later, the four raters met for a two-hour learning session, conducted by one researcher (JR), where the rating criteria were discussed and practised. One day after the education, the raters received 65 new video recordings which they were instructed to assess individually. Those ratings were used for the purpose of the study. For intra-rater reliability, the same video recordings (but with another random order) were sent to each one of the raters 10 to 14 days later. The raters were instructed to assess the movement deviations from the vertical alignment of the four body segments foot, knee, pelvic, and trunk. This was during three consecutive squats where all body segments were assessed at the same time. A deviation of a segment was scored as one point. No deviation was scored as zero points. The total score of the SLS ranges between 0 to 4 points.

4.6 Study IV

4.6.1 Study design

Study IV is a cross-sectional study and part of a longitudinal project which, among other things, aims to investigate the predictive value of the SLS in female soccer players. The study is preregistered at the United States National Library of Medicine, Clinical Trials Gov (clinical trials identifier: NCT05289284A). Twenty female soccer teams from the three highest divisions in the Swedish Soccer League (divisions 1-3) were invited to participate, 18 teams accepted the invitation, and 269 players were screened for demographics, previous injuries, biomechanical- and psychosocial factors.

4.6.2 Data collection

Before the data collection, two pilot tests were conducted for the measurement of hip strength and ankle dorsiflexion, one including 17 subjects and one including 12 subjects. All soccer players were screened from January to February 2022 by one researcher (JR), and the biomechanical tests and one questionnaire on previous injuries and demographics were performed at the local club. Questions on psychosocial factors were collected via a web-based survey (SurveyMonkey©) which was sent to the subjects directly after the screening at the local club.

4.6.3 The dependent variable

The SLS that was evaluated in Study III was used for the purpose of Study IV as a dependent variable but with a modification regarding the total score. In Study III, the total score ranged between 0 to 4 points but for Study IV it was changed to 0 to 1 point (pass/fail). That is, a total score of zero points if the subject passed the test (no fail in any segments), or one point if the subject failed the test (fail in one or more segments). No change was made for the assessment and rating of the separate segments (foot, knee, pelvic and trunk) which were dichotomously assessed as a pass (0 points) or fail (1 point).

4.6.4 The independent variables

4.6.4.1 Demographics

Demographic data on the participant's age, height, weight, soccer division, and leg dominance were collected with the questionnaire. The dominant leg was defined as the preferred kicking leg [193, 194] and the other leg was defined as the non-dominant leg.

4.6.4.2 Previous injuries

The subjects could register three kinds of injuries, a time-loss injury, a severe injury and an injury problem. Injury data was collected for the whole body, but only injuries located in the head, lower belly, lower back, pelvic or lower extremities were used for the analyses. For the questions asked about injury problems, a modified version from the Oslo Sports Trauma Research Centre Overuse Injury Questionnaire [195, 196] was used, this questionnaire is especially designed to capture overuse injuries. All questions regarding injuries were collected with questionnaires.

4.6.4.3 *Ankle dorsiflexion*

For the measurement of ankle dorsiflexion, the weight-bearing lunge test (WBLT) was used [197, 198]. This method calculates the ankle dorsiflexion (trigonometric angle, TA) by using a simple trigonometric function (TA=90- arctangent (GK/HW)). It is calculated from the two distances heel to the wall (HW distance) and ground to the knee (GK distance) where the knee is defined as the anterosuperior edge of the patellae. To measure the HW- and GK distance, two 70-centimetre (cm) tape measures were fixed perpendicular to each other on the ground and wall. Before the test, the subjects were informed of the standardised instructions and provided with a demonstration by the test leader (JR). The subjects were asked to place the foot to be assessed on the tape so that digitorum 1 and the centre of the heel were aligned on the tape, the non-tested foot was instructed to be placed behind them in a "fencing position". The position of the non-tested leg for Study IV differs from previous studies [197, 198] thus pilot testing on 29 persons showed that the subjects could increase their range of ankle dorsiflexion by putting the non-tested foot in different positions. The subjects were allowed to try to reach the wall three times before the final measure was done. In previous articles [197, 198] a mean value of three attempts was used. For Study IV only the maximum distance was recorded due to the practical time frame for screening a soccer team. The WBLT test has in previous studies shown an "almost perfect" intra-rater reliability, a Standard Error of Measurement of 0.6°-1.18° and a Minimal Detectable Change (MDC) of 1.7°-3.26° [197, 198].

4.6.4.4 *Hip strength*

The combined maximal hip abductor and external rotation strength was measured with a handheld dynamometer (MicroFET2^{TM} wireless, Hoggan Scientific, LLC. USA) with the player performing an isometric clamshell (CLAM) [199]. The subjects were placed side-lying on a treatment bunk with the hips in 45° flexion and 0° abduction/adduction, and the knees in 90° flexion. This position was fixated with a firm yoga block between the knees and two traction belts, one around the ankles and one around the knees. Before fixation, the distance between the proximal greater trochanter and the proximal lateral epicondyle of the femur was measured on each side for the tested leg. The handheld dynamometer was placed just proximal to the lateral epicondyle of the upper knee on the tested leg and was fixated under the traction belt. Three maximal efforts were made for 5 seconds with about 15-30 seconds rest in between the three sets, and a mean value for the three maximal efforts was calculated. The handheld dynamometer values were measured in Newton (N), and these were multiplied by the length of the femur (m) to calculate maximal peak torque values (Nm). Torque was then body size normalised by the subject's weight (kg) and height (m) [200, 201] and multiplied by 100 (Nm/(kg*m)*100). The CLAM test has an "almost perfect" test-retest reliability (ICC=0.97, 95 % CI 0.94-0.99) and good validity (Pearson's correlation coefficient=0.84) [199].

4.6.4.5 Perceived Stress Scale-14 items

A Swedish version of the Perceived Stress Scale-14 items **(**PSS-14) [202] was used to measure perceived stress in Study IV, PSS-14 was originally developed by Cohen et al. [203]. The scale is recommended to be used when screening people with and without known stress-related disorders [203], it contains 14 items and the total score ranges from 0-56, where 56 represents high stress [203]. The Swedish version of PSS-14 has shown satisfactory psychometric properties [202].

4.6.4.6 Pittsburgh Sleep Quality Index

Sleep quality was measured with the Pittsburgh Sleep Quality Index (PSQI) which has been widely used by researchers and clinicians in different settings, populations, and languages [204]. It aims to measure sleep in different dimensions, but it may also be used as a simple screening measure to identify good and poor sleepers [204, 205]. The PSQI contains seven different components, 19 items and a total global score that ranges from 0-21, where 21 represents poor sleep [204, 205]. The cut-off for poor sleepers has been set to six points with a sensitivity of 89.6%, a specificity of 86.5% [205] and an area under the curve of 0.999 [204]. A Swedish unpublished translation of the PSQI exists which has been used in clinical settings and medical research and development.

4.6.4.7 Generalized Anxiety Disorder-7 items

Anxiety was measured with the Generalized Anxiety Disorder-7 items (GAD-7) scale which has shown good reliability and validity [206]. It contains seven items and ranges from 0-21, where 21 represents high anxiety [206]. A score of 10 or greater represents a cut-off for identifying subjects with GAD. Cut-off points of 5, 10 and 15 have been interpreted as mild, moderate, and severe levels of anxiety [206]. The Swedish version was used [207].

4.6.4.8 Athletic Fear Avoidance Questionnaire

Fear of avoidance was measured with the Athletic Fear Avoidance Questionnaire (AFAQ) which has shown good psychometric properties [208]. The scale measures sport-injury-related fear avoidance in athletes and could be used to identify potential psychological barriers, for example to rehabilitation [208]. It contains 10 items and ranges from 10-50, where 50 represents a high fear of avoidance. A Swedish unpublished version was used for Study IV and the test-retest reliability was found adequate ($ICC_{2.1}$ 0.74) (unpublished data).

4.7 Data management

4.7.1 Study I

To deal with the problem of including multiple data from the same study in the meta-analysis three choices were made to get conformity in the data management. Firstly, when more than one reliability data was presented for the same rating, a mean value was calculated (e.g., dominant/non-dominant leg). Secondly, if different assessment methods were presented for the same test the assessment method that corresponded most to the other included studies was chosen. Thirdly when different types of kappa were presented, plain kappa was chosen in front of other measurement units (e.g., weighted kappa).

4.7.2 Study II

In Study II, the Qinematic™ data were used for both the reliability and validity studies, but for the validity study, only the SLS data for the "way down" was used as the "way up" showed poor reliability and a significant difference between the test occasions. Furthermore, data was excluded when the Qinematic™ ordered the subjects to perform an easy form ("easy mode") of the SLS, a less difficult SLS with the non-tested leg in the ground for balance.

For construct validity, Qinematic™ data was dichotomised into steps of two degrees, up to 20 degrees, of knee medial displacement, this data was then compared with the visually assessed SLS. The purpose was to find the Qinematic™ cut-off point that best correlated to the visually assessed SLS. A further detailed description of variables and data management can be read in Study II [209].

4.7.3 Study IV

The choice for the independent variables in Study IV was based on clinical experience and previous research [210]. However, some of the independent variables were for practical or statistical reasons modified. The variable age was due to the nonlinearity of the log odds [210, 211] divided into three categories (16-19 yr., 20-24 yr., and 25-39 yr.), the variable hip strength was for practical reasons multiplied with 100 and therefore expressed as Nm/(kg*m)*100, PSQI was dichotomised according to previous work (≤5 good sleepers, ≥6 poor sleepers) [204, 205] and GAD-7 was divided into three categories (no anxiety, ≥5 mild anxiety, and ≥10 moderate/severe anxiety) instead of the originally four categories. This was due to numerical problems [211] in the statistical calculations, meaning that less than five outcomes for the severe anxiety category were obtained. Regarding injuries, a time-loss injury previous 4 weeks and an injury problem previous 4 weeks were chosen to be included in the analysis before other injuries, this was due to the clinical aspect of knowing how a recent time-loss injury or injury problem would affect the outcome of the SLS. For the SLS for all segments (the total score), the variable a time-loss injury previous 4 weeks was replaced by the variable history of time-loss injuries season 2021, due to numerical problems in the statistical calculation [211].

4.8 Statistical analyses

All statistics used in this thesis, both descriptive and inferential are summarised in **Table 3.**

Table 3. Statistical analyses used in this thesis.

Statistics	**Study I**	**Study II**	**Study III**	**Study IV**	**Thesis**[a]
Descriptive:					
Frequency (n) and percentage (%)	X	X	X	X	
Min.-Max.	X			X	
Median		X		X	
Interquartile range (Q1 to Q3)		X			
Mean		X	X		
Standard deviation		X	X		
Inferential:					
Meta-analysis with Q-test and I^2	X				
Wilcoxon signed-rank test		X		X	
Wilcoxon rank sum test				X	
McNemar´s test				X	
Percent Agreement		X	X		
Cohen´s kappa (95% CI)		X	X		
$\text{Kappa}_{\text{max}}$			X		
Generalised kappa (95% CI)			X		
Prevalence index			X		
Bias index			X		
PABAK (95% CI)			X		
ICC 3.1 (95% CI)		X			
Spearman correlation coefficient		X			
Standard Error of Measurement		X			
Smallest Detectable Change		X			
Sensitivity (95% CI)		X			X
Specificity (95% CI)		X			X
Positive Predictive Value (95% CI)		X			X
Negative Predictive Value (95% CI)		X			X
ROC analysis		X			X
Bland and Altman plots		X			
Multivariate logistic regression				X	

[a]**Thesis:** Additional analyses.

Abbreviations: Q1 to Q3= 1st quartile (25%) to 3rd quartile (75%); 95% CI= 95% Confidence Interval; ROC= Receiver Operation Characteristics.

4.8.1 Study I

Two separate meta-analyses for inter-rater and intra-rater reliability were conducted and the reliability estimates ICC, kappa and AC1 together with sample size were extracted from each included study and transformed to Fisher's z scale [212-216]. The transformation was done to account for the non-normal distribution in correlation meta-analyses [212-216] but was then converted back to reliability estimate values after completed calculations, to assist in the interpretation of the results. The between-studies and total between-subgroup effect size heterogeneity were conducted due to expected heterogeneity from the clinical and/or methodological diversity of included studies. This was calculated with the Q test and expressed as I^2 statistics. The effect size was expressed as the pooled agreement of ICC, kappa and AC1 with 95% CI, the critical value to reject H0 was set to $p < 0.05$. Correlational statistics as kappa, ICC and AC1 were interpreted according to Landis and Koch's classification of strength of agreement [217]; < 0.00 = poor; 0.00-0.20 = slight; 0.21-0.40 = fair;0.41-0.60 = moderate; 0.61-0.80 = substantial; and 0.81-1.0 = almost perfect.

4.8.2 Study II

The normality of data was checked by comparing means and medians together with visual analyses of histograms, boxplots, and quantile-quantile (Q-Q) plots. Furthermore, the data was tested for skewness and kurtosis [218]. Not all data was normally distributed of which non-parametric statistics were used.

4.8.2.1 Test-retest reliability

Wilcoxon signed-rank test was used to test for significant differences between the two test occasions, and the level of significance was set to $p < 0.05$. Furthermore, both relative and absolute reliability were calculated. For relative reliability, the Spearman correlation coefficient in addition to the Intraclass Correlation Coefficient (ICC 3.1) was used as not all variables were normally distributed. For ICC 3.1, a two-way mixed effect model, absolute agreement, and single rater/measures were used. Standard error of measurement (SEM) and smallest detectable change (SDC) were used to calculate absolute reliability. The SEM is a measure of how far apart the outcomes of repeated measurements are and what the standard deviation around a single measurement is. SDC is a change in the construct that can be considered real, i.e. a change beyond the measurement error [149, 219]. As those parameters are expressed in the same unit as the original measurement and can be used on an individual level [149, 219], they should be as low as possible. The interpretations of the kappa and ICC estimates were similar to those in Study I. The Spearman correlation coefficient was interpreted as less than 0.3 low correlation, 0.3-0.5 fair correlation, 0.6-0.8 moderately strong correlation, at least 0.8 very strong [220, 221].

4.8.2.2 Construct validity

For construct validity, the visually assessed SLS was used as a reference standard and compared to the measures of Qinamatic™ which were dichotomised into 10 cut-off scores, each of 2 degrees, ranging from 2 to 20 degrees of knee medial displacement. The purpose was to find the Qinematic™ cut-off score that in the best way corresponded to the visually assessed SLS. To do so, three different calculations were conducted: First, the agreement between the visually assessed SLS and Qinematic™ was investigated by percent agreement and kappa. Second, diagnostic accuracy was assessed by calculating the area under the receiver operation characteristic (ROC) curve together with standard error and 95% CI, and third positive predictive value (PPV) and negative predictive value (NPV) were calculated.

The receiver operation characteristic curve is a plot of sensitivity against specificity that summarises the discriminative ability of a test across all cut-offs in one single measure, and as a rule of thumb, the following classification of the area under the curve has been suggested: > 0.9 = high accuracy; 0.7-0.9 = moderate accuracy; 0.5-0.7 = low accuracy; and 0.5 = a chance result [222, 223]. As sensitivity and specificity are two measures that evaluate the accuracy of a screening test in relation to a reference standard where the number of cases or non-cases are known, it is not suitable for clinical situations where the clinician must make inferences about the presence or absence of a disease [222]. Therefore, we also calculated the positive predictive value and negative predictive value as it is more suitable for decision-making in individuals [222, 224] since it is prevalence dependent, and provides information about the probability that a patient is affected or not with a given test result [222, 224].

4.8.3 Study III

Intra- and inter-rater reliability were in Study III investigated with Cohen´s kappa statistics together with percent agreement and 95% CI, both for each separate segment (foot, knee, pelvic and trunk) and as a merged rating for all segments together, the latter described as the variable “all segments”. For inter-rater reliability where multiple raters were compared, a generalised kappa coefficient presented by Fleiss was used [225, 226].

As the magnitude and interpretation of the kappa coefficient can be influenced by factors such as prevalence index (PI) and bias index (BI) we also calculated and presented this together with the prevalence-adjusted bias-adjusted kappa (PABAK) [227]. The PABAK coefficient is relatively uninformative on its own due to the hypothetical situation where no prevalence or bias exists, but when presented in addition to the kappa coefficient it shows the effects of prevalence and bias on the kappa measure [227]. Finally, and as a further help in the interpretation of kappa, the $\text{kappa}_{\text{max}}$ was calculated which displays the maximum value of kappa that could be obtained for a specific set of data [227]. The interpretations of the correlation estimates were similar to those in Study I.

4.8.4 Study IV

The normality of data was checked by comparing means and medians together with visual analyses of histograms, boxplots, and quantile-quantile (Q-Q) plots. Furthermore, the data was tested for skewness and kurtosis [218]. As not all data was normally distributed non-parametric statistics were used. Before any calculation, data concerning the left or right leg were categorised as a non-dominant leg (NDL) or dominant leg (DL). Regarding the outcome of the SLS, McNemar´s test was used to analyse the statistical difference between the dominant leg and the non-dominant leg. For the difference in hip strength and ankle dorsiflexion between the dominant leg and non-dominant leg, the Wilcoxon signed rank test for paired non-normally distributed data was used, and for the difference in hip strength and ankle dorsiflexion within the dominant leg and non-dominant leg, the Wilcoxon rank sum test for unpaired non-normally distributed data was used.

The dependent variable of Study IV was the SLS with a dichotomous outcome (pass/fail). Separately, for the dominant leg and non-dominant leg, two dependent variables were used in the statistical analysis: 1) the total score (pass/fail) of all segments, and 2) the pass/fail score for the separate knee segment. The choice of the independent variables was based on clinical experience and previous research [210]. The following continuous variables were included in the study: ankle dorsiflexion, hip strength, AFAQ, PSS-14, and the following categorical variables were included: age (16-19 yrs., 20-24 yrs., and 25-39 yrs.), soccer division (1-3), a severe injury (yes/no), a time-loss injury season 2021 (yes/no), a time-loss injury (yes/no), an injury problem (yes/no), PSQI (≤5 good sleepers/ ≥6 poor sleepers), and GAD-7 (no anxiety, ≥5 mild anxiety/≥10 moderate/severe anxiety).

All independent variables were tested with univariate logistic regression analyses. Separate models were constructed for the total score and the knee segment for both the dominant leg and non-dominant leg. A backward logistic regression analysis was used for the multivariate analyses that specified the significance level for the removal of eligible independent variables from the model at p ≥0.20. The results were expressed as odds ratio (OR) with a 95% confidence interval (95% CI). A stepwise logistic regression model might be justified when investigating a relatively new outcome, and when the importance of the covariates (independent variables) and their association with the outcome is not well understood [210, 211]. Stepwise regression might then be a fast and effective way to screen a large number of covariates [211]. Therefore, a stepwise logistic regression was chosen in Study IV as the association with most of the independent variables is unknown or has not previously been studied. It is recommended in multiple regression models that for every variable screened for association, there are at least 10 events [228]. However, this rule of thumb should not be applied categorically as other factors could affect the stability of a model [228], as well as there in some cases is evidence for reducing this rule to 5-9 events [229]. The final multivariate models were tested for adequacy by the Hosmer-Lemeshow Goodness of Fit test [230, 231] and by the "linktest" procedure in the statistical software program STATA 15.1. The remaining variables were also tested for possible interactions in an exploratory purpose. To ascertain that the basic assumptions for conducting logistic regression were met, data were checked for numerical limits, linearity of the log odds, multicollinearity, sample size, data independence, homogeneity, outlying and influential points [210, 211, 232, 233]. As the aim of the present study was to find associations with the outcome of the SLS, no adjustment for multiple comparisons (e.g., Bonferroni) was made as such an approach may inflate the risk of type II errors which makes it more difficult to identify associations [234].

4.8.5 Additional analysis

In the additional analysis sensitivity, specificity, positive predictive value, negative predictive value, and a non-parametric receiver operation characteristic analysis were used to investigate the SLS ability to discriminate subjects with a previous or present injury from those with no previous or present injury (discriminative validity and diagnostic accuracy). Data from Study IV were used and four different types of previous injuries and two different types of present injuries were analysed: "a severe injury", "a severe knee injury", "a present time-loss injury", "a time-loss injury previous four weeks", "a present injury problem" and "an injury problem previous four weeks". All separate segments of the SLS (foot, knee, pelvic and trunk) were analysed individually for the dominant leg and non-dominant leg. Furthermore, two different total scores were analysed; the dichotomous total score (pass/fail) used in Study IV and the total score with four categories (0-4 fail) which was used in Study III.

In Study I, the statistical analyses were completed using comprehensive meta-analysis V.3 [216], in Study II-IV and the additional analysis STATA version 15.1 was used for all statistical analyses. In Study III STATA was also used with the extension of the "kappaetc" command which handles all types of kappa presented in the article [235]. Microsoft Office Excel version 16 for Windows 10 was also used in the different studies. In Study II to plot the Bland-Altman plots, in Study III to calculate prevalence index (PI) and bias index (BI) and in Study IV to collect and organise the data before importing it to STATA. In Study III $\text{kappa}_{\text{max}}$ was calculated via a web calculator [236].

5 Results

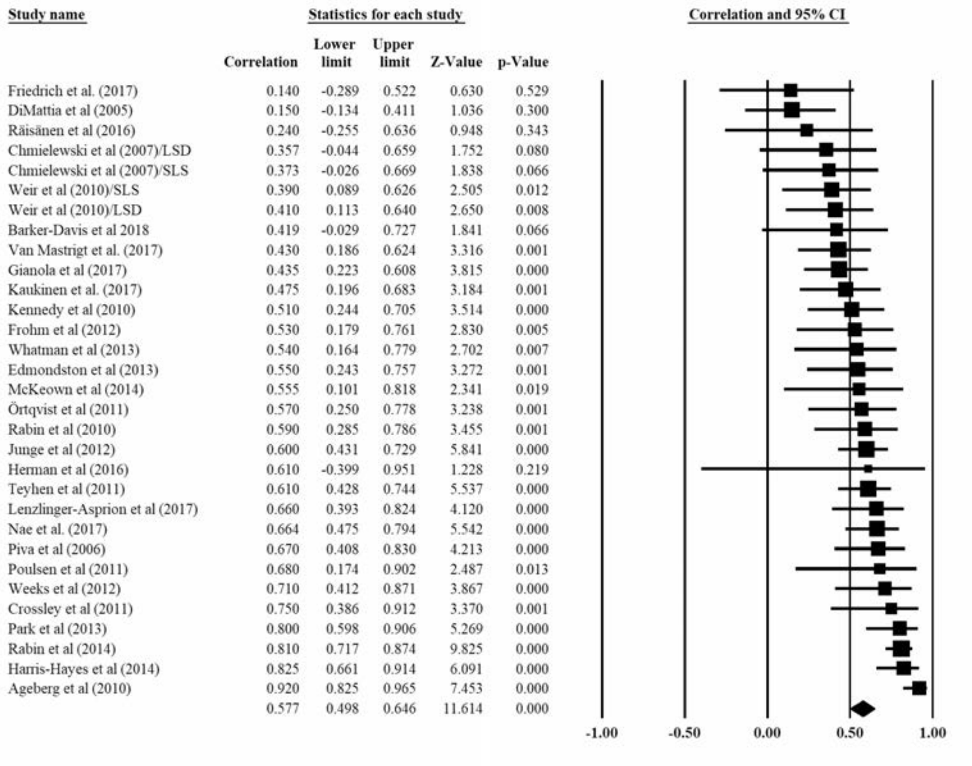
In this section, the overall results for Study I-IV and the additional analyses are presented.

5.1 Study I

The 31 included studies presented a variety of different SLS regarding their name, performance, and assessment criteria. In total, 29 studies were included for the synthesis of inter-rater reliability and 17 studies were included for the synthesis of intra-rater reliability.

5.1.1 Pooled agreement/synthesis of results

The pooled agreement for inter-rater reliability was 0.58 (95% CI 0.50-0.65), indicating a "moderate" agreement (**Figure 3**). The test for heterogeneity was significant ($Q=86.20$, $df=30$, $p<0.001$) and the I^2 statistics reported that 65% of the variability was attributed to heterogeneity.



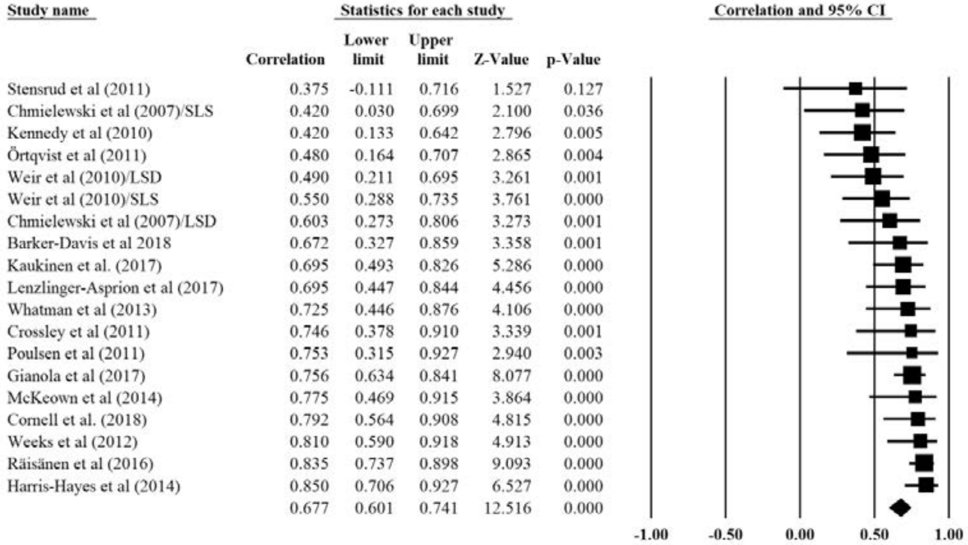
Study name	Statistics for each study				
	Correlation	Lower limit	Upper limit	Z-Value	p-Value
Friedrich et al. (2017)	0.140	-0.289	0.522	0.630	0.529
DiMattia et al (2005)	0.150	-0.134	0.411	1.036	0.300
Räisänen et al (2016)	0.240	-0.255	0.636	0.948	0.343
Chmielewski et al (2007)/LSD	0.357	-0.044	0.659	1.752	0.080
Chmielewski et al (2007)/SLS	0.373	-0.026	0.669	1.838	0.066
Weir et al (2010)/SLS	0.390	0.089	0.626	2.505	0.012
Weir et al (2010)/LSD	0.410	0.113	0.640	2.650	0.008
Barker-Davis et al 2018	0.419	-0.029	0.727	1.841	0.066
Van Mastrigt et al. (2017)	0.430	0.186	0.624	3.316	0.001
Gianola et al (2017)	0.435	0.223	0.608	3.815	0.000
Kaukinen et al. (2017)	0.475	0.196	0.683	3.184	0.001
Kennedy et al (2010)	0.510	0.244	0.705	3.514	0.000
Frohm et al (2012)	0.530	0.179	0.761	2.830	0.005
Whatman et al (2013)	0.540	0.164	0.779	2.702	0.007
Edmondston et al (2013)	0.550	0.243	0.757	3.272	0.001
McKeown et al (2014)	0.555	0.101	0.818	2.341	0.019
Örtqvist et al (2011)	0.570	0.250	0.778	3.238	0.001
Rabin et al (2010)	0.590	0.285	0.786	3.455	0.001
Junge et al (2012)	0.600	0.431	0.729	5.841	0.000
Herman et al (2016)	0.610	-0.399	0.951	1.228	0.219
Teyhen et al (2011)	0.610	0.428	0.744	5.537	0.000
Lenzlinger-Asprion et al (2017)	0.660	0.393	0.824	4.120	0.000
Nae et al. (2017)	0.664	0.475	0.794	5.542	0.000
Piva et al (2006)	0.670	0.408	0.830	4.213	0.000
Poulsen et al (2011)	0.680	0.174	0.902	2.487	0.013
Weeks et al (2012)	0.710	0.412	0.871	3.867	0.000
Crossley et al (2011)	0.750	0.386	0.912	3.370	0.001
Park et al (2013)	0.800	0.598	0.906	5.269	0.000
Rabin et al (2014)	0.810	0.717	0.874	9.825	0.000
Harris-Hayes et al (2014)	0.825	0.661	0.914	6.091	0.000
Ageberg et al (2010)	0.920	0.825	0.965	7.453	0.000
	0.577	0.498	0.646	11.614	0.000

Abbreviations: CI=Confidence Interval

Heterogeneity: Q=86.20, *df*=30, $p<0.001$; I^2=65%

Figure 3. Forest plot and the pooled agreement coefficient of studies on the agreement coefficient (ICC, kappa, AC1) for inter-rater reliability of the Single Leg Squat in a random effect model. **Nota Bene:** This Figure was previously published in Study I; no changes have been made, the article is licensed under a Creative Commons Attributions 4.0 International License, https://creativecommons.org/licenses/by/4.0/.

The pooled agreement for intra-rater reliability was 0.68 (95% CI 0.60-0.74), indicating a “substantial” agreement, **see Figure 4**. The test for heterogeneity was significant (Q=38.46, df=18, p=0.003) and the I^2 statistics reported that 53% of the variability was attributed to heterogeneity.



Study name	Correlation	Lower limit	Upper limit	Z-Value	p-Value
Stensrud et al (2011)	0.375	-0.111	0.716	1.527	0.127
Chmielewski et al (2007)/SLS	0.420	0.030	0.699	2.100	0.036
Kennedy et al (2010)	0.420	0.133	0.642	2.796	0.005
Örtqvist et al (2011)	0.480	0.164	0.707	2.865	0.004
Weir et al (2010)/LSD	0.490	0.211	0.695	3.261	0.001
Weir et al (2010)/SLS	0.550	0.288	0.735	3.761	0.000
Chmielewski et al (2007)/LSD	0.603	0.273	0.806	3.273	0.001
Barker-Davis et al 2018	0.672	0.327	0.859	3.358	0.001
Kaukinen et al. (2017)	0.695	0.493	0.826	5.286	0.000
Lenzlinger-Asprion et al (2017)	0.695	0.447	0.844	4.456	0.000
Whatman et al (2013)	0.725	0.446	0.876	4.106	0.000
Crossley et al (2011)	0.746	0.378	0.910	3.339	0.001
Poulsen et al (2011)	0.753	0.315	0.927	2.940	0.003
Gianola et al (2017)	0.756	0.634	0.841	8.077	0.000
McKeown et al (2014)	0.775	0.469	0.915	3.864	0.000
Cornell et al. (2018)	0.792	0.564	0.908	4.815	0.000
Weeks et al (2012)	0.810	0.590	0.918	4.913	0.000
Räisänen et al (2016)	0.835	0.737	0.898	9.093	0.000
Harris-Hayes et al (2014)	0.850	0.706	0.927	6.527	0.000
	0.677	0.601	0.741	12.516	0.000

Abbreviations: CI=Confidence Interval
Heterogeneity: Q=38.46, *df*=18, p<0.003; I^2=53%

Figure 4. Forest plot and the pooled agreement coefficient of studies on the agreement coefficient (ICC, kappa, AC1) for intra-rater reliability of the Single Leg Squat in a random effect model. **Nota Bene:** This Figure was previously published in Study I; no changes have been made, the article is licensed under a Creative Commons Attributions 4.0 International License, https://creativecommons.org/licenses/by/4.0/.

5.1.2 Subgroup analysis

When comparing the pooled inter-rater agreement for the approach of assessing one/two segments at a time **(**0.62, 95% CI 0.44-0.76) with the assessment of multiple segments (0.57, 95% CI 0.47-0.65) there were no significant differences (p=0.56) shown in the subgroup analysis. Nor were there any significant differences in the pooled agreement for intra-rater reliability (0.72, 95% CI 0.56-0.82 vs. 0.66, 95% CI 0.58-0.74, p=0.53).

When comparing the pooled inter-rater agreement for the approach of assessing ≤3-point rating scales (0.64, 95% CI 0.56-0.71) with the assessment of ≥4-point rating scales (0.47, 95% CI 0.33-0.58) subgroup analysis showed a significant difference between the two approaches (p=0.016). No significant difference was found for the pooled agreement of intra-rater reliability (0.71, 95% CI 0.62-0.77 vs. 0.60, 95% CI 0.44-0.73, p=0.18).

5.1.3 Methodological quality

Seven studies [130, 131, 157, 237-240] did not fulfil QAREL item 11, and one study [241] did not fulfil item 8. Sensitivity analysis on the importance of study quality showed that the pooled agreement for inter-rater reliability slightly increased to 0.60 (95% CI 0.51-0.67), while the intra-rater reliability decreased to 0.62 (95% CI 0.53-0.71) when those eight studies were eliminated from the meta-analyses. Furthermore, all studies were assessed as "uncertain" for one or more items.

5.2 Study II

5.2.1 Test-retest reliability

Altogether, 37 included subjects produced 296 Qinematic™ measures as the right and left leg were measured during the "way up" and the "way down" for both test occasions. After exclusion due to various reasons, 85% of the data were available. For the test-retest reliability, both relative and absolute reliability were measured, and the results are summarised in **Table 4.** One of the variables, "left knee up", showed a significant difference between the two test occasions (T1 6.34°, T2 0.66°, p =0.013) while the other three (left knee down, right knee up, and right knee down) did not. Those three variables reached "substantial reliability", with ICCs ranging from 0.64 to 0.69 and "moderately strong" (r=0.61-0.68) for the Spearman correlation coefficient. The Standard error of measurement and smallest detectable change were calculated as absolute reliability (**see Table 4**).

Table 4. Results from the test-retest reliability study.

	Data[1]			Differences[2]	Relative reliability[3]		Absolut Reliability		
	T1 Median (Q1, Q3)	**T2** Median (Q1, Q3)	**GRAND** Median (Q1, Q3)	**T1 vs. T2** p-value	**Spearman (r)** p-value	**ICC (3.1)** (95% CI)	**SEM**[4]	**SDC**[5]	Mean difference between T2 and T1 with 95% confidence interval (95% CI)[6]
Left knee up (n=64)	-6.34 (-16.93, 6.26)	0.66 (-8.40, 10.63)	-2.53 (-11.53, 9.20)	**0.013***	0.53 (0.002)	0.50 (0.17 to 0 .72)	10.66	29.55	**6.89 (1.98 to 11.81) ***
Left knee down (n=64)	1.07 (-6.50, 19.42)	0.80 (-9.50, 13.42)	1.07 (-6.76, 15.48)	0.5496	0.68 (<0.001)	0.69 (0.45 to 0.83)	9.85	27.30	-2.15 (-7.19 to 2.89)
Right knee up (n=61)	-6.48 (-16.59, 7.09)	-1.95 (-10.35, 7.87)	-3.20 (-12.06, 7.09)	0.2059	0.61 (<0.001)	0.69 (0.45 to 0.84)	9.04	25.06	2.83 (-1.91 to 7.56)
Right knee down (n=62)	0.59 (-11.71, 13.52)	2.75 (-10.04, 12.59)	2.27 (-10.04, 12.81)	0.5967	0.61 (<0.001)	0.64 (0.38 to 0.81)	9.09	25.20	1.24 (-3.53 to 6.00)

Nota Bene: Table 4 was previously published in Study II, no changes have been made, the article is licensed under a Creative Commons Attributions 4.0 International License, https://creativecommons.org/licenses/by/4.0/.

*Denotes a statistically significant change; n=denotes the number of measurements done by Qinematic™ for each variable; A negative value (-) denotes a medial displacement for the left knee and a lateral displacement for the right knee, contrariwise a positive value (+) denotes a medial displacement of the right knee and a lateral displacement for the left knee.

[1] Test occasion 1=T1, Test occasion 2=T2, Q1: 1st quartile (25%), Q3: 3rd (75%). Grand Median = median of all measures from test occasions 1 and 2. All data are in degrees.

[2] Wilcoxon signed-rank test (paired test). A p-value of $p<0.05$ was considered to be statistically significant*.

[3] Reliability based on a two-way mixed effect model, calculating the absolute agreement, based on single ratings. ICC (3.1):

$$ICC_{agreement} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_o^2 + \sigma_{residual}^2}, \sigma_{error}^2 = \sigma_o^2 + \sigma_{residual}^2$$

[4] SEM: Standard error of the measurement:

$$SEM_{agreement} = \sqrt{}(\sigma_o^2 + \sigma_{residual}^2)$$

[5] SDC: Smallest detectable change: $SDC = \pm 1.96 * \sqrt{2} * SEM_{agreement}$

[6] Mean difference between the two test occasions was calculated together with the 95% confidence interval (95% CI). *A 95% CI that does not include zero indicates a systematic change in the mean between T1 and T2.

5.2.2 Construct validity

For the validity study, the 37 subjects produced 148 video recordings and after exclusion due to various reasons 76% of the data were available. As there was a significant difference between test occasions one and two for the left leg on the "way up" (also confirmed as a systematic difference with Bland Altman plots), this movement was not used in the validity study. In other words, only the way down for the right and left leg was evaluated.

The largest area under the curve was reported when using a cut-off at 6° of medial displacement and showed a measure of 0.82 (standard error 0.04, 95% CI 0.74-0.90), which indicates a "moderate" accuracy [222, 223]. At the same cut-off point, the positive predictive value was 0.58 (95% CI 0.47-0.68), and the negative predictive value was 0.94 (95% CI 0.86-0.98).

5.3 Study III

Due to poor video quality, three of the 37 included subjects were excluded and a further three subjects could only be assessed for one leg. Hence, in total 65 video recordings and 34 test persons (24 women, 10 men) were included in the study. The test persons had a mean (±SD) age of 34 (12) years and about 80% of those were physically active two days or more per week.

5.3.1 Inter- and intra-rater reliability

The inter-rater and intra-rater reliability for each rater and each specific segment, together with a merged rating for all raters (rater 1-4) and all segments together are presented in **Table 5** and **Table 6.** For all raters together (rater 1-4), the variable all segments for inter-rater reliability obtained a generalised kappa coefficient of "moderate" agreement 0.52 (95% CI 0.43-0.61), while PABAK reached "substantial" agreement (0.70, 95% CI 0.65-0.76). For intra-rater reliability and the variable all segments, an overall average kappa was calculated for all raters (rater 1-4) which reached an "almost perfect" agreement ($\kappa = 0.82$, 95% CI 0.77-0.86), no important difference was seen between kappa and PABAK.

Table 5. Inter-rater reliability for experienced raters with >20 years of clinical experience and novice raters with ≤4 years of clinical experience.

Raters	PA[a]	Kappa[b] (CI 95%)	Kappa$_{max}$[c]	PI[d]	BI[e]	PABAK[f] (CI 95%)
Experienced Rater 1 vs. Rater 2						
Foot	1.0	1.00 (1.00-1.00)	1.0	0.91	0	1.00 (1.00-1.00)
Knee	0.71	0.42 (0.21-0.64)	0.73	0.09	-0.14	0.42 (0.19-0.64)
Pelvis	0.77	0.44 (0.22-0.66)	0.52	0.46	-0.20	0.54 (0.33-0.75)
Trunk	0.86	0.63 (0.40-0.85)	0.71	0.52	-0.11	0.72 (0.55-0.90)
All segments[g]	0.84	0.57 (0.46-0.68)	0.71	0.50	-0.11	0.67 (0.58-0.76)
Novice Rater 3 vs. Rater 4						
Foot	0.99	0.66 (0.02-1.00)	0.66	0.95	0.02	0.97 (0.91-1.00)
Knee	0.88	0.41 (0.10-0.72)	0.88	0.69	-0.03	0.69 (0.51-0.87)
Pelvis	0.88	0.44 (0.12-0.76)	0.58	0.75	0.09	0.75 (0.60-0.92)
Trunk	0.89	0.68 (0.46-0.90)	0.68	0.58	0.11	0.79 (0.63-0.94)
All segments[g]	0.90	0.55 (0.40-0.70)	0.79	0.75	0.05	0.80 (0.73-0.87)
All raters Rater 1-4	**PA[a]**	**Generalised kappa[h] (CI 95%)**				**PABAK[f] (CI 95%)**
All segments[g]	0.85	0.52 (0.43-0.61)				0.70 (0.65-0.76)

Nota Bene: This Table was previously published in Study III, no changes have been made, the article is licensed under a Creative Commons Attributions 4.0 International License, https://creativecommons.org/licenses/by/4.0/.

[a] **PA:** Percent agreement

[b] **Kappa:** Cohen's kappa, calculated by; $\kappa = \frac{P_o - P_c}{1 - P_c}$

Where; P_o (observed agreement) $= \frac{a+d}{n}$ and P_c (chance agreement) $= \frac{\left(\frac{f_1 x g_1}{n}\right)+\left(\frac{f_2 x g_2}{n}\right)}{n}$

[c] **Kappa$_{max}$:** Is calculated so that the proportions of positive and negative judgements by each rater (i.e., the marginal totals) are taken as fixed, and the distribution of paired ratings (i.e. the cell frequency a,b,c and d) is adjusted to represent the greatest possible agreement. That will say, the maximum possible agreement for either the presence or absence of the disease is the smaller of the marginal totals in each case [227].

[d] **PI:** Prevalence index, calculated by; $PI = \frac{a-d}{n}$

[e] **BI:** Bias index, calculated by; $BI = \frac{b-c}{n}$

[f] **PABAK:** Prevalence-adjusted bias-adjusted kappa, calculated by; $PABAK = 2P_0 - 1$

[g] **All segments:** Denotes a merged kappa coefficient for the inter-rater reliability of each of the segments together (foot, knee, pelvis and trunk).

[h] **Generalised kappa:** A generalisation of Scott´s pi presented by Fleiss to calculate the inter-rater reliability of multiple raters [225, 235].

Table 6. Intra-rater reliability for experienced raters with >20 years of clinical experience and novice raters with ≤4 years of clinical experience.

Raters	PA[a]	Kappa[b] (CI 95%)	Kappa$_{max}$[c]	PI[d]	BI[e]	PABAK[f] (CI 95%)
Experienced Rater 1						
Foot	1.0	1.0 (1.00-1.00)	1.0	0.91	0.00	1.00 (1.00-1.00)
Knee	0.99	0.97 (0.91-1.00)	0.97	-0.06	0.02	0.97 (0.91-1.00)
Pelvis	0.94	0.86 (0.73-1.00)	0.86	0.32	-0.06	0.88 (0.76-1.00)
Trunk	0.95	0.89 (0.77-1.00)	0.96	0.40	0.02	0.91 (0.80-1.00)
All segments[g]	0.97	0.93 (0.88-0.98)	0.98	0.39	-0.01	0.94 (0.90-0.98)
Experienced Rater 2						
Foot	1.0	1.0 (1.00-1.00)	1.0	0.91	0.00	1.00 (1.00-1.00)
Knee	0.86	0.71 (0.52-0.89)	0.97	0.25	-0.02	0.72 (0.55-0.90)
Pelvis	0.92	0.74 (0.51-0.96)	0.95	0.65	0.02	0.85 (0.71-0.98)
Trunk	0.99	0.95 (0.85-1.00)	0.95	0.62	0.02	0.97 (0.91-1.00)
All segments[g]	0.94	0.82 (0.73-0.91)	0.99	0.60	0.00	0.89 (0.83-0.94)
Novice Rater 3						
Foot	0.99	0.66 (0.02-1.00)	0.66	0.95	0.02	0.97 (0.91-1.00)
Knee	0.92	0.72 (0.48-0.96)	0.94	0.68	-0.02	0.85 (0.71-0.98)
Pelvis	0.89	0.17 (-0.22-0.55)	0.88	0.86	-0.02	0.79 (0.63-0.94)
Trunk	0.92	0.69 (0.43-0.95)	0.94	0.71	-0.02	0.85 (0.71-0.98)
All segments[g]	0.93	0.62 (0.45-0.78)	0.96	0.80	0.01	0.86 (0.80-0.92)
Novice Rater 4						
Foot	0.97	0.48 (-0.16-1.00)	1.0	0.94	0.00	0.94 (0.85-1.00)
Knee	0.91	0.70 (0.47-0.92)	0.70	0.63	0.09	0.82 (0.67-0.96)
Pelvis	0.91	0.69 (0.46-0.93)	0.90	0.63	0.03	0.82 (0.67-0.96)
Trunk	0.92	0.82 (0.66-0.97)	0.82	0.40	0.08	0.85 (0.71-0.98)
All segments[g]	0.93	0.75 (0.64-0.86)	0.83	0.65	0.05	0.85 (0.79-0.92)
Rater 1-4	**PA[a]**	**Overall kappa[h] (CI 95%)**	**Kappa$_{max}$[c]**	**PI[d]**	**BI[e]**	**PABAK[f] (CI 95%)**
All segments[g]	0.94	0.82 (0.77-0.86)	0.97	0.61	0.01	0.89 (0.86-0.91)
Nota Bene: This Table was previously published in Study III, no changes have been made, the article is licensed under a Creative Commons Attributions 4.0 International License, https://creativecommons.org/licenses/by/4.0/.						

[a] **PA:** Percent agreement

[b] **Kappa:** Cohens´ kappa, calculated by; $\kappa = \frac{P_o - P_c}{1 - P_c}$

Where; P_o (observed agreement) $= \frac{a+d}{n}$ and P_c (chance agreement) $= \frac{\left(\frac{f_1 x g_1}{n}\right) + \left(\frac{f_2 x g_2}{n}\right)}{n}$

[c] **Kappa$_{max}$:** Is calculated so that the proportions of positive and negative judgements by each rater (i.e., the marginal totals) are taken as fixed, and the distribution of paired ratings (i.e. the cell frequency a,b,c and d) is adjusted to represent the greatest possible agreement. That will say, the maximum possible agreement for either the presence or absence of the disease is the smaller of the marginal totals in each case [227].

[d] **PI:** Prevalence index, calculated by; $PI = \frac{a-d}{n}$

[e] **BI:** Bias index, calculated by; $BI = \frac{b-c}{n}$

[f] **PABAK:** Prevalence-adjusted bias-adjusted kappa, calculated by; $PABAK = 2P_0 - 1$

[g] **All segments:** Denotes a merged kappa coefficient for the intra-rater reliability of each segment together (foot, knee, pelvis and trunk).

[h] **Overall kappa:** Presents an overall average kappa for the variable all segments for all raters comparing test occasions one and two. Calculated with Cohens´kappa.

5.4 Study IV

A total of 254 players from soccer divisions 1-3 in Sweden were included in the study. Demographics, a previous severe injury, and biomechanical- and psychosocial factors stratified by division are described in **Table 7.**

Table 7: Subject characteristics for the total group and stratified by divisions 1-3.

Characteristics	Total group (n=254)	Division 1 (n=89)	Division 2 (n=51)	Division 3 (n=114)
Age, yr.				
Mdn (min-max)	22 (16-39)	23 (17-38)	23 (16-31)	19 (16-39)
Height, m				
Mdn (min-max)	1.70 (1.52-1.83)	1.71 (1.57-1.82)	1.68 (1.52-1.83)	1.69 (1.55-1.83)
Weight, kg				
Mdn (min-max)	63 (50-85)	64 (55-85)	62.5 (50-78)	63 (50-83)
Ankle dorsiflexion				
Mdn (min-max)				
Dominant leg	45° (32°-56°)	44° (32°-54°)	45° (32°-55°)	45° (36°-56°)
Non-dominant leg	45° (34°-58°)	46° (35°-53°)	44° (34°-53°)	45° (36°-58°)
Hip strength				
Mdn (min-max)				
Dominant leg	96* (48-196)	93 (60-172)	101 (62-196)	96 (48-144)
Non-dominant leg	98* (40-204)	97 (40-160)	102 (63-204)	97 (55-149)
AFAQ[a]				
Mdn (min-max)	23 (10-45)	23 (10-42)	25 (10-42)	21 (20-45)
PSS-14[b]				
Mdn (min-max)	32 (20-42)	31 (22-42)	33 (22-41)	32 (20-39)
PSQI[c]				
Mdn (min-max)	5 (0-15)	4 (0-15)	4 (0-15)	5 (1-14)
GAD-7[d]				
Mdn (min-max)	5 (0-20)	4 (0-20)	6 (0-16)	6 (0-17)
Severe injuries DL				
[e]Knee injuries, n (%)	40 (51%)	9 (39%)	16 (80%)	15 (43%)
[f]Other injuries, n (%)	38 (49%)	14 (61%)	4 (20%)	20 (47%)
Severe injuries NDL				
[e]Knee injuries, n (%)	36 (58%)	15 (52%)	10 (91%)	11 (50%)
[f]Other injuries, n (%)	26 (42%)	14 (48%)	1 (9%)	11 (50%)

[*]Denotes statistically significant differences between groups, p-values at p≤0.05;

[a]**AFAQ**: Athletic Fear Avoidance Questionnaire; [b]**PSS-14**: Perceived Stress Scale 14-item; [c]**PSQI**: Pittsburgh Sleep Quality Index; [d]**GAD-7:** Generalized Anxiety Disorder 7-item scale**;** [e]**Knee injuries;** contains fractures, ligament- and overuse injuries expressed in total numbers and percentage; [f]**Other injuries;** contains all other injuries in the lower back and lower extremity except for the knee, expressed in total numbers and percentage. **n:** denotes the number of subjects in the total group and each division; **Mdn:** median; **yr.**=years; **m**=metres; **kg**=kilograms; **DL**=dominant leg; **NDL**=non-dominant leg.

5.4.1 Dominant versus non-dominant leg

In the sample, 231 players were right-footed, and 23 players were left-footed. Generally, there were more cases (fail on the SLS) for the dominant leg compared to the non-dominant leg, both for the total score and the knee segment. For the total score, there were 176 cases on the dominant leg compared to 117 cases for the non-dominant leg (p<0.001) and for the knee segment, there were 102 cases for the dominant leg compared to 70 cases for the non-dominant leg (p<0.001). A significant difference was found between the dominant leg and non-dominant leg regarding hip strength (p=0.03) but not for ankle dorsiflexion (p=0.11), **see Table 7**. There was a difference in hip strength between those who passed the SLS and those who failed the SLS for the total score and the knee segment. Within the non-dominant leg, the difference between those who passed and failed the SLS was significant for the total score (p=0.02) and the knee segment (p=0.01), but not for the dominant leg (total score: p=0.06, knee segment: 0.32).

5.4.2 SLS for all segments, the total score

Regarding the univariate logistic regression analysis for the total score of all segments for the dominant leg, two variables were significantly associated with a failure on the test: ankle dorsiflexion and hip strength. For the non-dominant leg, four variables were significantly associated: soccer division, age, hip strength, and severe injury.

The multivariate models for the total scores are reported in **Table 8** and **Table 9**. The independent variables associated with the outcome of the SLS for the total score differed depending on which leg was tested, except for hip strength that was associated with both the dominant leg and the non-dominant leg (dominant leg: OR 0.99, 95% CI 0.98-0.99, p=0.04, non-dominant leg: OR 0.99, 95% CI 0.97-0.99, p=0.03).

Table 8. Multivariate analysis of failing on the total score for the dominant leg during the SLS.

	SLS for all segments: dominant leg		
Variables	**OR**[a]	**95 % CI**[b]	**p-value**[c]
PSS-14[d]			
(No stress/stress; 0-56)	0.91	0.83-0.98	**0.02***
GAD-7[e]			
No anxiety	1		
Mild anxiety ≥5	1.83	0.96-3.50	0.07
Moderate/severe anxiety ≥10	2.21	0.96-5.07	0.06
Ankle dorsiflexion[f]			
WBLT[f] measured in degrees (TA[f])	0.94	0.87-1.01	0.08
Hip strength[g]			
CLAM[g] measured in Nm/(kg*m) *100	0.99	0.98-0.99	**0.04***

***Denotes statistically significant p-values at p≤0.05.**

[a]**OR:** odds ratio; [b]**95 % CI:** 95 % confidence interval; [c]**p-value:** probability value; [d]**PSS-14:** Perceived Stress Scale 14-item instrument; [e]**GAD-7:** Generalized Anxiety Disorder 7-item scale; [f]**Ankle dorsiflexion:** measured with the Weight Bearing Dorsiflexion Lunge Test (WBLT) and calculated with a trigonometric dorsiflexion angle (TA). [g]**Hip strength:** Side-Lying Clamshell (CLAM).

Table 9. Multivariate analysis of failing on the total score for the non-dominant leg during the SLS.

	SLS for all segments: non-dominant leg		
Variables	**OR**[a]	**95 % CI**[b]	**p-value**[c]
Division			
Div. 1	1		
Div. 2	1.79	0.85-3.79	0.13
Div. 3	1.94	1.06-3.57	**0.03***
A previous severe injury[d]			
No	1		
Yes	0.38	0.19-0.77	**0.01***
An injury problem[e]			
No	1		
Yes	2.28	0.98-5.31	0.06
Hip strength[f]			
CLAM[e] measured in Nm/(kg*m) *100	0.99	0.97-0.99	**0.03***

***Denotes statistically significant p-values at p≤0.05.**

[a]**OR:** odds ratio; [b]**95 % CI:** 95 % confidence interval; [c]**p-value:** probability value; [d]**A previous severe injury**: One or more time-loss injuries during season 2021, or earlier, that lasted 3 months or more; [e]**An injury problem**: An injury problem located in the head, lower belly, lower back, pelvic or lower extremities that did not demand any time-loss from game or training during the four weeks before or during the test occasion; [f]**Hip strength:** Side-Lying Clamshell (CLAM).

5.4.3 SLS for the knee segment

Regarding the univariate logistic regression analysis of the knee segment for the dominant leg, four variables were significantly associated with a failure on the test: soccer division, age, PSS-14, and GAD-7 if the subject belonged to the category mild anxiety. For the non-dominant leg, five variables were significantly associated: soccer division, age, hip strength, an injury problem, and AFAQ.

The multivariate models for the knee segment are reported in **Table 10 and Table 11.** The independent variables associated with the outcome of the SLS for the knee segment differed depending on which leg was tested, except for division that was associated with both the dominant leg and non-dominant leg (dominant leg: div 2; OR 2.34, 95% CI 1.01-5.12, p=0.033. div 3; OR 3.07, 95% CI 1.61-5.85, p=0.001.non-dominant leg: div 2; OR 3.30, 95% CI 1.33-8.00, p=0.01. div 3; OR 3.05, 95% CI 1.44-6.43, p=0.003).

Table 10. Multivariate analysis of failing on the knee segment for the dominant leg during the Single Leg Squat test.

	SLS for the knee segment: dominant leg		
Variables	**OR**[a]	**95 % CI**[b]	**p-value**[c]
PSS-14[d]			
(No stress/stress; 0-56)	0.90	0.83-0.98	**0.01***
GAD-7[e]			
No anxiety	1		
Mild anxiety ≥5	1.95	1.04-3.66	**0.04***
Moderate/severe anxiety ≥10	1.52	0.68-3.39	0.31
Ankle dorsiflexion[f]			
WBLT[f] measured in degrees (TA[f])	0.95	0.88-1.01	0.11
Division			
Div. 1	1		
Div. 2	2.34	1.01-5.12	**0.03***
Div. 3	3.07	1.61-5.85	**0.00***

***Denotes statistically significant p-values at $p \leq 0.05$;**

[a]**OR:** odds ratio; [b]**95 % CI:** 95 % confidence interval; [c]**p-value:** probability value; [d]**PSS-14:** Perceived Stress Scale 14-item instrument; [e]**GAD-7:** Generalized Anxiety Disorder 7-item scale; [f]**Ankle dorsiflexion:** measured with the Weight Bearing Dorsiflexion Lunge Test (WBLT) and calculated with a trigonometric dorsiflexion angle (TA).

Table 11. Multivariate analysis of failing on the knee segment for the non-dominant leg during the Single Leg Squat test.

	SLS for the knee segment: non-dominant leg		
Variables	**OR[a]**	**95 % CI[b]**	**p-value[c]**
AFAQ[d] (No fear/fear; 10-50)	0.95	0.91-0.99	**0.01***
Hip strength[e] CLAM[e] measured in Nm/(kg*m) *100	0.98	0.96-0.99	**0.00***
Division			
Div. 1	1		
Div. 2	3.30	1.33-8.00	**0.01***
Div. 3	3.05	1.44-6.43	**0.00***
An injury problem[f]			
No	1		
Yes	3.11	1.25-7.76	**0.02***

*__Denotes statistically significant p-values at $p \leq 0.05$;__

[a]**OR:** odds ratio; [b]**95 % CI:** 95 % confidence interval; [c]**p-value:** probability value; [d]**AFAQ**: Athletic Fear Avoidance Questionnaire; [e]**Hip strength:** Side-Lying Clamshell (CLAM); [f]**An injury problem**: An injury problem located in the head, lower belly, lower back, pelvic or lower extremities that did not demand any time-loss from game or training during the four weeks before or during the test occasion.

5.5 Additional analysis

To investigate the discriminative validity of the SLS some additional analyses not published in the four studies were performed. All results from the additional analysis are summarised in **Appendix 1** (named Table 1-6)**.** Depending on the definition of injury and the period, the prevalence of injuries varied between 4.8% and 23.1%.

When interpreting the sensitivity, specificity, and predictive values in relation to each other for the different injuries in **Appendix 1** (Table 1-6), the results show that the SLS is not good enough to discriminate a player with any kind of previous or present injury compared to an uninjured player. This is regardless of if the SLS is assessed as a separate segment or as a total score for all segments (pass/fail or with four categories).

However, the separate foot and trunk segment had a specificity of >73% and a negative predictive value of >75% for all types of injuries, meaning that the SLS for those two segments in 75% of all cases could discriminate those who haven't had a previous or present injury compared to those reporting any kind of injury. Furthermore, the results for the foot and trunk segments showed a negative predictive value of over 90% for "a present time-loss injury" and "a time-loss injury previous 4 weeks".

6 Discussion

6.1 Main findings

The overall aim of this thesis was to develop and assess aspects of reliability and validity of the SLS among physically active people, and from a biopsychosocial perspective investigate factors associated with the SLS in a sample of female soccer players.

The systematic review and meta-analysis in **Study I** made it clear that several different versions of the SLS exist regarding its performance and rating criteria. When meta-analysed, the SLS showed moderate reliability across all types of SLS including the Forward Step-Down and Lateral Step-Down. The assessment with a ≤3-point rating scale showed a higher pooled agreement for inter-rater reliability compared with ≥4-point rating scales and the reliability was not affected by the number of observed body segments.

In Study II, a marker-less motion capture (MMC) system, the Qinematic™ was evaluated for its test-retest reliability and construct validity when measuring a SLS. The idea was to bring the Qinematic™ into future studies in this thesis, but as the Qinematic™ showed poor absolute reliability and a diagnostic accuracy that only partly was acceptable, the Qinematic™ wasn't further investigated.

To further investigate the visually assessed SLS, a standardised multi-segmental SLS was proposed and evaluated in **Study III**. Regardless of the raters' experience, and with a 2-hour education, the proposed multi-segmental SLS showed a "moderate" inter-rater reliability and an "almost perfect" intra-rater reliability for a combined assessment of all segments together (the variable "all segments").

In Study IV, this multi-segmental SLS was studied in a sample of female soccer players to investigate its association with demographics, previous injuries, and biomechanical and psychosocial factors. The study identified a variety of factors associated with the outcome of the SLS for both the total score and the separate knee segment. The results implicate that the clinician seemingly needs to consider several factors when assessing the SLS among female soccer players, such as leg dominance, division inherency, hip strength, and psychosocial factors. The results might also be of importance to consider in future prospective studies on the predictive value of the SLS for injury prevention in female soccer players.

Moreover, in an **additional analysis,** the diagnostic accuracy of the SLS to discriminate subjects with a previous or present injury from those with no previous or present injury was investigated. The analysis showed that the SLS was not good enough to discriminate a player with any kind of previous or present injury regardless of whether the SLS was assessed as a separate segment or as a total score for all segments.

However, analysis of the separate foot and trunk segments indicated that those segments could discriminate against those who haven't had a previous or present injury.

6.2 The SLS in clinical testing

Study I clearly showed that the performance and assessment of a visually assessed SLS vary in the clinical setting [242], which makes it difficult for the clinician to decide on what test is the best to use. The results from Study I suggest that even if the SLS is performed in various ways the SLS seemingly shows acceptable levels of reliability across all types of SLS including the Forward Step-Down and Lateral Step-Down. The results from Study I validated the possibility for a clinician to assess more than one segment at a time, and it brought some clearness in the choice of rating scales and confirmed the knowledge that the reliability of a visual assessment, in general, will improve with fewer classification categories [105, 109].

Since the publication of Study I (2019), three new studies with a multi-segmental approach for the SLS [162, 243, 244] and one narrative review for the Lateral Step-Down test [245] have been published. These studies support our findings that the SLS has acceptable levels of reliability across all types of SLS and that the clinician can assess more than one segment at a time [242]. The four studies report an inter-rater reliability ranging from moderate to almost perfect (kappa 0.55-0.93) [162, 243-245]. In a general perspective, previous and present findings and the results from Study I point out the need for a standardisation of the SLS regarding its performance and assessment criterion, as well as an improvement of the methodological quality of the studies investigating their measurement properties [105, 156, 242, 245]. With that in mind, the multi-segmental SLS investigated in Study III was developed according to previous findings [105, 156, 242, 245] and was evaluated with a more robust methodological standardisation. The multi-segmental SLS investigated in Study III was concluded to be reliable enough to be used in an active population. Interestingly, regardless of the rater's experience and following a 2-hour education, and without too many facilitating factors the results still showed acceptable kappa values. This is in contrast to some previous studies on the reliability of the SLS, which in general improves with a more extensive education of the raters [242, 246]. Some studies add more facilitating factors such as markers and poles which also affect the reliability positively [242]. Hence, the reliability of the SLS in Study III might be improved with longer education and with regular use of a clinician even if the results reported are within acceptable levels.

Different methods and classifications can be used to interpret kappa values [149], and the choice of a cut-off score might seem arbitrary. Parameters of reliability are sample dependent as it is easier to distinguish subjects in a heterogenous population than in a homogenous population [149]. In a skewed, homogenous, population the kappa value will usually be lower even if the value theoretical can reach one [149]. In the clinical context, the rater's experience and level of education most often vary. In addition, the subjects assessed and the context (environment and settings) in which the test is performed will most likely also vary. Taking the clinical context into consideration, it might be reasonable to consider a lower cut-off score (kappa >0.40) to be acceptable. This has been proposed in some previous studies investigating the reliability of different tests used in manual therapy [247-250]. On the contrary, others recommend a kappa of 0.60-0.75 and state that anything less might be useless [223].

6.3 Validity and reliability of the Qinematic™

When studying the validity and reliability of the Qinematic™, it turned out that three out of four variables in Study II (left knee down, right knee up, and right knee down) reached "substantial reliability", with an ICC ranging from 0.64 to 0.69 and "moderately strong" (r=0.61-0.68) for the Spearman correlation coefficient. However standard error of measurement (SEM: 9.09°-9.85°) and smallest detectable change (SDC: 26.06°-27.30°) were relatively high. This is in contraction to previous studies using Kinect data showing relatively small standard errors of measurement (SEM: 3.62°, 4.38°) [125] and smallest detectable changes (SDC: 4.1°) [126] values for the peak joint angle of different joints and different functional tests (the vertical drop jump, the SLS and double leg squat). This difference could be because the Qinematic™ registers segmental displacement during the whole movement and calculates a Net Trajectory Angle (NTA) which cannot be compared to studies that calculate the peak joint angle [122, 124-127]. Moreover, Grooten et al. [192] also showed poor measurement properties of the Qinematic™ in its ability to measure balance, posture, and side bending.

Still, Qinematic™ might be able to monitor a group over time but cannot be recommended to be used on an individual level as the absolute reliability (SEM and SDC) was too high. The high absolute reliability could, of course, have been due to the individual variation in performing the SLS between the two test occasions, but when analysing the video recordings with high NTA such as 15-30 degrees, it was clear that the poor absolute reliability was not a result of a large within-subject variation. Most likely this might have been an effect of the small medial and lateral displacements of the knee occurring during the movement which resulted in large angles of NTA which estimates the "line of best fit" during the whole SLS. No matter what, the NTA which captures the whole movement in one point seems to be more unreliable compared to capturing peak joint angles at one specific point during the movement.

To investigate the construct validity of the Qinematic™, data from the Qinematic™ was compared to a visual assessment of the SLS assessed as "a knee-medial-to-foot position" or "a knee-over-foot position". The largest area under the curve was reported to be 0.82 when using a cut-off at 6° of medial displacement which indicates a "moderate" accuracy [222, 223]. At the same cut-off point, the Positive Predictive Value was 0.58 (95% CI 0.47 to 0.68), and the Negative Predictive Value was 0.94 (95% CI 0.86 to 0.98). This indicates that the SLS has a probability of 58% to assess subjects as "a knee medial to foot position" when the Qinematic™ exceeded 6 degrees (positive predictive value) and a 94% probability to assess subjects as "a knee over foot position" when Qinematic™ does not exceed 6 degrees. From a clinical perspective, it can be debated if the use of Qinematic™ adds any new information as clinicians can visually assess a knee-over-foot position with good accuracy [129]. As a result of poor absolute reliability, the NTA was not recommended for the use to assess knee medial to foot position and the Qinematic™ wasn't used in further studies.

6.4 Discriminative validity of the SLS

The validity of a test is important to determine if the test measures the construct it purports to measure [8], and the discriminative validity to understand if a test for example can discriminate between injured and non-injured subjects. In a clinical context, the discriminative validity of the SLS could support a practitioner in the clinical decision-making process to decide whether a player should participate in a game or not, both if the player has a minor injury problem or in return to play after a time-loss injury.

When the prevalence of a condition in a sample is low, the Positive Predictive Value will decrease and the Negative Predictive Value will increase [251]. This statistical fact is also seen in the additional analysis in the present thesis, **see Appendix 1**. There are today no generally accepted thresholds for specificity, sensitivity, and predictive values [251, 252], even though there have been attempts to recommend qualitative descriptors [251]. Regardless of the acceptable threshold, it is evident that the results of diagnostic accuracy must be interpreted in relation to each other and not separately to eliminate over- or underestimation of the results [251].

When interpreting the sensitivity, specificity, and predictive values in relation to each other for the different injuries in the additional analysis, the results showed that the SLS could not discriminate a player with any kind of previous or present injury from a player not reporting an injury. This was evident regardless of whether the SLS was assessed as a separate segment or as a total score for all segments (pass/fail or with four categories). This finding is in line with Whatman et al. [244] who reported; that individuals with a history of a previous intra-articular knee injury (3-11 years ago) have no increased likelihood of failing on a visually assessed SLS. Whatman et al. [244] further discussed several reasons for their results and proposed that the time since the injury was one important factor as well as there might be no remaining deficits left with appropriate rehabilitation. Somewhat unexpectedly, the odds of failing on the SLS for the total score on the non-dominant leg in Study IV were significantly lower for those with a previous severe injury. A possible explanation for the decreased odds might be that 50 percent or more of the included subjects in Study IV who reported a previous severe injury, reported knee injuries (ligament injuries or fractures) that caused a time loss of at least three months. These subjects most likely underwent rehabilitation where knee control and thus the SLS were integrated. Interestingly, when looking more carefully at the statistics in the study by Whatman et al [244], one could see that the odds of failing (however not significant) on the vertical drop jump were lower if the subject had a history of a previous intra-articular knee injury (OR=0.91) compared to those with no previous intra-articular knee injury when adjusted for sex and age and that the odds ratio for the SLS was close to one (OR=1.04, non-significant). Perhaps these results, and the results from Study IV are an indication of not just successful rehabilitation but also improved movement quality.

In contrast to a previous injury that occurred some month or years ago, a time-loss injury or an injury problem that occurred four weeks ago/or is ongoing might be of greater interest as the player continues to play with pain and discomfort. In a study of non-arthritic hip pain patients, it was shown that the patients who passed the SLS reported less pain and greater levels of physical function in their activities in daily living and sport-related activities [243]. This is in contrast with our additional analysis but is in line with the findings from Study IV, which showed 2-3 times higher odds of failing on the total score and the knee segment in the non-dominant leg (but not the dominant leg) if the player had an ongoing injury problem or an injury problem that occurred previous four weeks.

One reason for the poor discriminative validity of the SLS, and no further injury association in Study IV, might be that we except for “a severe injury” in the knee didn’t analyse any other type of injuries from various anatomical locations. In other words, the outcome of the SLS was only analysed in relation to an injury in the lower extremity and trunk. Another reason might be the exclusion criteria of Study IV which excluded all players “*suffering from an ongoing injury that made it impossible to perform the physical tests without pain, or else if the player considered that participation meant an additional risk for injury”*. The players were informed that if they experienced pain during the test, a maximum limit of 3-4 on the Visual Analog Scale was acceptable. Pain during a functional movement is of interest and might be an important clinical sign [1, 88] as there is abundant research saying that proprioception will be affected by pain, swelling, trauma, and fatigue [253]. Even though the SLS is not a specific test of proprioception in a specific body part it involves both sensory and motor functions in the whole kinetic chain from the foot to the trunk.

However, the separate foot and trunk segment in the additional analysis had a specificity of >73% and a negative predictive value of >75% for all types of injuries, meaning that the SLS for those two segments in 75% of all cases could discriminate those who haven’t had a previous or present injury. Furthermore, the SLS for those two segments had a negative predictive value of >90% for “a present time-loss injury” and “a time-loss injury last 4 weeks”. However, due to the low prevalence of the number of cases in these segments, the results should be interpreted with caution. From a clinical perspective, it is important to know that previous injuries are a risk factor for a new injury, both in male and female soccer players [254-256]. In addition, research also reports that the recall accuracy is low in retrospective questionnaires regarding sports injuries (i.e. the athletes do not remember their injuries and the circumstances around them) [257, 258] and that the perception of pain and the concept of injury differ extensively between subjects. For example, it has been reported that athletes have consistently higher pain tolerance and different attitudes towards pain compared to non-athletes or active control subjects [12, 259]. If an intervention is costly, high specificity is preferred and for an athlete, costly often means time-consuming [74]. Hence, the knowledge of the high specificity and negative predictive value of the foot and trunk in discriminating those without a previous or present injury could perhaps be used by the clinician in the decision to reduce or facilitate the rehabilitation regarding some specific traits in the exercise prescription.

6.5 The SLS from a biopsychosocial perspective

In a biopsychosocial perspective and within the context of the Movement Continuum Theory (MCT) various factors could affect the outcome of a functional test.
According to the MCT, both external and internal factors affect the quantity and quality of a movement [7]. External factors can be divided into external (e.g. heat, rain) and social factors (e.g. politics, culture), and internal factors can be divided into physical (e.g. anatomy, physiology) and psychological factors (e.g. personality, emotions) [7]. Using the MCT as a theoretical framework highlights the importance of viewing the assessment of the SLS from a more holistic perspective in clinical practice as well as for research purposes. This is in line with the results from Study IV; that the outcome of the SLS was associated with various demographic, biomechanical and in addition psychosocial factors, both for the total score and for the knee segment. In a general perspective, and with small variations, the same independent variables turned out to be of importance in the multivariate models within the dominant leg and non-dominant leg, but they differed between the dominant leg and non-dominant leg.

6.5.1 Biomechanical and demographic factors

The SLS could be seen as an easy test, which needs to be complemented with other more challenging tests to provoke or elicit a movement quality deficit, especially in well-trained athletes. However, looking at the whole sample of female soccer players in Study IV, 46%-69% of the players failed the SLS when assessed as a total score and 28%-40% failed the SLS when assessed as a separate knee segment.

The results from Study IV also showed that regardless of whether the SLS was assessed as a total score for all segments, or as a separate knee segment, it was significantly more common to fail on the dominant leg than on the non-dominant leg. This finding has also been found in a cohort of 558 youth soccer players (boys and girls) where the authors suggested that this could have been due to an imbalance in knee control between the legs [260]. The better SLS performance for the non-dominant leg might not be surprising if one considers the nature of the sport where repeated soccer drills in a unipedal stance will modify proprioceptive factors, muscular control, and strength in the non-dominant leg [261-264]. In line with this finding, Study IV also found an increased odds of failing on the total score for the non-dominant leg, but not the dominant leg, for players in the lowest division compared to players from the highest division and 2-3 times higher odds of failing on the knee segment for both the dominant leg and non-dominant leg in players from a lower division. It could be argued that the players in the higher division, who are more skilled, also have a higher skill in controlling the weight-bearing leg on the soccer field, and therefore might be better in the performance of the SLS.

The hip strength in stud IV was significantly stronger in the non-dominant leg compared to the dominant leg for the whole sample of soccer players. The observed higher levels of hip strength for the non-dominant leg might not be surprising if considering the nature of the sport with repeated soccer drills in a unipedal stance. For the total score, we found significantly lower odds of failing on the SLS for higher levels of hip strength on both legs, however for the knee segment, this was only seen for the non-dominant leg. Overall, the results implicate that hip strength is important and associated with the outcome of the SLS. However, the confidence interval for the association with hip strength was close to one, and a significant test (Wilcoxon rank sum test for unpaired non-normally distributed data) showed that there were no differences in hip strength for those who failed and those who passed the SLS on the total score for the dominant leg (p=0.06). This might raise some questions about how strongly hip strength is associated with the SLS, especially for the dominant leg. Other important aspects of muscle function other than pure strength during the SLS are muscle activation and neuromuscular control [50-52], these features contribute with other qualities than strength which has been shown effective in correcting a knee valgus [86, 265]. Considering our findings, one could reason that the role of hip strength might be less important if other factors such as muscle activation and neuromuscular control are worse in the dominant leg compared to the non-dominant leg. In other words, the SLS might demand a certain amount of muscle strength, but if other functions such as muscle activation or neuromuscular control are superior to strength, the importance of strength might not be that decisive for a successful outcome of the SLS.

Ankle dorsiflexion was not significantly associated with the total score and the separate knee segment in any leg, and this was a little bit surprising as most research supports that decreased ankle dorsiflexion affects the SLS in a negative way [50, 54, 169-172, 266]. The findings in Study IV could have been related to the relatively good ankle mobility displayed by this sample of female soccer players, and the lack of contrast in our data as the range in data was small, which might have hampered the possibility of finding significant associations with the outcome of the SLS. Perhaps, females have an increased ankle dorsiflexion compared to men, and our findings are normal for a sample of female soccer. In that case, ankle dorsiflexion is not a factor when assessing the SLS in a female population. However, future studies must confirm or reject this finding in a similar population.

6.5.2 Psychosocial factors

Research has shown that psychosocial factors such as an athlete's thoughts, emotions, and actions are associated with the outcome of rehabilitation and that an athlete who can deal with his fear and anxiety will experience more positive rehabilitation results [176, 177, 267]. Psychosocial factors such as anxiety [185] and fear of movement [179, 268] could for example affect readiness to return to sport and return to the preinjury level after an anterior cruciate ligament injury. In addition, negative life-event stress and strong stress responsivity have been reported to be important risk factors for injuries among athletes [56], it is also possible that such an emotion could affect the rehabilitation and return to sport [267, 269]. It is therefore recommended to take these factors into consideration in the process of rehabilitation and return to sports [267]. Since studies are sparse on how psychosocial factors relate, affect or associate to an outcome of a functional movement test that is included in the rehabilitation, further research in this field is warranted. However, evidence exists that physical performance might be affected by psychosocial factors during the rehabilitation of an anterior cruciate ligament injury [59, 60].

Three out of four investigated psychosocial factors were significantly associated with the models in present thesis. Unexpectedly, small increases on the stress scale were associated with relatively large odds of passing the SLS on the dominant leg. Stress per se might not explain the outcome of the SLS. Instead, stress might be related to other non-measured variables directly associated with the outcome. Even though the associations were not seen for both legs and no causal direction can be given from Study IV, the results indicate that a clinician might consider psychosocial factors such as anxiety and fear of avoidance in the rehabilitation and test-retest situation when using the SLS.

The results might also be of importance to consider in future prospective studies on the predictive value of the SLS for injury prevention in female soccer players. Perhaps we can understand, or consider, these psychosocial factors not only as factors associated with the outcome but as factors affecting the pathway between a previous injury and the test situation. The impact of psychosocial factors on the outcome of a functional test has not yet been considered in sports medicine but might be of importance for future studies to better understand the outcome of a test.

6.6 Methodological considerations

6.6.1 Internal validity

Internal validity examines the extent to which the study design, conduct, and analysis can answer the research questions without any bias [270]. In other words, it concerns whether the design and conduct of a study eliminates other explanations for the results. Overall, the internal validity of the included studies in this thesis can be seen as good. However, there are limitations and some practical choices have been made which could have affected the internal validity, those limitations are important to keep in mind when interpreting the results.

6.6.1.1 Study I

The major strengths of Study I are the study design with an extensive literature search and the applied methodology which was performed according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (PRISMA statement) [188, 189]. The methodological study design strengthens the internal validity as bias is minimised in the methods of data collection and analyses of the results. Furthermore, the risk of bias within the included studies was checked with the Quality Appraisal of Reliability Studies Checklist (QAREL) [190] which is a reliable and specially designed checklist to assess the quality of studies of diagnostic reliability [191]. One drawback with Study I is that the methodological quality of the included studies was questioned which must be considered when interpreting the results. One specific issue was that seven studies didn't fulfil item 11 of QAREL (if appropriate statistics were used). However, sensitivity analysis on this matter showed that the results of pooled agreement stayed the same when those seven studies were excluded.

Another strength of Study I was the merge of different SLS and the inclusion of Lateral Step-Down and Forward Step-Down which allowed pooling more than 30 studies with similar performance but also to compare multiple results from different studies. On the contrary, one could question whether it is correct to compare different types of SLS and the Forward Step-Down and Lateral Step-Down as there is a biomechanical difference in their performance. Nevertheless, the sensitivity analysis when excluding the Forward Step-Down and Lateral Step-Down only gave a minor change in the results which confirms the robustness of the results. Finally, Study I reported a moderate heterogeneity (I2=53%-65%) which implicates a great variability across the included studies. The included studies varied in performance, assessment protocols, study population, and experience of the assessors, this suggests that a standardisation of the SLS is required.

Statistically, different correlation statistics were merged in the pooled analyses as many of the included studies in the meta-analysis used various kappa statistics and ICC models and sometimes did not report the ICC model used. This could of course have affected the pooled agreement estimates. To be able to perform the meta-analyses, some choices had to be made if more than one reliability measure was presented for the same rating if different assessment methods were presented in the same study, and regarding which kappa statistics to include if more than one option was given. Those choices were necessary for the data processing and this methodology has previously been reported [156]. Finally, in a systematic review and meta-analysis, there is always a risk that a study has been missed due to poor indexing.

6.6.1.2 Study II

A major strength of Study II regarding the internal validity is the methodological structure where the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) [147, 271] and the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) [272] were used. Furthermore, an adequate sample size with more than 50 measures was used as recommended for reliability and validity studies [149, 273]. The most obvious limitation with Study II was the lack of a 3D kinematic gold standard which would have been better than the used visual assessment. However, the use of a construct such as the visually assessed SLS was the most obvious and practical choice in the absence of a gold standard.

Statistically, as two variables (left knee down and right knee up at occasion one) were non-normally distributed Spearman correlation coefficient was calculated in addition to ICC. To calculate and assess ICC on non-normally distributed data is a limitation in Study IV and the ICC results must therefore be interpreted with some caution.

6.6.1.3 Study III

Study III had three major strengths; the use of a methodological standardisation based on the Quality Appraisal for Reliability Studies checklist (QAREL) [190], the use of different statistical computations, and that the evaluated SLS was based on the findings from previous studies investigating the reliability of the SLS. Furthermore, an adequate sample size with more than 50 measures was used as recommended for reliability and validity studies [149, 273]

Statistically, the magnitude of kappa is among others influenced by prevalence and bias, and a comparison of the strength of the kappa across other studies with different statistics could therefore be difficult [149, 227]. The interpretation of the magnitude of the kappa statistics in Study II with both kappa_{max} and PABAK can therefore be seen as a strength of the study.

6.6.1.4 Study IV

The major strengths of Study IV are the inclusion of a specific sample of both elite and sub-elite female soccer players, the recruitment of many players, and the inclusion of different associated factors. The analyses of the four models were based on valid and reliable instruments for collecting the data and a robust statistical analysis. In multiple regression models it is recommended that for every variable screened for association, there should be at least 10 events [228]. This was not an issue for the total score models and the knee model for the dominant leg but a little bit on the low side for the non-dominant leg in the knee model. However, this rule of thumb should not be applied categorically as other factors could affect the stability of a model [228], as well as there in some cases is evidence for reducing this rule to 5-9 events [229]. Further limitations that need to be considered when interpreting the results are that only one person (JR) performed all biomechanical tests, which could have rendered a systematic error in the assessment. Retrospective questions about previous injuries have been shown to have low recall accuracy which could have affected the data of previous injuries [257, 258]. In addition, the psychosocial questionnaires used in **Study IV** were not evaluated for its measurement properties in the exact same population.

Even though Study IV was based on a robust statistical analysis, it cannot be ruled out that other unmeasured or confounding factors might have affected the results. In addition, Study IV used a cross-sectional design, meaning that no conclusion can be made about a causal relationship between the dependent and independent variables.

6.6.2 External validity

External validity refers to whom the results can be applied, sampling and different design factors (contextual factors) are therefore threats to the generalisability of the results [270]. The study population in **Study II and III** was a convenience sample of both men (27%) and women (73%), with an average age of 34 years (SD ±12), who were relatively active, mostly with running/jogging and weightlifting. Therefore, no further generalisations to another population than an active non-injured population can be made from the findings in **Study II and III**. However, the population in **Study II and III** might be seen as an adequate population for the SLS to be applied for in general use as in primary care.

Optimally, the SLS that was used in **Study IV** should have been evaluated in the same population as it was used. Still, the subjects in **Study IV** (elite and sub-elite female soccer players) can be seen as an active population, which corresponds to the study population in **Study III**. On the other hand, in **Study III** there were also males included.

The proposed multi-segmental SLS in **Study IV** was built on the findings from **Study I** where 65 % of the included subjects were healthy active people including many athletes aged 18 to 37 years. Furthermore, the functional aspects of sport-related actions were considered when proposing the SLS.

A strength of **Study III** was the use of video recordings to observe and assess the SLS, recording was chosen to standardise the testing procedure so several raters could assess identical performances without the normal within-subject variety. On the other hand, this lowers the test's ecological validity if the clinician assesses the performance live in the clinical situation, which is most common. One further limitation of **Study III** is that no generalisation can be done across raters or clinicians from the four raters that were included. Finally, the results from **Study IV** might only be generalised to female soccer players of the same age and players at the same competition level (divisions 1-3) in Sweden.

6.7 Clinical implications

Movement quality has been identified as an independent attribute as it, unlike other quantitative measures such as power or strength, aims to capture other important aspects of a movement such as movement efficiency and the maintenance of a good posture, balance, and vertical alignment [3, 89, 100]. In this context, the present thesis contributes to further understanding and knowledge about the use of the SLS in the clinical setting and research.

The study of Qinematic™ indicates that this marker-less motion capture system might be able to monitor a group over time but is not recommended to be used on an individual level due to poor absolute reliability. Perhaps the use of Qinematic™ could be used as a pedagogic tool in the dialogue with the patient.

This thesis revealed that clinicians can use a multi-segmental approach to the SLS in a reliable way, preferably with a ≤3-point rating scale. However, even though the different SLS, the Forward Step-Down, and the Lateral Step-Down measure the same construct, not all SLS exhibit the same kinematics and kinetics. A clinician is therefore advised to be aware of those differences when choosing a SLS. Furthermore, not all SLS are developed and evaluated with a proper methodological quality, the clinician is therefore recommended to scrutinise the methodological setup for the chosen test. The standardised multi-segmental SLS in Study III is proposed to be reliable enough to be used in an active population. However, as the education of the raters is of importance there should be a minimum of a 2-hour education on the assessment criteria before its use.

The clinician needs to consider the following when using the SLS:

- Although the SLS sometimes is seen as less vigorous compared to other tests, the SLS should not be considered an easy test as 46%-69% of the female soccer players in study IV failed the SLS when assessed as a total score, and 28%-40% failed the SLS when assessed as a separate knee segment.

- It is important to obtain information about division inherency and leg dominance as it is more common to fail on the SLS on the dominant leg, and the odds of failing on the SLS were higher for those players in the lower divisions, both for the knee segment (non-dominant leg/dominant leg) and the total score (only the non-dominant leg).

- Since the SLS seems to be associated with hip strength, further examination of hip strength is recommended in those subjects who fail the SLS.

- The outcome of the SLS can be expressed as a total score but is in addition also recommended to be reported as a pass/fail for each separate segment. This is to give a more complete picture of the outcome of the SLS.

- The SLS assessed as a total score or as a separate knee segment is associated with previous injuries, biomechanical factors, and psychosocial factors but differs depending on whether the dominant leg or non-dominant leg is assessed.

- The SLS cannot discriminate players with any kind of previous or present injury in the lower back or lower extremities from non-injured players.

From a research perspective, it is of interest to further study the predictive value of the SLS in relation to various associated factors in a longitudinal design. The differences found in this study between the dominant leg and non-dominant leg indicate that data should be stratified, rather than adjusted, for leg dominance in the statistical analysis.

7 Conclusion

The SLS including the Forward Step-Down and Lateral Step-Down is a reliable movement quality test that could be used in the clinical setting, both with a uni-segmental and multi-segmental approach, preferably with a ≤3-point rating scale.

Due to poor absolute reliability, the use of the Qinematic™ net trajectory angle, which estimates the “line of best fit”, cannot be recommended to assess a knee medial to foot position.

The proposed standardised multi-segmental SLS is seen as reliable enough to be used in the clinical setting regardless of the rater’s experience and with a common 2-hour education.

When assessed as a total score or as a separate knee segment, the SLS is associated with a variety of factors that differ depending on whether the dominant leg or non-dominant leg is assessed. The clinician seemingly needs to consider several factors when assessing the SLS among female soccer players, such as leg dominance, division inherency, hip strength, and some psychosocial factors. These results might be of importance to consider in future prospective studies on the predictive value of the SLS for injury prevention in female soccer players.

The SLS does not seem to discriminate between injured and non-injured athletes.

8 Future research and directions

Future research on the SLS regarding its measurement properties should be carried out with good methodological quality and performed in different populations and for raters with different experiences. Furthermore, a general standardisation of the SLS performance and assessment criteria would be desirable. Until today, it seems that the SLS has not been evaluated for its responsiveness which is of importance if the test is to be used as an evaluation tool for progress during the rehabilitation.

A completely different aspect of the SLS would be to understand how the test person senses or apprehends the SLS when it is performed. Does the test person understand what the test is all about, why it is important to perform the test, and is the test experienced as easy or difficult? Such qualitative aspects could contribute to a further understanding and development of the test.

The predictive validity of the SLS still needs to be evaluated, and therefore were the female soccer players in Study IV followed during the soccer season 2022 and registered for time-loss injuries and/or injury problems every four weeks. It will be interesting to investigate this data and explore the SLS in an injury-preventive aspect. Moreover, data regarding hormonal factors such as menstruation, pain when bleeding and hormonal contraceptives were collected at baseline. The purpose is to investigate how the SLS is affected by those gender-specific hormonal factors.

Finally, to get an objective measure of movement quality, there is still a need for a reliable and valid marker-less motion capture system in the clinical setting that could quantify the movement.

9 Acknowledgement

First of all, I would like to send my gratitude to **all participants** who chose to participate in Study II-IV, especially all female soccer players in Study IV. I have never met a group of people with such warm and friendly treatment.

Words cannot explain the gratitude I am feeling towards my supervisors. I have always said that **Eva Rasmussen-Barr** and **Wim Grooten** are a perfect complement to each other, so great that a mathematical paradox is created: 1+1=3, and that is quality!!

To my principal supervisor, **Eva Rasmussen-Barr**, thank you for everything! Especially for picking me up under your wings sometime in the spring of 2014 when I was a master's student, you never hesitated. During our years together you always had a positive attitude and a rational approach, even when things were not getting our way, you answered all my questions and emails with incredible patience, and you always wanted the best for me. You gave me so much of your time, and you made me grow as a person, not just as a researcher. For this, I am deeply grateful, and I will never forget. Thank you!

To my co-supervisor, **Wim Grooten**, thank you for being who you are as a person and especially thank you for all your time. Time is all we got, and you gave me more than necessary. Your energy and interest in small and big things, whatever they might concern, is remarkable. I never met someone with such a positive attitude and interest in teaching and helping others, and if you have not understood it yet, you are very good at it, simply the best.

To my mentor, **Anna Frohm**, during my years as a half-time doctoral student, you have been a tremendous support in just being there at our work, Idrottsmedicinska Kliniken Bosön. Chatting about patients, functional movement tests, and movement quality in the real clinical setting, but above all you have been my creative collaborator regarding life. For your consideration and care, I am grateful, and I know that we have a lot of discussions in front of us.

I would also like to thank the rest of the crew on **Idrottsmedicinska Kliniken Bosön** for your support and understanding on those mornings and days that I have been a little bit absent or faraway.

To my co-author, **Philip Von Rosen**, it has been a pleasant acquaintance to get to know you and I would like to thank you for all your support during Study IV. You always take your time to answer a question thoughtfully and with great interest. I hope that we can continue our collaboration in the future.

Finally, to my friend, bellowed and dear wife, **Towe Ressman**, you know me better than I know myself. I do not always hear or understand what you say, but you are my guiding star and safe existence in life. Due to circumstances, those years have been a hectic time for us both. I think we have learned a lot and managed well, and we did it together. Thank you for believing in us, I Love You!

10 References

1. Cook G, Burton L, Hoogenboom B. Pre-participation screening: the use of fundamental movements as an assessment of function - part 1. N Am J Sports Phys Ther, 2006. **1**(2): p. 62-72.
2. Mottram S, Comerford M. A new perspective on risk assessment. Phys Ther Sport, 2008. **9**(1): p. 40-51. https://doi.org/10.1016/j.ptsp.2007.11.003
3. McCunn R, Aus der Funten K, Fullagar HH, McKeown I, Meyer T. Reliability and Association with Injury of Movement Screens: A Critical Review. Sports Med, 2016. **46**(6): p. 763-81. https://doi.org/10.1007/s40279-015-0453-1
4. Bennett H, Arnold J, Norton K, Davison K. Are we really "screening" movement? The role of assessing movement quality in exercise settings. J Sport Health Sci, 2020. **9**(6): p. 489-492. https://doi.org/10.1016/j.jshs.2020.08.002
5. Piva SR, Fitzgerald K, Irrgang JJ, et al. Reliability of measures of impairments associated with patellofemoral pain syndrome. BMC Musculoskelet Disord, 2006. **7**: p. 33. https://doi.org/10.1186/1471-2474-7-33
6. Herman G, Nakdimon O, Levinger P, Springer S. Agreement of an Evaluation of the Forward-Step-Down Test by a Broad Cohort of Clinicians With That of an Expert Panel. Journal of sport rehabilitation, 2016. **25**(3): p. 227-232.
7. Cott C, Finch E, Gasner D, Yoshida K, Thomas S, Verrier M. The movement continuum theory of physical therapy. Physiotherapy Canada, 1995. **47**(2): p. 87-96.
8. Mokkink LB, Terwee CB, Patrick DL, et al. 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**(7): p. 737-45. https://doi.org/10.1016/j.jclinepi.2010.02.006
9. Hislop HJ. Tenth Mary McMillan lecture. The not-so-impossible dream. Phys Ther, 1975. **55**(10): p. 1069-80. https://doi.org/10.1093/ptj/55.10.1069
10. Engel GL. The need for a new medical model: a challenge for biomedicine. Science, 1977. **196**(4286): p. 129-36. https://doi.org/10.1126/science.847460
11. van Mechelen W, Hlobil H, Kemper HC. Incidence, severity, aetiology and prevention of sports injuries. A review of concepts. Sports Med, 1992. **14**(2): p. 82-99. https://doi.org/10.2165/00007256-199214020-00002
12. Hainline B, Turner JA, Caneiro JP, Stewart M, Lorimer Moseley G. Pain in elite athletes-neurophysiological, biomechanical and psychosocial considerations: a narrative review. Br J Sports Med, 2017. **51**(17): p. 1259-1264. https://doi.org/10.1136/bjsports-2017-097890
13. Crossley KM, Patterson BE, Culvenor AG, Bruder AM, Mosler AB, Mentiplay BF. Making football safer for women: a systematic review and meta-analysis of injury prevention programmes in 11 773 female football (soccer) players. Br J Sports Med, 2020. **54**(18): p. 1089-1098. https://doi.org/10.1136/bjsports-2019-101587
14. Zech A, Hollander K, Junge A, et al. Sex differences in injury rates in team-sport athletes: A systematic review and meta-regression analysis. J Sport Health Sci, 2022. **11**(1): p. 104-114. https://doi.org/10.1016/j.jshs.2021.04.003
15. Dick RW. Is there a gender difference in concussion incidence and outcomes? Br J Sports Med, 2009. **43 Suppl 1**: p. i46-50. https://doi.org/10.1136/bjsm.2009.058172
16. Prien A, Grafe A, Rössler R, Junge A, Verhagen E. Epidemiology of Head Injuries Focusing on Concussions in Team Contact Sports: A Systematic Review. Sports Med, 2018. **48**(4): p. 953-969. https://doi.org/10.1007/s40279-017-0854-4

17. Boling M, Padua D, Marshall S, Guskiewicz K, Pyne S, Beutler A. Gender differences in the incidence and prevalence of patellofemoral pain syndrome. Scand J Med Sci Sports, 2010. **20**(5): p. 725-30. https://doi.org/10.1111/j.1600-0838.2009.00996.x
18. Horan D, Büttner F, Blake C, Hägglund M, Kelly S, Delahunt E. Injury incidence rates in women's football: a systematic review and meta-analysis of prospective injury surveillance studies. Br J Sports Med, 2022. https://doi.org/10.1136/bjsports-2021-105177
19. Ekstrand J, Hagglund M, Walden M. Injury incidence and injury patterns in professional football: the UEFA injury study. Br J Sports Med, 2011. **45**(7): p. 553-8. https://doi.org/10.1136/bjsm.2009.060582
20. Faude O, Junge A, Kindermann W, Dvorak J. Injuries in female soccer players: a prospective study in the German national league. Am J Sports Med, 2005. **33**(11): p. 1694-700. https://doi.org/10.1177/0363546505275011
21. Jacobson I, Tegner Y. Injuries among Swedish female elite football players: a prospective population study. Scand J Med Sci Sports, 2007. **17**(1): p. 84-91. https://doi.org/10.1111/j.1600-0838.2006.00524.x
22. Crossley KM, Stefanik JJ, Selfe J, et al. 2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 1: Terminology, definitions, clinical examination, natural history, patellofemoral osteoarthritis and patient-reported outcome measures. Br J Sports Med, 2016. **50**(14): p. 839-43. https://doi.org/10.1136/bjsports-2016-096384
23. Crossley KM, van Middelkoop M, Callaghan MJ, Collins NJ, Rathleff MS, Barton CJ. 2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 2: recommended physical interventions (exercise, taping, bracing, foot orthoses and combined interventions). Br J Sports Med, 2016. **50**(14): p. 844-52. https://doi.org/10.1136/bjsports-2016-096268
24. Foss KD, Myer GD, Magnussen RA, Hewett TE. Diagnostic Differences for Anterior Knee Pain between Sexes in Adolescent Basketball Players. J Athl Enhanc, 2014. **3**(1). https://doi.org/10.4172/2324-9080.1000139
25. Glaviano NR, Kew M, Hart JM, Saliba S. DEMOGRAPHIC AND EPIDEMIOLOGICAL TRENDS IN PATELLOFEMORAL PAIN. Int J Sports Phys Ther, 2015. **10**(3): p. 281-90.
26. Walden M, Hagglund M, Werner J, Ekstrand J. The epidemiology of anterior cruciate ligament injury in football (soccer): a review of the literature from a gender-related perspective. Knee Surg Sports Traumatol Arthrosc, 2011. **19**(1): p. 3-10. https://doi.org/10.1007/s00167-010-1172-7
27. Fältström A, Kvist J, Gauffin H, Hägglund M. Female Soccer Players With Anterior Cruciate Ligament Reconstruction Have a Higher Risk of New Knee Injuries and Quit Soccer to a Higher Degree Than Knee-Healthy Controls. Am J Sports Med, 2019. **47**(1): p. 31-40. https://doi.org/10.1177/0363546518808006
28. Meeuwisse WH. Assessing Causation in Sport Injury: A Multifactorial Model. Clinical Journal of Sport Medicine, 1994. **4**(3): p. 166-170.
29. Meeuwisse WH, Tyreman H, Hagel B, Emery C. A dynamic model of etiology in sport injury: the recursive nature of risk and causation. Clin J Sport Med, 2007. **17**(3): p. 215-9. https://doi.org/10.1097/JSM.0b013e3180592a48
30. Bahr R, Krosshaug T. Understanding injury mechanisms: a key component of preventing injuries in sport. Br J Sports Med, 2005. **39**(6): p. 324-9. https://doi.org/10.1136/bjsm.2005.018341
31. Bittencourt NFN, Meeuwisse WH, Mendonca LD, Nettel-Aguirre A, Ocarino JM, Fonseca ST. Complex systems approach for sports injuries: moving from risk factor

identification to injury pattern recognition-narrative review and new concept. Br J Sports Med, 2016. **50**(21): p. 1309-1314. https://doi.org/10.1136/bjsports-2015-095850
32. Hamill J, Knutzen K, Derrick TR, *Biomechanical basis of human movement*. 2015, Philadelphia, Pa.: Wolters Kluwer Health.
33. Hewett TE, Torg JS, Boden BP. Video analysis of trunk and knee motion during non-contact anterior cruciate ligament injury in female athletes: lateral trunk and knee abduction motion are combined components of the injury mechanism. Br J Sports Med, 2009. **43**(6): p. 417-22. https://doi.org/10.1136/bjsm.2009.059162
34. Lucarno S, Zago M, Buckthorpe M, et al. Systematic Video Analysis of Anterior Cruciate Ligament Injuries in Professional Female Soccer Players. Am J Sports Med, 2021. **49**(7): p. 1794-1802. https://doi.org/10.1177/03635465211008169
35. Hewett TE, Myer GD, Ford KR, et al. Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: a prospective study. Am J Sports Med, 2005. **33**(4): p. 492-501. https://doi.org/10.1177/0363546504269591
36. Weiss K, Whatman C. Biomechanics Associated with Patellofemoral Pain and ACL Injuries in Sports. Sports Med, 2015. **45**(9): p. 1325-1337. https://doi.org/10.1007/s40279-015-0353-4
37. Herrington L. Knee valgus angle during single leg squat and landing in patellofemoral pain patients and controls. Knee, 2014. **21**(2): p. 514-7. https://doi.org/10.1016/j.knee.2013.11.011
38. Cronstrom A, Creaby MW, Nae J, Ageberg E. Gender differences in knee abduction during weight-bearing activities: A systematic review and meta-analysis. Gait Posture, 2016. **49**: p. 315-28. https://doi.org/10.1016/j.gaitpost.2016.07.107
39. Della Villa F, Buckthorpe M, Grassi A, et al. Systematic video analysis of ACL injuries in professional male football (soccer): injury mechanisms, situational patterns and biomechanics study on 134 consecutive cases. Br J Sports Med, 2020. **54**(23): p. 1423-1432. https://doi.org/10.1136/bjsports-2019-101247
40. Powers CM. The influence of altered lower-extremity kinematics on patellofemoral joint dysfunction: a theoretical perspective. Journal of Orthopaedic & Sports Physical Therapy, 2003. **33**(11): p. 639-646.
41. Powers CM. The influence of abnormal hip mechanics on knee injury: a biomechanical perspective. J Orthop Sports Phys Ther, 2010. **40**(2): p. 42-51. https://doi.org/10.2519/jospt.2010.3337
42. Jimenez-Del-Barrio S, Mingo-Gomez MT, Estebanez-de-Miguel E, Saiz-Cantero E, Del-Salvador-Miguelez AI, Ceballos-Laita L. Adaptations in pelvis, hip and knee kinematics during gait and muscle extensibility in low back pain patients: A cross-sectional study. J Back Musculoskelet Rehabil, 2020. **33**(1): p. 49-56. https://doi.org/10.3233/bmr-191528
43. Shamsi MB, Sarrafzadeh J, Jamshidi A. Comparing core stability and traditional trunk exercise on chronic low back pain patients using three functional lumbopelvic stability tests. Physiother Theory Pract, 2015. **31**(2): p. 89-98. https://doi.org/10.3109/09593985.2014.959144
44. Milner CE, Hamill J, Davis IS. Distinct hip and rearfoot kinematics in female runners with a history of tibial stress fracture. J Orthop Sports Phys Ther, 2010. **40**(2): p. 59-66. https://doi.org/10.2519/jospt.2010.3024
45. Aderem J, Louw QA. Biomechanical risk factors associated with iliotibial band syndrome in runners: a systematic review. BMC Musculoskelet Disord, 2015. **16**: p. 356. https://doi.org/10.1186/s12891-015-0808-7

46. Botha N, Warner M, Gimpel M, Mottram S, Comerford M, Stokes M. Movement patterns during a small knee bend test in academy footballers with femoroacetabular impingement (FAI). Health Sciences Working Papers, 2014. **1**(10): p. 1-24.
47. Nakagawa TH, Maciel CD, Serrão FV. Trunk biomechanics and its association with hip and knee kinematics in patients with and without patellofemoral pain. Man Ther, 2015. **20**(1): p. 189-93. https://doi.org/10.1016/j.math.2014.08.013
48. Stickler L, Finley M, Gulgin H. Relationship between hip and core strength and frontal plane alignment during a single leg squat. Phys Ther Sport, 2015. **16**(1): p. 66-71. https://doi.org/10.1016/j.ptsp.2014.05.002
49. Willson JD, Ireland ML, Davis I. Core strength and lower extremity alignment during single leg squats. Med Sci Sports Exerc, 2006. **38**(5): p. 945-52. https://doi.org/10.1249/01.mss.0000218140.05074.fa
50. Cronstrom A, Creaby MW, Nae J, Ageberg E. Modifiable Factors Associated with Knee Abduction During Weight-Bearing Activities: A Systematic Review and Meta-Analysis. Sports Med, 2016. **46**(11): p. 1647-1662. https://doi.org/10.1007/s40279-016-0519-8
51. Wilczyński B, Zorena K, Ślęzak D. Dynamic Knee Valgus in Single-Leg Movement Tasks. Potentially Modifiable Factors and Exercise Training Options. A Literature Review. Int J Environ Res Public Health, 2020. **17**(21). https://doi.org/10.3390/ijerph17218208
52. Neamatallah Z, Herrington L, Jones R. An investigation into the role of gluteal muscle strength and EMG activity in controlling HIP and knee motion during landing tasks. Phys Ther Sport, 2020. **43**: p. 230-235. https://doi.org/10.1016/j.ptsp.2019.12.008
53. Dix J, Marsh S, Dingenen B, Malliaras P. The relationship between hip muscle strength and dynamic knee valgus in asymptomatic females: A systematic review. Phys Ther Sport, 2019. **37**: p. 197-209. https://doi.org/10.1016/j.ptsp.2018.05.015
54. Lima YL, Ferreira V, de Paula Lima PO, Bezerra MA, de Oliveira RR, Almeida GPL. The association of ankle dorsiflexion and dynamic knee valgus: A systematic review and meta-analysis. Phys Ther Sport, 2018. **29**: p. 61-69. https://doi.org/10.1016/j.ptsp.2017.07.003
55. Martikainen P, Bartley M, Lahelma E. Psychosocial determinants of health in social epidemiology. Int J Epidemiol, 2002. **31**(6): p. 1091-3. https://doi.org/10.1093/ije/31.6.1091
56. Ivarsson A, Johnson U, Andersen MB, Tranaeus U, Stenling A, Lindwall M. Psychosocial Factors and Sport Injuries: Meta-analyses for Prediction and Prevention. Sports Med, 2017. **47**(2): p. 353-365. https://doi.org/10.1007/s40279-016-0578-x
57. Slimani M, Bragazzi NL, Znazen H, Paravlic A, Azaiez F, Tod D. Psychosocial predictors and psychological prevention of soccer injuries: A systematic review and meta-analysis of the literature. Phys Ther Sport, 2018. **32**: p. 293-300. https://doi.org/10.1016/j.ptsp.2018.05.006
58. Brink MS, Visscher C, Arends S, Zwerver J, Post WJ, Lemmink KA. Monitoring stress and recovery: new insights for the prevention of injuries and illnesses in elite youth soccer players. Br J Sports Med, 2010. **44**(11): p. 809-15. https://doi.org/10.1136/bjsm.2009.069476
59. Paterno MV, Flynn K, Thomas S, Schmitt LC. Self-Reported Fear Predicts Functional Performance and Second ACL Injury After ACL Reconstruction and Return to Sport: A Pilot Study. Sports Health, 2018. **10**(3): p. 228-233. https://doi.org/10.1177/1941738117745806
60. Kvist J, Silbernagel KG. Fear of Movement and Reinjury in Sports Medicine: Relevance for Rehabilitation and Return to Sport. Phys Ther, 2022. **102**(2). https://doi.org/10.1093/ptj/pzab272

61. McCunn R, Aus der Fünten K, Govus A, Julian R, Schimpchen J, Meyer T. THE INTRA- AND INTER-RATER RELIABILITY OF THE SOCCER INJURY MOVEMENT SCREEN (SIMS). Int J Sports Phys Ther, 2017. **12**(1): p. 53-66.
62. McCall A, Carling C, Nedelec M, et al. Risk factors, testing and preventative strategies for non-contact injuries in professional football: current perceptions and practices of 44 teams from various premier leagues. Br J Sports Med, 2014. **48**(18): p. 1352-7. https://doi.org/10.1136/bjsports-2014-093439
63. Meurer MC, Silva MF, Baroni BM. Strategies for injury prevention in Brazilian football: Perceptions of physiotherapists and practices of premier league teams. Phys Ther Sport, 2017. **28**: p. 1-8. https://doi.org/10.1016/j.ptsp.2017.07.004
64. Bakken A, Targett S, Bere T, et al. The functional movement test 9+ is a poor screening test for lower extremity injuries in professional male football players: a 2-year prospective cohort study. Br J Sports Med, 2017. https://doi.org/10.1136/bjsports-2016-097307
65. McCunn R, Aus der Funten K, Whalan M, Sampson JA, Meyer T. Soccer Injury Movement Screen (SIMS) Composite Score Is Not Associated With Injury Among Semiprofessional Soccer Players. J Orthop Sports Phys Ther, 2018. **48**(8): p. 630-636. https://doi.org/10.2519/jospt.2018.8037
66. Trinidad-Fernandez M, Gonzalez-Sanchez M, Cuesta-Vargas AI. Is a low Functional Movement Screen score (≤ 14/21) associated with injuries in sport? A systematic review and meta-analysis. BMJ Open Sport & Exercise Medicine, 2019. **5**(1): p. e000501.
67. Whittaker JL, Booysen N, de la Motte S, et al. Predicting sport and occupational lower extremity injury risk through movement quality screening: a systematic review. Br J Sports Med, 2017. **51**(7): p. 580-585. https://doi.org/10.1136/bjsports-2016-096760
68. Fältström A, Hägglund M, Hedevik H, Kvist J. Poor Validity of Functional Performance Tests to Predict Knee Injury in Female Soccer Players With or Without Anterior Cruciate Ligament Reconstruction. Am J Sports Med, 2021. **49**(6): p. 1441-1450. https://doi.org/10.1177/03635465211002541
69. Krosshaug T, Steffen K, Kristianslund E, et al. The vertical drop jump is a poor screening test for ACL injuries in female elite soccer and handball players: a prospective cohort study of 710 athletes. The American journal of sports medicine, 2016. **44**(4): p. 874-883.
70. Petushek E, Nilstad A, Bahr R, Krosshaug T. Drop Jump? Single-Leg Squat? Not if You Aim to Predict Anterior Cruciate Ligament Injury From Real-Time Clinical Assessment: A Prospective Cohort Study Involving 880 Elite Female Athletes. J Orthop Sports Phys Ther, 2021. **51**(7): p. 372-378. https://doi.org/10.2519/jospt.2021.10170
71. Alrayani H, Herrington L, Liu A, Jones R. Frontal plane projection angle predicts patellofemoral pain: Prospective study in male military cadets. Phys Ther Sport, 2023. **59**: p. 73-79. https://doi.org/10.1016/j.ptsp.2022.12.004
72. Plisky PJ, Rauh MJ, Kaminski TW, Underwood FB. Star Excursion Balance Test as a predictor of lower extremity injury in high school basketball players. J Orthop Sports Phys Ther, 2006. **36**(12): p. 911-9. https://doi.org/10.2519/jospt.2006.2244
73. Eckard T, Padua D, Mauntel T, et al. Association between double-leg squat and single-leg squat performance and injury incidence among incoming NCAA Division I athletes: A prospective cohort study. Phys Ther Sport, 2018. **34**: p. 192-200. https://doi.org/10.1016/j.ptsp.2018.10.009

74. Bahr R. Why screening tests to predict injury do not work-and probably never will...: a critical review. Br J Sports Med, 2016. **50**(13): p. 776-80. https://doi.org/10.1136/bjsports-2016-096256
75. Hewett TE. Response to: 'Why screening tests to predict injury do not work-and probably never will...: a critical review'. Br J Sports Med, 2016. https://doi.org/10.1136/bjsports-2016-096388
76. Verhagen E, van Dyk N, Clark N, Shrier I. Do not throw the baby out with the bathwater; screening can identify meaningful risk factors for sports injuries. Br J Sports Med, 2018. **52**(19): p. 1223-1224. https://doi.org/10.1136/bjsports-2017-098547
77. Dunlop G, Ardern CL, Andersen TE, et al. Return-to-Play Practices Following Hamstring Injury: A Worldwide Survey of 131 Premier League Football Teams. Sports Med, 2019. https://doi.org/10.1007/s40279-019-01199-2
78. Webster KE, Hewett TE. Meta-analysis of meta-analyses of anterior cruciate ligament injury reduction training programs. J Orthop Res, 2018. **36**(10): p. 2696-2708. https://doi.org/10.1002/jor.24043
79. Bizzini M, Dvorak J. FIFA 11+: an effective programme to prevent football injuries in various player groups worldwide-a narrative review. Br J Sports Med, 2015. **49**(9): p. 577-9. https://doi.org/10.1136/bjsports-2015-094765
80. Impellizzeri FM, Bizzini M, Dvorak J, Pellegrini B, Schena F, Junge A. Physiological and performance responses to the FIFA 11+ (part 2): a randomised controlled trial on the training effects. J Sports Sci, 2013. **31**(13): p. 1491-502. https://doi.org/10.1080/02640414.2013.802926
81. Hägglund M, Waldén M, Atroshi I. Preventing knee injuries in adolescent female football players - design of a cluster randomized controlled trial [NCT00894595]. BMC Musculoskelet Disord, 2009. **10**: p. 75. https://doi.org/10.1186/1471-2474-10-75
82. Mandelbaum BR, Silvers HJ, Watanabe DS, et al. Effectiveness of a neuromuscular and proprioceptive training program in preventing anterior cruciate ligament injuries in female athletes: 2-year follow-up. Am J Sports Med, 2005. **33**(7): p. 1003-10. https://doi.org/10.1177/0363546504272261
83. Renstrom P, Ljungqvist A, Arendt E, et al. Non-contact ACL injuries in female athletes: an International Olympic Committee current concepts statement. Br J Sports Med, 2008. **42**(6): p. 394-412. https://doi.org/10.1136/bjsm.2008.048934
84. Seyedi M, Zarei M, Daneshjoo A, et al. Effects of FIFA 11 + warm-up program on kinematics and proprioception in adolescent soccer players: a parallel-group randomized control trial. Sci Rep, 2023. **13**(1): p. 5527. https://doi.org/10.1038/s41598-023-32774-3
85. Jeong J, Choi DH, Shin CS. Core Strength Training Can Alter Neuromuscular and Biomechanical Risk Factors for Anterior Cruciate Ligament Injury. Am J Sports Med, 2021. **49**(1): p. 183-192. https://doi.org/10.1177/0363546520972990
86. Mozafaripour E, Seidi F, Minoonejad H, Bayattork M, Khoshroo F. The effectiveness of the comprehensive corrective exercise program on kinematics and strength of lower extremities in males with dynamic knee valgus: a parallel-group randomized wait-list controlled trial. BMC Musculoskelet Disord, 2022. **23**(1): p. 700. https://doi.org/10.1186/s12891-022-05652-8
87. Cook G, Burton L, Hoogenboom B. Pre-participation screening: the use of fundamental movements as an assessment of function - part 2. N Am J Sports Phys Ther, 2006. **1**(3): p. 132-9.
88. Frohm A, Heijne A, Kowalski J, Svensson P, Myklebust G. A nine-test screening battery for athletes: a reliability study. Scand J Med Sci Sports, 2012. **22**(3): p. 306-15. https://doi.org/10.1111/j.1600-0838.2010.01267.x

89. McGill S, Frost D, Andersen J, Crosby I, Gardiner D. Movement quality and links to measures of fitness in firefighters. Work, 2013. **45**(3): p. 357-66. https://doi.org/10.3233/wor-121538
90. Bennett H, Arnold J, Martin M, Norton K, Davison K. A randomised controlled trial of movement quality-focused exercise versus traditional resistance exercise for improving movement quality and physical performance in trained adults. J Sports Sci, 2019. **37**(24): p. 2806-2817. https://doi.org/10.1080/02640414.2019.1665234
91. Bennett H, Davison K, Arnold J, Slattery F, Martin M, Norton K. Multicomponent Musculoskeletal Movement Assessment Tools: A Systematic Review and Critical Appraisal of Their Development and Applicability to Professional Practice. J Strength Cond Res, 2017. **31**(10): p. 2903-2919. https://doi.org/10.1519/jsc.0000000000002058
92. Sahrmann S, *Diagnosis and treatment of movement impairment syndromes*. 2002, St. Louis, Mo.: Mosby.
93. Hollman JH, Galardi CM, Lin IH, Voth BC, Whitmarsh CL. Frontal and transverse plane hip kinematics and gluteus maximus recruitment correlate with frontal plane knee kinematics during single-leg squat tests in women. Clin Biomech (Bristol, Avon), 2014. **29**(4): p. 468-74. https://doi.org/10.1016/j.clinbiomech.2013.12.017
94. Hollman JH, Ginos BE, Kozuchowski J, Vaughn AS, Krause DA, Youdas JW. Relationships between knee valgus, hip-muscle strength, and hip-muscle recruitment during a single-limb step-down. J Sport Rehabil, 2009. **18**(1): p. 104-17.
95. Chmielewski TL, Hodges MJ, Horodyski M, Bishop MD, Conrad BP, Tillman SM. Investigation of clinician agreement in evaluating movement quality during unilateral lower extremity functional tasks: a comparison of 2 rating methods. J Orthop Sports Phys Ther, 2007. **37**(3): p. 122-9. https://doi.org/10.2519/jospt.2007.2457
96. He X, Chow MCS, Qiu J, et al. Knee wobbling during the single-leg-squat-and-hold test reflects dynamic knee instability in patients with anterior cruciate ligament injury. Res Sports Med, 2022: p. 1-12. https://doi.org/10.1080/15438627.2022.2113879
97. Ford KR, Myer GD, Smith RL, Vianello RM, Seiwert SL, Hewett TE. A comparison of dynamic coronal plane excursion between matched male and female athletes when performing single leg landings. Clin Biomech (Bristol, Avon), 2006. **21**(1): p. 33-40. https://doi.org/10.1016/j.clinbiomech.2005.08.010
98. McLean SG, Lipfert SW, van den Bogert AJ. Effect of gender and defensive opponent on the biomechanics of sidestep cutting. Med Sci Sports Exerc, 2004. **36**(6): p. 1008-16. https://doi.org/10.1249/01.mss.0000128180.51443.83
99. Fukuda W, Kawamura K, Yokoyama S, et al. A cross-sectional study to assess variability in knee frontal plane movement during single leg squat in patients with anterior cruciate ligament injury. J Bodyw Mov Ther, 2021. **28**: p. 144-149. https://doi.org/10.1016/j.jbmt.2021.07.016
100. Frost D, Andersen J, Lam T, Finlay T, Darby K, McGill S. The relationship between general measures of fitness, passive range of motion and whole-body movement quality. Ergonomics, 2013. **56**(4): p. 637-49. https://doi.org/10.1080/00140139.2011.620177
101. van Melick N, van Cingel RE, Brooijmans F, et al. Evidence-based clinical practice update: practice guidelines for anterior cruciate ligament rehabilitation based on a systematic review and multidisciplinary consensus. Br J Sports Med, 2016. **50**(24): p. 1506-1515. https://doi.org/10.1136/bjsports-2015-095898
102. Dos'Santos T, McBurnie A, Comfort P, Jones PA. The Effects of Six-Weeks Change of Direction Speed and Technique Modification Training on Cutting Performance and

Movement Quality in Male Youth Soccer Players. Sports (Basel), 2019. **7**(9). https://doi.org/10.3390/sports7090205

103. Munro A, Herrington L, Carolan M. Reliability of 2-dimensional video assessment of frontal-plane dynamic knee valgus during common athletic screening tasks. J Sport Rehabil, 2012. **21**(1): p. 7-11.
104. Bernhardt J, Bate PJ, Matyas TA. Accuracy of observational kinematic assessment of upper-limb movements. Phys Ther, 1998. **78**(3): p. 259-70. https://doi.org/10.1093/ptj/78.3.259
105. Whatman C, Hume P, Hing W. The reliability and validity of visual rating of dynamic alignment during lower extremity functional screening tests: a review of the literature. Physical Therapy Reviews, 2015. **20**(3): p. 210-224. https://doi.org/10.1179/1743288x15y.0000000006
106. Carlsson H, Rasmussen-Barr E. Clinical screening tests for assessing movement control in non-specific low-back pain. A systematic review of intra- and inter-observer reliability studies. Man Ther, 2013. **18**(2): p. 103-10. https://doi.org/10.1016/j.math.2012.08.004
107. Granstrom HMR, Ang BOPR, Rasmussen-Barr EPR. Movement control tests for the lumbopelvic complex. Are these tests reliable and valid? Physiother Theory Pract, 2017. **33**(5): p. 386-397. https://doi.org/10.1080/09593985.2017.1318422
108. Monnier A, Heuer J, Norman K, Ang BO. Inter- and intra-observer reliability of clinical movement-control tests for marines. BMC Musculoskelet Disord, 2012. **13**: p. 263. https://doi.org/10.1186/1471-2474-13-263
109. Knudson D. What can professionals qualitatively analyze? Journal of Physical Education, Recreation & Dance, 2000. **71**(2): p. 19-23.
110. Plisky P, Schwartkopf-Phifer K, Huebner B, Garner MB, Bullock G. Systematic Review and Meta-Analysis of the Y-Balance Test Lower Quarter: Reliability, Discriminant Validity, and Predictive Validity. Int J Sports Phys Ther, 2021. **16**(5): p. 1190-1209. https://doi.org/10.26603/001c.27634
111. Herrington L, Alenezi F, Alzhrani M, Alrayani H, Jones R. The reliability and criterion validity of 2D video assessment of single leg squat and hop landing. J Electromyogr Kinesiol, 2017. **34**: p. 80-85. https://doi.org/10.1016/j.jelekin.2017.04.004
112. Puig-Diví A, Escalona-Marfil C, Padullés-Riu JM, Busquets A, Padullés-Chando X, Marcos-Ruiz D. Validity and reliability of the Kinovea program in obtaining angles and distances using coordinates in 4 perspectives. PLoS One, 2019. **14**(6): p. e0216448. https://doi.org/10.1371/journal.pone.0216448
113. Thewlis D, Bishop C, Daniell N, Paul G. Next-generation low-cost motion capture systems can provide comparable spatial accuracy to high-end systems. J Appl Biomech, 2013. **29**(1): p. 112-7. https://doi.org/10.1123/jab.29.1.112
114. Lopes TJA, Ferrari D, Ioannidis J, Simic M, Mícolis de Azevedo F, Pappas E. Reliability and Validity of Frontal Plane Kinematics of the Trunk and Lower Extremity Measured With 2-Dimensional Cameras During Athletic Tasks: A Systematic Review With Meta-analysis. J Orthop Sports Phys Ther, 2018. **48**(10): p. 812-822. https://doi.org/10.2519/jospt.2018.8006
115. Lally EM, Ericksen H, Earl-Boehm J. Measurement Properties of Clinically Accessible Movement Assessment Tools for Analyzing Single-Leg Squats and Step-Downs: A Systematic Review. J Sport Rehabil, 2022. **31**(4): p. 476-489. https://doi.org/10.1123/jsr.2021-0287
116. Lam WWT, Tang YM, Fong KNK. A systematic review of the applications of markerless motion capture (MMC) technology for clinical measurement in rehabilitation. J Neuroeng Rehabil, 2023. **20**(1): p. 57. https://doi.org/10.1186/s12984-023-01186-9

117. Mauntel TC, Cameron KL, Pietrosimone B, Marshall SW, Hackney AC, Padua DA. Validation of a Commercially Available Markerless Motion-Capture System for Trunk and Lower Extremity Kinematics During a Jump-Landing Assessment. J Athl Train, 2021. **56**(2): p. 177-190. https://doi.org/10.4085/1062-6050-0023.20
118. Shotton J, Fitzgibbon A, Cook M, et al. *Real-time human pose recognition in parts from single depth images*. in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. 2011. Ieee.
119. Yang Y, Pu F, Li Y, Li S, Fan Y, Li D. Reliability and validity of Kinect RGB-D sensor for assessing standing balance. IEEE Sensors Journal, 2014. **14**(5): p. 1633-1638.
120. Scott B, Seyres M, Philp F, Chadwick EK, Blana D. Healthcare applications of single camera markerless motion capture: a scoping review. PeerJ, 2022. **10**: p. e13517. https://doi.org/10.7717/peerj.13517
121. Clark RA, Mentiplay BF, Hough E, Pua YH. Three-dimensional cameras and skeleton pose tracking for physical function assessment: A review of uses, validity, current developments and Kinect alternatives. Gait Posture, 2019. **68**: p. 193-200. https://doi.org/10.1016/j.gaitpost.2018.11.029
122. Eltoukhy M, Kelly A, Kim CY, Jun HP, Campbell R, Kuenze C. Validation of the Microsoft Kinect(R) camera system for measurement of lower extremity jump landing and squatting kinematics. Sports Biomech, 2016. **15**(1): p. 89-102. https://doi.org/10.1080/14763141.2015.1123766
123. Eltoukhy M, Kuenze C, Oh J, Wooten S, Signorile J. Kinect-based assessment of lower limb kinematics and dynamic postural control during the star excursion balance test. Gait Posture, 2017. **58**: p. 421-427. https://doi.org/10.1016/j.gaitpost.2017.09.010
124. Kotsifaki A, Whiteley R, Hansen C. Dual Kinect v2 system can capture lower limb kinematics reasonably well in a clinical setting: concurrent validity of a dual camera markerless motion capture system in professional football players. BMJ Open Sport Exerc Med, 2018. **4**(1): p. e000441. https://doi.org/10.1136/bmjsem-2018-000441
125. Mentiplay BF, Hasanki K, Perraton LG, Pua YH, Charlton PC, Clark RA. Three-dimensional assessment of squats and drop jumps using the Microsoft Xbox One Kinect: Reliability and validity. J Sports Sci, 2018. **36**(19): p. 2202-2209. https://doi.org/10.1080/02640414.2018.1445439
126. Schmitz A, Ye M, Boggess G, Shapiro R, Yang R, Noehren B. The measurement of in vivo joint angles during a squat using a single camera markerless motion capture system as compared to a marker based system. Gait Posture, 2015. **41**(2): p. 694-8. https://doi.org/10.1016/j.gaitpost.2015.01.028
127. Wochatz M, Tilgner N, Mueller S, et al. Reliability and validity of the Kinect V2 for the assessment of lower extremity rehabilitation exercises. Gait Posture, 2019. **70**: p. 330-335. https://doi.org/10.1016/j.gaitpost.2019.03.020
128. Qinematic™: Company website. 2018 [Accessed 14 August 2019]; Available from: https://www.qinematic.com/
129. Ageberg E, Bennell KL, Hunt MA, Simic M, Roos EM, Creaby MW. Validity and inter-rater reliability of medio-lateral knee motion observed during a single-limb mini squat. BMC Musculoskelet Disord, 2010. **11**: p. 265. https://doi.org/10.1186/1471-2474-11-265
130. Crossley KM, Zhang WJ, Schache AG, Bryant A, Cowan SM. Performance on the single-leg squat task indicates hip abductor muscle function. Am J Sports Med, 2011. **39**(4): p. 866-73. https://doi.org/10.1177/0363546510395456
131. Edmondston S, Leo Y, Trant B, Vatna R, Kendell M, Smith A. Symmetry of trunk and femoro-pelvic movement responses to single leg loading tests in asymptomatic females. Man Ther, 2013. **18**(3): p. 231-6. https://doi.org/10.1016/j.math.2012.10.010

132. Stensrud S, Myklebust G, Kristianslund E, Bahr R, Krosshaug T. Correlation between two-dimensional video analysis and subjective assessment in evaluating knee control among elite female team handball players. Br J Sports Med, 2011. **45**(7): p. 589-95. https://doi.org/10.1136/bjsm.2010.078287
133. Khuu A, Foch E, Lewis CL. NOT ALL SINGLE LEG SQUATS ARE EQUAL: A BIOMECHANICAL COMPARISON OF THREE VARIATIONS. Int J Sports Phys Ther, 2016. **11**(2): p. 201-11.
134. Olivier B, Quinn SL, Benjamin N, Green AC, Chiu J, Wang W. Single-Leg Squat Delicacies-The Position of the Nonstance Limb is an Important Consideration. J Sport Rehabil, 2019. **28**(4): p. 318-324. https://doi.org/10.1123/jsr.2018-0181
135. Lewis CL, Foch E, Luko MM, Loverro KL, Khuu A. Differences in lower extremity and trunk kinematics between single leg squat and step down tasks. PloS one, 2015. **10**(5): p. e0126258.
136. Martonick NJ, McGowan CP, Baker RT, Larkins LW, Seegmiller JG, Bailey JP. Comparison of Three Single Leg Weightbearing Tasks with Statistical Parametric Mapping. Biomechanics, 2022. **2**(4): p. 591-600.
137. Batty LM, Feller JA, Damasena I, et al. Single-Leg Squat After Anterior Cruciate Ligament Reconstruction: An Analysis of the Knee Valgus Angle at 6 and 12 Months. Orthop J Sports Med, 2020. **8**(8): p. 2325967120946328. https://doi.org/10.1177/2325967120946328
138. Hall MP, Paik RS, Ware AJ, Mohr KJ, Limpisvasti O. Neuromuscular Evaluation With Single-Leg Squat Test at 6 Months After Anterior Cruciate Ligament Reconstruction. Orthop J Sports Med, 2015. **3**(3): p. 2325967115575900. https://doi.org/10.1177/2325967115575900
139. Yamazaki J, Muneta T, Ju YJ, Koga H, Morito T, Sekiya I. The kinematic analysis of female subjects after double-bundle anterior cruciate ligament reconstruction during single-leg squatting. J Orthop Sci, 2013. **18**(2): p. 284-9. https://doi.org/10.1007/s00776-012-0350-5
140. Morita Â K, Tavella Navega M. Women with patellofemoral pain show changes in trunk and lower limb sagittal movements during single-leg squat and step-down tasks. Physiother Theory Pract, 2023: p. 1-9. https://doi.org/10.1080/09593985.2023.2228396
141. Zawadka M, Smolka J, Skublewska-Paszkowska M, et al. Sex-dependent differences in single-leg squat kinematics and their relationship to squat depth in physically active individuals. Sci Rep, 2020. **10**(1): p. 19601. https://doi.org/10.1038/s41598-020-76674-2
142. Baldon Rde M, Lobato DF, Carvalho LP, Santiago PR, Benze BG, Serrao FV. Relationship between eccentric hip torque and lower-limb kinematics: gender differences. J Appl Biomech, 2011. **27**(3): p. 223-32.
143. Weeks BK, Carty CP, Horan SA. Effect of sex and fatigue on single leg squat kinematics in healthy young adults. BMC Musculoskelet Disord, 2015. **16**: p. 271. https://doi.org/10.1186/s12891-015-0739-3
144. Zeller BL, McCrory JL, Kibler WB, Uhl TL. Differences in kinematics and electromyographic activity between men and women during the single-legged squat. Am J Sports Med, 2003. **31**(3): p. 449-56. https://doi.org/10.1177/03635465030310032101
145. Graci V, Van Dillen LR, Salsich GB. Gender differences in trunk, pelvis and lower limb kinematics during a single leg squat. Gait Posture, 2012. **36**(3): p. 461-6. https://doi.org/10.1016/j.gaitpost.2012.04.006

146. Construct Validity-Definition, Types, & Examples. Accessed 15 November 2023]; Available from: https://www.scribbr.com/methodology/construct-validity/
147. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res, 2010. **19**(4): p. 539-49. https://doi.org/10.1007/s11136-010-9606-8
148. Higgins JPT TJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane, 2023. [cited 2023 9 September]; Available from: Available from www.training.cochrane.org/handbook.
149. de Vet HCWd, Terwee CB, Mokkink LB, Knol DL, *Measurement in medicine : a practical guide*. 2011, Cambridge ;: Cambridge University Press.
150. Hopkins WG. Measures of reliability in sports medicine and science. Sports Med, 2000. **30**(1): p. 1-15.
151. Alenezi F, Herrington L, Jones P, Jones R. The reliability of biomechanical variables collected during single leg squat and landing tasks. J Electromyogr Kinesiol, 2014. **24**(5): p. 718-21. https://doi.org/10.1016/j.jelekin.2014.07.007
152. Jamaludin NI, Sahabuddin FNA, Rasudin NS, Shaharudin S. The Concurrent Validity and Reliability of Single Leg Squat Among Physically Active Females with and without Dynamic Knee Valgus. Int J Sports Phys Ther, 2022. **17**(4): p. 574-584. https://doi.org/10.26603/001c.35706
153. Levinger P, Gilleard W, Coleman C. Femoral medial deviation angle during a one-leg squat test in individuals with patellofemoral pain syndrome. Physical Therapy in sport, 2007. **8**(4): p. 163-168.
154. Whatman C, Hing W, Hume P. Kinematics during lower extremity functional screening tests--are they reliable and related to jogging? Phys Ther Sport, 2011. **12**(1): p. 22-9. https://doi.org/10.1016/j.ptsp.2010.10.006
155. Whatman C, Hume P, Hing W. The reliability and validity of physiotherapist visual rating of dynamic pelvis and knee alignment in young athletes. Phys Ther Sport, 2013. **14**(3): p. 168-74. https://doi.org/10.1016/j.ptsp.2012.07.001
156. Nae J, Creaby MW, Cronstrom A, Ageberg E. Measurement properties of visual rating of postural orientation errors of the lower extremity - A systematic review and meta-analysis. Phys Ther Sport, 2017. https://doi.org/10.1016/j.ptsp.2017.04.003
157. DiMattia MA, Livengood AL, Uhl TL, Mattacola CG, Malone TR. What are the validity of the single-leg-squat test and its relationship to hip-abduction strength. J Sport Rehabil, 2005. **14**(2): p. 108-123.
158. Fox AS, Bonacci J, Saunders N. The relationship between performance of a single-leg squat and leap landing task: moving towards a netball-specific anterior cruciate ligament (ACL) injury risk screening method. Sports Biomech, 2020. **19**(4): p. 493-509. https://doi.org/10.1080/14763141.2018.1498535
159. Horan SA, Watson SL, Carty CP, Sartori M, Weeks BK. Lower-limb kinematics of single-leg squat performance in young adults. Physiother Can, 2014. **66**(3): p. 228-33. https://doi.org/10.3138/ptc.2013-09
160. Jones D, Tillman SM, Tofte K, et al. Observational ratings of frontal plane knee position are related to the frontal plane projection angle but not the knee abduction angle during a step-down task. journal of orthopaedic & sports physical therapy, 2014. **44**(12): p. 973-978.
161. Mauntel TC, Frank BS, Begalle RL, Blackburn JT, Padua DA. Kinematic differences between those with and without medial knee displacement during a single-leg squat. J Appl Biomech, 2014. **30**(6): p. 707-12. https://doi.org/10.1123/jab.2014-0003

162. Perrott MA, Pizzari T, Opar MS, Cook J. Athletes with a clinical rating of good and poor lumbopelvic stability have different kinematic variables during single leg squat and dip test. Physiother Theory Pract, 2019: p. 1-10. https://doi.org/10.1080/09593985.2019.1655823
163. Rabin A, Portnoy S, Kozol Z. THE ASSOCIATION BETWEEN VISUAL ASSESSMENT OF QUALITY OF MOVEMENT AND THREE-DIMENSIONAL ANALYSIS OF PELVIS, HIP, AND KNEE KINEMATICS DURING A LATERAL STEP DOWN TEST. Journal of Strength and Conditioning Research, 2016. **30**(11): p. 3204-3211. https://doi.org/10.1519/jsc.0000000000001420
164. Gomes DA, da Costa GV, Martins EC, et al. Are visual assessments of the single-leg squat valid to be used in clinical practice? A systematic review of measurement properties based on the COSMIN guideline. Phys Ther Sport, 2023. **63**: p. 118-125. https://doi.org/10.1016/j.ptsp.2023.07.009
165. O'Connor S, McCaffrey N, Whyte EF, Moran KA. Can a Standardized Visual Assessment of Squatting Technique and Core Stability Predict Injury? J Strength Cond Res, 2020. **34**(1): p. 26-36. https://doi.org/10.1519/jsc.0000000000003262
166. McGovern RP, Martin RL, Christoforetti JJ, Kivlan BR. EVIDENCE-BASED PROCEDURES FOR PERFORMING THE SINGLE LEG SQUAT AND STEP-DOWN TESTS IN EVALUATION OF NON-ARTHRITIC HIP PAIN: A LITERATURE REVIEW. Int J Sports Phys Ther, 2018. **13**(3): p. 526-536.
167. da Costa GV, de Castro MP, Sanchotene CG, Ribeiro DC, de Brito Fontana H, Ruschel C. Relationship between passive ankle dorsiflexion range, dynamic ankle dorsiflexion range and lower limb and trunk kinematics during the single-leg squat. Gait Posture, 2021. **86**: p. 106-111. https://doi.org/10.1016/j.gaitpost.2021.03.015
168. Park KM, Cynn HS, Choung SD. Musculoskeletal predictors of movement quality for the forward step-down test in asymptomatic women. J Orthop Sports Phys Ther, 2013. **43**(7): p. 504-10. https://doi.org/10.2519/jospt.2013.4073
169. Rabin A, Kozol Z. Measures of range of motion and strength among healthy women with differing quality of lower extremity movement during the lateral step-down test. J Orthop Sports Phys Ther, 2010. **40**(12): p. 792-800. https://doi.org/10.2519/jospt.2010.3424
170. Rabin A, Kozol Z, Moran U, Efergan A, Geffen Y, Finestone AS. Factors associated with visually assessed quality of movement during a lateral step-down test among individuals with patellofemoral pain. J Orthop Sports Phys Ther, 2014. **44**(12): p. 937-46. https://doi.org/10.2519/jospt.2014.5507
171. Grindstaff TL, Dolan N, Morton SK. Ankle dorsiflexion range of motion influences Lateral Step Down Test scores in individuals with chronic ankle instability. Phys Ther Sport, 2017. **23**: p. 75-81. https://doi.org/10.1016/j.ptsp.2016.07.008
172. Rabin A, Kozol Z, Spitzer E, Finestone A. Ankle dorsiflexion among healthy men with different qualities of lower extremity movement. J Athl Train, 2014. **49**(5): p. 617-23. https://doi.org/10.4085/1062-6050-49.3.14
173. Carroll LA, Kivlan BR, Martin RL, Phelps AL, Carcia CR. The Single Leg Squat Test: A "Top-Down" or "Bottom-Up" Functional Performance Test? Int J Sports Phys Ther, 2021. **16**(2): p. 360-370. https://doi.org/10.26603/001c.21317
174. Karimi K, Seidi F, Mousavi SH, Alghosi M, Morad NH. Comparison of postural sway in individuals with and without dynamic knee valgus. BMC Sports Sci Med Rehabil, 2023. **15**(1): p. 75. https://doi.org/10.1186/s13102-023-00686-4
175. Weldring T, Smith SM. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health Serv Insights, 2013. **6**: p. 61-8. https://doi.org/10.4137/hsi.S11093

176. Forsdyke D, Smith A, Jones M, Gledhill A. Psychosocial factors associated with outcomes of sports injury rehabilitation in competitive athletes: a mixed studies systematic review. Br J Sports Med, 2016. **50**(9): p. 537-44. https://doi.org/10.1136/bjsports-2015-094850
177. van de Wouw A. Advocating a holistic approach for sport injury prevention and rehabilitation. Br J Sports Med, 2023. https://doi.org/10.1136/bjsports-2022-105695
178. Ardern CL, Taylor NF, Feller JA, Webster KE. Fear of re-injury in people who have returned to sport following anterior cruciate ligament reconstruction surgery. J Sci Med Sport, 2012. **15**(6): p. 488-95. https://doi.org/10.1016/j.jsams.2012.03.015
179. Kvist J, Ek A, Sporrstedt K, Good L. Fear of re-injury: a hindrance for returning to sports after anterior cruciate ligament reconstruction. Knee Surg Sports Traumatol Arthrosc, 2005. **13**(5): p. 393-7. https://doi.org/10.1007/s00167-004-0591-8
180. Flanigan DC, Everhart JS, Pedroza A, Smith T, Kaeding CC. Fear of reinjury (kinesiophobia) and persistent knee symptoms are common factors for lack of return to sport after anterior cruciate ligament reconstruction. Arthroscopy, 2013. **29**(8): p. 1322-9. https://doi.org/10.1016/j.arthro.2013.05.015
181. Rasmussen-Barr E, Lindqvist C, Östhols S, Boström C. Are patient reported outcome measures (PROMs) useful in low back pain? Experiences of physiotherapists in primary health care in Sweden. Musculoskelet Sci Pract, 2021. **55**: p. 102414. https://doi.org/10.1016/j.msksp.2021.102414
182. Östhols S, Boström C, Rasmussen-Barr E. Clinical assessment and patient-reported outcome measures in low-back pain - a survey among primary health care physiotherapists. Disabil Rehabil, 2019. **41**(20): p. 2459-2467. https://doi.org/10.1080/09638288.2018.1467503
183. Rowland DL, Van Lankveld JJ. Anxiety and performance in sex, sport, and stage: Identifying common ground. Frontiers in psychology, 2019. **10**: p. 1615.
184. Rice SM, Gwyther K, Santesteban-Echarri O, et al. Determinants of anxiety in elite athletes: a systematic review and meta-analysis. Br J Sports Med, 2019. **53**(11): p. 722-730. https://doi.org/10.1136/bjsports-2019-100620
185. Burland JP, Toonstra JL, Howard JS. Psychosocial Barriers After Anterior Cruciate Ligament Reconstruction: A Clinical Review of Factors Influencing Postoperative Success. Sports Health, 2019. **11**(6): p. 528-534. https://doi.org/10.1177/1941738119869333
186. WMA DECLARATION OF HELSINKI – ETHICAL PRINCIPLES FOR MEDICAL RESEARCH INVOLVING HUMAN SUBJECTS. Accessed 4 November 2023]; Available from: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/
187. U.S National Library of Medicine, ClinicalTrials.gov. Accessed 17 January 2023]; Available from: https://clinicaltrials.gov/
188. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. Bmj, 2009. **339**: p. b2700. https://doi.org/10.1136/bmj.b2700
189. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS med, 2009. **6**(7): p. e1000097.
190. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). J Clin Epidemiol, 2010. **63**(8): p. 854-61. https://doi.org/10.1016/j.jclinepi.2009.10.002

191. Lucas N, Macaskill P, Irwig L, et al. The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). BMC medical research methodology, 2013. **13**(1): p. 111.
192. Grooten WJA, Sandberg L, Ressman J, Diamantoglou N, Johansson E, Rasmussen-Barr E. Reliability and validity of a novel Kinect-based software program for measuring posture, balance and side-bending. BMC Musculoskelet Disord, 2018. **19**(1): p. 6. https://doi.org/10.1186/s12891-017-1927-0
193. Gabbard C, Hart S. A question of foot dominance. J Gen Psychol, 1996. **123**(4): p. 289-96. https://doi.org/10.1080/00221309.1996.9921281
194. McGrath TM, Waddington G, Scarvell JM, et al. The effect of limb dominance on lower limb functional performance--a systematic review. J Sports Sci, 2016. **34**(4): p. 289-302. https://doi.org/10.1080/02640414.2015.1050601
195. Clarsen B, Bahr R, Myklebust G, et al. Improved reporting of overuse injuries and health problems in sport: an update of the Oslo Sport Trauma Research Center questionnaires. Br J Sports Med, 2020. https://doi.org/10.1136/bjsports-2019-101337
196. Clarsen B, Myklebust G, Bahr R. Development and validation of a new method for the registration of overuse injuries in sports injury epidemiology: the Oslo Sports Trauma Research Centre (OSTRC) overuse injury questionnaire. Br J Sports Med, 2013. **47**(8): p. 495-502.
197. Howe LP, Bampouras TM, North JS, Waldron M. WITHIN-SESSION RELIABILITY FOR INTER-LIMB ASYMMETRIES IN ANKLE DORSIFLEXION RANGE OF MOTION MEASURED DURING THE WEIGHT-BEARING LUNGE TEST. Int J Sports Phys Ther, 2020. **15**(1): p. 64-73.
198. Langarika-Rocafort A, Emparanza JI, Aramendi JF, Castellano J, Calleja-González J. Intra-rater reliability and agreement of various methods of measurement to assess dorsiflexion in the Weight Bearing Dorsiflexion Lunge Test (WBLT) among female athletes. Phys Ther Sport, 2017. **23**: p. 37-44. https://doi.org/10.1016/j.ptsp.2016.06.010
199. Aramaki H, Katoh M, Hiiragi Y, Kawasaki T, Kurihara T, Ohmi Y. Validity and reliability of isometric muscle strength measurements of hip abduction and abduction with external hip rotation in a bent-hip position using a handheld dynamometer with a belt. J Phys Ther Sci, 2016. **28**(7): p. 2123-7. https://doi.org/10.1589/jpts.28.2123
200. Bazett-Jones DM, Squier K. Measurement properties of hip strength measured by handheld dynamometry: Reliability and validity across the range of motion. Phys Ther Sport, 2020. **42**: p. 100-106. https://doi.org/10.1016/j.ptsp.2020.01.005
201. Bazett-Jones DM, Tylinksi T, Krstic J, Stromquist A, Sparks J. PEAK HIP MUSCLE TORQUE MEASUREMENTS ARE INFLUENCED BY SAGITTAL PLANE HIP POSITION. Int J Sports Phys Ther, 2017. **12**(4): p. 535-542.
202. Eklund M, Bäckström M, Tuvesson H. Psychometric properties and factor structure of the Swedish version of the Perceived Stress Scale. Nord J Psychiatry, 2014. **68**(7): p. 494-9. https://doi.org/10.3109/08039488.2013.877072
203. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav, 1983. **24**(4): p. 385-96.
204. Dietch JR, Taylor DJ, Sethi K, Kelly K, Bramoweth AD, Roane BM. Psychometric Evaluation of the PSQI in U.S. College Students. J Clin Sleep Med, 2016. **12**(8): p. 1121-9. https://doi.org/10.5664/jcsm.6050
205. Buysse DJ, Reynolds CF, 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res, 1989. **28**(2): p. 193-213. https://doi.org/10.1016/0165-1781(89)90047-4

206. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med, 2006. **166**(10): p. 1092-7. https://doi.org/10.1001/archinte.166.10.1092
207. fbanken, Generalized Anxiety Disorder 7-item scale GAD-7. Accessed 17 January 2023]; Available from: https://fbanken.se/form/218/generalised-anxiety-disorder-7-item-scale
208. Dover G, Amar V. Development and Validation of the Athlete Fear Avoidance Questionnaire. J Athl Train, 2015. **50**(6): p. 634-42. https://doi.org/10.4085/1062-6050-49.3.75
209. Ressman J, Rasmussen-Barr E, Grooten WJA. Reliability and validity of a novel Kinect-based software program for measuring a single leg squat. BMC Sports Sci Med Rehabil, 2020. **12**: p. 31. https://doi.org/10.1186/s13102-020-00179-8
210. Stoltzfus JC. Logistic regression: a brief primer. Acad Emerg Med, 2011. **18**(10): p. 1099-104. https://doi.org/10.1111/j.1553-2712.2011.01185.x
211. Hosmer DW, Lemeshow S, Sturdivant RX, *Applied logistic regression*. 2013, Hoboken, N.J.: Wiley.
212. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR, *Introduction to Meta-Analysis*. 2011: Wiley.
213. Botella J, Suero M, Gambara H. Psychometric inferences from a meta-analysis of reliability and internal consistency coefficients. Psychol Methods, 2010. **15**(4): p. 386-97. https://doi.org/10.1037/a0019626
214. Cuchna JW, Hoch MC, Hoch JM. The interrater and intrarater reliability of the functional movement screen: A systematic review with meta-analysis. Phys Ther Sport, 2016. **19**: p. 57-65. https://doi.org/10.1016/j.ptsp.2015.12.002
215. Ottenbacher KJ, Hsu Y, Granger CV, Fiedler RC. The reliability of the functional independence measure: a quantitative review. Arch Phys Med Rehabil, 1996. **77**(12): p. 1226-32.
216. Pierce CA, *Software Review: Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2006). Comprehensive Meta-Analysis (Version 2.2.027) [Computer software]. Englewood, NJ: Biostat*. 2008: Los Angeles, CA. p. 188-191.
217. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics, 1977. **33**(1): p. 159-74.
218. Dagostino RB, Belanger A, Dagostino RB. A SUGGESTION FOR USING POWERFUL AND INFORMATIVE TESTS OF NORMALITY. American Statistician, 1990. **44**(4): p. 316-321. https://doi.org/10.2307/2684359
219. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res, 2005. **19**(1): p. 231-40. https://doi.org/10.1519/15184.1
220. Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med, 2018. **18**(3): p. 91-93. https://doi.org/10.1016/j.tjem.2018.08.001
221. Chan YH. Biostatistics 104: correlational analysis. Singapore Med J, 2003. **44**(12): p. 614-9.
222. Fischer JE, Bachmann LM, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. Intensive Care Med, 2003. **29**(7): p. 1043-51. https://doi.org/10.1007/s00134-003-1761-8
223. Streiner DL, Norman GR, Cairney J, *Health measurement scales : a practical guide to their development and use*. 2015, Oxford: Oxford University Press.
224. Trevethan R. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. Front Public Health, 2017. **5**: p. 307. https://doi.org/10.3389/fpubh.2017.00307

225. Fleiss JL. Measuring nominal scale agreement among many raters. Psychological bulletin, 1971. **76**(5): p. 378.
226. Myunghee Cho. Paik P, Joseph L, *Statistical Methods for Rates and Proportions*. 3rd ed. ed. Wiley series in probability and statistics. 2003, US: Wiley-Interscience.
227. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther, 2005. **85**(3): p. 257-68.
228. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE, *Regression Methods in Biostatistics [Elektronisk resurs] Linear, Logistic, Survival, and Repeated Measures Models*. 2012, Boston, MA: Springer US.
229. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol, 2007. **165**(6): p. 710-8. https://doi.org/10.1093/aje/kwk052
230. Hosmer DW, Lemeshow S, Klar J. Goodness-of-fit testing for the logistic regression model when the estimated probabilities are small. Biometrical Journal, 1988. **30**(8): p. 911-924.
231. Lemeshow S, Hosmer Jr DW. A review of goodness of fit statistics for use in the development of logistic regression models. American journal of epidemiology, 1982. **115**(1): p. 92-106.
232. Wakkee M, Hollestein LM, Nijsten T. Multivariable analysis. J Invest Dermatol, 2014. **134**(5): p. 1-5. https://doi.org/10.1038/jid.2014.132
233. Vittinghoff E, *Regression methods in biostatistics [Elektronisk resurs] linear, logistic, survival, and repeated measures models*. 2005, New York: Springer.
234. Armstrong RA. When to use the Bonferroni correction. Ophthalmic Physiol Opt, 2014. **34**(5): p. 502-8. https://doi.org/10.1111/opo.12131
235. Klein D. Implementing a general framework for assessing interrater agreement in Stata. Stata Journal, 2018. **18**(4): p. 871-901.
236. VassarStats: Website for Statistical Computation- Kappa as a Measure of Concordance in Categorical Sorting. [cited 2020 June 1]; Available from: http://vassarstats.net/kappa.html
237. McKeown I, Taylor-McKeown K, Woods C, Ball N. Athletic ability assessment: a movement assessment protocol for athletes. International journal of sports physical therapy, 2014. **9**(7): p. 862.
238. Raisanen A, Pasanen K, Krosshaug T, Avela J, Perttunen J, Parkkari J. Single-Leg Squat as a Tool to Evaluate Young Athletes' Frontal Plane Knee Control. Clinical Journal of Sport Medicine, 2016. **26**(6): p. 478-482.
239. Weeks BK, Carty CP, Horan SA. Kinematic predictors of single-leg squat performance: a comparison of experienced physiotherapists and student physiotherapists. BMC Musculoskelet Disord, 2012. **13**: p. 207. https://doi.org/10.1186/1471-2474-13-207
240. Cornell DJ, Ebersole KT. Intra-Rater Test-Retest Reliability and Response Stability of the Fusioneticstm Movement Efficiency Test. International Journal of Sports Physical Therapy, 2018. **13**(4): p. 618-632.
241. Barker-Davies RM, Roberts A, Bennett AN, Fong DTP, Wheeler P, Lewis MP. Single leg squat ratings by clinicians are reliable and predict excessive hip internal rotation moment. Gait Posture, 2018. **61**: p. 453-458. https://doi.org/10.1016/j.gaitpost.2018.02.016
242. Ressman J, Grooten WJA, Rasmussen Barr E. Visual assessment of movement quality in the single leg squat test: a review and meta-analysis of inter-rater and intrarater reliability. BMJ Open Sport Exerc Med, 2019. **5**(1): p. e000541. https://doi.org/10.1136/bmjsem-2019-000541

243. McGovern RP, Christoforetti JJ, Martin RL, Phelps AL, Kivlan BR. Evidence for Reliability and Validity of Functional Performance Testing in the Evaluation of Nonarthritic Hip Pain. J Athl Train, 2019. **54**(3): p. 276-282. https://doi.org/10.4085/1062-6050-33-18
244. Whatman C, Toomey C, Emery C. Visual rating of movement quality in individuals with and without a history of intra-articular knee injury. Physiother Theory Pract, 2019: p. 1-7. https://doi.org/10.1080/09593985.2019.1703229
245. Silva RLE, Pinheiro YT, Lins CAA, de Oliveira RR, Scattone Silva R. Assessment of quality of movement during a lateral step-down test: Narrative review. J Bodyw Mov Ther, 2019. **23**(4): p. 835-843. https://doi.org/10.1016/j.jbmt.2019.05.012
246. Ressman J, Grooten WJA, Rasmussen-Barr E. Visual assessment of movement quality: a study on intra- and interrater reliability of a multi-segmental single leg squat test. BMC Sports Sci Med Rehabil, 2021. **13**(1): p. 66. https://doi.org/10.1186/s13102-021-00289-x
247. Fjellner A, Bexander C, Faleij R, Strender LE. Interexaminer reliability in physical examination of the cervical spine. J Manipulative Physiol Ther, 1999. **22**(8): p. 511-6.
248. Jonsson A, Rasmussen-Barr E. Intra- and inter-rater reliability of movement and palpation tests in patients with neck pain: A systematic review. Physiother Theory Pract, 2018. **34**(3): p. 165-180. https://doi.org/10.1080/09593985.2017.1390806
249. Pool JJ, Hoving JL, de Vet HC, van Mameren H, Bouter LM. The interexaminer reproducibility of physical examination of the cervical spine. J Manipulative Physiol Ther, 2004. **27**(2): p. 84-90. https://doi.org/10.1016/j.jmpt.2003.12.002
250. Stochkendahl MJ, Christensen HW, Hartvigsen J, et al. Manual examination of the spine: a systematic critical literature review of reproducibility. J Manipulative Physiol Ther, 2006. **29**(6): p. 475-85, 485.e1-10. https://doi.org/10.1016/j.jmpt.2006.06.011
251. Lange RT, Lippa SM. Sensitivity and specificity should never be interpreted in isolation without consideration of other clinical utility metrics. Clin Neuropsychol, 2017. **31**(6-7): p. 1015-1028. https://doi.org/10.1080/13854046.2017.1335438
252. Owoeye OBA, Wiley JP, Walker REA, Palacios-Derflingher L, Emery CA. Diagnostic Accuracy of a Self-report Measure of Patellar Tendinopathy in Youth Basketball. J Orthop Sports Phys Ther, 2018. **48**(10): p. 758-766. https://doi.org/10.2519/jospt.2018.8088
253. Röijezon U, Clark NC, Treleaven J. Proprioception in musculoskeletal rehabilitation. Part 1: Basic science and principles of assessment and clinical interventions. Man Ther, 2015. **20**(3): p. 368-77. https://doi.org/10.1016/j.math.2015.01.008
254. Alahmad TA, Kearney P, Cahalan R. Injury in elite women's soccer: a systematic review. Phys Sportsmed, 2020. **48**(3): p. 259-265. https://doi.org/10.1080/00913847.2020.1720548
255. Hägglund M, Waldén M, Ekstrand J. Previous injury as a risk factor for injury in elite football: a prospective study over two consecutive seasons. Br J Sports Med, 2006. **40**(9): p. 767-72. https://doi.org/10.1136/bjsm.2006.026609
256. Sonesson S, Lindblom H, Hägglund M. Higher age and present injury at the start of the season are risk factors for in-season injury in amateur male and female football players-a prospective cohort study. Knee Surg Sports Traumatol Arthrosc, 2023. **31**(10): p. 4618-4630. https://doi.org/10.1007/s00167-023-07517-6
257. 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**(6): p. 545-7. https://doi.org/10.1136/bjsm.37.6.545
258. Junge A, Dvorak J. Influence of definition and data collection on the incidence of injuries in football. Am J Sports Med, 2000. **28**(5 Suppl): p. S40-6. https://doi.org/10.1177/28.suppl_5.s-40

259. Tesarz J, Schuster AK, Hartmann M, Gerhardt A, Eich W. Pain perception in athletes compared to normally active controls: a systematic review with meta-analysis. Pain, 2012. **153**(6): p. 1253-1262. https://doi.org/10.1016/j.pain.2012.03.005
260. Räisänen AM, Arkkila H, Vasankari T, et al. Investigation of knee control as a lower extremity injury risk factor: A prospective study in youth football. Scand J Med Sci Sports, 2018. **28**(9): p. 2084-2092. https://doi.org/10.1111/sms.13197
261. Barone R, Macaluso F, Traina M, Leonardi V, Farina F, Di Felice V. Soccer players have a better standing balance in nondominant one-legged stance. Open Access J Sports Med, 2010. **2**: p. 1-6. https://doi.org/10.2147/oajsm.S12593
262. Matsuda S, Demura S, Nagasawa Y. Static one-legged balance in soccer players during use of a lifted leg. Percept Mot Skills, 2010. **111**(1): p. 167-77. https://doi.org/10.2466/05.23.26.27.Pms.111.4.167-177
263. Matsuda S, Demura S, Uchiyama M. Centre of pressure sway characteristics during static one-legged stance of athletes from different sports. J Sports Sci, 2008. **26**(7): p. 775-9. https://doi.org/10.1080/02640410701824099
264. Snyder N, Cinelli M. Comparing Balance Control Between Soccer Players and Non-Athletes During a Dynamic Lower Limb Reaching Task. Res Q Exerc Sport, 2020. **91**(1): p. 166-171. https://doi.org/10.1080/02701367.2019.1649356
265. Ford KR, Nguyen AD, Dischiavi SL, Hegedus EJ, Zuk EF, Taylor JB. An evidence-based review of hip-focused neuromuscular exercise interventions to address dynamic lower extremity valgus. Open Access J Sports Med, 2015. **6**: p. 291-303. https://doi.org/10.2147/oajsm.S72432
266. Crowell KR, Nokes RD, Cosby NL. Weak Hip Strength Increases Dynamic Knee Valgus in Single-Leg Tasks of Collegiate Female Athletes. J Sport Rehabil, 2021. **30**(8): p. 1220-1223. https://doi.org/10.1123/jsr.2021-0043
267. Covassin T, Beidler E, Ostrowski J, Wallace J. Psychosocial aspects of rehabilitation in sports. Clin Sports Med, 2015. **34**(2): p. 199-212. https://doi.org/10.1016/j.csm.2014.12.004
268. Ardern CL, Taylor NF, Feller JA, Webster KE. Fifty-five per cent return to competitive sport following anterior cruciate ligament reconstruction surgery: an updated systematic review and meta-analysis including aspects of physical functioning and contextual factors. Br J Sports Med, 2014. **48**(21): p. 1543-52. https://doi.org/10.1136/bjsports-2013-093398
269. Ardern CL, Kvist J, Webster KE. Psychological Aspects of Anterior Cruciate Ligament Injuries. Operative Techniques in Sports Medicine, 2016. **24**(1): p. 77-83. https://doi.org/https://doi.org/10.1053/j.otsm.2015.09.006
270. Carter R, Lubinsky J, Domholdt E, *Rehabilitation research : principles and applications*. 2011, Philadelphia, Pa.: Saunders.
271. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res, 2018. **27**(5): p. 1171-1179. https://doi.org/10.1007/s11136-017-1765-4
272. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol, 2003. **3**: p. 25. https://doi.org/10.1186/1471-2288-3-25
273. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Quality of Life Research, 2012. **21**(4): p. 651-657. https://doi.org/10.1007/s11136-011-9960-1

Appendix 1

Table 1. Discriminative validity/diagnostic accuracy for the SLS test regarding all types of previous serious injuries.

Assessment		Cases that failed the SLS test*	Non-cases that passed the SLS test**	Sensitivity	Specificity	PPV	NPV
Foot	**NDL**	1/4 (25%)	192/243 (79%)	1.9 (0.2-3.6)	98.5 (96.9-100.0)	25.0 (19.6-30.4)	79.0 (73.9-84.1)
	DL	1/5 (20%)	186/242 (77%)	1.6 (0.1-3.4)	97.9 (96.1-99.7)	20.0 (15.0-25.0)	76.9 (71.6-82.1)
Knee	**NDL**	9/69 (13%)	135/178 (76%)	17.3 (12.6-22.0)	69.2 (63.5 -75.0)	13.0 (8.8-17.2)	75.8 (70.5-81.2)
	DL	22/98 (22%)	114/149 (77%)	38.6 (32.5-44.7)	60.0 (53.9-66.1)	22.5 (17.3-27.7)	76.5 (71.2-81.8)
Pelvic	**NDL**	6/53 (11%)	148/194 (76%)	11.5 (7.6-15.5)	75.9 (70.6-81.2)	11.3 (7.4-15.3)	76.3 (71.0-81.6)
	DL	25/111 (23%)	104/136 (77%)	43.9 (37.7-50.1)	54.7 (48.5-60.9)	22.5 (17.3-27.7)	76.5 (71.2-81.8)
Trunk	**NDL**	5/54 (9%)	146/193 (76%)	9.6 (5.9-13.3)	74.9 (69.5-80.3)	9.3 (5.6-12.9)	75.7 (70.3-81.0)
	DL	17/62 (27%)	145/185 (78%)	29.8 (24.1-35.5)	76.3 (71.0-81.6)	27.4 (21.9-32.0)	78.4 (73.2-83.5)
Total score Pass/fail	**NDL**	14/114 (12%)	95/133 (71%)	26.9 (21.4-32.5)	48.7 (42.5-55.0)	12.3 (8.2-77.1)	71.4 (65.8-77.1)
	DL	40/170 (24%)	60/77 (78%)	70.2 (64.5-75.9)	31.6 (25.8-37.4)	23.5 (18.2-28.8)	77.9 (72.8-83.1)
Assessment		**Sensitivity**	**Specificity**	**Correctly classified**		**AUC**	
Total score 4 categories **NDL**	**≥0**	100.0	0.00	21.1		0.37 (0.30-0.44)	
	≥1	26.9	48.7	44.1			
	≥2	7.7	78.0	63.2			
	≥3	3.85	92.8	74.1			
	>3	0.00	100.0	79.0			
DL	**≥0**	100.0	0.00	23.1		0.51 (0.43-0.60)	
	≥1	71.2	31.6	40.5			
	≥2	38.6	65.8	59.5			
	≥3	5.3	91.6	71.7			
	>3	0.00	100.0	76.9			

*Number (%) of subjects with serious injuries that failed on the SLS test. ** Number (%) of subjects without serious injuries that passed the SLS test. **Cases:** A previous serious injury lasting for ≥3 month located in the lower extremity and trunk, central lumbar/trunk pain and head injuries are counted for both legs; **NDL**: n=247, injury prevalence 21.1% (n=52); **DL:** n=247, injury prevalence 23.1% (n=57).

Table 2. Discriminative validity/diagnostic accuracy for the SLS test regarding a previous serious knee injury.

Assessment		Cases that failed the SLS test*	Non-cases that passed the SLS test**	Sensitivity	Specificity	PPV	NPV
Foot	**NDL**	0/4 (0%)	220/250 (88%)	0.0 (0.0-0.0)	98.2 (96.6-99.8)	0.0 (0.0-0.0)	88.0 (84.0-92.0)
	DL	1/6 (17%)	214/248 (86%)	2.9 (0.8-4.9)	97.7 (95.9-99.6)	16.7 (12.1-21.3)	86.3 (82.1-91.0)
Knee	**NDL**	3/70 (4%)	157/184 (85%)	10.0 (6.3-13.7)	70.1 (64.5-75.7)	4.3 (1.8-6.8)	85.3 (81.0-89.7)
	DL	11/102 (11%)	128/152 (84%)	31.4 (25.7-37.1)	58.5 (52.4-64.5)	10.8 (7.0-14.6)	84.2 (79.7-88.7)
Pelvic	**NDL**	2/56 (4%)	170/198 (86%)	6.7 (3.6-9.7)	75.9 (70.6-81.2)	3.6 (1.3-5.9)	85.9 (81.6-90.1)
	DL	19/113 (17%)	125/141 (89%)	54.3 (48.2-60.4)	57.1 (51.0-63.2)	16.8 (12.2-21.4)	88.7 (84.8-92.6)
Trunk	**NDL**	3/56 (5%)	171/198 (86%)	10.0 (6.3-13.7)	76.3 (71.-81.6)	5.4 (2.6-8.1)	86.4 (82.1-90.6)
	DL	9/65 (14%)	163/189 (86%)	25.7 (20.3-31.1)	74.4 (69.1-79.8)	13.9 (9.6-18.1)	86.2 (82.0-90.5)
Total score Pass/fail	**NDL**	7/117 (6%)	114/137 (83%)	23.3 (18.1-28.5)	50.9 (44.7-57.0)	6.0 (3.1-8.9)	83.2 (78.6-87.1)
	DL	23/176 (13%)	66/78 (85%)	65.7 (59.9-71.6)	30.1 (24.5-35.8)	13.1 (8.9-17.2)	84.6 (80.2-89.1)

Assessment			Sensitivity	Specificity	Correctly classified	AUC
Total score 4 categories	**NDL**	**≥0**	100.0	0.0	11.8	0.35 (0.3-0.4)
		≥1	23.3	50.9	47.6	
		≥2	3.3	79.0	70.1	
		≥3	0.0	92.9	81.9	
		>3	0.0	100.0	88.2	
	DL	**≥0**	100.0	0.0	13.8	0.50 (0.4-0.6)
		≥1	65.7	30.1	35.0	
		≥2	40.0	64.8	61.4	
		≥3	8.6	92.7	81.1	
		>3	0.0	100.0	86.2	

*****Number (%) of subjects with serious knee injuries that failed on the SLS test. ****** Number (%) of subjects without serious knee injuries that passed the SLS test. **Cases:** A previous serious knee injury lasting for ≥3 month located in the lower extremity and trunk, central lumbar/trunk pain and head injuries are counted for both legs; **NDL**: n=254, injury prevalence 11.8% (n=30); **DL:** n=254, injury prevalence 13.8% (n=35).

Table 3. Discriminative validity/diagnostic accuracy for the SLS test regarding a present time-loss injury.

Assessment		Cases that failed the SLS test*	Non-cases that passed the SLS test**	Sensitivity	Specificity	PPV	NPV
Foot	**NDL**	0/4 (0%)	226/246 (92%)	0.0 (0.0-0.0)	98.3 (96.6-99.9)	0.0 (0.0-0.0)	91.9 (88.5-95.3)
	DL	1/6 (17%)	227/244 (93%)	5.6 (2.7-8.4)	97.8 (96.0-99.6)	16.7 (12.1-21.3)	93.0 (89.9-96.2)
Knee	**NDL**	5/69 (7%)	166/181 (92%)	25.0 (19.6-30.4)	72.2 (66.6-77.7)	7.3 (4.0-10.5)	91.7 (88.3-95.1)
	DL	6/100 (6%)	138/150 (92%)	33.3 (27.5-39.2)	59.5 (53.4-65.6)	6.0 (3.1-8.9)	92.0 (88.6-95.4)
Pelvic	**NDL**	5/55 (9%)	180/195 (92%)	25.00 (19.6-30.4)	78.3 (73.2-83.4)	9.1 (5.5-12.7)	92.3 (89.0-95.6)
	DL	6/111 (5%)	127/139 (91%)	33.3 (27.5-39.2)	54.7 (48.6-60.9)	5.4 (2.6-8.2)	91.4 (87.9-94.9)
Trunk	**NDL**	4/55 (7%)	179/195 (92%)	20.0 (15.0-25.0)	77.8 (72.7-83.0)	7.3 (4.1-10.5)	91.8 (88.4-95.2)
	DL	2/64 (3%)	170/186 (91%)	11.1 (7.2-15.0)	73.3 (67.8-78.8)	3.1 (1.0-5.3)	91.4 (87.9-94.9)
Total score Pass/fail	**NDL**	8/115 (7%)	123/135 (9%)	40.0 (33.9-46.1)	53.5 (47.3-59.7)	7.0 (3.8-10.1)	91.1 (87.6-94.6)
	DL	10/174 (6%)	68/76 (90%)	55.6 (49.4-61.7)	29.3 (23.7-35.0)	5.8 (2.9-8.6)	89.5 (85.7-93.3)

Assessment		Sensitivity	Specificity	Correctly classified	AUC
NDL **Total score 4 categories**	**≥0**	100.0	0.0	8.0	0.46 (0.4- 0.6)
	≥1	40.0	53.5	52.4	
	≥2	15.0	80.9	75.6	
	≥3	5.0	93.5	86.4	
	>3	0.0	100.0	92.0	
DL	**≥0**	100.0	0.0	7.2	0.40 (0.3-0.5)
	≥1	55.6	29.3	31.2	
	≥2	22.2	63.4	60.4	
	≥3	5.6	92.7	86.4	
	>3	0.0	100.0	92.8	

*Number (%) of subjects with a present time-loss injury that failed the SLS test. ** Number (%) of subjects without a present time-loss injury that passed the SLS test. **Cases:** A present time-loss injury that caused a time-loss from training or competition the day the SLS test was performed, injuries located in the lower extremity and trunk, central lumbar/trunk pain and head injuries are counted for both legs. **NDL**: n=250, injury prevalence 8.0 % (n=20); **DL:** n=250, injury prevalence 7.2% (n=18).

Table 4. Discriminative validity/diagnostic accuracy for the SLS test regarding a time-loss injury previous 4 weeks.

Assessment		Cases that failed the SLS test*	Non-cases that passed the SLS test**	Sensitivity	Specificity	PPV	NPV
Foot	**NDL**	0/4 (0%)	234/246 (95%)	0 (0.0-0.0)	98.3 (96.7-99.9)	0 (0.0-0.0)	95.1 (92.5-97.8)
	DL	1/6 (17%)	232/244 (95%)	7.7 (4.4-11.0)	97.9 (96.1-99.7)	16.7 (12.1-21.3)	95.1 (92.4-97.8)
Knee	**NDL**	6/69 (9%)	175/181 (97%)	50.0 (43.8-56.2)	73.5 (68.1-79.0)	8.7 (5.2-12.2)	96.7 (94.5-98.9)
	DL	6/100 (6%)	143/150 (95%)	46.2 (40.0-52.3)	60.3 (54.3-66.4)	6.0 (3.1-8.9)	95.3 (92.7-98.0)
Pelvic	**NDL**	7/55 (13%)	190/195 (97%)	58.3 (52.2-64.4)	79.8 74.9-84.8)	12.7 (8.6-16.9)	97.4 (95.5-99.4)
	DL	4/111 (4%)	130/139 (94%)	30.8 (25.1-36.5)	54.9 (48.7-61.0)	3.6 (1.3-5.9)	93.5 (90.5-96.6)
Trunk	**NDL**	6/55 (11%)	189/195 (97%)	50.0 (43.8-56.2)	79.4 (74.4-84.4)	10.9 (7.0-14.8)	96.6 (94.8-99.1)
	DL	4/64 (6%)	177/186 (95%)	30.8 (25.1-36.5)	74.7 (69.3-80.1)	6.3 (3.3-9.3)	95.2 (92.5-97.8)
Total score Pass/fail	**NDL**	8/115 (7%)	131/135 (97%)	66.7 (60.8-72.5)	55.0 (48.9-61.2)	7.0 (3.8-10.1)	97.0 (94.9-99.1)
	DL	8/174 (5%)	71/76 (93%)	61.5 (55.5-67.6)	30.0 (25.3-35.6)	4.6 (2.0-7.2)	93.4 (90.4-96.5)

Assessment		Sensitivity	Specificity	Correctly classified	AUC
Total score 4 categories **NDL**	**≥0**	100.0	0.0	4.8	0.67 (0.50-0.85)
	≥1	66.7	55.0	55.6	
	≥2	50.0	82.8	81.2	
	≥3	25.0	94.5	91.2	
	>3	0.0	100.0	95.2	
DL	**≥0**	100.0	0.0	5.2	0.50 (0.32-0.68)
	≥1	61.5	30.0	31.6	
	≥2	38.5	64.7	63.2	
	≥3	15.4	93.3	89.2	
	>3	0.0	100.0	94.8	

*****Number (%) of subjects with a time-loss injury previous 4 weeks that failed the SLS test. ****** Number (%) of subjects without a time-loss injury previous 4 weeks that passed the SLS test. **Cases:** A time-loss injury previous 4 weeks that caused a time-loss from training or competition, injuries located in the lower extremity and trunk, central lumbar/trunk pain and head injuries are counted for both legs. **NDL**: n=250, injury prevalence 4.8 % (n=12); **DL:** n=250, injury prevalence 5.2% (n=13).

Table 5. Discriminative validity/diagnostic accuracy for the SLS test regarding a present injury problem.

Assessment		Cases that failed the SLS test*	Non-cases that passed the SLS test**	Sensitivity	Specificity	PPV	NPV
Foot	**NDL**	1/4 (25%)	212/250 (85%)	2.6 (0.6-4.5)	98.6 (97.2-100.0)	25.0 (19.7-30.3)	84.8 (80.4-89.2)
	DL	2/6 (33%)	202/248 (82%)	4.2 (1.7-6.6)	98.1 (96.4-99.8)	33.3 (27.5-39.1)	81.5 (76.7-86.2)
Knee	**NDL**	12/70 (17%)	157/184 (85%)	30.8 (25.1-36.5)	73.0 (67.6-78.5)	17.1 (12.5-21.8)	85.3 (81.0-89.7)
	DL	18/102 (18%)	122/152 (80%)	37.5 (31.6-43.5)	59.2 (53.2-65.3)	17.7 (13.0-22.3)	80.3 (75.4-85.2)
Pelvic	**NDL**	12/56 (21%)	171/198 (86%)	30.8 (25.1-36.5)	79.5 (74.6-84.5)	21.4 (16.4-26.5)	86.4 (82.1-90.6)
	DL	20/113 (88%)	113/141 (80%)	41.7 (35.6-47.7)	54.9 (48.7-61.0)	17.7 (13.0-22.4)	80.1 (75.2-85.1)
Trunk	**NDL**	14/56 (25%)	173/198 (87%)	36.0 (30.1-41.8)	80.54 (75.6-85.3)	25.0 (19.7-30.3)	87.4 (83.3-91.5)
	DL	12/65 (18%)	153/189 (81%)	25.0 (19.7-30.3)	74.3 (68.9-79.7)	18.5 (13.7-23.2)	81.0 (76.1-85.8)
Total score Pass/fail	**NDL**	22/117 (19%)	120/137 (88%)	56.4 (50.3-62.5)	55.8 (49.7-61.9)	18.8 (14.0-23.6)	87.6 (83.5-91.7)
	DL	33/176 (19%)	63/78 (81%)	68.8 (63.1-74.5)	30.6 (24.9-36.3)	18.8 (14.0-23.6)	80.8 (75.9-85.6)

Assessment			Sensitivity	Specificity	Correctly classified	AUC
Total score 4 categories	**NDL**	**≥0**	100.0	0.0	15.4	0.57 (0.48-0.67)
		≥1	56.4	55.8	55.9	
		≥2	25.6	82.3	73.6	
		≥3	12.8	94.9	82.3	
		>3	0.0	100.0	84.7	
	DL	**≥0**	100.0	0.0	18.9	0.48 (0.40-0.57)
		≥1	68.6	30.6	37.8	
		≥2	31.2	63.1	57.1	
		≥3	8.3	92.7	76.8	
		>3	0.0	100.0	81.1	

*****Number (%) of subjects with a present injury problem that failed the SLS test. ****** Number (%) of subjects without a present injury problem that passed the SLS test. **Cases:** An injury problem that did not cause any time-loss from training or competition, injury problems located in the lower extremity and trunk, central lumbar/trunk and head are counted for both legs. **NDL**: n=254, injury prevalence 15.4% (n=39); **DL:** n=254, injury prevalence 18.9% (n=48).

Table 6. Discriminative validity/diagnostic accuracy for the SLS test regarding an injury problem previous 4 weeks.

Assessment		**Cases that failed the SLS test***	**Non-cases that passed the SLS test****	**Sensitivity**	**Specificity**	**PPV**	**NPV**
Foot	**NDL**	0/4 (0%)	220/250 (88%)	0.0 (0.0-0.0)	98.2 (96.6-99.8)	0.0 (0.0-0.0)	88.0 (84.0-92.0)
	DL	1/6 (17%)	209/248 (84%)	2.5 (0.6-4.4)	97.7 (95.8-99.5)	16.7 (12.1-21.3)	84.3 (79.8-88.8)
Knee	**NDL**	13/70 (19%)	167/184 (91%)	43.3 (37.2-49.4)	74.6 (69.2-79.9)	18.6 (13.8-23.4)	90.8 (87.2-94.3)
	DL	17/102 (17%)	129/152 (85%)	42.5 (36.4-48.6)	60.3 (54.3-66.3)	16.7 (12.1-21.3)	84.9 (80.5-89.3)
Pelvic	**NDL**	10/56 (18%)	178/198 (90%)	33.3 (27.6-39.1)	79.5 (74.5-84.4)	17.9 (13.2-22.6)	89.9 (86.2-93.6)
	DL	19/113 (17%)	120/141 (85%)	47.5 (41.4-53.6)	56.1 (50.0-62.2)	16.8 (12.2-21.4)	85.1 (80.7-89.5)
Trunk	**NDL**	7/56 (13%)	175/198 (88%)	23.3 (18.1-28.5)	78.1 (73.0-83.2)	12.5 (8.4-16.6)	88.4 (84.4-92.3)
	DL	8/65 (12%)	157/189 (83%)	20.0 (15.1-24.9)	73.4 (67.9-78.8)	12.3 (8.3-16.4)	83.1 (78.5-87.7)
Total score Pass/fail	**NDL**	18/117 (15%)	125/137 (89%)	60.0 (53.0-66.0)	55.8 (49.7-61.9)	15.4 (11.0-19.8)	91.2 (87.8-94.7)
	DL	29/176 (17%)	67/78 (86%)	72.5 (67.0-78.0)	31.3 (25.6-37.0)	16.5 (11.9-21.0)	86.0 (81.6-90.2)
Assessment		**Sensitivity**	**Specificity**	**Correctly classified**		**AUC**	
Total score 4 categories **NDL**	**≥0**	100.0	0.0	11.8		0.58 (0.48-0.68)	
	≥1	60.0	55.8	56.3			
	≥2	26.7	82.1	75.6			
	≥3	6.7	93.8	83.5			
	>3	0.0	100.0	88.2			
DL	**≥0**	100.0	0.0	15.8		0.50 (0.41-0.59)	
	≥1	72.5	31.3	37.8			
	≥2	32.5	63.6	58.7			
	≥3	7.5	92.5	79.1			
	>3	0.0	100.0	84.3			

*Number (%) of subjects with an injury problem previous 4 weeks that failed the SLS test. ** Number (%) of subjects without an injury problem previous 4 weeks that passed the SLS test. **Cases:** An injury problem previous 4 weeks that did not cause any time-loss from training or competition, injury problems located in the lower extremity and trunk, central lumbar/trunk and head are counted for both legs. **NDL**: n=254, injury prevalence 11.8% (n=30); **DL:** n=254, injury prevalence 15.8% (n=40).