

**Helene Kværne Vileid**

**Interrater reliability of tests evaluating  
anatomical risk factors for ACL injury in  
elite female football and handball players**

**Master thesis in Sports Physiotherapy**  
Department of Sports Medicine  
Norwegian School of Sport Sciences, 2013



## Abstract

**Study design:** Methodological study

**Background:** A cohort was set up at the Oslo Sports Trauma Research Centre in 2007 aiming at investigating risk factors for anterior cruciate ligament (ACL) injuries in elite female football and handball players. The reliability and measurement error of the tests included in the cohort has not been established for this sample.

**Objective:** To test the interrater reliability of a series of tests evaluating anatomical risk factors for ACL injury in elite female football and handball players. The anatomical risk factors are generalized joint laxity (GJL), genu recurvatum, knee laxity, hip anteversion and 12 anthropometric measures.

**Methods:** A protocol was designed for establishing interrater reliability of the following anatomical measures: generalized joint laxity (GJL), genu recurvatum, knee laxity, hip anteversion and 11 anthropometric measures of the lower limbs and height. The players were tested and retested on two separate days with intervals varying between 2–51 days. Reliability was assessed using a set of statistical tests; intraclass correlation coefficients ( $ICC_{2,1}$ ), root mean square error (RMSE), paired t-test, Wilcoxon signed rank test and Bland-Altman plots.

**Results:** 42 female athletes (22 football and 20 handball players) were included. Mean  $\pm$  SD age, football players:  $21.1 \pm 3.2$  and handball players:  $21.9 \pm 4.4$ . The interrater reliability was moderate for measures of knee laxity on the right side, ASIS width, and bilateral tibial and femoral widths ( $ICC_{2,1}$  0.55–0.65). Low interrater reliability was found for knee laxity on the left side, bilateral genu recurvatum, hip anteversion on the left side, bilateral tibial and femoral widths and femoral-tibial ratio on the right side ( $ICC_{2,1}$  0.29–0.44). Femoral-tibial ratio on the left side and hip anteversion on the right side showed little if any correlation ( $ICC_{2,1}$  0.17–0.22). Measuring height had the highest correlation ( $ICC_{2,1}$  0.99). The GJL showed no difference between the two days of testing.

**Conclusion:** The interrater reliability of the anatomical tests included in this study varied from little to very high reliability. A greater variance of the sample, continuous variables, double-blinding, intrarater testing prior to the study, better piloting and better training of the testers are alterations that might have improved the results of this study.

# Contents

<b>Abstract</b>	<b>3</b>
<b>Acknowledgements</b>	<b>7</b>
<b>1. Theory chapter</b>	<b>8</b>
<b>1.1 Background</b>	<b>8</b>
<b>1.2 Epidemiology of ACL injuries</b>	<b>8</b>
<b>1.3 Anatomy and biomechanics of the ACL</b>	<b>9</b>
<b>1.4 Risk factors</b>	<b>10</b>
1.4.1 Why identify risk factors?	10
1.4.2 Extrinsic risk factors	11
1.4.3 Intrinsic risk factors	12
1.4.4 Potential anatomical risk factors	13
1.4.5 Generalized joint laxity	13
1.4.6 Genu recurvatum	14
1.4.7 Knee laxity	14
1.4.8 Hip anteversion	15
1.4.9 Anthropometric measures	15
<b>1.5 Effects of validity and reliability of screening tests</b>	<b>15</b>
1.5.1 Validity	16
1.5.2 Reliability	17
<b>1.6 Sources of measurement error</b>	<b>18</b>
1.6.1 Systematic and random error	19
<b>1.7 Statistical measures of reliability</b>	<b>20</b>
<b>1.8 Screening tests</b>	<b>22</b>
1.8.1 Generalised joint laxity	22
1.8.2 Genu recurvatum	22
1.8.3 Knee laxity	23
1.8.4 Hip anteversion	23
1.8.5 Anthropometric measures	23
<b>1.9 Objective</b>	<b>24</b>
<b>1.10 Hypothesis</b>	<b>24</b>



<b>2. Methods</b>	<b>26</b>
<b>2.1 Study design</b>	<b>26</b>
<b>2.2 Participants</b>	<b>26</b>
<b>2.3 Testers</b>	<b>29</b>
<b>2.4 Collection of data</b>	<b>29</b>
<b>2.5 Screening tests</b>	<b>30</b>
2.5.1 Tests in this reliability study	30
2.5.2 Generalized joint laxity (GJL)	30
2.5.3 Genu recurvatum	31
2.5.4 Knee laxity	32
2.5.5 Hip anteversion	34
2.5.6 Anthropometric measures of lower extremities and height	34
<b>2.6 Remaining tests of the cohort</b>	<b>36</b>
<b>2.7 Test order</b>	<b>36</b>
<b>2.8 Statistics</b>	<b>36</b>
<b>2.9 Ethics approval</b>	<b>37</b>
<b>3. Results</b>	<b>38</b>
<b>3.1 Descriptive data of the players</b>	<b>38</b>
<b>3.2 Differences between tests on day one and two</b>	<b>39</b>
<b>3.3 Interrater reliability</b>	<b>40</b>
<b>3.4 Generalized joint laxity</b>	<b>41</b>
<b>3.5 Actual measurement error of the tests</b>	<b>41</b>
<b>4. Discussion</b>	<b>45</b>
<b>4.1 Reliability of the different tests</b>	<b>45</b>
4.1.1 Generalized joint laxity	45
4.1.2 Genu recurvatum	46
4.1.3 Knee laxity	46
4.1.4 Hip anteversion	49
4.1.5 Anthropometric measures of lower extremities and height	51

<b>4.2 Methods</b>	<b>52</b>
4.2.1 Participants	52
4.2.2 Testers	52
4.2.3 Data collection	53
4.2.4 Study design	54
<b>4.3 Statistics</b>	<b>55</b>
<b>4.4 Clinical relevance</b>	<b>57</b>
4.4.1 Perspectives	58
<b>5. Conclusion</b>	<b>59</b>
<b>References</b>	<b>60</b>
<b>List of tables</b>	<b>69</b>
<b>Table of figures</b>	<b>69</b>
<b>Acronyms</b>	<b>70</b>
<b>Attachments</b>	<b>71</b>

## Acknowledgements

What makes us happy? I feel that achievement makes me happy – trying something new, taking on a challenge, finishing something off that I can be proud of. It doesn't have to be something as big as writing a thesis; baking a nice cake or going for a run on a rainy day when I really have to push myself to go outside, can also give great satisfaction. Well, I feel happy and proud now, because I have finished this project. However, I did not do this all on my own.

First of all, I owe a great thanks to all the handball and football players who participated in the voluntary testing. All of the testers also deserve a big thank you.

Secondly, I would like to thank my supervisors, Agnethe Nilstad and Thor Einar Andersen, for guidance through this project. Their availability and quick feedback have been very much appreciated. I would also like to thank Ingar Holme and Mok Kam Ming for statistical advice and discussions.

Thank you to the Oslo Sport Trauma Research Center for letting us carry out this research and particularly for financially supporting the second round of testing in order for us to get a sample size large enough.

Thank you to friends and family. I believe they already know how much they mean to me. Also, a special thanks to my friend Sissel Sandve for helping out with the layout.

Thank you to the girls that I have shared 'an office' with, the girls I have had lunch with and the guys across the corridor in the biomechanics' room. You've made this period funnier and less lonely.

A huge thanks to Joar Harøy, my friend and colleague. I am so glad we have been in this together. Your sense of humor got us through.

Looking back at this research process, I realize that I have learnt a lot.

*Helene Kværne Vileid, Oslo,*

*21<sup>st</sup> May 2013*

# 1. Theory chapter

## 1.1 Background

An ACL injury represents a major area within sports medicine and they cause the most time lost from competition in football (Myklebust & Bahr, 2005; Lohmander, Englund, Dahl, & Roos, 2007; Meunier, Odensten, & Good, 2007). An ACL injury causes morbidity, long disability time and a devastating influence on activity levels and quality of life, high economic costs and potential long-term effects such as early development of osteoarthritis (Elliot, Goldberg, & Kuehl, 2010).

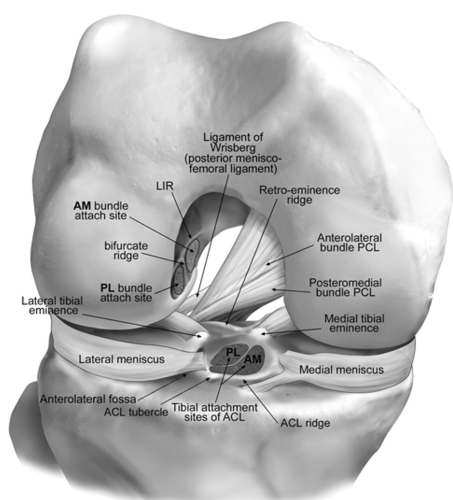
## 1.2 Epidemiology of ACL injuries

The highest risk of an ACL injury is among top-level female athletes who compete in football and handball (Bahr & Engebretsen, 2009). The incidence of all injuries in elite female football has been reported in several studies, and it varies from 12.6–23.6 per 1000 game hours (Giza, Mithofer, Farrell, Zarins, & Gill, 2005; Tegnander, Olsen, Moholdt, Engebretsen, & Bahr, 2008). Reviews show an ACL injury rate of 0.28 per 1000 athlete-exposures in female football (Renstrom et al., 2008) and 0.82 ACL injuries per 1000 playing hours in female elite handball (Myklebust, Maehlum, Holm, & Bahr, 1998; Prodromos, Han, Rogowski, Joyce, & Shi, 2007). It has been demonstrated that female athletes who participate in pivoting and jumping sports have a 4–6 times higher risk of suffering ACL injuries than male athletes in the same sports (Ferretti, Papandrea, Conteduca, & Mariani, 1992; Arendt & Dick, 1995; Myklebust, Maehlum, Engebretsen, Strand, & Solheim, 1997). A more recent review concluded that female football players have a 2–3 times higher ACL injury risk compared to male players (Walden, Hagglund, Werner, & Ekstrand, 2011). Studies, both in Norwegian elite handball (Myklebust, Maehlum et al. 1997; Myklebust, Maehlum et al. 1998) and in German (Faude, Junge, Kindermann, & Dvorak, 2006) and Norwegian elite football (Tegnander et al., 2008) showed that 5–10 % of the players sustain an ACL injury each season. In a typical league that equals to one entire team each season.

Of patients who have suffered an ACL injury, 50 % of them displayed radiological signs of osteoarthritis (OA) eight years after injury (Myklebust & Bahr, 2005). After 15 years the occurrence of OA increases to 80 %. These findings are regardless of whether the ACL was reconstructed or not (Myklebust and Bahr 2005).

### 1.3 Anatomy and biomechanics of the ACL

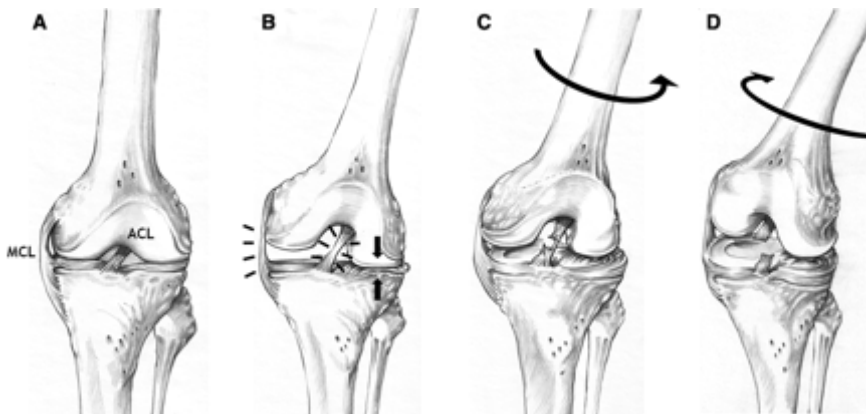
Anatomic and biomechanical studies have shown that the ACL consists of 2 major bundles, the posterolateral bundle (PL) and the anteromedial bundle (AM) (Jarvela & Jarvela, 2013). Both bundles originate from the posteromedial side of the lateral femoral condyle and insert just anteriorly to the intercondylar tibial eminence (Siegel, Vandenaeker-Albanese, & Siegel, 2012). The role of the ACL is to resist anterior tibial translation and inhibit extreme ranges of tibial rotation (Amis, 2012). The difference in biomechanical behaviour of the two bundles could explain partial ACL ruptures. The AM bundle plays a greater role in restraining the tibia from anterior translation when the knee is flexed at 45 degrees or more, while the PL restrains more towards full extension (Colombet, Dejour, Panisset, & Siebold, 2010).



**Figure 1:** An illustration of the tibial and femoral ACL bundle attachments in the native knee. AM, anteromedial bundle; PL, posterolateral bundle; LIR, lateral intercondylar ridge; PCL, posterior cruciate ligament. Mechanisms of injury (Ziegler et al., 2011)

Research so far suggests a series of elements determining the mechanisms of ACL injuries. First of all, most ACL injuries (70–84 %) happens when the player is in no physical contact with other players, thus termed non-contact ACL injuries (Boden, Dean, Feagin, Jr., & Garrett, Jr., 2000; Fauno & Wulff, 2006). Several theories for ACL injury mechanisms have been proposed, and the actual mechanism is still a matter of controversy (Koga et al., 2010). The mechanisms involved in most cases of non-contact ACL injuries are when the player performs a sudden change of direction or a cutting manoeuvre when decelerating, or during landing from a jump with the knee close to full

extension with the foot planted (Boden, Dean et al. 2000; Olsen, Myklebust et al. 2004; Krosshaug, Slauterbeck et al. 2007; Koga, Nakamae et al. 2010). Koga et al. (2010) have combined their study with previous findings and proposed a hypothesis for injury mechanism. They used a model-based- image-matching method analysing ten anterior cruciate ligament injury video sequences. Their hypothesis is seen in **Figure 2**. There seems to be general agreement that mechanisms involved are knee valgus, increased internal rotation of the hip and external rotation of the tibia (Boden et al., 2000; Olsen, Myklebust, Engebretsen, & Bahr, 2004; Krosshaug, Slauterbeck, Engebretsen, & Bahr, 2007). Hewett et al. (2005) also found high dynamic knee valgus to be predicting a possible future ACL injury (Hewett et al., 2005).



**Figure 2:** ACL injury mechanism. **A:** the knee in an unloaded position **B:** valgus loading leading to compression of the lateral side and tightening of the medial collateral ligament **C:** quadriceps contraction and the compressive load causing the lateral femoral condyle to shift posteriorly and internal rotation and anterior translation of the tibia leading to ACL rupture **D:** after the tear there is a displacement of the medial femoral condyle and the tibia rotates externally (Koga et al., 2010)

## **1.4 Risk factors**

### **1.4.1 Why identify risk factors?**

To prevent ACL injuries, an important step is to identify risk factors, then modify or eliminate these risk factors in order to ultimately reduce the number of ACL injuries (Alentorn-Geli, Myer et al. 2009). Understanding the risk factors and the mechanisms of injury is important in order to identify people at higher risk and target specifically developed methods for preventing injuries at them (Bahr & Krosshaug, 2005; Renstrom

et al., 2008). Most preventive programs contain a combination of proprioceptive, plyometric and neuromuscular exercises, stretching and core training. Such multi-component studies have shown better results than single component programs. Studies on preventive programs have resulted in significant decrease in the number of non-contact ACL injuries both within football and handball (Caraffa, Cerulli, Projetti, Aisa, & Rizzo, 1996; Mandelbaum et al., 2005; Petersen et al., 2005; Myklebust et al., 2007; Gilchrist et al., 2008). It seems like the preventive programs have a direct benefit on decreasing the number of non-contact ACL injuries, but due to multi-component programs, it is difficult to distinguish the effect of each single component on the non-contact ACL injury risk (Alentorn-Geli, Myer et al. 2009). With compliance rates as low as 28 % to the preventive programmes, this is a limiting factor itself for a higher success on the reduction of ACL injuries (Myklebust, Engebretsen et al. 2003). Most research has studied isolated risk factors, thus the consequence of clustering of risk factors is unclear, and ACL injuries are most probably a result of several factors appearing together (Renstrom et al., 2008). The following section will briefly describe how the risk factors can be divided into different categories. However, the main focus will be on some of the anatomical risk factors like generalized joint laxity (GJL), knee laxity, genu recurvatum, hip anteversion and some antropometric measures. In this project, the reliability of these tests have been investigated. The preventive potential of these factors is fairly small, since anatomy is difficult to change (Alentorn-Geli et al., 2009a).

#### **1.4.2 Extrinsic risk factors**

Risk factors for ACL injuries are categorized as extrinsic or intrinsic. Extrinsic factors are external to the athlete. Little is known about these extrinsic risk factors such as footwear, level and type of activity, playing surface, protective equipment or meteorological conditions (Renstrom et al., 2008). The existing evidence is mainly based on American or Australian Football (Scranton, Jr. et al., 1997; Orchard, Seward, McGivern, & Hood, 1999; Orchard, Chivers, Aldous, Bennell, & Seward, 2005). Although, Myklebust et al. (2003) found that female elite handball players are at higher risk of suffering an ACL injury during competition compared to training.

### **1.4.3 Intrinsic risk factors**

The intrinsic risk factors are inherent to the athlete (Smith et al., 2012a). Several intrinsic risk factors have been proposed, but there is lacking evidence linking these factors to the actual ACL injuries, except from a few factors as will be described further in this chapter (Griffin et al., 2006; Renstrom et al., 2008; Alentorn-Geli et al., 2009a). The intrinsic risk factors can supplementary be divided into modifiable and non-modifiable factors.

#### ***Modifiable factors***

Some biomechanical and neuromuscular factors, such as muscle strength, power and activation patterns, are considered as modifiable (Alentorn-Geli et al., 2009b; Myer, Brent, Ford, & Hewett, 2011). As an example, female athletes often demonstrate quadriceps dominance, which is an imbalance between knee extensors and flexor strength, recruitment and coordination (Myer et al., 2009).

#### ***Non-modifiable factors***

Non-modifiable factors are sex, genetics, hormonal and anatomical factors (Alentorn-Geli, Myer et al. 2009; Myer, Brent et al. 2011). Being a female is a known risk factor as female athletes have a greater risk of injuring their ACL compared with males when taking part in the same sports at similar levels of exposure (Smith et al., 2012b). The reason behind this sex discrepancy is still unknown, but research is trying to identify differences and risk factors. Sex-based anatomic differences, neuromuscular control variations and sex-hormones have been investigated. There are some studies that have claimed that females are at a greater risk of ACL injury during the preovulatory phase of the menstrual cycle (Wojtys, Huston, Boynton, Spindler, & Lindenfeld, 2002; Myklebust et al., 2003), while a recent review reported that these findings must be considered with caution until we have more accurate ways of determining menstrual status at the time of injury (Smith et al., 2012b).

Other risk factors being investigated are previous ACL reconstruction, genetics and neurocognitive function. Swanik et al (2007) found that ACL injured patients had significantly slower reaction time, slower processing speed, lower visual and verbal memory scores when comparing them with matched controls using neurocognitive tests (Swanik, Covassin, Stearne, & Schatz, 2007).



#### **1.4.4 Potential anatomical risk factors**

Anatomical variables that are considered as potential risk factors are ACL size, knee geometry, anterior-posterior (AP) knee laxity, generalized joint laxity, body mass index and static alignment of the lower limbs (Smith et al., 2012a). Researches hypothesize that anatomic variations could partly explain the large difference in ACL injury incidence between the sexes, as differences between men and women are well documented. Women have a smaller ACL than men when comparing the length, cross-sectional area and volume (Chandrashekar, Slauterbeck, & Hashemi, 2005). ACL volume is related to intercondylar notch width and is found to be smaller in women. Smith et al. (2012) stated in their review that findings of different studies of femoral intercondylar notch width are difficult to compare due to different measurement techniques. Still they claimed that the majority of studies showed that an increase of the risk of ACL injury is observed as the intercondylar notch width decreases. With regards to knee geometry they further referred to eight studies that have correlated measures of tibial geometry with the risk of ACL injury (Smith et al., 2012a). When comparing ACL injured individuals with controls, there were findings identified using MRI measures. These were increased posterior slope of the lateral and medial tibia plateaus, reduced condylar depth on the medial tibial plateau and an anterior medial ridge on the intercondylar notch (Shultz et al., 2010).

#### **1.4.5 Generalized joint laxity**

The term generalized joint laxity (GJL) indicates a range of motion (ROM) which is greater than the mean ROM in the general population (Kim, Kumar, & Kim, 2010). Research has reported that increased generalized joint laxity in the female athlete may be a risk factor for an ACL injury (Uhorchak et al., 2003). These results were further confirmed by Ramesh and colleagues (Ramesh, Von, Azzopardi, & Schranz, 2005). They also concluded that knee hyperextension in particular is a risk factor. Women have greater GJL than men. Children and certain racial groups have increased prevalence of GJL (Remvig, Jensen, & Ward, 2007). GJL also seem to have familiar tendencies and declines with age (Boyle, Witt, & Riegger-Krugh, 2003; Beighton, Solomon, & Soskolne, 1973).

#### **1.4.6 Genu recurvatum**

Genu recurvatum is defined as knee extension greater than five degrees (Loudon, Goist, & Loudon, 1998). During hyperextension, the ACL restrains anterior tibial migration. Increased knee hyperextension has been theorized to stress the ACL due to its role in limiting the knee to go into this position (Trimble, Bishop, Buckley, Fields, & Rozea, 2002). Sports with a lot of pivoting, landing or deceleration set high demands to dynamic stability of the knee. Increased knee laxity, with less tight ligaments and tendons stabilizing the knee, may result in decreased dynamic knee stability and may be an important contributor to increased risk of ACL injury, particularly a non-contact injury (Myer, Ford, Paterno, Nick, & Hewett, 2008). Ramesh et al. (2005) have suggested a pathway of injury mechanism; as the athlete's foot lands on the ground there is a sequence of movements which eventually leads to eccentric contraction of the quadriceps, which further leads to hyperextension of the knee with increased anterior translation of the tibia. This may lead to the ACL hitting the tibial notch thereby slitting itself (Ramesh et al., 2005). Still, they recognise that there is a multifactorial interplay, with regards to landing technique degree of tibial rotation as well as extrinsic factors for such an event to happen.

#### **1.4.7 Knee laxity**

Another potential risk factor for an ACL injury in the female athlete seems to be increased anterior-posterior (AP) laxity of the knee (Uhorchak et al., 2003). Increased AP knee laxity can be broadly defined as abnormal displacement of the tibia with respect to the femur. In the unloaded state, ligaments, capsule and other soft tissues will provide the stability, while in the loaded state there is an interaction between ligaments, condylar geometry and tibiofemoral contact forces generated by gravitational forces and muscle activity. Investigation has shown that adolescent girls have greater knee laxity and increased generalized joint laxity compared to adolescent boys (Myer et al., 2008). Evidence has suggested a gender difference in the occurrence of ACL injuries in the post-pubertal, but not in the pre-pubertal athlete (Myer et al., 2008; Shea, Grimm, Ewing, & Aoki, 2011). Furthermore, systematic reviews have concluded that menstrual status and changes in female hormonal concentrations may have an effect on AP knee ligament laxity and thus increase ACL injury risk in female athletes (Hewett, Zazulak, & Myer, 2007; Zazulak, Paterno, Myer, Romani, & Hewett, 2006; Park, Stefanyshyn, Ramage, Hart, & Ronsky, 2009).

#### **1.4.8 Hip anteversion**

Hip anteversion, also referred to as femoral anteversion, is defined as the angle between an imaginary transverse line that runs medially to laterally through the knee joint and an imaginary transvers line running through the neck and head of femur (Cibulka, 2004). An increase in this angle causes a decreased internal moment arm and thus may lead to an inefficiency of the gluteus medius. Hip anteversion may be one of the lower extremity alignments explaining the discrepancy of ACL injury rate between men and women, since women seem to have a larger femoral anteversion angle compared to men (Shultz, Nguyen, & Schmitz, 2008; Medina McKeon & Hertel, 2009). At birth the average of the hip anteversion angle ranges from 30-40°, but decrease progressively to about 15° at skeletal maturation (Souza & Powers, 2009; Shultz et al., 2008). The lack of normative data in the literature makes it difficult to define a 'normal' range, but McKeon et al. (2009) did a study where they reported representative values of lower limb alignments from a sample of athletes. They found a mean of 8,3° in men while in women the mean was 11,5° (Medina McKeon & Hertel, 2009). Hip anteversion greater than 30° may be considered 'excessive' and less than 8° is referred to as retroversion (Souza & Powers, 2009).

#### **1.4.9 Anthropometric measures**

Anthropometry are measurements of the physical dimensions and properties of the body (Roche, Heymsfield, & Lohman, 1996). In medicine, anthropometric measurements of height, weight and more specific body measurements like girths and skin folds can be used to trace trends and to determine human growth and development (Lohman, Roche, & Martorell, 1988). Measurements are also taken in preparation for three dimensional (3D) motion analysis. The 3D analysis is a method for screening dynamic neuromuscular control and knee joint load, for example during jump-landing movement tasks. Hewett et al. (2005) recommended such measurements should be a part of future screening of female athletes. They found an increased risk of ACL injury for the female athletes with increased dynamic valgus and high abduction loads (Hewett et al., 2005).

### **1.5 *Effects of validity and reliability of screening tests***

Validity and reliability are at the core of what is accepted as scientific proof (Thomas, Silverman et al. 2011). They are indicators of the quality of a measuring instrument (Kimberlin & Winterstein, 2008). Unreliable tests will produce inaccurate estimates of

variables. This can result in less powerful statistical tests and may lead to inappropriate conclusions in research, for example that variables being studied are unrelated or that the treatments implemented are ineffective (Arnold, Gansneder, & Perrin, 2005). To assure the quality of the results of the on-going cohort, reliability and validity need to be assessed.

### **1.5.1 Validity**

The validity of research is to what extent the conclusions are believable and useful (Carter, Lubinsky, & Domholdt, 2011). Validity of a test is to what extent the test accurately measures what it is supposed to (Berg, Latin, & Berg, 2004). It is the appropriateness, meaningfulness and usefulness of the tests (Carter et al., 2011). There are several different types of validity:

**Internal validity** is the extent to which the results of a study are caused by the treatments used in the study (Thomas, Silverman, & Nelson, 2005). It refers to the cause and effect relationship.

**Construct validity** is the degree to which a test measures a hypothetical construct; usually established by indirect measures (Thomas et al., 2005). It is concerned with the meaning of variables within the study (Carter et al., 2011). To look at the validity of more abstract measures like for example proprioception, a sophisticated understanding of the nature and definition is needed (Arnold et al., 2005).

**External validity** is the generalizability of the results of the study (Thomas et al., 2005). It describes to whom, in what settings and at what times can the results of the research be generalized (Carter et al., 2011).

**Statistical conclusion validity** is to what extent the statistical tools are used correctly to analyse the data (Carter et al., 2011).

The concept of validity is less intuitive and more complex than the concept of reliability, but the relationship between the two is unidirectional (Portney & Watkins, 2009). Ideally the results of a research would show high validity and high reliability, so the results would be repeatable and accurately reflect what was intended. But validity

and reliability do not necessarily go hand in hand (Berg et al., 2004). A test can be highly reliable, but have low validity. An example is that our tape measure could be wrongly marked, so that the readings would be consistently one cm less than the actual length. The reliability could be high, but it would not be a valid measure of height. It is also possible to have both low reliability and low validity, but it is not possible to have low reliability and high validity (Portney & Watkins, 2009).

### **1.5.2 Reliability**

Reliability testing within physiotherapy might be done for different purposes. It is used to describe the reliability of the instruments, the measurements or the testers. Although the tests would be the same in each case, the interpretation could be different. The procedure must be reliable in order for us to know whether an observed change on a reassessment is within the boundaries of assessment error or whether there has been a true change (Sole, Hamren et al. 2007).

*Test-retest reliability*, also called repeatability, instrument reliability or stability is whether repeated measures using a single instrument will give similar results (Arnold et al., 2005; Kimberlin & Winterstein, 2008). There is a need to standardize the measurement procedures in order to show acceptable consistency and precision of the measurements (Shultz, Nguyen et al. 2006).

*Intrarater reliability* is whether multiple measures by a single tester give similar results, thus looking at the degree of stability found when a test is performed two or more times under identical conditions (Thomas, Silverman et al. 2011).

*Interrater reliability* is whether two or more independent testers will find the same results, thus looking at the degree of consistency of several different testers taking the same measure under identical conditions (Thomas et al., 2005).

Reliability is about consistency or to what extent the measurements are free from error. It is often discussed in terms of observed score, true score and error score (Thomas et al., 2005). An observed score is when a measurement is taken from an individual, but whether this represents the true score of this individual is unknown. The true score is the score the individual would have gotten had the measurements been taken by a

perfect measuring instrument under ideal conditions (Portney & Watkins, 2009). Any mistake that was made in the measurement process is measurement error (Berg et al., 2004). The difference between the true score and the observed score is measurement error and can be summarized as the equation:

$$\textit{Observed score} = \textit{true score} \pm \textit{measurement error}$$

Researchers should aim to remove error to yield the true score and increase confidence in accurate measurements. The estimate of how much the measurements is attributable to error and how much represents the accurate reading is reflected by the degree of the reliability coefficient (Berg et al., 2004; Thomas et al., 2005). If there are big differences between the measurements, the reliability can be questioned, while if the difference is small there is no need to worry about the reliability (Portney & Watkins, 2009).

## **1.6 Sources of measurement error**

Measurement error is contamination or ‘noise’ of the true score. The sources of this error can be traced to the *instrument*, to the *testing procedures* or to the *subject* (Berg et al., 2004). This will then affect different components of reliability like *instrument*, *intrarater*, *interrater* and *intrasubject* reliability. These can be difficult to completely separate, and most reliability research will reflect a combination of the components (Carter et al., 2011). Careful planning, training, clear instructions and testing of equipment can minimize the potential for error, thereby improving reliability. Pilot testing is one useful way of detecting possible mistakes in the data collection and also to practice and review testing skills. Still, research is faced with error that cannot be controlled or predicted. Further, there will be a brief description of the threats of error for the different components.

*Instrument reliability* may be assessed in itself. Measurement error due to instrumentation includes obvious causes like inaccuracy or lack of calibration or whether it is the right tool for the job. Mechanical instruments will also have some level of background noise. Instrument error also refers to the difficulty of scoring tests or the inadequacy of a test to discriminate between abilities, for example when performing manual muscle testing or categorizing a gait pattern. The reliability of observational

scales uses the tester as the instrument and the reliability will also be linked to determining intrarater and interrater reliability (Carter et al., 2011).

**Intrarater reliability** may be affected because of learning or fatigue either by the tester or by the individual subject's inconsistency (Carter et al., 2011).

**Interrater reliability** can vary with the skills of the testers, inconsistencies in test procedures or changes in the variable being tested (Arnold et al., 2005). Even with detailed protocols and the same skills, different testers can disagree about the scores of a test. Before comparing testers, intrarater reliability should be established for each individual tester (Portney & Watkins, 2009).

**Intrasubject reliability** is associated with many factors like mood, motivation, fatigue, health, fluctuations in memory, change in performance, previous practice and familiarity with the test items (Thomas et al., 2005).

### **1.6.1 Systematic and random error**

Measurement errors are further divided into two types; **systematic** and **random errors**.

**Systematic errors** are predictable and occur in one direction, either overestimating or underestimating the true score. They can be detected, corrected or readjusted for. Systematic errors are not a problem for reliability as they are constant, but rather a concern of validity as the test values may not truly represent the quantity being measured (Portney & Watkins, 2009).

**Random errors** are due to unpredictable factors such as fatigue, inattention, mechanical inaccuracy or mistakes. It is often referred to as 'sampling error' (Lexell & Downham, 2005). It is assumed that if a sufficient number of measurements are taken, random errors will eventually be offset. The average score will then be close estimate of the true score (Portney & Watkins, 2009).

## **1.7 Statistical measures of reliability**

A comprehensive set of statistical methods is required to address the reliability of measurements (Lexell & Downham, 2005). Looking at changes in the mean of measurements from two test occasions is one of the methods. A difference between the two tests can be calculated; the mean difference (MD) and a 95 % confidence interval for the MD can show a true systematic difference. If zero is not included in the 95 % confidence interval, it is an indication of a systematic change. The paired t-test is a method for detecting systematic errors between groups of measurements (Carter et al., 2011). Reliability is often quantified in two ways; *relative* or *absolute*.

*Relative reliability* is the degree to which individuals maintain their position in a sample over repeated measurements (Carter et al., 2011). It is measured with correlation coefficients which give information about association between the variables and not necessarily about their proximity. Different correlation coefficients are used with different types of data. Pearson's correlation coefficient (Pearson's  $r$ ) may be used to quantify the reliability when the data is continuous and there are two repeated measures to be compared (Lexell & Downham, 2005). The intraclass correlation coefficient (ICC) is the most widely used test. Correlation is expressed as values between 0.00 – 1.00. A correlation coefficient of 1.00 indicates a perfect association between the repeated tests (Portney & Watkins, 2009). A disadvantage is that there is no universal agreement about the cut-off values on how to interpret the ICC-values, but a general guideline is that values above 0.75 are considered to be good reliability and values below 0.50 as low (Portney & Watkins, 2009; Lexell & Downham, 2005). Carter et al. (2011) use a system to determine the strength of reliability. However, they too, state the importance of interpreting these values in the context that they are used and not to be used as absolute standards (Carter et al., 2011). In this study their descriptions for the strength of correlation coefficients will be used as guidelines and are seen in **Table 1**. Acceptable reliability values may differ depending on whether one is using the measurement to make judgments about individual change, as when in clinical practice and requiring higher values, rather than group change as in group research design (Carter et al., 2011).



**Table 1:** Strength of correlation coefficients (Carter et al., 2011)

Value	Descriptive terms
0.00–0.25	Little, if any correlation
0.26–0.49	Low correlation
0.50–0.69	Moderate correlation
0.70–0.89	High correlation
0.90–1.00	Very high correlation

A difficulty with the ICC that has generated confusion is that there are at least six different formulas. Model 2 is in the majority of cases the most appropriate one for establishing interrater reliability. It is further divided into ICC<sub>2,1</sub> and ICC<sub>2,k</sub>, and they are used for a generalization across different testers. The ICC<sub>2,k</sub> includes the mean of k measures taken at every test time, where k is the number of trials (Arnold et al., 2005). Which model that is used is important to consider when comparing results from different research. ICC-based on single ratings will generate lower correlation values than those based on mean ratings (Portney & Watkins, 2009).

**Absolute reliability** is to what extent repeated measures vary. It is expressed either in the actual units of the measurements or as a proportion of the measurement values (Carter et al., 2011) Bland-Altman (BA) plots, coefficient of variation (CV), standard error of measurement (SEM) and the root mean squared error (RMSE) are all examples of measures for absolute reliability. BA plots are used to present the data graphically. The differences of the two tests are plotted against the mean of the two tests for each subject (Lexell & Downham, 2005). The 95 % CI can be included to look for systematic errors or outliers. The RMSE is measuring the average magnitude of the error. Lower values are better (Willmott & K, 2005).

## **1.8 Screening tests**

A cohort at the Oslo Sport Trauma Research Center (OSTRC) is aiming to investigate several of the above mentioned risk factors by including anatomical, genetic, biomechanical and neuromuscular measures. Injury registration and recording of injury mechanisms will enable researchers of the cohort to look for correlation between the results of the tests and the ACL ruptures registered over the years, thus maybe in the future allowing identification of risk factors. The next sections will give a description of the methods of measurement used in this cohort for collecting the previously mentioned anatomical risk factors.

### **1.8.1 Generalised joint laxity**

The Beighton and Horan joint mobility index (BHJMI) is an effective and the most commonly used way of screening for GJL is (Boyle et al., 2003). This method is a good clinical tool as it can be carried out easily within a minute, without any equipment and it applies a dichotomous principle (Kim et al., 2010). The criteria for assessing GJL were described by Carter and Wilkinson already in 1964, but were further modified by Beighton and Horan in 1969 and revised in 1973 (Boyle et al., 2003; Kim et al., 2010). The index is looking at both upper and lower limbs, assessing four bilateral movements and one unilateral. It is a nine point scale, so a score between 0–9 is calculated and is often placed into one of three categories: 0–2, 3–5, 6–9, where the latter is indicative of GJL (Beighton et al., 1973). There is no universal agreement for this cut-off level, although several studies have used the method described by Beighton (Kim et al., 2010). Boyle et al. (2003) conducted an intrarater and interrater reliability study of the BHJMI where they concluded that the measure has good to excellent reliability for females from 15-45 years of age.

### **1.8.2 Genu recurvatum**

Genu recurvatum can be measured with a goniometer. In the literature it is done either in the standing position or a supine non-weight-bearing position (Ramesh et al., 2005; Shultz et al., 2006). There are also variations whether the measurement is recorded in a passive, relaxed position, or with maximal active extension by the individual or with the examiner applying a posteriorly directed force to the anterior of the knee until passive resistance is noted (Trimble et al., 2002; Shultz et al., 2006; Ramesh et al., 2005).

### **1.8.3 Knee laxity**

The ability to quantify AP knee laxity has been a goal in the clinical and research setting in order to objectively measure persons with ACL injuries about the knee. An ACL rupture results in a larger AP movement of the tibia with regard to the femur (AP laxity) and the patients will complain of the knee feeling unstable (Wiertsema, van Hooff, Migchelsen, & Steultjens, 2008). The KT1000 Knee Ligament Arthrometer was first used in 1982, and has become the most frequently used testing device for assessing AP knee laxity. Numerous studies have been published on the reliability of the KT1000 showing contradicting results. The developers of the instrument showed high interrater reliability (Malcom, Daniel, Stone, & Sachs, 1985; Daniel et al., 1985). They established parameters for both normal and diagnostic laxities for ACL deficiency. There is a wide range of AP displacements in the normal knee, and 88% of the normal population show a right-left difference of less than 2 mm (Daniel et al., 1985). When a 3 mm displacement difference between the two knees is detected, the diagnostic accuracy of ACL rupture is very high (Daniel, Stone, Sachs, & Malcom, 1985).

### **1.8.4 Hip anteversion**

The Craig's test or the trochanteric prominence angle test (TPAT) is to date the only clinical method described to measure hip anteversion (Souza & Powers, 2009). A goniometer or an inclinometer is used as the measurement instrument (Ruwe, Gage, Ozonoff, & DeLuca, 1992).

### **1.8.5 Anthropometric measures**

Clinical techniques using tape measures and calipers are used in the collection of anthropometric measurements. The measurements will be used to estimate inertia parameter with Yeadon's method where 40 geometric solids based on 95 anthropometric measurements models the human body (Yeadon, 1990b). Shultz et al. (2006) have described that some measurements lack the required precision and that several reports have looked at the reliability of specific lower extremity anatomic measurements, but not a comprehensive collection. Therefore they investigated intratester and intertester reliability of clinical measures of lower extremity anatomic characteristics (Shultz et al., 2006).

For some of the tests included, the reliability has already been investigated on other samples as presented in **Table 2**. The studies included are those that have investigated healthy subjects or studies on patients with knee injuries, mainly ACL ruptures. When the cohort that this study is a part of was set up, a pilot study was done looking at the reliability of each of the tests. Therefore a larger study was needed to assess the reliability of the cohort in greater detail. This study aims to prove high quality of the cohort. The two most important criteria for judging the quality of research methods are looking at the reliability and validity (Arnold et al., 2005). The main purpose of performing reliability testing of the test in this cohort is to ensure generalizability. We wished to look at the interrater reliability of some of the tests being used as several testers are engaged in this on-going cohort. We also wanted to investigate the magnitude of interrater reliability for this particular sample.

### **1.9 Objective**

This project aims to test the interrater reliability of a series of tests evaluating anatomical risk factors for ACL injury in elite female football and handball players. The anatomical risk factors are GJL, genu recurvatum, knee laxity, hip anteversion and 12 anthropometric measures.

### **1.10 Hypothesis**

The following hypothesis is based on the already existing findings, as well as a reliability pilot study done in 2007 when starting the cohort.

***Anatomical characteristics will show moderate to high interrater reliability***

**Table 2: Studies investigating interrater reliability of the anatomical measures**

Test	Author	Journal	Title	Subjects	Results
<b>GJJL</b>	Boyle et al. (2003)	Journal of Athletic Training	Intrater and Interrater Reliability of the Beighton and Horan Joint Mobility Index	36 female students, mean age 25.4	Spearman Rho 0.87
	Shultz et al. (2006)	Clinical Journal of Sports Medicine	Intrater and Interrater Reliability of Clinical Measures of Lower Extremity Anatomic Characteristics: Implications for multicenter Studies	7 men and 9 women, mean age 25.6	ICC <sub>2,1</sub> 0.48–0.75
<b>Genu recurvatum</b>	Trimble et al. (2002)	Clinical Biomechanics	The relationship between clinical measurements of lower extremity posture and tibial translation	16 men and 27 women, mean age 23.3	ICC 0.95
	Wiertsema et al. (2008)	The Knee	Reliability of the KT1000 arthrometer and the Lachman test in patients with an ACL rupture	14 men and 6 women, mean age 30 All had total ACL rupture	ICC <sub>2,1</sub> 0.14
<b>Knee laxity</b>	Hanten & Pace (1987)	Physical Therapy Journal of the American Physical Therapy Association	Reliability of Measuring Anterior Laxity of the Knee Joint Using a Knee Ligament Arthrometer	43 men, mean age 18.5	ICC <sub>2,k</sub> 0.92
	Forster et al. (1989)	The Journal of Bone and Joint Surgery	Is the KT1000 Knee Ligament Arthrometer Reliable?	10 men, age 15–45 6 awaiting ACL surgery	Not reliable or reproducible!
	Schuser et al. (2004)	The American Journal of Sports Medicine	A New Mechanical Testing Device for Measuring Anteroposterior Knee Laxity	16 healthy knees and 36 knee injuries; ACL, PCL, MCL, menisci	Weighted statistics: 0.52 Moderate agreement
	Collette et al. (2012)	Knee Surg Sports Traumatol Arthrosc	Objective evaluation of anterior knee laxity; comparison of the KT-1000 and GNRB arthrometers	15 students, age 19–27	Intrater: significant difference
<b>Hip anteversion</b>	Shultz et al. (2006)	Clinical Journal of Sports Medicine	Intrater and Interrater Reliability of Clinical Measures of Lower Extremity Anatomic Characteristics: Implications for Multicenter Studies	7 men and 9 women, mean age 25.6	ICC <sub>2,1</sub> 0.48–0.74
	Souza et al. (2009)	Journal of Orthopaedic Sports Physical Therapy	Concurrent Criterion-Related Validity and Reliability of a Clinical Test to Measure Femoral Anteversion	9 men and 9 women, mean age 25.4	ICC <sub>2,3</sub> 0.83
<b>Anthropometrics</b>	Burkhart et al. (2008)	Journal of Biomechanics	Reliability of upper and lower extremity anthropometric measurements and the effect on tissue mass predictions	25 men and 25 women, mean age 22.5	ICC <sub>2,1</sub> 0.27–0.99  Only 6 measures with ICC < 0.70
	Shultz et al. (2006)	Clinical Journal of Sports Medicine	Intrater and Interrater Reliability of Clinical Measures of Lower Extremity Anatomic Characteristics: Implications for multicenter Studies	7 men and 9 women, mean age 25.6	ICC <sub>2,1</sub> 0.80–0.98

## **2. Methods**

### **2.1 Study design**

This project is a methodological study involving test-retest measures of GJL, genu recurvatum, AP knee laxity, hip anteversion and 12 anthropometric measures. A large cohort was set up at the Oslo Sports Trauma Research Center (OSTRC) aiming at investigating risk factors for ACL injuries in elite female football and handball players. The handball players have participated in annual screening tests of new players since the start of the cohort in June 2007, while the football players have been included since February 2009. Several authors have recommended large prospective studies to identify ACL risk factors (Shultz et al., 2006; Myklebust, Skjølberg, & Bahr, 2013; Padua, 2010). This study, looking at the reliability of some of the tests being used, is only a small, but rather important, piece of the large cohort.

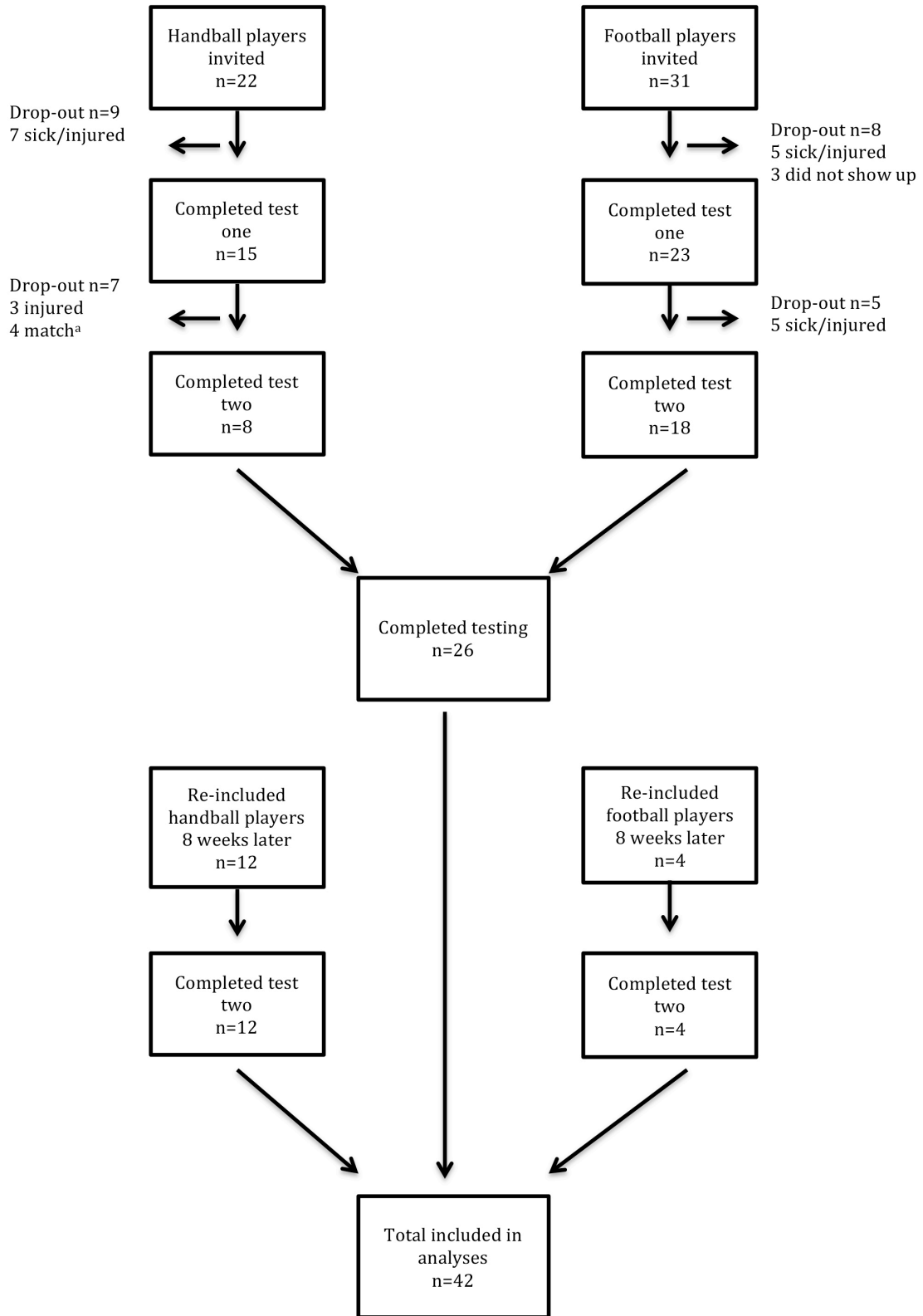
### **2.2 Participants**

All new elite handball players who are expected to play in the elite league the following season were invited to participate in the on-going cohort study. We contacted the coaches for the twelve elite handball teams in Postenligaen and were given contact details for their new players. Of these, 22 players were encouraged to participate in retesting 3–5 days after the initial testing. They represented the teams located geographically near to where the testing was based. One team agreed to participate in the reliability study with 21 players. Dates for testing and re-testing were arranged with the players by e-mail or telephone.

The new elite football players had already been enrolled earlier that season, thus we invited a team of female football players playing in the 1<sup>st</sup> division to participate in a test-retest session in August 2012 together with the handball players. We aimed to recruit players that were as close to the ones in the cohort as possible. This was to ensure reliability for both sports and increase the generalizability of the study. In total we had arranged for 31 football players for the reliability study, recruiting 24 players from one team.

The tests were carried out at The Norwegian School of Sport Sciences in Oslo from the 10<sup>th</sup> to the 20<sup>th</sup> of August 2012. Due to a high drop-out rate (n=12) during the initial

testing in August, re-testing was arranged for some of the players six weeks later (29<sup>th</sup> of September to 2<sup>nd</sup> of October). To ensure a sufficient number of players, they were given incentives of NOK 1480 to participate. In addition, we encouraged some of the players who had already been tested in August to participate for a new retest session, but we were unable to recruit enough players. Therefore, four female football players from a local elite team were also included in this second test-round, thus having to complete both the test and the retest (*Figure 3*).



*Figure 3: An overview of the players included and the drop-outs in this project*



All together 26 players were tested and retested with 2–10 days between the two tests, while it took 7 weeks between test one and two for 16 of the players (*Table 3*).

**Table 3:** Dates and numbers of days between testing

Days	Test 1	Test 2	Handball (n)	Football (n)
2	30.09.2012	02.10.2012		4
4	14.08.2012	18.08.2012		2
5	14.08.2012	19.08.2012	8	
6	12.08.2012	18.08.2012		8
10	08.08.2012	18.08.2012		4
40	20.08.2012	29.09.2012	6	
45	15.08.2012	29.09.2012	4	
45	18.08.2012	02.10.2012		1
46	15.08.2012	30.09.2012	2	
49	12.08.2012	30.09.2012		1
51	12.08.2012	02.10.2012		2
<b>TOTAL:</b>			<b>20</b>	<b>22</b>

### 2.3 Testers

The test team consisted of 11 physiotherapists (five had a MSc in sports physiotherapy, and another five were MSc students), five from sports sciences and one was a student. Most of the testers had been a part of the test team during the previous data collections, and were therefore familiar with the test procedures. In order for the testers to organise the equipment needed and to familiarize themselves with the procedures of the tests, the 6<sup>th</sup> and 7<sup>th</sup> of August 2012 were booked for pilot testing. Most testers only had to refresh their skills. The new testers had individual teaching and practice on these two days.

### 2.4 Collection of data

During the first round of testing in August, all the testers were thoroughly informed about the procedure of interrater reliability testing. We also ensured that different testers conducted test one and test two. Prior to testing on day one, each player were given an information sheet about the testing and signed a consent form (*attatchment 1, 2 and 3*).

## 2.5 Screening tests

The handball players completed several questionnaires and gave a blood sample, before going through the different test, as this was a part of the overall test procedures of the cohort.

### 2.5.1 Tests in this reliability study

This study focused on the anatomical tests and investigated reliability of the following tests; GJL, genu recurvatum, anterior-posterior knee laxity, hip anteversion and some anthropometric measures. A detailed description of the test procedures will follow.

### 2.5.2 Generalized joint laxity (GJL)

The Beighton and Horan Joint Mobility Index was used to assess GJL (CARTER & WILKINSON, 1964; Beighton et al., 1973). The tester demonstrated each step of the test as described in *Table 4*, having the player repeating it. A score  $\geq 4$  on this scale from 0–9, indicated increased joint laxity.



*Figure 4: Measuring generalized joint laxity*

**Table 4:** *Beighton and Horan joint mobility index*

Joint	Finding	Points
Left little (fifth) finger	Passive dorsiflexion beyond 90°	1
	Passive dorsiflexion ≤ 90°	0
Right little (fifth) finger	Passive dorsiflexion beyond 90°	1
	Passive dorsiflexion ≤ 90°	0
Left thumb	Passive dorsiflexion to the flexor aspect of the forearm	1
	Unable to passively dorsiflex thumb to flexor aspect of the forearm	0
Right thumb	Passive dorsiflexion to the flexor aspect of the forearm	1
	Unable to passively dorsiflex thumb to flexor aspect of the forearm	0
Left elbow	Hyperextends beyonds 10°	1
	Extends ≤ 10	0
Right elbow	Hyperextends beyonds 10°	1
	Extends ≤ 10	0
Left knee	Hyperextends beyonds 10°	1
	Extends ≤ 10	0
Right knee	Hyperextends beyonds 10°	1
	Extends ≤ 10	0
Forward flexion of trunk with knees full extended	Palms and hands can rest flat on the floor	1
	Palms and hands cannot rest flat on the floor	0

### **2.5.3 Genu recurvatum**

Genu recurvatum was measured with the player lying on a bench in a supine position with the distal aspect of the lower leg resting on a 15 centimetre high box. In this case, the box included with the KT1000 gear was used. A mark was placed on the lateral joint line by palpating the anterior and posterior portions of the knee, holding the index finger and the thumb on each portion and then eyeballing half the distance between the two. The most prominent point of the lateral malleolus and the greater trochanter were also palpated and marked with a pen. The player was lying relaxed with arms on the stomach. Genu recurvatum was measured to the nearest degree on the goniometer when aligning it on the lateral point of the knee joint line with the axis in line with the lateral malleolus and the greater trochanter.



*Figure 5: Measuring genu recurvatum*

#### **2.5.4 Knee laxity**

The KT1000 arthrometer (MEDmetric Corp, San Diego, California) was used to measure anterior-posterior (AP) knee laxity. With the player on a bench in a supine position, the medial knee joint line was marked. A thigh support box, which was a part of the equipment, was placed under the distal femur, so the player would keep the knee in a comfortably, flexed angle of between 20–35°. By strapping the thighs to the plinth external hip rotation was controlled. A foot support platform was also used as a part of the equipment to maintain proper leg orientation and encourage relaxation. The player was asked to relax and to keep the hands on the stomach.

A trial run was performed before the KT1000 was placed on the player's lower leg in order for the player to become familiar with the movement, and thus allow better relaxation. This was a manual demonstration of the movement where the tibia is displaced anteriorly relative to the femur, similar to the Lachmann's Test (Torg, Conrad, & Kalen, 1976).

The KT1000 arthrometer was strapped to the anterior side of the leg. During the measurement the patellar support was fixed with one hand. The other hand gave posterior directed force by pushing the force sensor handle until an audio tone signal, and the tester would reinitialize the arthrometer. Then an anterior directed force (134 N)



by pulling the force sensor handle was given, and a measure of the displacement in millimetres was read by the tester at the correct audio tone signal. A measurement of a manual pull was also taken. The hand was then placed on the posterior aspect of the proximal leg as when performing a Lachmann's test, instead of on the handle. Each leg was tested, starting with the right one. When the player had a previous ACL injury, the healthy leg was tested first.



*Figure 6: Measuring knee laxity*

### 2.5.5 Hip anteversion

Hip anteversion was measured using the trochanteric prominence angle test (TPAT). The player was in a prone position on the bench with the pelvic fixed with a belt, the knee at the edge of the bench and passively flexed at 90°. The hip was passively rotated internally and externally until the most prominent lateral portion of the greater trochanter was palpable. In this position, the angle between a vertical line and the shaft of the tibia is measured with a goniometer as a positive angle. The variable for the hip anteversion is the number of degrees measured.



*Figure 7: Measuring hip anteversion*

### 2.5.6 Anthropometric measures of lower extremities and height

In preparation for the three-dimensional (3D) motion analyses, we collected anthropometric measures of the players. The measures were used to estimate inertia parameters with Yeadon's method (Yeadon, 1990a; Yeadon, 1990b). Out of these, twelve measures were chosen in this interrater reliability study (*Table 5*).

**Table 5:** *The anthropometric measurements included in this interrater reliability study, including a description of the standardized procedures*

Measurement	Procedure
1. Height	The player was placed barefoot with heels against the wall. A point on the wall was made using a level from the top of the head. The distance from the floor to the point was measured with a tape measure
2. Tibial width right/left leg	The distance between the lateral bony aspects of the tibial condyles measured with a caliper
3. Femoral width right/left leg	The distance between the lateral bony aspects of the femoral condyles measured with a caliper
4. Anterior Superior Iliac Spine (ASIS) distance	The most anterior point of the iliac crest is palpated and marked bilaterally and the distance between them measured with a tape measure
5. Femoral length right/left leg	<p>The length was calculated by adding three different distances between four specific landmarks.</p> <ul style="list-style-type: none"> <li>• <b>The hip joint centre:</b> <i>estimated as the distance between the ASIS x 0.33, directed vertically below the ASIS level</i></li> <li>• <b>The crotch:</b> <i>an anterior mark on the thigh measured in a horizontal line from the crotch</i></li> <li>• <b>Mid-thigh:</b> <i>the maximum perimeter of the thigh</i></li> <li>• <b>The knee joint centre:</b> <i>the joint line, in the middle of the lateral femoral condyle and the lateral tibial condyle</i></li> </ul>
6. Tibial length right/left leg	<p>The length was calculated by adding three different lengths between four specific landmarks.</p> <ul style="list-style-type: none"> <li>• <b>The knee joint centre:</b> <i>the joint line, in the middle of the lateral femoral condyle and the lateral tibial condyle</i></li> <li>• <b>Maximum calf perimeter:</b> <i>measured as a point anteriorly</i></li> <li>• <b>Minimum calf perimeter:</b> <i>measured as a point anteriorly</i></li> <li>• <b>Ankle joint centre:</b> <i>the lateral malleolus</i></li> </ul>
7. Femoral-tibial ratio right/left leg	Using the measurements from the hip joint center (##) to the ankle joint center (##) calculating the femoral length / the tibial length

The caliper that was used for measuring the width of the femoral and tibial condyles was placed on both sides of the condyles and a slight pressure was given in order to ensure measurements of the bony aspects of the condyles. When measuring the perimeter the tape measure was placed close to the skin without giving any extra pressure. All measurements were completed with the player in standing position and at the end of expiration phase. The variable for anthropometrics is the number of centimetres (cm) measured to the nearest 0.5 cm.

## **2.6 Remaining tests of the cohort**

The players went through all of the tests already described, but also conducted tests measuring strength, balance, flexibility and movement patterns. Reliability will also be assessed for the following tests in separate studies; leg extensor strength, the star excursion balance test, hamstrings flexibility and 3D motion analysis.

## **2.7 Test order**

The tests were carried out in a pre-defined order according to the player lists provided. The female elite handball players spent seven hours for their first day of testing (*attachment 4*), and three hours on day two (*attachment 5*). The seven hours on the first day was only because they were organized in pairs and had to complete all tests. The football players spent three hours on testing both on day one and on day two. Testing was organised at seven different stations for the handball players, while the subjective station was not included in the reliability testing. All anatomical measurements were conducted at one test station in the following order; the star excursion balance test, the Beighton and Horan score for GJL, knee laxity with the KT1000, hip abductor strength with the dynamometer, hamstring flexibility, genu recurvatum and hip anteversion. Both testers and players had a timetable to follow to make sure no players would miss any test stations. In this timetable the players were organised in a pre-defined random order with 30 minutes on each station. When ahead of schedule, or if players had not turned up, the next available player was tested.

## **2.8 Statistics**

A sample size of 40 players was needed (20 football and 20 handball players) to detect a moderate to strong correlation or significant difference between test one and test two. The data were plotted into Excel (Microsoft Excel 2010) and were analysed using PASW Statistics 18. Descriptive data of demographic variables are presented with the mean, standard deviation, minimum and maximum of the variables.

We tested the dependant variables for a normal distribution using Shapiro Wilk test. For the parametric techniques we were planning to use, it was assumed that the variables would be normally distributed. Most of the statistical techniques are reasonably robust and tolerant of violations of this rule. With a sample size larger than 30, it should not



cause any major problems even when data were not normally distributed (Pallant, 2010). Therefore, we used the statistical procedures as planned.

Paired t-tests were used to assess systematic bias between-tests, that is if values obtained from one test day systematically differed from the other. The Wilcoxon signed ranks test was used when analysing the Beighton and Horan joint mobility index, as the non-parametric equivalent of the paired t-test. The sub-tests of the Beighton and Horan joint mobility index were analysed with the Mc Nemar test as the variables were measured as a nominal scale. To assess interrater reliability, intraclass correlation coefficients (ICC) with the corresponding 95 % confidence intervals (CI) were calculated, as well as Bland-Altman plots. The ICC model used was model 2, where the ANOVA is performed as a two-way random effects model (ICC<sub>2,1</sub>) with absolute agreement. The paired t-tests gave the mean difference (MD) with 95% CI. For the Bland-Altman plots the difference of the scores of the two days were plotted against the mean for each player's two scores. Four horizontal reference lines were superimposed on the scatterplots, one solid centre line representing zero difference, one line for the mean difference between the measurements, along with lines marking two standard deviations above and below the mean ( $\pm 1.96$ ).

The root mean square error (RMSE) was calculated. The error is the amount by which the value implied by the repeated measures differs from the quantity to be estimated. Therefore the RMSE is a useful measure and it also shows the calculated results on the same scale as the data (Vihinen, 2012). For all measurements a p-value at or below 0.05 was considered statistically significant.

## **2.9 Ethics approval**

The Regional Committee for Medical Research Ethics and the Norwegian Social Science Data Services approved the cohort study, which this specific study is a part of (*attachment 6*). Hence, there was no need for further approvals. All included participants were asked to sign a consent form prior to testing. They were also informed about the data being confidentially treated, and about their rights to withdraw from the project at any point.

### 3. Results

#### 3.1 Descriptive data of the players

In total, 42 players (20 handball and 22 football players) were included in both test one and test two. Descriptive data of their age, height, weight and number of years playing at an elite level are listed in Table 6.

*Table 6: Descriptive data of the players in the study*

Variables	Sport	N	Mean	± SD	Minimum	Maximum
Age (years)	Handball	20	21.9	4.4	17.0	34.0
	Football	22	21.1	3.2	17.0	28.0
Height (cm)	Handball	20	170.0	6.4	158.0	181.0
	Football	22	167.3	3.0	160.0	171.5
Weight (kg)	Handball	20	68.0	9.0	55.0	90.0
	Football	17	63.4	5.0	51.0	71.0
Elite level (years)	Handball	20	3.4	4.2	0.0	15.0
	Football	20	3.3	3.2	0.0	12.0

*N= number of participants, SD= standard deviation*

### 3.2 Differences between tests on day one and two

The mean values for all anatomical tests on the two sessions are presented in *Table 7*.

There were significant differences between the values of day one and day two on 11/18 of the tests.

**Table 7:** Anatomical tests on day one and two. Mean values( $\pm$  SD) is presented, followed by the mean difference ( $\pm$  SD) between the two days and the significance of the difference

TEST (n=42)	Day one: Mean $\pm$ SD	Day two: Mean $\pm$ SD	MD $\pm$ SD of MD	p-value
Genu Recurvatum Right ( $^{\circ}$ )	178.00 $\pm$ 3.40	180.00 $\pm$ 3.50	-1.83 $\pm$ 3.60	0.002*
Genu Recurvatum Left ( $^{\circ}$ )	178.00 $\pm$ 3.80	180.00 $\pm$ 3.70	-1.95 $\pm$ 3.80	0.002*
AP Knee Laxity Right (mm)	4.50 $\pm$ 1.50	4.20 $\pm$ 1.40	0.30 $\pm$ 1.20	0.10
AP Knee Laxity Left (mm)	5.50 $\pm$ 1.50	3.90 $\pm$ 1.20	1.55 $\pm$ 1.40	p $\leq$ 0.001*
Hip Anteversion Right ( $^{\circ}$ )	9.90 $\pm$ 2.80	7.50 $\pm$ 2.40	2.50 $\pm$ 3.10	p $\leq$ 0.001*
Hip Anteversion Left ( $^{\circ}$ )	9.20 $\pm$ 2.60	7.90 $\pm$ 3.20	1.30 $\pm$ 3.40	0.018*
Height (cm)	168.40 $\pm$ 5.20	168.50 $\pm$ 5.20	-0.02 $\pm$ 0.70	0.81
Tibial Width Right (cm)	9.30 $\pm$ 0.60	9.40 $\pm$ 0.50	-0.12 $\pm$ 0.60	0.22
Tibial Width Left (cm)	9.30 $\pm$ 0.60	9.40 $\pm$ 0.50	-0.13 $\pm$ 0.60	0.18
Femoral Width Right (cm)	9.80 $\pm$ 0.50	9.90 $\pm$ 0.50	0.02 $\pm$ 0.60	0.80
Femoral Width Left (cm)	9.80 $\pm$ 0.60	9.90 $\pm$ 0.50	-0.04 $\pm$ 0.70	p $\leq$ 0.001*
ASIS Width (cm)	24.80 $\pm$ 2.20	24.40 $\pm$ 1.70	0.35 $\pm$ 1.90	0.23
Tibial Length Right (cm)	39.50 $\pm$ 2.60	40.50 $\pm$ 2.00	-0.95 $\pm$ 1.70	p $\leq$ 0.001*
Tibial Length Left (cm)	39.30 $\pm$ 2.10	40.10 $\pm$ 2.00	-0.89 $\pm$ 1.40	p $\leq$ 0.001*
Femoral Length Right (cm)	41.20 $\pm$ 2.30	40.70 $\pm$ 2.30	0.50 $\pm$ 1.90	0.09
Femoral Length Left (cm)	41.20 $\pm$ 2.60	40.20 $\pm$ 2.40	0.96 $\pm$ 1.90	p $\leq$ 0.001*
Femoral-Tibial Ratio Right (cm)	1.00 $\pm$ 0.06	1.00 $\pm$ 0.06	0.04 $\pm$ 0.05	p $\leq$ 0.001*
Femoral-Tibial Ratio Left (cm)	1.05 $\pm$ 0.06	1.00 $\pm$ 0.05	0.46 $\pm$ 0.07	p $\leq$ 0.001*

SD= standard deviation, MD= mean difference, \*p  $\leq$  0.05

### 3.3 Interrater reliability

**Table 8** lists the interrater reliability estimates for all tests. The ICC for KT1000 shows a large difference between the left and right leg. Hip anteversion, tibial width, femoral width and femoral-tibial ratio all have low ICC values. The mean differences with negative values were genu recurvatum, height, bilateral tibia width, femoral width left and bilateral tibia length. These measurements will therefore have had higher values in cm or degrees on test day two.

**Table 8:** Intraclass correlation coefficients (ICC) with 95% CIs for the anatomical tests, followed by mean difference with 95% CI

TEST (n=42)	ICC <sub>2,1</sub>	95 % CI	MD	95 % CI
Genu Recurvatum Right (°)	0.41	(0.12–0.64)	-1.83	(-2.9–0.72)
Genu Recurvatum Left (°)	0.44	(0.14–0.66)	-1.95	(-3.1–0.80)
AP Knee Laxity Right (mm)	0.65	(0.42–0.79)	0.30	(-0.06–0.68)
AP Knee Laxity Left (mm)	0.29	(-0.08–0.59)	1.55	(1.11–1.98)
Hip Anteversion Right (°)	0.22	(-0.05–0.48)	2.50	(1.51–3.44)
Hip Anteversion Left (°)	0.29	(0.01–0.54)	1.30	(0.23–2.34)
Height (cm)	0.99	(0.99–1.00)	-0.02	(-0.23–0.18)
Tibial Width Right (cm)	0.33	(0.04–0.58)	-0.12	(-0.31–0.08)
Tibial Width Left (cm)	0.35	(0.06–0.59)	-0.13	(-0.32–0.06)
Femoral Width Right (cm)	0.35	(0.05–0.59)	0.02	(-0.16–0.21)
Femoral Width Left (cm)	0.30	(-0.00–0.56)	-0.03	(-0.25–0.18)
ASIS Width (cm)	0.55	(0.30–0.73)	0.35	(-0.24–0.93)
Tibial Length Right (cm)	0.68	(0.39–0.83)	-0.95	(-1.49–0.41)
Tibial Length Left (cm)	0.71	(0.38–0.86)	-0.89	(-1.32–0.46)
Femoral Length Right (cm)	0.66	(0.46–0.80)	0.50	(-0.09–1.08)
Femoral Length Left (cm)	0.68	(0.41–0.86)	0.96	(0.39–1.54)
Femoral-Tibial Ratio Right (cm)	0.40	(0.08–0.64)	0.04	(0.02–0.06)
Femoral-Tibial Ratio Left (cm)	0.17	(-0.08–0.42)	0.46	(0.02–0.07)

ICC= Intraclass correlation coefficient, CI ICC= Confidence interval of the intraclass correlation coefficient, MD= Mean Difference

### 3.4 Generalized joint laxity

There was no difference ( $p= 0.36$ ) between day one and day two of testing when looking at the Beighton Horan joint mobility index. Further, there were no differences between the two test days when analysing all of the sub tests of the Beighton and Horan joint mobility index separately.

### 3.5 Actual measurement error of the tests

Femoral length on the right side had an error of 2.1 cm, which was the largest RMSE of the variables measured in cm (*Table 9*). The genu recurvatum on the left side had an error of  $4.2^\circ$ , which was the largest of the variables measured in degrees.

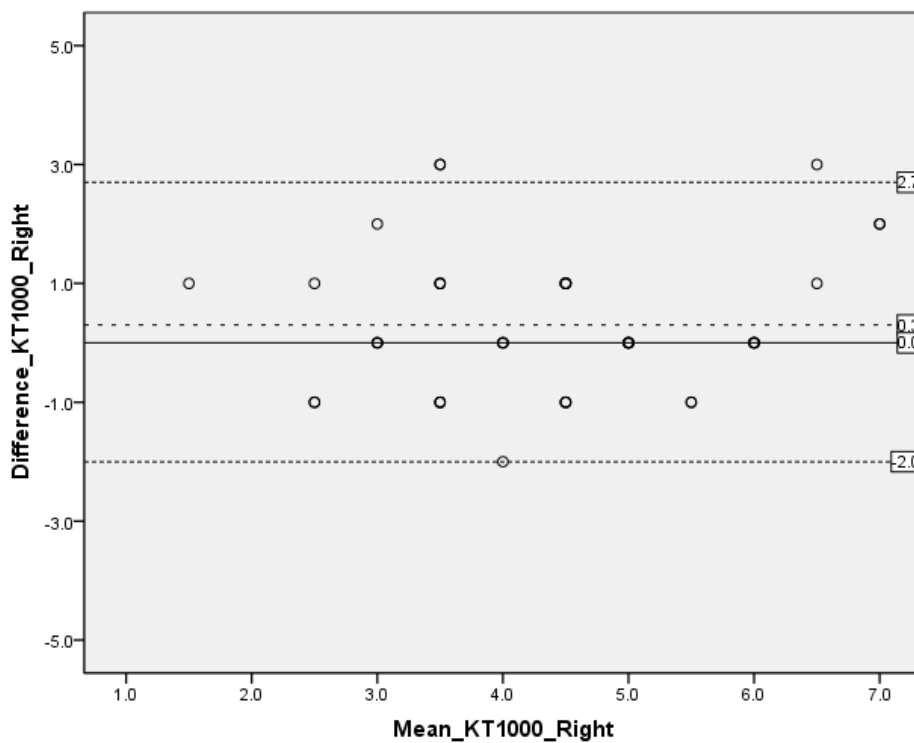
*Table 9: Root mean square error (RMSE) of anatomical test*

TEST (n=42)	RMS error
Genu Recurvatum Right ( $^\circ$ )	4.0
Genu Recurvatum Left ( $^\circ$ )	4.2
AP Knee Laxity Right (mm)	1.2
AP Knee Laxity Left (mm)	2.1
Hip Anteversion Right ( $^\circ$ )	3.9
Hip Anteversion Left ( $^\circ$ )	3.6
Height (cm)	0.7
Tibial Width Right (cm)	0.6
Tibial Width Left (cm)	0.6
Femoral Width Right (cm)	0.6
Femoral Width Left (cm)	0.7
ASIS Width (cm)	1.9
Tibial Length Right (cm)	2.0
Tibial Length Left (cm)	1.6
Femoral Length Right (cm)	1.9
Femoral Length Left (cm)	2.1
Femoral-Tibial Ratio Right (cm)	0.1
Femoral-Tibial Ratio Left (cm)	0.1

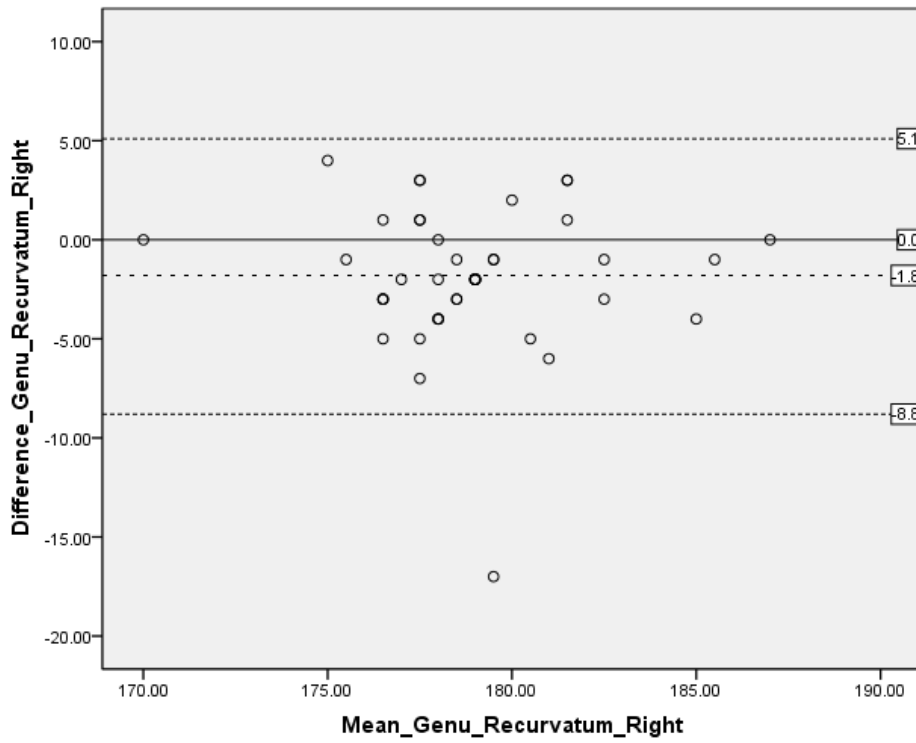
Bland-Altman plots illustrate the difference between day one and day two tests plotted against the means of the two test days (**Figure 8**). Only some of the tests are shown here. The plots presented illustrate different findings. **Plot 1** shows a horizontal band and demonstrates that assumptions for linear regression have been met. The systematic drift is only minor and almost all of the measurements are within  $\pm 1.96$  SD. The second plot has fairly good measurements, but there is one outlier. One extreme value has significantly altered the statistical description of the data. **Plot 3** shows a systematic bias with the measurements of day two being lower on average. **Plot 4** and **5** show grid patterns due to the resolution of the variables.

**Figure 8:** Bland-Altman plots showing individual differences and averages of the two days of testing. The solid centre line represents zero difference. The mean difference is represented by the centre dashed line, while the upper and lower dashed lines represent  $\pm$  SD

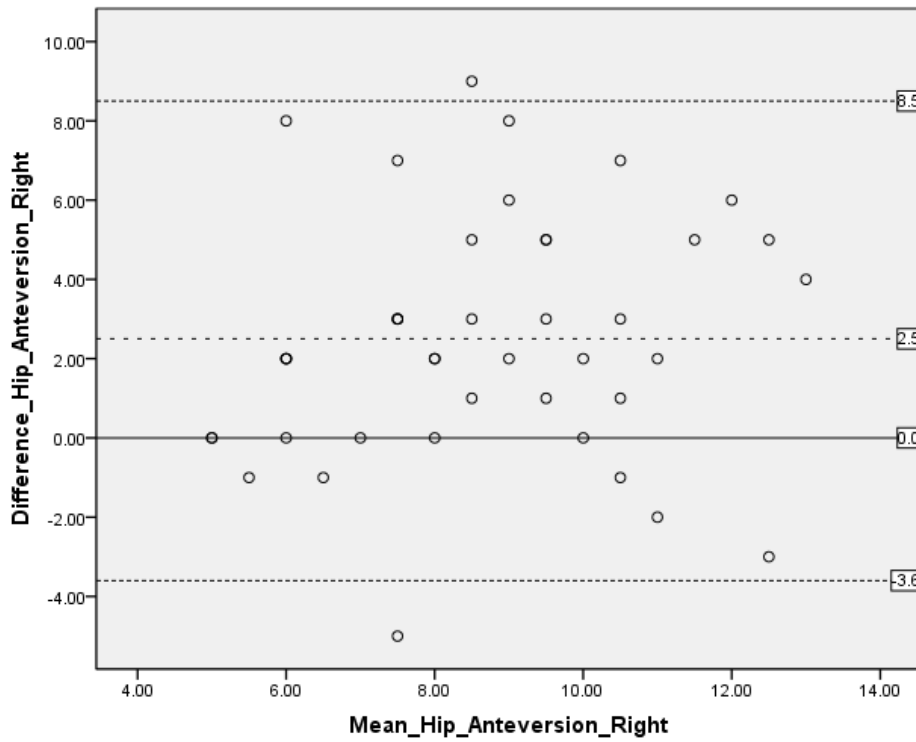
**Plot 1:** AP Knee laxity measured with KT1000 on the right leg



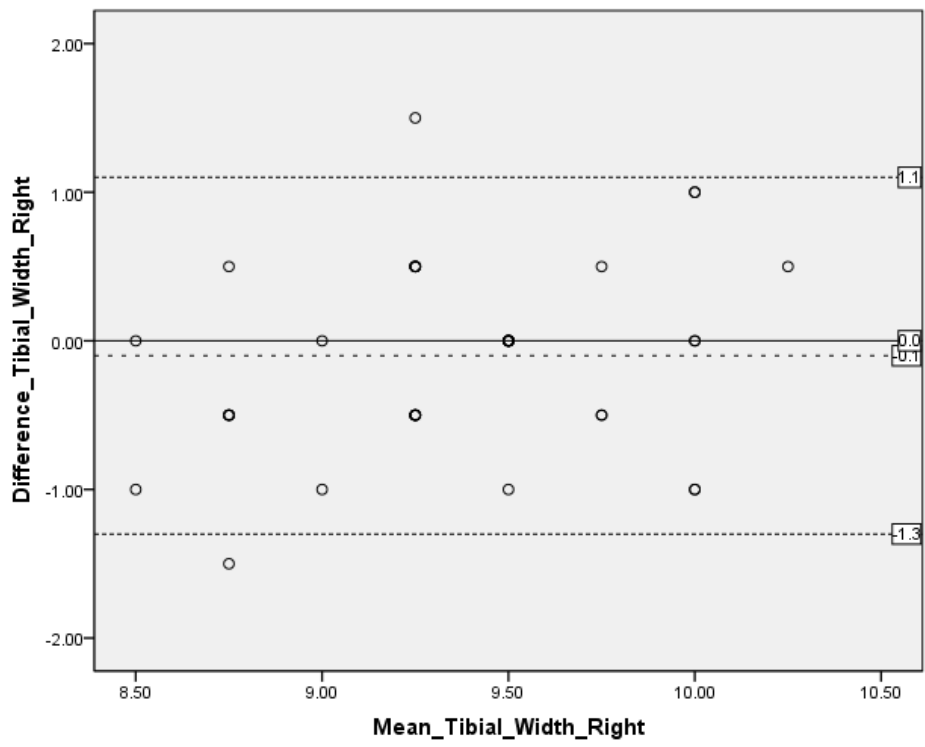
**Plot 2:** Genu recurvatum measured with the goniometer on the right leg



**Plot 3:** Hip anteversion right side



**Plot 4:** Tibial width measured with a caliper on the right leg





## 4. Discussion

### 4.1 *Reliability of the different tests*

In the current study we have investigated the reliability of anatomical tests used in a larger cohort aiming to establish risk factors for ACL rupture in elite female football and handball players. The hypothesis was that the interrater reliability for anatomical characteristics would be moderate to high. The main findings were that the measurements of generalized joint laxity has an excellent reliability, there was a large difference in correlation coefficients of knee laxity between the right and the left leg, the anthropometric measurements of segment lengths of the lower extremities showed moderate to high correlation while measurements of lower extremity widths low to moderate correlation values. In addition, more than half of the tests showed a significant difference between the two days of testing.

Although the different tests in this study are commonly used for varying clinical conditions, the reliability of the test for this cohort in particular have not yet been clearly established. A test is not reliable in itself, but is closely linked to the sample and the conditions it has proven reliability for (Streiner & Norman, 2006). So, for example, the Beighton tests showed good to excellent reproducibility for generalized joint hypermobility (GJH) and benign joint hypermobility syndrome (BJHS) in a study looking at patients with Ehlers-Danlos Syndrome or BJHS, but it would still need to be proven reliable for young female athletes in order for the results of the cohort to be believable (Juul-Kristensen, Rogind, Jensen, & Remvig, 2007).

#### 4.1.1 **Generalized joint laxity**

Measuring GJL with Beighton and Horan joint mobility index showed substantial agreement between the two days of testing in our study. Sources of error, which may have contributed to less than perfect reliability may have included tester error, for example variability in verbal instructions. Delayed onset of muscle soreness (DOMS) can have the effect of reducing range of movement (Zainuddin, Newton, Sacco, & Nosaka, 2005). As the players had no restrictions in the days prior to the testing, DOMS could possibly have reduced elbow extension and forward flexion of the trunk and thus have affected the measurements. On the other hand, the handball players were in the beginning of their season and the football players were mid-season at the time of

testing, so any excessive strength training leading to DOMS is unlikely. The results are in agreement with previous research where BHJMI has reported good to excellent reliability on healthy females (Boyle et al., 2003).

#### **4.1.2 Genu recurvatum**

The interrater reliability of genu recurvatum was low (ICC<sub>2,1</sub> of 0.41 on left leg and 0.44 on right leg). The reasons for this might be difficulties with palpation of landmarks and aligning the axis of the goniometer precisely to measure the number of degrees with the goniometer. This might have to do with testers' experience and precision. Some anatomical landmarks are not clearly identifiable discrete points, but rather relatively large and curved areas, like for example the greater trochanter, and palpation discrepancies can lead to angular variations. This is supported by Moriguchi et al. (2009). Our results are in contrast with the findings of Trimble et al. (2002). They had an ICC value of 0.95. There are slight variations in measurement methods, as they did their measurements while the subjects were standing and contracting their quadriceps muscles, while our players were lying in a relaxed position. Also they used a goniometer where the arms were lengthened. They measured three times and used the average of three measurements (Trimble et al., 2002). Using a mean of three measurements, thus using a ICC<sub>2,k</sub> model, will generally result in higher values (Portney & Watkins, 2009).

#### **4.1.3 Knee laxity**

Knee laxity measurements with the KT1000 showed moderate reliability on the right leg (ICC<sub>2,1</sub> 0.65), while on the left leg the reliability was low (ICC<sub>2,1</sub>0.29). This substantial difference between the ICC values of the two legs is not described in other studies. Reasons for this could be that maintaining contact between the flat patellar sensor pad and the rounded surface of the patella is awkward. A further explanation could be the difference in the dominant and weaker side of the examiner. Both testers in our study were right-hand-dominant physiotherapists. Still, when testing on the left knee, the left hand performed the pull. No comments are made in the protocol for the positioning of the testers or what hand to use. The less experienced tester in our study supports these explanations and reports testing on the left leg to be more demanding. This weakness has also been pointed out by Sernert et al. (2007). They did a study looking at knee laxity measurements examined by a left-hand and a right-hand-

dominant physiotherapist in patients with ACL injuries and healthy controls. The left-handed tester performed all the KT1000 measurements pulling with the left hand and the right-handed with the right (Sernert, Helmers, Kartus, Ejerhed, & Kartus, 2007). They found that the right-handed testers measured significantly higher values in the right knee while the left-handed tester measured significantly higher values in the left knee.

Other factors that could possibly contribute to the differences in measurements between testers are height, strength and hand size. Ballantyne et al. (1995) investigated these factors assuming that gender would influence the measurements. They found that testers experience was a more important factor influencing reliability than gender, which is in agreement with our results as both of our testers were female and one of our testers had never used the KT1000 prior to the testing in this study.

The less amount of experience with the use of KT1000 for one of our testers is believed to be an explanation for not obtaining higher ICC-values. One tester had performed the testing with the KT1000 in the cohort in previous seasons, while the other tester tried the instrument first time during practise prior to the pilot. Collette et al. (2011) showed significant differences in the results between testers. They had 15 physiotherapists participating as testers, and seven of them had used the KT1000 for more than six months on many different patients while the other six only had practice before the testing (Collette, Courville, Forton, & Gagniere, 2012). They showed that the KT1000 was less reproducible in the hands of an inexperienced tester than with an experienced. This is supported by other research (Pugh, Mascarenhas, Arneja, Chin, & Leith, 2009; Konig, Rutt, Kumm, & Breidenbach, 1998). In contrast to these results, Forster et al. (1989) showed inconsistency between KT1000 measurements by different examiners, and experience did not make the testers more consistent in their measurements. They queried the accuracy and reliability of the instrument. However, a weakness of their study is that the sample size was small (Forster, Warren-Smith, & Tew, 1989; Schuster, McNicholas, Wachtl, McGurty, & Jakob, 2004). Still, it would be advisable for the testers of the cohort to practise a greater amount prior to testing to aim for more reliable results.

We used the foot support platform in our study to maintain proper leg orientation, but this still allows variations of rotation and could possibly explain measurement errors. Positioning the knee in the same way with the same amount of rotation can be difficult between two testers (Collette et al., 2012; Daniel et al., 1985; Hanten & Pace, 1987). The knee should be in a neutral rotation during testing, because anterior displacement is reduced by internal rotation and increased by external rotation of the tibia (Guskiewicz et al., 1995). Different devices have been made over the years in attempt to control the rotation of the tibia (Mayr et al., 2011). Mayr et al. (2011) have developed a laximeter used in combination with the KT1000. Their results showed very high interrater reliability on healthy subjects (Conbach's alpha of 0.95). The testers in our study attempted to keep proper leg orientation through the test with the KT1000. However, it is advisable for the future training of testers, to give extra emphasis to leg orientation and how a difference in positioning the feet may affect the result.

In our study the players were encouraged to relax while AP knee laxity was measured, but the degree of relaxation was only subjectively registered by the testers. The lack of muscle relaxation has been implicated as a cause of measurement error, especially the hamstring muscle as it can decrease anterior translation (Hanten & Pace, 1987; Collette et al., 2012). On the other hand this might not be an explanation as two studies found no significant difference in values obtained awake and under anaesthesia (Forster et al., 1989; Sernert, Kartus, Kohler, Ejerhed, & Karlsson, 2001). Collette et al. (2012) placed surface (EMG) electrodes on the hamstrings to control relaxation, but still observed a significant difference between two testers.

A weakness of our study was using the KT1000 as an objective instrument to measure anterior translation of the tibia quantifying the movements in mm. KT1000 has been validated as a diagnostic tool for ACL injury, but it is also used as a continuous variable to compare the results of reconstruction techniques. Arneja & Leith (2009) reviewed the validity of the KT1000 arthrometer using it as a tool to compare the derived scores as a continuous variable. They concluded that it should be used as a dichotomous diagnostic tool and evaluations with continuous variables may be inappropriate (Arneja & Leith, 2009). A weakness of their review is that it was only based on four studies, where three had limitations in methodology. However, when discussing our results with a statistician (Ingar Holme, OSTRC) he commented on the exact same problem. The

variables for KT1000 are not continuous as such, as the variance of the values is low, while the variance between each possible measurement is large. This may affect the ICC values. Portney & Watkins (2009) demonstrated how the reliability coefficient decreases as the true variance in a set of scores decreases. The between subject variance of the scores of knee laxity in our data is small. Reliability is based on the proportion of the total observed variance that is attributable to error. Reliability will improve as the total variance increases. When the total between subject variance gets bigger, the proportion of the error component will be less accounted for (Portney & Watkins, 2009). A lack of variability may occur when the samples are homogeneous, the scoring is strict or the range of the rating system is very restricted (Portney & Watkins, 2009). This can be checked, by looking for significance of the between subject variance in the ANOVA. Our results show significant differences in these tests, so the sample should be heterogeneous according to these analysis. Still, with the sample being female football and handball players within the age range of 17–34 years, the sample can be said to be homogenous. Further, the ranges of the rating system are restricted with the KT1000 being measured to the nearest mm. In a future reliability study, these problems can be solved by including individuals that have a wider range of scores, for example men and women. Also, considering statistical analysis for variables on a nominal scale might improve the results.

#### **4.1.4 Hip anteversion**

There was low interrater reliability of hip anteversion in our study, with ICC<sub>2,1</sub> values of 0.22 of the right hip and 0.29 of the left. The reasons behind such low values could have several explanations, such as again the experience of the testers, the statistical approach for calculating reliability and the sample studied.

Being able to palpate the greater trochanter is critical for attaining an accurate measurement, and deciding the exact position when the greater trochanter is most prominent can be difficult. One of our testers was unfamiliar with the test prior to this study and inexperience of the testers may be a source of error. However, both testers had anatomy knowledge and experience with palpation as both were physiotherapists. Palpation of the greater trochanter has shown large interrater discrepancies (Moriguchi et al., 2009). Souza et al. (2009) suggested that their way of training their testers prior to the testing may have contributed to their high level of reliability for hip anteversion

(ICC<sub>2,3</sub> of 0.83). Magnetic resonance imaging (MRI) measures were revealed to the testers and they could retest on a small pilot group. They acknowledged the need to investigate the influence training has on clinical measures (Souza & Powers, 2009). They too, based their results on an average of three measurements.

It seems like measuring hip anteversion is difficult on overweight individuals. Our sample only included females, and there are large differences in body composition between the sexes (Stevens, Katz, & Huxley, 2010). Fat distribution differs, with women having a greater hip circumference. When measuring hip anteversion a likely source of error is the soft tissue overlying the greater trochanter. Souza et al. (2009) found that the largest errors in their study were present in the subjects with the highest BMI. This is supported by findings done by Piva et al. (2006). They tested on subjects with patellofemoral pain and initially reported a ICC<sub>2,1</sub> of 0.45. Interestingly, their reliability improved to 0.81 after excluding individuals with a BMI of 25 or greater. This could be an explanation for our low results. However, it is less likely considering that this study was done on athletic females with BMI of the football players of 19 and BMI of the handball players of 23. It is also considered unlikely that muscle bulk could lead to errors when palpating as the greater trochanter lies fairly superficially.

In the literature, several investigations have tested interrater reliability of hip anteversion measured clinically (Ruwe et al., 1992; Piva et al., 2006; Souza & Powers, 2009; Shultz et al., 2006; Sutlive et al., 2004). The reliability coefficients of these studies vary considerably (0.17–0.83). The majority of these studies show a low to moderate reliability which is in agreement with our results (Ruwe et al., 1992; Sutlive et al., 2004; Piva et al., 2006). Souza et al (2009) tested on healthy adults with an average body mass index (BMI) of 23 and had an ICC<sub>2,3</sub> value of 0.83 and their average difference was only 1.7°. Some of the studies are done on other samples; subjects with patellofemoral pain and children with cerebral palsy as the latter is a population with high prevalence of excessive hip anteversion (Ruwe et al., 1992; Sutlive et al., 2004). When new testers are included in the next test rounds of the cohort, it is recommended that a greater amount of supervision and practise is carried out prior to testing. Also using an ICC<sub>2k</sub> value could improve results of future reliability testing.

#### **4.1.5 Anthropometric measures of lower extremities and height**

Tibial and femoral width, the measurements measured with a caliper, showed low reliability (ICC<sub>2,1</sub> of 0.30–0.35). Again we need to consider how ICC values may be influenced by the restricted variance of our sample as discussed with the KT1000. The rating system here is restricted to the nearest 0.5 cm, so the variable is not continuous as such. This was also visible as a grid pattern seen in the Bland-Altman plots. The SD of the mean on the two days of these measurements showed values close to the nearest half cm. Still femoral width on the left side had a significant difference between the means of the two days of testing. The reason for this is unclear. The RMSE is also fairly low on these caliper measurements (0.6–0.7 cm) considering that the measurements are to the nearest 0.5 cm. In order to improve ICC values, measurements could possibly be taken in mm in future studies. This would on the other hand complicate the readings of the measurements.

A source of error could be a slight variation of landmark identification. One point is used for calculating several different anthropometric measures. For example the ASIS or knee joint centre in our study as described in *Table 5*. Variations can therefore lead to differences in outcomes (Moriguchi et al., 2009). A difference of three cm was found in a study where they compared waist circumference measured at four different, yet closely located points (Wang et al., 2003).

A strength of our study is that measurements were performed bilaterally. Shultz et al. (2006) only tested on the right side and side-to-side symmetry of anatomic measurements cannot be assumed.

Significant differences were found between the two days of testing in half of the anthropometric measures that were tested for reliability. Burkhart et al. (2008) looked at the reliability of 48 segments measurements on healthy adults to report the effect of measurement differences on tissue mass predictions. They also reported significant differences between testers in more than half of their measurements sites, but these differences were relatively small in general (75–80 % were < 1 cm) (Burkhart, Arthurs, & Andrews, 2008). More than 80 % of their measurement sites showed good to excellent interrater reliability.

## **4.2 Methods**

### **4.2.1 Participants**

The more the results can be generalized, the higher the degree of external validity (Berg et al., 2004). This study is done on a very homogenous group and it could therefore be argued that the results can only be generalised to a similar population. All the handball players are a part of the larger cohort, so doing reliability testing directly on them will produce trustworthy reliability results for the cohort. Using a first division team in football also makes the results generalizable to elite female football.

The players had no restrictions on what to do with regards to training in the days or hours prior to the testing in our study. Excessive training during this time could be a confounding factor. As we have described previously this may have affected GJL and genu recurvatum, due to DOMS.

### **4.2.2 Testers**

The tests included in this study were not so dependent on the player's ability to perform, but rather the testers' abilities. The testers' clinical experience seems to consistently affect the results of the reliability studies as have been discussed previously. It is recommended that intrarater reliability is established for each tester before comparing the testers to each other (Portney & Watkins, 2009). Some of the studies referring to high reliability, have described more thorough ways of training their testers than what we carried out (Souza & Powers, 2009; Shultz et al., 2006). This may for sure have affected the low ICC-values of our results. However, one of the testers on each of the test stations had experience from previous test rounds of the cohort.

A weakness of our study is how the pilot study was carried out. Time and days were set for the testing, and all testers were encouraged to bring a family member or a friend for the pilot testing. Only one person turned up, but the testers were still practising the testing procedures. Although most testers had experience from prior testing sessions, the pilot could preferably be carried out in a more similar way to the actual test days. This would possibly increase skills and confidence of the new inexperienced testers and refresh the expertise of the testers who had participated earlier.



Despite standardised procedures and protocols, the testers' variability of verbal instructions may be a source of error. Their voice, mood, motivation and interaction with the players can affect the results. For example with the BHJMI, Boyle et al. (2003) noted differences in the ability to press the thumb towards the forearm depending on where on the thumb the pressure was put. In our study we did not give instructions into such detail.

Some of the days the number of players to be tested would vary considerably, from between 6-18 players. The increased workload could lead to a more stressful testing setting and potentially less inaccurate instructions and thus less inaccurate measurements.

Expectation errors or testing bias could have operated in our study. For example when the tester had tested hip anteversion on one side, a similar result would be expected on the opposite side. The testers report that this was the case for all of the tests included in this study. This could possibly be avoided in a future study by testing one side of the players first and then the other. However, this would be more complicated and time consuming to organize.

#### **4.2.3 Data collection**

In this study systematic errors can be observed where the means of the tests are higher on one of the days of testing. This is also easily visible on the BA plots, where the line of the mean is either above or below the zero line. Unfortunately, we cannot separate the measurements to each tester and this does not allow us to detect whether one of the testers is a source of systematic error. It was not consistently one tester testing all the players on the first day and the second tester testing all the players on the second day. Tester one did some of the first tests and so did tester two, and next time they changed the order. However, this random order is more likely to reduce tester errors like fatigue and learning effect. Shultz et al. (2006) were able to reveal that one of their six testers measured significantly different to the others. When the measurements done by this tester were removed from the analysis, the ICC values improved. As mentioned previously, this advocates that testers differ in their techniques and that training is important. Ideally, the interrater reliability testing in our study should have been done with at least three testers in order to be able to exclude a potentially inaccurate tester

making systematic errors. However, this would have been more time consuming and costly.

The variation of intervals of days between the two tests might have affected the results. Anatomical measurements are not likely to change within any of the test intervals in our study, but the testers and the players would more easily remember the results when the days of testing are closer in time. Also, one problem occurred for the four players with only two days between test one and two. The majority of pen marks from day one were still visible when they arrived for testing on day two, which could have affected some of the measurements. The fact that the players were not all tested in the same order could potentially have affected the results too.

On a majority of the tests the two testers were present and assisting each other both on test day one and on test day two. This could possibly be a source of error. A double-blind setup where the players would not know their measurements, and tester two would not know what measurements tester one registered, would be ideal (Portney & Watkins, 2009). Still, the testers of this study reported that they tested such a large number of players and each tester did several tests, so the chances of measurement recall were minor.

The experimental mortality might have been a confounding factor, but this is beyond the direct control of the researcher (Thomas et al., 2005). Dropouts in our study are believed to have happened for several reasons like boredom, sickness, inconvenience, injuries and holidays. This could have led to that the remaining players may be unique from the standpoint of health, interest, motivation or other factors (Berg et al., 2004). They may not represent a realistic sample of elite female handball and football players, although we believe that they do.

#### **4.2.4 Study design**

In this methodological study, as described previously, there are several sources of variability when doing repeated measurements (Carter et al., 2011). Differences may be attributed to the instrument, intrarater, interrater and intrasubject components. The degree of standardization is determined by the purpose of the study. The purpose of the larger cohort is to identify risk factors for ACL ruptures. The design of this study is a

test retest of the set protocol of the cohort. The protocol could be said to be partly standardised as there are sources of variability that we are not controlling. It could be debated that aiming for a highly standardized approach, thus trying to control as many of the possible sources of variability, would determine the upper limits of reliability of the components (Carter et al., 2011). Each of the tests has several sources of error, and a stricter standardization could have limited these errors. For example, standardizing the positioning of tester and which hand to pull with when performing the knee laxity test with the KT1000. A positive aspect of a study design with a partially standardized approach is that even though it does not describe reliability as it is in a typical clinical setting, as a non-standardized approach does, it can still be achievable in a clinical setting (Carter et al., 2011). Our study design can therefore be said to have good generalizability.

### **4.3 Statistics**

The *relative reliability* and the low ICC values have already been discussed for each of the tests.

*Absolute reliability* was reported as RMSE in this study. A couple of the studies referred to have reported their results in SEM (Piva et al., 2006; Shultz et al., 2006; Ballantyne et al., 1995). The formulas they have used to calculate SEM are based on the ICC value. As our ICC values were lower than what we had hoped for, we used RMSE.

AP knee laxity measured with the KT1000 showed a RMSE of 1.2 mm on the right leg and 2.1 mm on the left leg. Recall that a side-to-side difference is regarded as being the best way to report an ACL rupture as absolute values are difficult to use due to their wide range (Sernert, Kartus, Jr., Ejerhed, & Karlsson, 2004; Daniel et al., 1985). A difference larger than 3 mm is the way of distinguishing between the injured and non-injured knee indicating ACL rupture (Daniel et al., 1985). A measurement error of 2.1 mm must then be said to be large. A side-to-side difference of knee laxity has also been found both in normal subjects, in patients with ACL rupture and in patients with reconstructed ACL (Sernert et al., 2004; Forster et al., 1989). Not all studies reported which side had the highest laxity measurements, and other studies reported varying results, some with increased laxity of the left knee and some of the right knee (Daniel et al., 1985; Forster et al., 1989; Sernert et al., 2004). However, when looking at the raw

data of the measurements on the left knee, the difference of measurements between the two days was larger than 3 mm in 9/42 (21%) players. Six of the players had a side-to-side difference of more than 3 mm on test day one. On test day two the side-to-side difference was more consistent, and only two players had a difference larger than 3 mm. Some of these players could possibly have had ACL reconstruction, but interestingly, the players with side-to-side differences of more than 3 mm were not the same on test day one and test day two. Clinical tests, like the Lachman test and anterior drawer test, have still shown superior accuracy to the KT1000 when diagnosing ACL rupture (Graham, Johnson, Dent, & Fairclough, 1991).

The measurements done with goniometer, genu recurvatum and hip anteversion, had an RMSE of approximately 4°. When the actual ROM of hip anteversion is small in the first place, an error of 4° will make a large difference. Even Souza et al. (2006), whom had an ICC value of 0.83, question the clinical utility of the clinical test for assessing femoral anteversion. The goniometer as a measurement tool is a simple and easy instrument to use, but interrater reliability to ensure ROM has been poor (Chapleau, Canet, Petit, Laflamme, & Rouleau, 2011). It is seen as an inaccurate tool for measuring ROM in other joints too, even when the ROM is larger like in the elbow and knee (Brosseau et al., 1997; Chapleau et al., 2011). Chapleau et al. (2011) stated that it is a useful tool in a clinical setting, but it should not be used in a research protocol. Relatively subtle differences in measurement technique could result in erroneous interpretations. When the measurement tool is not reliable or valid, classifications of players with excessive anteversion or retroversion would only be possible in extreme cases.

Measurements with the tape measure were rounded to the nearest 0.5 cm and the RMSE was 2.1 cm at the most. From a clinical aspect, an error of 2.1 cm does not seem large, particularly when the mean length of tibia in this study is approximately 40 cm. Difficulty with palpation of bony landmarks with variation of individuals soft tissue, placing the proximal and the distal ends of tape measure one cm differently from one test to the next seems reasonable. When looking at the raw data of the tibial length on the right leg, it can be counted that 14/42 (33%) players had a measurement difference larger than 2.0 cm, while only 7/42 (17%) had a difference larger than 2.5 cm. On the

left leg 7/42 (17%) had a difference larger than 2 cm and only 3/42 (7%) had a difference larger than 2.5 cm. The biggest difference measured was 4 cm.

#### **4.4 Clinical relevance**

One of the major challenges as a researcher is knowing what differences are and are not important (Arnold et al., 2005). Statistics are only tools to make logical choices and the researcher or clinician have to make judgements whether the statistical findings are of importance. To make clinical judgements as to whether the magnitude of measurement error is acceptable for this study is difficult. Selecting a cut-off value should be done prior to testing in order to see how many measurements deviate from this value. As demonstrated in the previous chapter, the error changes drastically when the cut-off value is only changed by 0.5 cm.

Ideally, to attain the most valid measurements the gold standards for each measurement should be chosen. There are also practical, ethical and financial considerations when choosing tests. Should the players be exposed to radiation, would the drop-outs be larger if they had to turn up in a radiology department? How would this affect the budget? The tests chosen in this cohort for the measurements of the anatomic characteristics are also the most commonly used.

By choosing the KT1000 arthrometer instead of other instruments for measuring AP knee laxity, it allows for comparison to other studies. Also, Pugh et al. (2009) concluded in their review on instruments for testing AP knee laxity that the KT1000 and the Rolimeter provide the best results. The latter is a lightweight arthrometer that is compact and may be easier to integrate into clinical use (Pugh et al., 2009). However, the GNRB® is a new arthrometer developed to alleviate the difficulties of using the KT1000, and was tested for reliability and reproducibility and validated against the KT1000 (Robert, Nouveau, Gageot, & Gagniere, 2009). The reproducibility was significantly better, irrespectively of testers' experience. It maintains good control of the limb position in rotation and hamstring relaxation, so accuracy improved. This study had a great conflict of interest with the authors being co-investors of the GNRB®. Collette et al. (2012) did a similar study and supported Robert et al.'s findings, showing superior intra and interrater reproducibility of the GNRB®.

Different imaging techniques have been used to measure hip anteversion, such as radiographs, computed tomography (CT) and MRI (Souza & Powers, 2009). The latter shown to be highly reliable ( $r = 0.97$ ) (Tomczak et al., 1997). A high correlation is also found between measuring hip anteversion both with CT and MRI (Botser et al., 2012; Tomczak et al., 1997). When comparing clinical measurements with MRI results only moderate agreement was found (ICC 0.67–0.69) and the clinical utility of the clinical test should be questioned (Souza & Powers, 2009).

#### **4.4.1 Perspectives**

If the anatomical tests, despite low reliability, identify some essential anatomical risk factors for ACL rupture in elite female handball and football players, how would this affect the aim to prevent ACL ruptures? When these risk-factors are non-modifiable, what is the clinical importance of identifying these? Should there be a screening at young age? Should some players be recommended against playing? Should the preventive programmes be more directly aimed at the players with these risk factors? These are only thoughts that need greater consideration, but that remains for the cohort or other studies to discuss.

## 5. Conclusion

Interrater reliability was high for measuring GJL and height. Moderate values were found for measures of AP knee laxity on the right side, ASIS width and bilateral tibial and femoral lengths. More than half of the tests had low interrater reliability. They were measures of AP knee laxity of the left side, bilateral genu recurvatum, hip anteversion left side, bilateral tibial and femoral widths and femoral-tibial ratio right side. Little if any interrater reliability was found for hip anteversion right side and femoral-tibial ratio left side.

The utility of the tests for identifying risk factors need to be queried, and the findings of the tests with low reliability need to be interpreted with some caution.

A greater variance of the sample, continuous variables, double-blinding, intrarater testing prior to the study, better piloting and better training of the testers are alterations that might have improved the results of this study.

## Reference List

- Alentorn-Geli, E., Myer, G. D., Silvers, H. J., Samitier, G., Romero, D., Lazaro-Haro, C. et al. (2009a). Prevention of non-contact anterior cruciate ligament injuries in soccer players. Part 1: Mechanisms of injury and underlying risk factors. *Knee.Surg.Sports Traumatol.Arthrosc.*, *17*, 705-729.
- Alentorn-Geli, E., Myer, G. D., Silvers, H. J., Samitier, G., Romero, D., Lazaro-Haro, C. et al. (2009b). Prevention of non-contact anterior cruciate ligament injuries in soccer players. Part 2: a review of prevention programs aimed to modify risk factors and to reduce injury rates. *Knee.Surg.Sports Traumatol.Arthrosc.*, *17*, 859-879.
- Amis, A. A. (2012). The functions of the fibre bundles of the anterior cruciate ligament in anterior drawer, rotational laxity and the pivot shift. *Knee.Surg.Sports Traumatol.Arthrosc.*, *20*, 613-620.
- Arendt, E. & Dick, R. (1995). Knee injury patterns among men and women in collegiate basketball and soccer. NCAA data and review of literature. *Am.J.Sports Med.*, *23*, 694-701.
- Arneja, S. & Leith, J. (2009). Review article: Validity of the KT-1000 knee ligament arthrometer. *J.Orthop.Surg.(Hong.Kong.)*, *17*, 77-79.
- Arnold, B. L., Gansneder, B. M., & Perrin, D. H. (2005). *Research methods in athletic training*. Philadelphia, Pa.: F.A. Davis.
- Bahr, R. & Krosshaug, T. (2005). Understanding injury mechanisms: a key component of preventing injuries in sport. *Br.J.Sports Med.*, *39*, 324-329.
- Bahr, R. & Engebretsen, L. (2009). *Sports injury prevention*. Chichester, UK: Wiley-Blackwell.
- Ballantyne, B. T., French, A. K., Heimsoth, S. L., Kachingwe, A. F., Lee, J. B., & Soderberg, G. L. (1995). Influence of examiner experience and gender on interrater reliability of KT-1000 arthrometer measurements. *Phys.Ther.*, *75*, 898-906.
- Beighton, P., Solomon, L., & Soskolne, C. L. (1973). Articular mobility in an African population. *Ann.Rheum.Dis.*, *32*, 413-418.
- Berg, K. E., Latin, R. W., & Berg, K. E. (2004). *Essentials of research methods in health, physical education, exercise science, and recreation*. (vols. 2nd ed.) Philadelphia: Lippincott Williams & Wilkins.
- Boden, B. P., Dean, G. S., Feagin, J. A., Jr., & Garrett, W. E., Jr. (2000). Mechanisms of anterior cruciate ligament injury. *Orthopedics*, *23*, 573-578.
- Botser, I. B., Ozoude, G. C., Martin, D. E., Siddiqi, A. J., Kuppaswami, S., & Domb, B. G. (2012). Femoral anteversion in the hip: comparison of measurement by



computed tomography, magnetic resonance imaging, and physical examination. *Arthroscopy*, 28, 619-627.

- Boyle, K. L., Witt, P., & Riegger-Krugh, C. (2003). Intrarater and Interrater Reliability of the Beighton and Horan Joint Mobility Index. *J.Athl.Train.*, 38, 281-285.
- Brosseau, L., Tousignant, M., Budd, J., Chartier, N., Duciaume, L., Plamondon, S. et al. (1997). Intratester and intertester reliability and criterion validity of the parallelogram and universal goniometers for active knee flexion in healthy subjects. *Physiother.Res.Int.*, 2, 150-166.
- Burkhart, T. A., Arthurs, K. L., & Andrews, D. M. (2008). Reliability of upper and lower extremity anthropometric measurements and the effect on tissue mass predictions. *J.Biomech.*, 41, 1604-1610.
- Caraffa, A., Cerulli, G., Progetti, M., Aisa, G., & Rizzo, A. (1996). Prevention of anterior cruciate ligament injuries in soccer. A prospective controlled study of proprioceptive training. *Knee.Surg.Sports Traumatol.Arthrosc.*, 4, 19-21.
- Carter, C. & Wilkinson, J. (1964). Persistent joint laxity and congenital dislocation of the hip. *J.Bone Joint Surg.Br.*, 46, 40-45.
- Carter, R. E., Lubinsky, J., & Domholdt, E. (2011). *Rehabilitation research*. (vols. 4th ed.) St. Louis, Miss.: Elsevier Saunders.
- Chandrashekar, N., Slaughterbeck, J., & Hashemi, J. (2005). Sex-based differences in the anthropometric characteristics of the anterior cruciate ligament and its relation to intercondylar notch geometry: a cadaveric study. *Am.J.Sports Med.*, 33, 1492-1498.
- Chapleau, J., Canet, F., Petit, Y., Laflamme, G. Y., & Rouleau, D. M. (2011). Validity of goniometric elbow measurements: comparative study with a radiographic method. *Clin.Orthop.Relat Res.*, 469, 3134-3140.
- Cibulka, M. T. (2004). Determination and significance of femoral neck anteversion. *Phys.Ther.*, 84, 550-558.
- Collette, M., Courville, J., Forton, M., & Gagniere, B. (2012). Objective evaluation of anterior knee laxity; comparison of the KT-1000 and GNRB(R) arthrometers. *Knee.Surg.Sports Traumatol.Arthrosc.*, 20, 2233-2238.
- Colombet, P., Dejour, D., Panisset, J. C., & Siebold, R. (2010). Current concept of partial anterior cruciate ligament ruptures. *Orthop.Traumatol.Surg.Res.*, 96, S109-S118.
- Daniel, D. M., Malcom, L. L., Losse, G., Stone, M. L., Sachs, R., & Burks, R. (1985). Instrumented measurement of anterior laxity of the knee. *J.Bone Joint Surg.Am.*, 67, 720-726.
- Daniel, D. M., Stone, M. L., Sachs, R., & Malcom, L. (1985). Instrumented measurement of anterior knee laxity in patients with acute anterior cruciate ligament disruption. *Am.J.Sports Med.*, 13, 401-407.

- Elliot, D. L., Goldberg, L., & Kuehl, K. S. (2010). Young women's anterior cruciate ligament injuries: an expanded model and prevention paradigm. *Sports Med.*, *40*, 367-376.
- Faude, O., Junge, A., Kindermann, W., & Dvorak, J. (2006). Risk factors for injuries in elite female soccer players. *Br.J.Sports Med.*, *40*, 785-790.
- Fauno, P. & Wulff, J. B. (2006). Mechanism of anterior cruciate ligament injuries in soccer. *Int.J.Sports Med.*, *27*, 75-79.
- Ferretti, A., Papandrea, P., Conteduca, F., & Mariani, P. P. (1992). Knee ligament injuries in volleyball players. *Am.J.Sports Med.*, *20*, 203-207.
- Forster, I. W., Warren-Smith, C. D., & Tew, M. (1989). Is the KT1000 knee ligament arthrometer reliable? *J.Bone Joint Surg.Br.*, *71*, 843-847.
- Gilchrist, J., Mandelbaum, B. R., Melancon, H., Ryan, G. W., Silvers, H. J., Griffin, L. Y. et al. (2008). A randomized controlled trial to prevent noncontact anterior cruciate ligament injury in female collegiate soccer players. *Am.J.Sports Med.*, *36*, 1476-1483.
- Giza, E., Mithofer, K., Farrell, L., Zarins, B., & Gill, T. (2005). Injuries in women's professional soccer. *Br.J.Sports Med.*, *39*, 212-216.
- Graham, G. P., Johnson, S., Dent, C. M., & Fairclough, J. A. (1991). Comparison of clinical tests and the KT1000 in the diagnosis of anterior cruciate ligament rupture. *Br.J.Sports Med.*, *25*, 96-97.
- Griffin, L. Y., Albohm, M. J., Arendt, E. A., Bahr, R., Beynon, B. D., Demaio, M. et al. (2006). Understanding and preventing noncontact anterior cruciate ligament injuries: a review of the Hunt Valley II meeting, January 2005. *Am.J.Sports Med.*, *34*, 1512-1532.
- Guskiewicz, K. M., Perrin, D. H., Martin, D. E., Kahler, D. M., Gansneder, B. M., & McCue, F. C. (1995). Effect of ACL Reconstruction and Tibial Rotation on Anterior Knee Laxity. *J.Athl.Train.*, *30*, 243-246.
- Hanten, W. P. & Pace, M. B. (1987). Reliability of measuring anterior laxity of the knee joint using a knee ligament arthrometer. *Phys.Ther.*, *67*, 357-359.
- Hewett, T. E., Myer, G. D., Ford, K. R., Heidt, R. S., Jr., Colosimo, A. J., McLean, S. G. et al. (2005). 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.*, *33*, 492-501.
- Hewett, T. E., Zazulak, B. T., & Myer, G. D. (2007). Effects of the menstrual cycle on anterior cruciate ligament injury risk: a systematic review. *Am.J.Sports Med.*, *35*, 659-668.
- Jarvela, T. & Jarvela, S. (2013). Double-bundle versus single-bundle anterior cruciate ligament reconstruction. *Clin.Sports Med.*, *32*, 81-91.

- Juul-Kristensen, B., Rogind, H., Jensen, D. V., & Remvig, L. (2007). Inter-examiner reproducibility of tests and criteria for generalized joint hypermobility and benign joint hypermobility syndrome. *Rheumatology.(Oxford)*, *46*, 1835-1841.
- Kim, S. J., Kumar, P., & Kim, S. H. (2010). Anterior cruciate ligament reconstruction in patients with generalized joint laxity. *Clin.Orthop.Surg.*, *2*, 130-139.
- Kimberlin, C. L. & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *Am.J.Health Syst.Pharm.*, *65*, 2276-2284.
- Koga, H., Nakamae, A., Shima, Y., Iwasa, J., Myklebust, G., Engebretsen, L. et al. (2010). Mechanisms for noncontact anterior cruciate ligament injuries: knee joint kinematics in 10 injury situations from female team handball and basketball. *Am.J.Sports Med.*, *38*, 2218-2225.
- Konig, D. P., Rutt, J., Kumm, D., & Breidenbach, E. (1998). [Diagnosis of anterior knee instability. Comparison between the Lachman test, the KT-1,000 arthrometer and the ultrasound Lachman test]. *Unfallchirurg*, *101*, 209-213.
- Krosshaug, T., Slauterbeck, J. R., Engebretsen, L., & Bahr, R. (2007). Biomechanical analysis of anterior cruciate ligament injury mechanisms: three-dimensional motion reconstruction from video sequences. *Scand.J.Med.Sci.Sports*, *17*, 508-519.
- Lexell, J. E. & Downham, D. Y. (2005). How to assess the reliability of measurements in rehabilitation. *Am.J.Phys.Med.Rehabil.*, *84*, 719-723.
- Lohman, T. G., Roche, A. F., & Martorell, R. (1988). *Anthropometric standardization reference manual*. Champaign, Ill.: Human Kinetics Books.
- Lohmander, L. S., Englund, P. M., Dahl, L. L., & Roos, E. M. (2007). The long-term consequence of anterior cruciate ligament and meniscus injuries: osteoarthritis. *Am.J.Sports Med.*, *35*, 1756-1769.
- Loudon, J. K., Goist, H. L., & Loudon, K. L. (1998). Genu recurvatum syndrome. *J.Orthop.Sports Phys.Ther.*, *27*, 361-367.
- Malcom, L. L., Daniel, D. M., Stone, M. L., & Sachs, R. (1985). The measurement of anterior knee laxity after ACL reconstructive surgery. *Clin.Orthop.Relat Res.*, 35-41.
- Mandelbaum, B. R., Silvers, H. J., Watanabe, D. S., Knarr, J. F., Thomas, S. D., Griffin, L. Y. et al. (2005). 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.*, *33*, 1003-1010.
- Mayr, H. O., Hoell, A., Bernstein, A., Hube, R., Zeiler, C., Kalteis, T. et al. (2011). Validation of a measurement device for instrumented quantification of anterior translation and rotational assessment of the knee. *Arthroscopy*, *27*, 1096-1104.
- Medina McKeon, J. M. & Hertel, J. (2009). Sex differences and representative values for 6 lower extremity alignment measures. *J.Athl.Train.*, *44*, 249-255.

- Meunier, A., Odensten, M., & Good, L. (2007). Long-term results after primary repair or non-surgical treatment of anterior cruciate ligament rupture: a randomized study with a 15-year follow-up. *Scand.J.Med.Sci.Sports*, *17*, 230-237.
- Moriguchi, C. S., Carnaz, L., Silva, L. C., Salazar, L. E., Carregaro, R. L., Sato, T. O. et al. (2009). Reliability of intra- and inter-rater palpation discrepancy and estimation of its effects on joint angle measurements. *Man.Ther.*, *14*, 299-305.
- Myer, G. D., Brent, J. L., Ford, K. R., & Hewett, T. E. (2011). Real-time assessment and neuromuscular training feedback techniques to prevent ACL injury in female athletes. *Strength.Cond.J.*, *33*, 21-35.
- Myer, G. D., Ford, K. R., Barber Foss, K. D., Liu, C., Nick, T. G., & Hewett, T. E. (2009). The relationship of hamstrings and quadriceps strength to anterior cruciate ligament injury in female athletes. *Clin.J.Sport Med.*, *19*, 3-8.
- Myer, G. D., Ford, K. R., Paterno, M. V., Nick, T. G., & Hewett, T. E. (2008). The effects of generalized joint laxity on risk of anterior cruciate ligament injury in young female athletes. *Am.J.Sports Med.*, *36*, 1073-1080.
- Myklebust, G. & Bahr, R. (2005). Return to play guidelines after anterior cruciate ligament surgery. *Br.J.Sports Med.*, *39*, 127-131.
- Myklebust, G., Engebretsen, L., Braekken, I. H., Skjølberg, A., Olsen, O. E., & Bahr, R. (2003). Prevention of anterior cruciate ligament injuries in female team handball players: a prospective intervention study over three seasons. *Clin.J.Sport Med.*, *13*, 71-78.
- Myklebust, G., Engebretsen, L., Braekken, I. H., Skjølberg, A., Olsen, O. E., & Bahr, R. (2007). Prevention of noncontact anterior cruciate ligament injuries in elite and adolescent female team handball athletes. *Instr.Course Lect.*, *56*, 407-418.
- Myklebust, G., Maehlum, S., Engebretsen, L., Strand, T., & Solheim, E. (1997). Registration of cruciate ligament injuries in Norwegian top level team handball. A prospective study covering two seasons. *Scand.J.Med.Sci.Sports*, *7*, 289-292.
- Myklebust, G., Maehlum, S., Holm, I., & Bahr, R. (1998). A prospective cohort study of anterior cruciate ligament injuries in elite Norwegian team handball. *Scand.J.Med.Sci.Sports*, *8*, 149-153.
- Myklebust, G., Skjølberg, A., & Bahr, R. (2013). ACL injury incidence in female handball 10 years after the Norwegian ACL prevention study: important lessons learned. *Br.J.Sports Med.*
- Olsen, O. E., Myklebust, G., Engebretsen, L., & Bahr, R. (2004). Injury mechanisms for anterior cruciate ligament injuries in team handball: a systematic video analysis. *Am.J.Sports Med.*, *32*, 1002-1012.
- Orchard, J., Seward, H., McGivern, J., & Hood, S. (1999). Rainfall, evaporation and the risk of non-contact anterior cruciate ligament injury in the Australian Football League. *Med.J.Aust.*, *170*, 304-306.

- Orchard, J. W., Chivers, I., Aldous, D., Bennell, K., & Seward, H. (2005). Rye grass is associated with fewer non-contact anterior cruciate ligament injuries than bermuda grass. *Br.J.Sports Med.*, *39*, 704-709.
- Padua, D. A. (2010). Executing a collaborative prospective risk-factor study: findings, successes, and challenges. *J.Athl.Train.*, *45*, 519-521.
- Pallant, J. (2010). *SPSS survival manual : a step by step guide to data analysis using SPSS*. (vols. 4th ed.) Crows Nest: Allen & Unwin.
- Park, S. K., Stefanyshyn, D. J., Ramage, B., Hart, D. A., & Ronsky, J. L. (2009). Relationship between knee joint laxity and knee joint mechanics during the menstrual cycle. *Br.J.Sports Med.*, *43*, 174-179.
- Petersen, W., Braun, C., Bock, W., Schmidt, K., Weimann, A., Drescher, W. et al. (2005). A controlled prospective case control study of a prevention training program in female team handball players: the German experience. *Arch.Orthop.Trauma Surg.*, *125*, 614-621.
- Piva, S. R., Fitzgerald, K., Irrgang, J. J., Jones, S., Hando, B. R., Browder, D. A. et al. (2006). Reliability of measures of impairments associated with patellofemoral pain syndrome. *BMC.Musculoskelet.Disord.*, *7*, 33.
- Portney, L. G. & Watkins, M. P. (2009). *Foundations of clinical research*. (vols. 3rd ed.) Upper Saddle River: Pearson Prentice Hall.
- Prodromos, C. C., Han, Y., Rogowski, J., Joyce, B., & Shi, K. (2007). A meta-analysis of the incidence of anterior cruciate ligament tears as a function of gender, sport, and a knee injury-reduction regimen. *Arthroscopy*, *23*, 1320-1325.
- Pugh, L., Mascarenhas, R., Arneja, S., Chin, P. Y., & Leith, J. M. (2009). Current concepts in instrumented knee-laxity testing. *Am.J.Sports Med.*, *37*, 199-210.
- Ramesh, R., Von, A. O., Azzopardi, T., & Schranz, P. J. (2005). The risk of anterior cruciate ligament rupture with generalised joint laxity. *J.Bone Joint Surg.Br.*, *87*, 800-803.
- Remvig, L., Jensen, D. V., & Ward, R. C. (2007). Epidemiology of general joint hypermobility and basis for the proposed criteria for benign joint hypermobility syndrome: review of the literature. *J.Rheumatol.*, *34*, 804-809.
- Renstrom, P., Ljungqvist, A., Arendt, E., Beynon, B., Fukubayashi, T., Garrett, W. et al. (2008). Non-contact ACL injuries in female athletes: an International Olympic Committee current concepts statement. *Br.J.Sports Med.*, *42*, 394-412.
- Robert, H., Nouveau, S., Gageot, S., & Gagniere, B. (2009). A new knee arthrometer, the GNRB: experience in ACL complete and partial tears. *Orthop.Traumatol.Surg.Res.*, *95*, 171-176.
- Roche, A. F., Heymsfield, S., & Lohman, T. G. (1996). *Human body composition*. Champaign, Ill.: Human Kinetics.

- Ruwe, P. A., Gage, J. R., Ozonoff, M. B., & DeLuca, P. A. (1992). Clinical determination of femoral anteversion. A comparison with established techniques. *J.Bone Joint Surg.Am.*, *74*, 820-830.
- Schuster, A. J., McNicholas, M. J., Wachtl, S. W., McGurty, D. W., & Jakob, R. P. (2004). A new mechanical testing device for measuring anteroposterior knee laxity. *Am.J.Sports Med.*, *32*, 1731-1735.
- Scranton, P. E., Jr., Whitesel, J. P., Powell, J. W., Dormer, S. G., Heidt, R. S., Jr., Losse, G. et al. (1997). A review of selected noncontact anterior cruciate ligament injuries in the National Football League. *Foot Ankle Int.*, *18*, 772-776.
- Sernert, N., Helmers, J., Kartus, C., Ejerhed, L., & Kartus, J. (2007). Knee-laxity measurements examined by a left-hand- and a right-hand-dominant physiotherapist, in patients with anterior cruciate ligament injuries and healthy controls. *Knee.Surg.Sports Traumatol.Arthrosc.*, *15*, 1181-1186.
- Sernert, N., Kartus, J., Kohler, K., Ejerhed, L., & Karlsson, J. (2001). Evaluation of the reproducibility of the KT-1000 arthrometer. *Scand.J.Med.Sci.Sports*, *11*, 120-125.
- Sernert, N., Kartus, J. T., Jr., Ejerhed, L., & Karlsson, J. (2004). Right and left knee laxity measurements: a prospective study of patients with anterior cruciate ligament injuries and normal control subjects. *Arthroscopy*, *20*, 564-571.
- Shea, K. G., Grimm, N. L., Ewing, C. K., & Aoki, S. K. (2011). Youth sports anterior cruciate ligament and knee injury epidemiology: who is getting injured? In what sports? When? *Clin.Sports Med.*, *30*, 691-706.
- Shultz, S. J., Nguyen, A. D., & Schmitz, R. J. (2008). Differences in lower extremity anatomical and postural characteristics in males and females between maturation groups. *J.Orthop.Sports Phys.Ther.*, *38*, 137-149.
- Shultz, S. J., Nguyen, A. D., Windley, T. C., Kulas, A. S., Botic, T. L., & Beynnon, B. D. (2006). Intratester and intertester reliability of clinical measures of lower extremity anatomic characteristics: implications for multicenter studies. *Clin.J.Sport Med.*, *16*, 155-161.
- Shultz, S. J., Schmitz, R. J., Nguyen, A. D., Chaudhari, A. M., Padua, D. A., McLean, S. G. et al. (2010). ACL Research Retreat V: an update on ACL injury risk and prevention, March 25-27, 2010, Greensboro, NC. *J.Athl.Train.*, *45*, 499-508.
- Siegel, L., Vandenakker-Albanese, C., & Siegel, D. (2012). Anterior cruciate ligament injuries: anatomy, physiology, biomechanics, and management. *Clin.J.Sport Med.*, *22*, 349-355.
- Smith, H. C., Vacek, P., Johnson, R. J., Slauterbeck, J. R., Hashemi, J., Shultz, S. et al. (2012a). Risk factors for anterior cruciate ligament injury: a review of the literature - part 1: neuromuscular and anatomic risk. *Sports Health*, *4*, 69-78.
- Smith, H. C., Vacek, P., Johnson, R. J., Slauterbeck, J. R., Hashemi, J., Shultz, S. et al. (2012b). Risk factors for anterior cruciate ligament injury: a review of the

literature-part 2: hormonal, genetic, cognitive function, previous injury, and extrinsic risk factors. *Sports Health*, 4, 155-161.

- Souza, R. B. & Powers, C. M. (2009). Concurrent criterion-related validity and reliability of a clinical test to measure femoral anteversion. *J.Orthop.Sports Phys.Ther.*, 39, 586-592.
- Stevens, J., Katz, E. G., & Huxley, R. R. (2010). Associations between gender, age and waist circumference. *Eur.J.Clin.Nutr.*, 64, 6-15.
- Streiner, D. L. & Norman, G. R. (2006). "Precision" and "accuracy": two terms that are neither. *J.Clin.Epidemiol.*, 59, 327-330.
- Sutlive, T. G., Mitchell, S. D., Maxfield, S. N., McLean, C. L., Neumann, J. C., Swiecki, C. R. et al. (2004). Identification of individuals with patellofemoral pain whose symptoms improved after a combined program of foot orthosis use and modified activity: a preliminary investigation. *Phys.Ther.*, 84, 49-61.
- Swanik, C. B., Covassin, T., Stearne, D. J., & Schatz, P. (2007). The relationship between neurocognitive function and noncontact anterior cruciate ligament injuries. *Am.J.Sports Med.*, 35, 943-948.
- Tegnander, A., Olsen, O. E., Moholdt, T. T., Engebretsen, L., & Bahr, R. (2008). Injuries in Norwegian female elite soccer: a prospective one-season cohort study. *Knee.Surg.Sports Traumatol.Arthrosc.*, 16, 194-198.
- Thomas, J. R., Silverman, S. J., & Nelson, J. K. (2005). *Research methods in physical activity*. (vols. 5th ed.) Champaign, Ill.: Human Kinetics.
- Tomczak, R. J., Guenther, K. P., Rieber, A., Mergo, P., Ros, P. R., & Brambs, H. J. (1997). MR imaging measurement of the femoral antetorsional angle as a new technique: comparison with CT in children and adults. *AJR Am.J.Roentgenol.*, 168, 791-794.
- Torg, J. S., Conrad, W., & Kalen, V. (1976). Clinical diagnosis of anterior cruciate ligament instability in the athlete. *Am.J.Sports Med.*, 4, 84-93.
- Trimble, M. H., Bishop, M. D., Buckley, B. D., Fields, L. C., & Rozea, G. D. (2002). The relationship between clinical measurements of lower extremity posture and tibial translation. *Clin.Biomech.(Bristol, Avon.)*, 17, 286-290.
- Uhorchak, J. M., Scoville, C. R., Williams, G. N., Arciero, R. A., St, P. P., & Taylor, D. C. (2003). Risk factors associated with noncontact injury of the anterior cruciate ligament: a prospective four-year evaluation of 859 West Point cadets. *Am.J.Sports Med.*, 31, 831-842.
- Vihinen, M. (2012). How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis. *BMC.Genomics*, 13 Suppl 4, S2.
- Walden, M., Hagglund, M., Werner, J., & Ekstrand, J. (2011). 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.*, 19, 3-10.

- Wang, J., Thornton, J. C., Bari, S., Williamson, B., Gallagher, D., Heymsfield, S. B. et al. (2003). Comparisons of waist circumferences measured at 4 sites. *Am.J.Clin.Nutr.*, 77, 379-384.
- Wiertsema, S. H., van Hooff, H. J., Migchelsen, L. A., & Steultjens, M. P. (2008). Reliability of the KT1000 arthrometer and the Lachman test in patients with an ACL rupture. *Knee.*, 15, 107-110.
- Willmott, C and Matsuura and K (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance.  
[http://climate.geog.udel.edu/~climate/publication\\_html/Pdf/WM\\_CR\\_05.pdf](http://climate.geog.udel.edu/~climate/publication_html/Pdf/WM_CR_05.pdf)
- Wojtys, E. M., Huston, L. J., Boynton, M. D., Spindler, K. P., & Lindenfeld, T. N. (2002). The effect of the menstrual cycle on anterior cruciate ligament injuries in women as determined by hormone levels. *Am.J.Sports Med.*, 30, 182-188.
- Yeadon, M. R. (1990a). The simulation of aerial movement--I. The determination of orientation angles from film data. *J.Biomech.*, 23, 59-66.
- Yeadon, M. R. (1990b). The simulation of aerial movement--II. A mathematical inertia model of the human body. *J.Biomech.*, 23, 67-74.
- Zainuddin, Z., Newton, M., Sacco, P., & Nosaka, K. (2005). Effects of massage on delayed-onset muscle soreness, swelling, and recovery of muscle function. *J.Athl.Train.*, 40, 174-180.
- Zazulak, B. T., Paterno, M., Myer, G. D., Romani, W. A., & Hewett, T. E. (2006). The effects of the menstrual cycle on anterior knee laxity: a systematic review. *Sports Med.*, 36, 847-862.
- Ziegler, C. G., Pietrini, S. D., Westerhaus, B. D., Anderson, C. J., Wijdicks, C. A., Johansen, S. et al. (2011). Arthroscopically pertinent landmarks for tunnel positioning in single-bundle and double-bundle anterior cruciate ligament reconstructions. *Am.J.Sports Med.*, 39, 743-752.



## List of tables

<b>Table 1:</b> Strength of correlation coefficients (Carter et al., 2011).....	21
<b>Table 2:</b> Studies investigating interrater reliability of the anatomical measures.....	25
<b>Table 3:</b> Dates and numbers of days between testing.....	29
<b>Table 4:</b> Beighton and Horan joint mobility index.....	31
<b>Table 5:</b> The anthropometric measurements; a description of the standardized procedures .....	35
<b>Table 6:</b> Descriptive data of the players in the study.....	38
<b>Table 7:</b> Anatomical tests on day one and two. Mean values( $\pm$ SD), mean difference ( $\pm$ SD) between the two days and the significance of the difference.....	39
<b>Table 8:</b> Intraclass correlation coefficients (ICC) with 95% CIs for the anatomical tests, mean difference with 95% CI .....	40
<b>Table 9:</b> Root mean square error (RMSE) of anatomical test .....	41

## Table of figures

<b>Figure 1:</b> An illustration of the tibial and femoral ACL bundle attachments in the native knee.....	9
<b>Figure 2:</b> ACL injury mechanism.....	10
<b>Figure 3:</b> An overview of the players included and the drop-outs in this project .....	28
<b>Figure 4:</b> Measuring generalized joint laxity.....	30
<b>Figure 5:</b> Measuring genu recurvatum.....	32
<b>Figure 6:</b> Measuring knee laxity .....	33
<b>Figure 7:</b> Measuring hip anteversion.....	34
<b>Figure 8:</b> Bland-Altman plots.....	42

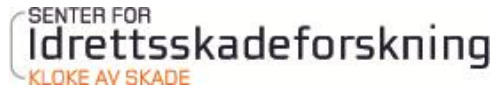
## Acronyms

<b>ACL</b>	Anterior cruciate ligament
<b>AM</b>	Anteromedial
<b>ASIS</b>	Anterior superior iliac spine
<b>AP</b>	Anterior posterior
<b>BA</b>	Bland-Altman
<b>BHJMI</b>	Beighton and Horan joint mobility index
<b>CI</b>	Confidence interval
<b>Cm</b>	Centimeter
<b>CV</b>	Coefficient of variation
<b>CT</b>	Computed tomography
<b>GJL</b>	Generalized joint laxity
<b>ICC</b>	Intracorrelation coefficient
<b>Kg</b>	Kilogram
<b>MD</b>	Mean difference
<b>Mm</b>	Millimeter
<b>MRI</b>	Magnetic resonance imaging
<b>OA</b>	Osteoarthritis
<b>OSTRC</b>	Oslo Sport and Trauma Research Centre
<b>RMSE</b>	Root mean square error
<b>ROM</b>	Range of motion
<b>SD</b>	Standard deviation
<b>SEM</b>	Standard error measurement
<b>TPAT</b>	Trochanteric prominence angle test
<b>3D</b>	Three dimensional

## **Attachments**

	<b>Page</b>
<b>Attachment 1</b> Information sheet for the players	72
<b>Attachment 2</b> Written consent form	73
<b>Attachment 3</b> Written consent form under 18	75
<b>Attachment 4</b> Test lists for 14 players	77
<b>Attachment 5</b> Test lists for 6 players	78
<b>Attachment 6</b> Approval from REK	79

# Attachment 1



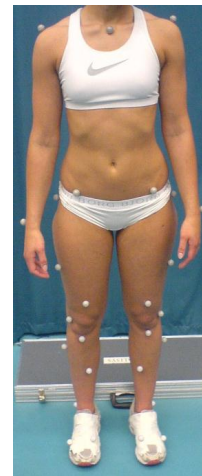
## Forskningsprosjekt blant fotballspillere i Toppserien 2012

Senter for idrettsskedeforskning ved Norges idrettshøgskole gjennomfører et forskningsprosjekt der vi undersøker risikofaktorer for korsbåndskader blant kvinnelige elitefotballspillere. Vi har derfor hver sesong siden 2009 testet alle spillere i Toppserien, og har nå totalt testet 320 spillere. Spillerne følges deretter opp de kommende sesongene i form av å registrere eventuelle korsbåndskader som oppstår.

Vi har nå satt av tid til testing av spillere fra **Stabæk onsdag 15. februar kl. 15.00**. Testingen foregår på Norges idrettshøgskole, og dere kan møte opp i resepsjonen ved hovedinngangen. Vi vil da ha et kort informasjonsmøte først, hvor vi også ber alle om å skrive under på en samtykkeerklæring for prosjektdeltakelsen.

Vi har totalt 7 teststasjoner som innebærer 3D bevegelsesanalyse av finter/vendinger og fallhopp/spenst, styrketester av forside/bakside lår og hofter, bevegelighet, balansetester, anatomiske målinger og en blodprøve. Testingen vil totalt ta ca. 6-7 timer, og dere vil selvfølgelig få en pause og mat og drikke underveis.

Under testingen har dere på treningstøy og de skoene dere vanligvis bruker til innetrening. For å gjøre testingen lettere bør dere bruke en shorts og t-skjorte. To av testene krever at hofter/hoftekam er tilgjengelig for markører (se bilde), så ta gjerne på en boksershorts, bikinitruse eller eventuelt en kort sykkelshorts til disse testene. Markørene vi bruker til bevegelsesanalysen festes med teip - **unngå derfor å bruke bodylotion** på testdagen.



For å se bilder fra testingen, kan dere finne dette på hjemmesiden til Senter for idrettsskedeforskning under følgende link;  
<http://www.klokavskade.no/no/Nyhetsarkiv/Nyhetsarkiv-2009/Hvorfor-skader-fotballjenter-fremre-korsband/>

Vi ser frem til å møte dere 15. februar.

Dersom dere har spørsmål i mellomtiden kan dere ta kontakt på telefon (99 22 44 69) eller e-post (agnethe.nilstad@nih.no).

Vennlig hilsen

Agnethe Nilstad  
Fysioterapeut MSc, PhD-kandidat  
Prosjektleder

## Attachment 2



### FORESPØRSEL OM DELTAKELSE I PROSJEKTET: *”Risikofaktorer for fremre korsbåndskader hos kvinnelige elitehåndball- og fotballspillere - En prospektiv kohortstudie”*

#### Bakgrunn for undersøkelsen

Korsbåndsskader i fotball og håndball har i det siste vært et svært aktuelt tema, både i media og i forskningssammenheng. Dette skyldes først og fremst den relativt store hyppigheten av denne alvorlige skaden, spesielt blant kvinnelige utøvere, som ser ut til å skade seg 3-7 ganger hyppigere enn menn. Problemet så langt er imidlertid at vi vet for lite om risikofaktorene og skademekanismene for korsbåndskader. Denne informasjonen er viktig når vi forsøker å forebygge skader, både for å kunne vite hvem som vil ha størst glede av forebyggende trening og for å kunne utvikle mest mulig effektive treningsmetoder.

Senter for idrettsskedeforskning er en forskningsgruppe bestående av fysioterapeuter, kirurger og biomekanikere med kunnskap innen idrettsmedisin. Vår hovedmålsetting er å forebygge skader i norsk idrett, med spesiell satsning på fotball, håndball, ski og snowboard. Denne studien er en viktig brikke i arbeidet med å finne ut hvorfor noen får en korsbåndskade. Vi ønsker nå å undersøke ulike mulige risikofaktorer for korsbåndskader, for deretter å kartlegge hvem som får korsbåndskader de påfølgende sesongene.

#### Gjennomføring av undersøkelsen

Vi ønsker at du som elitespiller deltar i denne studien, og deltakelsen er frivillig. Testingen vil finne sted på Norges idrettshøgskole. I løpet av en dag vil vi gjennomføre ulike styrke-, balanse- og bevegelighetstester, anatomiske målinger, samt gjennomføre en bevegelsesanalyse av hvordan du finter, vender, hopper og lander. Undersøkelsen starter med en kort oppvarming, deretter får du festet små refleksmarkører på kroppen (35 stk totalt). Du vil så bli bedt om å gjennomføre tre finter/vendinger og tre fallhopp. Under disse øvelsene vil det være 8 infrarøde kamera som filmer markørene, samtidig som kreftene fra underlaget blir målt. Dataene fra markører, kraftplattform og anatomiske mål benyttes i en matematisk modell som gir ut leddkrefter og momenter. Disse kreftene/momentene gir oss informasjon om hvordan muskler og passive strukturer som leddbånd belastes.

Bevegelsesanalysen vil ta ca. 1,5 time, inkludert anatomiske målinger og påsetting av markører. De andre testene gjennomføres resten av tiden laget er på NIH, og totalt vil testene ta om lag åtte timer. I tillegg til disse testene vil du få utdelt et skjema, der vi spør om trenings erfaring, tidligere skader, skade i familien, treningsmengde, menstruasjonsstatus og knefunksjon. Spørreskjemaet besvares i løpet av testdagen, og det vil ta ca. 30 min.

#### Behandling av testresultatene

Vi vil de neste tre sesongene følge opp alle lag og spillere som har deltatt på testing hos oss for å registrere alle korsbåndskader som oppstår.

Vi er også interessert i å kunne kontakte deg senere med tanke på oppfølgingsstudier. Dette kan f.eks. skje ved at du får tilsendt et spørreskjema. Av den grunn vil vi lagre resultatene fra testene og svarene på spørreskjemaet fram til 1.6.2017. Etter dette vil dataene bli anonymisert. Dataene vil bli behandlet konfidensielt, og kun i forskningssammenheng. Alle som utfører testingen og forskere som benytter dataene er underlagt taushetsplikt. Dersom du ikke ønsker å være med på etterundersøkelser, kan du reservere deg mot dette i samtykkeerklæringen. I så fall vil alle dine data bli anonymisert etter fire år.

Vi vil underveis i testingen ta videoopptak av dere som vi senere kan ønske å bruke i undervisnings- og formidlingssammenheng. Opptakene inkluderer situasjoner der dere kun har på shorts og sports-BH. Dersom dere ikke vil at deres opptak skal være aktuelle for slik bruk krysser dere av for det i samtykkeerklæringen.

#### Hva får du ut av det?

Vi kan ikke tilby noe honorar for oppmøtet, men vil dekke eventuelle reise- og matutgifter. I tillegg vil du få kopi av dine resultater fra styrketestene som gjennomføres i løpet av testdagen.

#### Angrer du?

Du kan selvfølgelig trekke deg fra forsøket når som helst uten å måtte oppgi noen grunn. Alle data som angår deg vil uansett bli anonymisert.

#### Spørsmål?

Ring gjerne til Tron Krosshaug, tlf.: 45 66 00 46 hvis du har spørsmål om prosjektet, eller send e-post til [tron.krosshaug@nih.no](mailto:tron.krosshaug@nih.no).

*”Risikofaktorer for fremre korsbåndskader hos kvinnelige elitehåndball- og fotballspillere - En prospektiv kohortstudie”*

## SAMTYKKEERKLÆRING

Jeg har mottatt skriftlig og muntlig informasjon om studien *Risikofaktorer for fremre korsbåndskader hos kvinnelige elitehåndball- og fotballspillere - En prospektiv kohortstudie*. Jeg er klar over at jeg kan trekke meg fra undersøkelsen på et hvilket som helst tidspunkt.

- Jeg ønsker ikke å bli kontaktet etter endt karriere med tanke på oppfølgingsstudier
- Jeg ønsker ikke at video av meg skal brukes i undervisningssammenheng

Sted

Dato

.....

.....

.....  
Underskrift

.....  
Navn med blokkbokstaver

.....  
Adresse

.....  
Mobiltelefon

.....  
E-postadresse

# Attachment 3



## FORESPØRSEL OM DELTAKELSE I PROSJEKTET: *”Risikofaktorer for fremre korsbåndskader hos kvinnelige elitehåndball- og fotballspillere - En prospektiv kohortstudie”*

### Bakgrunn for undersøkelsen

Korsbåndsskader i fotball og håndball har i det siste vært et svært aktuelt tema, både i media og i forsknings sammenheng. Dette skyldes først og fremst den relativt store hyppigheten av denne alvorlige skaden, spesielt blant kvinnelige utøvere, som ser ut til å skade seg 3-7 ganger hyppigere enn menn. Problemet så langt er imidlertid at vi vet for lite om risikofaktorene og skademekanismene for korsbåndskader. Denne informasjonen er viktig når vi forsøker å forebygge skader, både for å kunne vite hvem som vil ha størst glede av forebyggende trening og for å kunne utvikle mest mulig effektive treningsmetoder.

Senter for idrettsskedeforskning er en forskningsgruppe bestående av fysioterapeuter, kirurger og biomekanikere med kunnskap innen idrettsmedisin. Vår hovedmålsetting er å forebygge skader i norsk idrett, med spesiell satsning på fotball, håndball, ski og snowboard. Denne studien er en viktig brikke i arbeidet med å finne ut hvorfor noen får en korsbåndskade. Vi ønsker nå å undersøke ulike mulige risikofaktorer for korsbåndskader, for deretter å kartlegge hvem som får korsbåndskader de påfølgende sesongene.

### Gjennomføring av undersøkelsen

Vi ønsker at du som elitespiller deltar i denne studien, og deltakelsen er frivillig. Testingen vil finne sted på Norges idrettshøgskole. I løpet av en dag vil vi gjennomføre ulike styrke-, balanse- og bevegelighetstester, anatomiske målinger, samt gjennomføre en bevegelsesanalyse av hvordan du finter, vender, hopper og lander. Undersøkelsen starter med en kort oppvarming, deretter får du festet små refleksmarkører på kroppen (35 stk totalt). Du vil så bli bedt om å gjennomføre tre finter/vendinger og tre fallhopp. Under disse øvelsene vil det være 8 infrarøde kamera som filmer markørene, samtidig som kreftene fra underlaget blir målt. Dataene fra markører, kraftplattform og anatomiske mål benyttes i en matematisk modell som gir ut leddkrefter og momenter. Disse kreftene/momentene gir oss informasjon om hvordan muskler og passive strukturer som leddbånd belastes.

Bevegelsesanalysen vil ta ca. 1,5 time, inkludert anatomiske målinger og påsetting av markører. De andre testene gjennomføres resten av tiden laget er på NIH, og totalt vil testene ta om lag åtte timer. I tillegg til disse testene vil du få utdelt et skjema, der vi spør om trenings erfaring, tidligere skader, skade i familien, treningsmengde, menstruasjonsstatus og knefunksjon. Spørreskjemaet besvares i løpet av testdagen, og det vil ta ca. 30 min.

### Behandling av testresultatene

Vi vil de neste tre sesongene følge opp alle lag og spillere som har deltatt på testing hos oss for å registrere alle korsbåndskader som oppstår.

Vi er også interessert i å kunne kontakte deg senere med tanke på oppfølgingsstudier. Dette kan f.eks. skje ved at du får tilsendt et spørreskjema. Av den grunn vil vi lagre resultatene fra testene og svarene på spørreskjemaet fram til 1.6.2017. Etter dette vil dataene bli anonymisert. Dataene vil bli behandlet konfidensielt, og kun i forskningssammenheng. Alle som utfører testingen og forskere som benytter dataene er underlagt taushetsplikt. Dersom du ikke ønsker å være med på etterundersøkelser, kan du reservere deg mot dette i samtykkeerklæringen. I så fall vil alle dine data bli anonymisert etter fire år.

Vi vil underveis i testingen ta videoopptak av dere som vi senere kan ønske å bruke i undervisnings- og formidlings sammenheng. Opptakene inkluderer situasjoner der dere kun har på shorts og sports-BH. Dersom dere ikke vil at deres opptak skal være aktuelle for slik bruk krysser dere av for det i samtykkeerklæringen.

### Hva får du ut av det?

Vi kan ikke tilby noe honorar for oppmøtet, men vil dekke eventuelle reise- og matutgifter. I tillegg vil du få kopi av dine resultater fra styrketestene som gjennomføres i løpet av testdagen.

### Angrer du?

Du kan selvfølgelig trekke deg fra forsøket når som helst uten å måtte oppgi noen grunn. Alle data som angår deg vil uansett bli anonymisert.

### Spørsmål?

Ring gjerne til Tron Krosshaug, tlf.: 45 66 00 46 hvis du har spørsmål om prosjektet, eller send e-post til [tron.krosshaug@nih.no](mailto:tron.krosshaug@nih.no).

*”Risikofaktorer for fremre korsbåndskader hos kvinnelige elitehåndball- og fotballspillere - En prospektiv kohortstudie”*

## SAMTYKKEERKLÆRING

Jeg har mottatt skriftlig og muntlig informasjon om studien *Risikofaktorer for fremre korsbåndskader hos kvinnelige elitehåndball- og fotballspillere - En prospektiv kohortstudie*. Jeg er klar over at jeg kan trekke meg fra undersøkelsen på et hvilket som helst tidspunkt.

- Jeg ønsker ikke å bli kontaktet etter endt karriere med tanke på oppfølgingsstudier
- Jeg ønsker ikke at video av meg skal brukes i undervisningssammenheng

Sted

Dato

.....

.....

.....  
Underskrift spiller

.....  
Underskrift foresatt

.....  
Navn med blokkbokstaver

.....  
Adresse

.....  
Mobiltelefon

.....  
E-postadresse



# Attachment 4

Date and team: .....

Time	Marker placement	Time	3D	Time	Anthropometric	Time	Isokinetic
15:00	Player 1	15:00		15:00	Player 11	15:00	Player 9
	Player 2		Player 1		Player 12		Player 10
15:55	Player 3	15:55	Player 2	15:55	Player 13	15:55	Player 11
	Player 4		Player 3		Player 14		Player 12
16:50	Player 5	16:50	Player 4	16:50	Player 1	16:50	Player 13
17:17	BREAK		Player 5		Player 2		Player 14
17:40	Player 6	17:45	BREAK	17:45	BREAK	17:45	BREAK
18:05	Player 7	18:05	Player 6	18:05	Player 3	18:05	Player 1
	Player 8		Player 7		Player 4		Player 2
19:00	Player 9	19:00	Player 8	19:00	Player 5	19:00	Player 3
	Player 10		Player 9		Player 6		Player 4
19:55	Player 11	19:55	Player 10	19:55	Player 7	19:55	Player 5
	Player 12		Player 11		Player 8		Player 6
20:50	Player 13	20:50	Player 12	20:50	Player 9	20:50	Player 7
	Player 14		Player 13		Player 10		Player 8
		21:45	Player 14	21:45		21:45	

Subjective	Time	Anatomic	Time	Legpress/balance
Player 7	15:00	Player 5	15:00	Player 3
Player 8		Player 6		Player 4
Player 9	15:55	Player 7	15:55	Player 5
Player 10		Player 8		Player 6
Player 11	16:50	Player 9	16:50	Player 7
Player 12		Player 10		Player 8
BREAK	17:45	BREAK	17:45	BREAK
Player 13	18:05	Player 11	18:05	Player 9
Player 14		Player 12		Player 10
Player 1	19:00	Player 13	19:00	Player 11
Player 2		Player 14		Player 12
Player 3	19:55	Player 1	19:55	Player 13
Player 4		Player 2		Player 14
Player 5	20:50	Player 3	20:50	Player 1
Player 6		Player 4		Player 2
	21:45		21:45	

## Attachment 5

Date and team: .....

Time	Marker placement	Time	3D	Time	Anthropometric
09:00	Player A	09:00		09:00	Player B
09:30	Player B	09:30	Player A	09:30	Player C
10:00	Player C	10:00	Player B	10:00	Player D
10:30	Player D	10:30	Player C	10:30	Player E
11:00	Player E	11:00	Player D	11:00	Player F
11:30	Player F	11:30	Player E	11:30	Player A
12:00		12:00	Player F	12:00	

Time	Balance	Time	Anatomic	Time	Legpress
09:00		09:00	Player E	09:00	Player D
09:30	Player D	09:30	Player F	09:30	Player E
10:00	Player E	10:00	Player A	10:00	Player F
10:30	Player F	10:30	Player B	10:30	Player A
11:00	Player A	11:00	Player C	11:00	Player B
11:30	Player B	11:30	Player D	11:30	Player C
12:00	Player C	12:00		12:00	

## Attachment 6



### UNIVERSITETET I OSLO DET MEDISINSKE FAKULTET

Forsker dr.scient. Tron Krosshaug  
Norges idrettshøgskole  
Pb. 4014 Ullevål Stadion  
0806 Oslo

**Regional komité for medisinsk og helsefaglig  
forskningsetikk Sør-Øst A (REK Sør-Øst A)**

Postboks 1130 Blindern  
NO-0318 Oslo

**Dato:** 15.12.08  
**Deres ref.:**  
**Vår ref.:** S-07078a

Telefon: 22 84 46 66  
Telefaks: 22 85 05 90  
E-post: jorgen.hardang@medisin.uio.no

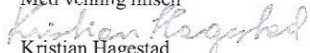
**S-07078a Risikofaktorer for fremre korsbåndskader hos kvinnelige elitehåndballspillere - en  
prospektiv kohortstudie [2.2007.511]**


Vi viser til skjema for protokolltillegg og endringer datert 3.12.08 vedlagt revidert informasjonsskriv.

Prosjektleder ønsker å utvide prosjektpopulasjonen til kvinnelige elitefotballspillere fra Toppserien i Norge (ca 240 spillere).

Komiteen godkjenner endringen slik den er beskrevet i skjema for protokolltillegg og endringer og videresender kopi av informasjonsskriv, endringsskjema samt komiteens vedtak til Helsedirektoratet for behandling av endring av biobanken.

Med vennlig hilsen

  
Kristian Hagestad  
Fylkeslege cand.med., spes. i samf.med  
Leder

  
Jørgen Hardang  
Sekretær

Kopi: Helsedirektoratet, Postboks 7000, St. Olavs plass, 0130 Oslo

