

Kandidat: Joachim Spyckerelle Raastad



*Intra-tester og inter-tester reliabilitet av senetykkelsesmåling med diagnostisk ultralyd hos pasienter med skuldertendinopati*

Manuellterapi teori: MANT395, 4. semester, 2016

Masterprogram i helsefag – Klinisk masterstudium i manuellterapi for fysioterapeuter

Institutt for global helse og samfunnsmedisin, Universitetet i Bergen

Antall ord: 10 963

## Forord

Arbeidet med denne masteroppgaven har vært krevende, men lærerikt. Det har utfordret meg til å tenke kritisk, reflekterende og metodisk. Erfaringene jeg har tilegnet meg i gjennomføringen av denne studien er erfaringer jeg vil bringe med meg videre.

Jeg vil gjerne rette en stor takk til alle som har bidratt til denne oppgaven:

- Hovedveileder Jan Magnus Bjordal
- Ingvill Fjell Naterstad, for veiledning
- Jon Joensen, for innsamling av data og gjennomføring av reliabilitetstesting
- Sturla Haslerud, for bruk av datamateriale fra doktorgradsarbeid
- Medstudent Martin Haagensen, for gode diskusjoner
- Familien min som har støttet meg gjennom dette arbeidet

Ganddal, 30.10.16

*Joachim Raastad*

## Innholdsfortegnelse

Sammendrag .....	1
Abstract .....	2
Tabellindex .....	3
Figurindex .....	3
1. Introduksjon .....	4
1.1 Presentasjon av problemområdet og bakgrunn for valg av tema .....	4
1.2 Sentral teori på området .....	5
1.2.1 Generelt om tendinopati .....	5
1.2.2 Diagnostisk ultralyd .....	6
1.2.3 Reliabilitet .....	7
1.3 Tidligere forskning på området .....	11
2. Hensikt og problemstilling .....	13
2.1 Hensikt .....	13
2.2 Problemstilling .....	13
3. Metode .....	13
3.1 Valg av forskningsdesign .....	13
3.2 Utvalg .....	14
3.2.1 Inklusjonskriterier .....	14
3.2.2 Eksklusjonskriterier .....	14
3.3 Etske hensyn .....	15
3.4 Måleinstrumenter .....	15
3.5 Datainnsamling .....	16
3.5.1 Datainnsamling til doktorgradsarbeid .....	16
3.5.2 Datainnsamling til mastergradsoppgave .....	17
3.6 Analyse .....	18
3.6.1 Blinding .....	18
3.6.2 Statistisk metode .....	18
4. Resultater .....	20
4.1 Demografiske data .....	20
4.2 Normalfordeling av data .....	20
4.3 Intra-tester reliabilitet .....	25

4.3.1 Relativ reliabilitet biceps brachii caput longum.....	25
4.3.2 Relativ reliabilitet supraspinatus .....	26
4.3.3 Absolutt reliabilitet.....	27
4.3.4 Spredningsplot og Pearsons korrelasjonskoeffisient.....	28
4.4 Inter-tester reliabilitet .....	30
4.4.1 Relativ reliabilitet biceps brachii caput longum.....	30
4.4.2 Relativ reliabilitet supraspinatus .....	31
4.4.3 Absolutt reliabilitet .....	32
4.4.4 Spredningsplot og Pearsons korrelasjonskoeffisient.....	33
5. Diskusjon.....	36
5.1 Metode.....	36
5.1.1 Valg av forskningsdesign .....	36
5.1.2 Utvalg og ekstern validitet .....	39
5.1.3 Testere .....	41
5.1.4 Variabilitet, standardisering og målemetoder.....	41
5.2 Valg av statistiske metoder.....	43
5.2.1 Normalfordeling av data .....	44
5.2.2 Relativ reliabilitet.....	44
5.2.3 Absolutt reliabilitet .....	45
5.3 Resultater .....	46
5.3.1 Intra-tester reliabilitet .....	46
5.3.2 Inter-tester reliabilitet .....	47
5.4 Kliniske implikasjoner .....	48
5.5 Videre forskning.....	49
6. Konklusjon.....	50
7. Referanser.....	51
Vedlegg 1: Godkjenning REK.....	58
Vedlegg 2: Forespørsel om deltakelse .....	59
Vedlegg 3: Mal til innsamling av UL data .....	63

## Sammendrag

**Bakgrunn og hensikt:** Tendinopati i skuldere er en av de vanligste årsakene til skuldersmerter og er en stor samfunnsøkonomisk belastning. Diagnostisk ultralyd er stadig hyppigere brukt i primærhelsetjenesten. Det er viktig å kartlegge reliabiliteten av kliniske måleinstrumenter for optimal bruk i klinisk praksis. Hensikten med denne studien var å undersøke inter-, og intra-tester reliabiliteten i måling av senetykkelsen av supraspinatus og biceps brachii på samme ultralydbildet hos pasienter med skuldertendinopati.

**Problemstilling:** Hvordan er inter-tester, og intra-tester reliabiliteten ved måling av senetykkelsen til biceps brachii caput longum og supraspinatus med ultralydsavbildning hos pasienter med skuldertendinopati?

**Metode:** En kvantitativ reliabilitetsstudie med et ikke-eksperimentelt design ble gjennomført. Tykkelsesmålinger med diagnostisk ultralyd av supraspinatus og biceps brachii caput longum ble gjennomført på deltakere med skuldertendinopati ( $N = 27$ ). Analysene av inter-tester reliabilitet ble gjennomført på grunnlag av tykkelsesmålinger foretatt av to testere. Intra-tester reliabilitet ble analysert på bakgrunn av to datasett fra en tester. Relativ reliabilitet ble beregnet ved bruk av intraklasse korrelasjonskoeffisientene  $ICC_{1,1}$  og  $ICC_{3,1}$ . Absolutt reliabilitet ble beregnet ved bruk av innen-subjekt standard deviasjon.

**Resultater:** Relativ reliabilitet for *biceps brachii* var meget god ved fem av syv analyser (range ICC: 0,81 – 0,94) og god ved to av syv (range ICC: 0,64 – 0,65). Resultatene for *supraspinatus* viste moderat reliabilitet ved to av syv analyser (range ICC: 0,52 – 0,54), fire indikerte en god reliabilitet (range ICC: 0,63 – 0,79) og én avdekket meget god reliabilitet (ICC: 0,90).

**Konklusjon:** Resultatene indikerer en meget god reliabilitet for biceps brachii og en god reliabilitet for supraspinatus i målingen av senetykkelse.

**Nøkkelord:** Tendinopati, skulder, diagnostisk ultralyd, reliabilitet, tykkelsesmåling

## Abstract

**Background and purpose:** Shoulder tendinopathy is a common reason for shoulder pain and represents a great socioeconomic burden. Ultrasonography is used increasingly in primary care. In order to apply clinical measurement instruments optimally, it is important to investigate their reliability. The purpose of this study was to investigate inter-, and intra-tester reliability of ultrasonographic measurements of tendon thickness in supraspinatus and biceps brachii.

**Research question:** What is the inter-tester, and intra-tester reliability of ultrasonographic measurements of tendon thickness in supraspinatus and biceps brachii caput longum in a population with shoulder tendinopathy?

**Methods:** A quantitative reliability study with a non-experimental design was performed. Ultrasonographic measurement of the tendon thickness in supraspinatus and biceps brachii was undertaken on participants with shoulder tendinopathy (N = 27). Analysis of inter-tester reliability was based on tendon thickness measurements from two raters. Intra-tester reliability analysis was performed on two sets of data from one rater. Intraclass correlation coefficients ICC<sub>1,1</sub> and ICC<sub>3,1</sub> was utilized for analysis of relative reliability. Within-subject standard deviation was utilized in analysis of absolute reliability.

**Results:** Relative reliability for *biceps brachii* was very good in five of seven analyses (range ICC: 0,81 – 0,94), and good in two of seven (range ICC: 0,64 – 0,65). The results for *supraspinatus* showed moderate reliability in two of seven analyses (range ICC: 0,52 – 0,54), four indicated good reliability (range ICC: 0,63 – 0,79) and one showed very good reliability (ICC: 0,90).

**Conclusion:** The results indicate a very good reliability for biceps brachii and a good reliability for supraspinatus in measuring tendon thickness with ultrasonography.

**Keywords:** Tendinopathy, shoulder, ultrasonography, reliability, tendon thickness

## Tabellindex

Tabell 1: Beskrivelse av grad av reliabilitet ved ulike ICC .....	19
Tabell 2: Oppsummering av demografiske data av inkluderte deltakere .....	20
Tabell 3: Oppsummering av test for normalfordeling av data for biceps .....	21
Tabell 4: Oppsummering av test for normalfordeling av data for supraspinatus .....	23
Tabell 5: Resultater av intra-tester relativ reliabilitet for biceps .....	26
Tabell 6: Resultater av intra-tester relativ reliabilitet for supraspinatus .....	27
Tabell 7: Resultater av intra-tester absolutt reliabilitet for biceps .....	27
Tabell 8: Resultater av intra-tester absolutt reliabilitet for supraspinatus .....	28
Tabell 9: Resultater av inter-tester relativ reliabilitet for biceps .....	31
Tabell 10: Resultater av inter-tester relativ reliabilitet for supraspinatus .....	32
Tabell 11: Resultater av inter-tester absolutt reliabilitet for biceps .....	33
Tabell 12: Resultater av inter-tester absolutt reliabilitet for supraspinatus .....	33

## Figurindex

Figur 1: Ultralydapparat fra General Electric, modell Logiq e .....	16
Figur 2: Tykkelsesmåling supraspinatus .....	17
Figur 3: Tykkelsesmåling biceps .....	17
Figur 4: Normalfordeling av datasett A <sup>1</sup> for biceps .....	22
Figur 5: Normalfordeling av datasett A <sup>2</sup> for biceps .....	22
Figur 6: Normalfordeling av datasett B for biceps .....	23
Figur 7: Normalfordeling av datasett A <sup>1</sup> for supraspinatus .....	24
Figur 8: Normalfordeling av datasett A <sup>2</sup> for supraspinatus .....	24
Figur 9: Normalfordeling av datasett B for supraspinatus .....	25
Figur 10: Spredningsplot for datasett A <sup>1</sup> og A <sup>2</sup> for biceps .....	29
Figur 11: Spredningsplot for datasett A <sup>1</sup> og A <sup>2</sup> for supraspinatus .....	30
Figur 12: Spredningsplot for datasett B og gjennomsnittet av A <sup>1</sup> og A <sup>2</sup> for biceps .....	34
Figur 13: Spredningsplot for datasett B og gjennomsnittet av A <sup>1</sup> og A <sup>2</sup> for supraspinatus .....	35

# 1. Introduksjon

## 1.1 Presentasjon av problemområdet og bakgrunn for valg av tema

Skuldersmerter er en vanlig muskel- og skjelettlidelse. Livstidsprevalensen av skuldersmerter har blitt estimert til mellom 7% - 67% og antas å være den tredje vanligste muskel- og skjelettlidelsen, bak rygg- og nakkesmerter (Luime et al., 2004). En av de hyppigste årsakene til skuldersmerter er tendinopati av skuldereener (Lewis, 2010). På tross av at skuldertendinopati forekommer hyppig, er ikke alle underliggende etiologiske og patofysiologiske faktorer fullstendig forstått (Rio et al., 2014; Oliva et al., 2016).

Tendinopati er et område innen helseforskning som det har blitt forsket mye på, og noen patoanatomiske mekanismer har blitt foreslått (Maffulli et al., 2010). Studier har videre indikert en assosiasjon mellom en fortykkelse av senen og tendinopati (Allen and Wilson, 2001; Joensen et al., 2009; Hodgson et al., 2014; Khan et al., 2014).

Diagnostisk ultralyd (UL) er et stadig mer populært klinisk verktøy for manuellterapeuter i primærhelsetjenesten. Stadig flere manuellterapeuter bruker UL som et diagnostisk verktøy og som et utfallsmål for behandlingseffekt. I klinisk praksis anses UL som et hensiktsmessig og valid verktøy i vurderingen av pasienter med tendinopati (Fredberg and Stengaard-Pedersen, 2008; Joensen et al., 2009; Levine et al., 2012).

En av fordelene med UL er at man kan utføre dynamiske undersøkelser og sammenlikne med kontralateral side, som kan bidra til en mer presis diagnostikk (Fredberg and Stengaard-Pedersen, 2008). Videre er UL en kostnadseffektiv og raskere undersøkelse sammenliknet med andre billeddiagnostiske undersøkelsesmetoder som magnetresonanstomografi (MR) og computertomografi (CT) (Moosmayer and Smith, 2005; Sunde and Tariq, 2008). Undersøkelsen er i tillegg ikke-invasiv og strålingsfri som gir en økt trygghet for pasienten (Leong et al., 2012). Potensielle ulemper med UL er at det tar lang tid å lære seg, og at man ofte vil få falske positive eller falske negative svar (Ferrari et al., 2002). Med andre ord at man finner patologi der det reelt ikke foreligger, eller motsatt.

I klinisk sammenheng og i forskning er det hensiktsmessig at måleinstrumenter er reliable og valide. Med dette menes at måleinstrumentene måler det de er tenkt og måle, og at man kan stole



på resultatene fra målingene (Carter et al., 2011). Pasienter med skuldertendinopatier er en stor samfunnsøkonomisk belastning (Hopkins et al., 2016). For å unngå at plagene blir langvarige vil det derfor være gunstig med presise måleverktøy for å iverksette hensiktsmessig behandling tidlig.

Det foreligger få studier som har undersøkt reliabiliteten av UL ved måling av senetykkelse i skulderen (Bjordal et al., 2003; Temes et al., 2014; McCreesh et al., 2015a). Endring i senetykkelse er blitt assosiert med tendinopatiske forandringer av skuldereener (Allen and Wilson, 2001; Joensen et al., 2009; Hodgson et al., 2014; Khan et al., 2014). Det vil derfor være interessant å undersøke reliabiliteten av tykkelsesmålinger med UL på skuldereener.

## **1.2 Sentral teori på området**

### **1.2.1 Generelt om tendinopati**

Begrepet tendinopati er en generell betegnelse for patologi og smerte fra en sene, ofte ved belastning (Murtaugh and Ihm, 2013; Riggins et al., 2015). Patofysiologien og etiologien av tendinopatier er fortsatt ikke fullstendig forstått (Abate et al., 2009; Bjordal, 2010; Rio et al., 2014). Tendinopati kan grovt sett deles inn i to hovedkategorier (Hodgson et al., 2014). I den første kategorien finnes begrepene tendinose og tendinopati, som innebærer en affeksjon av senen forårsaket av mekaniske og/eller degenerative forhold. I den andre kategorien brukes begrepet tendinitt, som insinuerer en inflammatorisk prosess i senen (Hodgson et al., 2014). I denne oppgaven vil begrepet tendinopati brukes.

Ved en betydelig økning i den mekaniske belastningen kan det oppstå repeterende mikrotraumer i vevet (Rees et al., 2009). Dersom denne belastningen vedvarer kan dette medføre tendinopatiske endringer i senen (Nicholas, 1990; Tallon et al., 2001). På samme måte kan en akutt overbelastning føre til seneforandringer som er histopatologisk sammenliknbare med tendinopati ved overbruk (Maganaris et al., 2004; Khan et al., 2014). Det er fortsatt ikke forstått hvorfor noen utvikler tendinopati ved slike belastninger, mens andre ikke gjør det (Oliva et al., 2016). Genetikk er foreslått som en mulig predisponerende faktor for utviklingen av tendinopatier (Harvie et al., 2004; Magra and Maffulli, 2008; Kambouris et al., 2012; Oliva et al., 2014). Etiologien til tendinopatier er trolig multifaktorell, men stammer sannsynligvis fra både eksterne og interne faktorer (Abate et al., 2009; Khan et al., 2014; Riggins et al., 2015).

Rotatorcuffen i skulderen er en av de vanligste anatomiske lokasjonene for tendinopati (Van der Windt et al., 1995). Supraspinatus er hyppig affisert ved tendinopatier i rotatorcuffen (Lundgreen et al., 2011). I tillegg indikerer litteratur at affeksjon av biceps brachii caput longum skjer hyppig ved tendinopati av rotatorcuffen, og at disse kan arte seg likt klinisk (Martinoli et al., 2003; Chen et al., 2005). I forlengelsen av dette fremstår det som hensiktsmessig å inkludere både supraspinatus og biceps brachii caput longum ved skulderundersøkelse med UL.

### 1.2.2 Diagnostisk ultralyd

UL er lydbølger med frekvens mellom 1-20 MHz som sendes og mottas av en transducer (Jacobson, 2012). Lydbølgene blir fortløpende omgjort til bilder på bakgrunn av en pulsekkomåling og fremstilles på en monitor (Ibid). Stadig flere manuellterapeuter bruker UL som et supplement til den kliniske undersøkelsen. Systematiske oversiktsartikler og meta-analyser har vist at UL, sammenliknet med MR, har tilsvarende spesifisitet og sensitivitet i diagnostiseringen av skuldertendinopati (de Jesus et al., 2009; Ottenheim et al., 2010; Smith et al., 2011). Måleverktøyets spesifisitet og sensitivitet innebærer i hvilken grad det kan påpeke fraværet av patologi indikert av en negativ test, og tilstedeværelsen av patologi indikert av en positiv test, respektivt (Carter et al., 2011, s. 183). Videre har UL har blitt validert for å vurdere forandringer i senen, sammenliknet med MR, kirurgiske og histologiske funn (Rutten et al., 2010; Hodgson et al., 2014). Disse faktorene har bidratt til at UL blir hyppigere anvendt i diagnostiseringen av skuldertendinopati.

I litteraturen brukes begrepene hyper-, og hypoekkoiske områder om i hvilken grad lydbølgene reflekteres tilbake fra vevet (Jacobson, 2012). Dette omtales som vevets ekkogenitet. Et hypoekkoisk område i senen vil fremstå som mørkere enn et hyperekkkoisk, og har tidligere blitt assosiert med langvarig tendinopati (Thain and Adler, 1999). Andre studier beskriver nedsatt ekkogenitet grunnet tap av normal fiberstruktur i senen ved tendinopati, liknende funn har i midlertid blitt gjort på friske individer (Bjordal et al., 2003; Hodgson et al., 2014). Overnevnte tegn bør derfor tolkes med forsiktighet som et kriterie i diagnostiseringen av tendinopatier. På en annen side er lokal fortykkelse av senen, sammenliknet med kontralateral side, blitt assosiert med tendinopatiske endringer (Allen and Wilson, 2001; Joensen et al., 2009; Hodgson et al., 2014). Dette kan indikere at tykkelsesmåling av senen kan være et hensiktsmessig alternativ i diagnostiseringen av tendinopati (Joensen et al., 2009).

Vinkelen på transduceren i forhold til vevet man undersøker vil i stor grad avgjøre om vevet fremstår som hyper-, eller hypoekkoisk. Det vil bli økt billedkvalitet dersom transduceren blir plassert perpendikulært på orienteringen av senens fiberretning. Vinkelen på transduceren er forbundet med potensielle målefeil, da små vinkelendringer kan gi et falskt inntrykk av ekkogenitet og dermed bli mistolket som patologi (Hodgson et al., 2014). Dette fenomenet omtales som anisotropi (Allen and Wilson, 2001).

### 1.2.3 Reliabilitet

Med reliabilitet menes i hvilken grad en målemetode er fri for målefeil og produserer presise og konsistente målinger av samme variabel under identiske forhold (American Educational Research Association, 1999; Portney and Watkins, 2000; Clark, 2001). Med andre ord indikerer reliabiliteten i hvilken grad målemetoden er pålitelig. Ved en klinisk måling er det sannsynlig at en viss grad av målefeil vil forekomme. I forlengelsen av dette kan reliabilitet anses som den aksepterte graden av målefeil for hensiktsmessig og praktisk bruk av en målemetode (Atkinson and Nevill, 1998).

Valg av statistisk metode for analysen av reliabilitet blir blant annet bestemt av om datamaterialet er homoskedastisk eller heteroskedastisk. Ved et heteroskedastisk datamateriale er forandringer i nivå av målte verdier systematisk lavere eller høyere nivå på den målte variabiliteten. Foreligger det ingen samsvar mellom målt variabilitet og størrelse på måleverdiene, omtales datamaterialet som homoskedastisk. Parametrisk analyse kan benyttes dersom datamaterialet er homoskedastisk og normalfordelt (Atkinson and Nevill, 1998).

Denne studien fokusere på intra-, og inter-tester reliabilitet. Intra-tester reliabilitet, eller test-retest reliabilitet, innebærer i hvilken grad det er samsvar mellom målingene gjennomført av samme tester på samme måleobjekt, ved to separate målinger (Carter et al., 2011). Inter-tester reliabilitet omhandler i hvilken grad det er samsvar mellom målingene til to individuelle målere på samme måleobjekt (ibid).

Reliabilitet blir gjerne vurdert med bakgrunn i to hovedteorier, den klassiske måleteorien og generaliserbarhetsteorien (Carter et al., 2011, s. 238). Den klassiske måleteorien baseres på antagelsen at det foreligger en sann verdi og en feilkilde eller ved hver måling. Ytterligere vil hver person ha en sann verdi for den målingen som blir utført. Siden man ikke kan vite den

sanne verdien, vil målefeilen avgjøres av variasjonen mellom gjentatte målinger. Målefeilen stammer, i denne teorien, primært fra variabiliteten mellom de ulike målingene.

Generaliserbarhetsteorien stammer fra den klassiske måleteorien, men tar høyde for at det kan foreligge andre kilder for variabilitet mellom målingene (Carter et al., 2011). Disse kildene for målefeil deles inn i ulike grupper for variabilitet. Potensielle kilder kan være testerens erfaring i bruk av måleinstrumentet, biologiske variasjoner hos testpersonen, eller presisjonen av måleinstrumentet. Generaliserbarhetsteorien muliggjør at man i større grad kan kvantifisere potensielle kilder til målefeil som er relevante i en klinisk hverdag (Carter et al., 2011, s. 238).

### ***1.2.3.1 Variasjon og målefeil***

Resultatene av kvantitative mål består hovedsaklig av to deler; feilkilder og måleverdier (Carter et al., 2011). Reliabilitetsteori har som mål å anslå målefeil og hvordan man kan minimere disse. I en reliabilitetsstudie vil man derfor undersøke hvorvidt måleresultatene er fri for tilfeldige og systematiske målefeil. Ved systematiske målefeil blir målingene uniformt påvirket i en positiv eller negativ retning (Atkinson and Nevill, 1998; Portney and Watkins, 2000). En slik feil vil være mulig å forutsi og identifisere. Et eksempel på systematiske målefeil kan være feilkalibrering av måleinstrumentet.

En annen kilde til variasjon er tilfeldige målefeil. Tilfeldige målefeil omtales som “støy” i målingene (Streiner et al., 2014). I motsetning til systematiske målefeil, vil tilfeldige målefeil være uforutsigbare og kunne affisere resultatet fra gang til gang. Graden av tilfeldige målefeil er normalt sett større enn systematiske målefeil (Atkinson and Nevill, 1998). Eksempler på tilfeldige målefeil kan være biologiske variasjoner hos testpersonene, plottingsfeil, manglende standardisering av måleprotokoll, eller mekaniske variasjoner av måleinstrumentet. Målinger kan være utsatt for en kombinasjon av testerfeil, instrumentfeil og variasjon av målersubjektet. Det kan være utfordrende å skille i hvilken grad de ulike variablene påvirker målefeilen (Portney and Watkins, 2000; Carter et al., 2011). Det er derfor viktig å standardisere måleprotokollen i størst mulig grad (Batterham and George, 2003). Reliabilitet kan kvantifiseres på to måter, relativ og absolutt reliabilitet (Carter et al., 2011, s. 239).

### 1.2.3.2 Relativ reliabilitet

Relativ reliabilitet undersøker forholdet, og graden av overenstemmelse, mellom to eller flere serier av repeterte målinger (Carter et al., 2011, s. 239-240). Dette innebærer at dersom et måleinstrument er reliabelt vil individuelle målinger innen en gruppe beholde sin posisjon i gruppen, selv ved gjentatte målinger. For eksempel vil en person som skårer nær toppen i en tenkt distribusjon ved første måling, forventes å holde seg nær toppen av distribusjonen selv om deres faktiske skår endres over tid (ibid).

Relativ reliabilitet kan bli målt ved å bruke en korrelasjonskoeffisient (Carter et al., 2011). En korrelasjonskoeffisient brukes hyppig i reliabilitetsstudier er intraklasse korrelasjonskoeffisient (ICC) som baseres på variansanalyse (ANOVA) (Rankin and Stokes, 1998; Carter et al., 2011). ICC beskrives som en univariat reliabilitetskoeffisient som er hensiktsmessig å bruke ved to eller flere sett av målinger av samme variabel (Atkinson and Nevill, 1998). En definisjon på ICC er: *“The ratio between the true variance and the total variance, where the true variance is the difference between the total variance and the variance due to error of measurement”* (Moenilssen, 1998). Dette vil si at ICC omtales som forholdet mellom den totale variansen som kan tilskrives målefeil og den sanne variansen. Videre indikerer ICC i hvilken grad det foreligger enighet og relasjon mellom målingene, og kan ytterligere reflektere måle metodens evne til å differensiere mellom individer (Stratford and Goldsmith, 1997; Portney and Watkins, 2000). ICC egner seg å bruke som korrelasjonskoeffisient ved måling av senetykkelse fordi variansen og variablene (mm) måles i samme metriske målestokk.

Den relative naturen av ICC gjør at størrelsen av ICC være avhengig av variasjonen mellom de målte deltakerne. Dersom målingene blir gjort på en homogen gruppe (lite variasjon) kan størrelsen av ICC være liten, selv om test-retest målingene er relativt like. Videre kan størrelsen av ICC være stor selv om test-retest målingene er relativt ulike, dersom målingene blir gjort på en heterogen gruppe (stor variasjon) (Weir, 2005).

Det finnes flere variasjoner av analyse med ICC (Shrout and Fleiss, 1979). Hvilken modell av ICC man bruker, avhenger av studiens hensikt, design og type målinger man skal gjennomføre. Det skilles mellom enveis og toveis ICC modeller, basert på om det foreligger en eller to kilder til systematisk varians (ibid).

ICC<sub>1,1</sub> beskrives som en “one way random” i analyseprogrammet Statistical Package for the Social Sciences (SPSS), og tar utgangspunkt i at målefeil er eneste grunn til variasjon i målingene til den enkelte deltaker (Shrout and Fleiss, 1979). I min studie vil deltakerne være en av de uavhengige faktorene. En toveismodell åpner for muligheten til å ytterligere undersøke to eller flere uavhengige faktorer for samme avhengige faktor. I inter-testerdelen av min studie er testerne den andre faktoren, og testrundene vil være den andre faktoren ved intra-testerdelen.

En annen modell av ICC er ICC<sub>3,1</sub>. Denne modellen tar høyde for at variabiliteten kan tilskrives en læringseffekt, uten antagelsen at målefeil kan tilskrives systematiske feil. Dette tilsier at dersom det ikke foreligger en læringseffekt vil ICC<sub>1,1</sub> være tilnærmet lik ICC<sub>3,1</sub> (Moe-Nilssen, 1998). ICC<sub>3,1</sub> tar utgangspunkt i at deltakerne blir ansett som uavhengig faktorer mens testerne er anses som avhengig faktorer, og modellen kalles derfor “two way mixed” i SPSS. Med denne modellen er man kun interessert i et bestemt sett med testere og man derfor ikke generalisere resultatene utover disse testerne (Shrout and Fleiss, 1979).

Graden av enighet mellom målingene kan uttrykkes som “agreement” eller “consistency”. Ved “agreement” er systematiske målefeil mellom testerne/målingene tatt høyde for, i motsetning til “consistency” da dette ikke blir gjort. Videre kan ICC oppgis som “average measures” eller “single measures”. “Average measures” baserer reliabiliteten på gjennomsnittet av  $k$  målinger, mens “single measures” baseres kun på enkeltmålinger (McGraw and Wong, 1996). Resultatene for ICC presenteres gjerne som et nummer mellom 0,00 og 1,00, der 1,00 indikerer perfekt reliabilitet, mens 0,00 indikerer ingen reliabilitet (Carter et al., 2011).

I reliabilitetsstudier anbefales det at man undersøker og presenterer både relativ og absolutt reliabilitet (Rankin and Stokes, 1998; Kottner et al., 2011). Dette med bakgrunn i at ICC ikke indikerer presisjonen i målingene i en gitt måleenhet og at resultatene kan bli påvirket av variasjonen i utvalget (Carter et al., 2011). Samtidig kan absolutt reliabilitet presentere reliabilitetsresultater i en måleenhet som har større praktisk nytteverdi (Rankin and Stokes, 1998; Carter et al., 2011).

### **1.2.3.3 Absolutt reliabilitet**

Absolutt reliabilitet indikerer i hvilken grad det er variabilitet for skåringer fra måling til måling (Carter et al., 2011). Men andre ord hvor stor forskjell det er for gjentatte målinger på samme individ. Resultatene for absolutt reliabilitet presenteres i den samme måleenheten som blir benyttet i målingen (Stratford and Goldsmith, 1997). Dette gjør resultatene ( $X \pm SD$ ) mer anvendbare i klinisk praksis. Dess mindre standardavviket (SD) er, dess bedre vil reliabiliteten av målingene være (Atkinson and Nevill, 1998). Et lite SD betyr ikke nødvendigvis at korrelasjonskoeffisienten er høy, og vice versa. Det er derfor hensiktsmessig å presentere den absolutte reliabiliteten og SD i en reliabilitetsstudie (Altman, 2000; Carter et al., 2011).

En metode for målingen av absolutt reliabilitet er “standard error of measurement” (SEM) (Rankin and Stokes, 1998; Weir, 2005). SEM er også kjent som “innen-subjekt standard deviasjon” og vil heretter bli omtalt som Sw (Bland and Altman, 1996). Generelt sett kan man si at dess lavere Sw dess bedre reliabilitet av målingene. På bakgrunn i variasjonen av måleenheter ved forskjellige målemetoder, finnes det ingen entydig konsensus om hva Sw må være for at målemetoden anses som reliabel (Atkinson and Nevill, 1998; Streiner et al., 2014).

Det kreves to eller flere målinger for å beregne Sw. For 95% av tilfellene, forventes det at differansen mellom den sanne verdien og målingen ligger under  $1,96 \times Sw$ . I forlengelsen av dette, må forskjellen ligge utenfor  $\sqrt{2} \times 1,96 \times Sw$  (eller  $2,77 \times Sw$ ) for den samme testpersonen, for å kunne fastslå om en reell målbar endring har funnet sted (Bland and Altman, 1996). Dette omtales som minste målbare endring. Dersom endringene er mindre enn den minste målbare endringen, kan dette bety at en reell endring ikke har funnet sted (De Vet et al., 2011, s. 243). Dette indikerer at dess mindre Sw, dess mindre blir størrelsen på den klinisk relevante endringen.

## **1.3 Tidligere forskning på området**

For å finne tidligere forskning ble det utført søk i elektroniske databaser med relevante søkeord relatert til tendinopati, skulder, reliabilitet og UL som ble logisk kombinert med boolske operatører. Det ble deretter utført et håndstøk av relevante referanser. Det foreligger en del studier vedrørende intra-, og inter-tester reliabiliteten for målinger med UL av den acromiohumorale

avstanden (McCreesh et al., 2015b). Færre studier har blitt gjennomført vedrørende reliabiliteten av UL ved tykkelsesmålinger av skuldereener.

Bjordal og kolleger undersøkte i 2003 blant annet intra-tester reliabiliteten ved måling med UL av senetykkelsen til supraspinatus (Bjordal et al., 2003). Studien ble gjennomført på friske individer, og ultralydmålingen ble gjennomført av en erfaren fysioterapeut. Studiens resultater indikerer god intra-tester reliabilitet av tykkelsesmåling, med en oppgitt absolutt reliabilitet på 0,24 mm mellom målingene. Gjennomsnittstykkelsen av m. supraspinatus ble i studien anslått til å være 6,5 mm.

I 2012 undersøkte Leong et al. blant annet intra-tester reliabiliteten på tykkelsesmålinger med UL på supraspinatussenen, hos symptomfrie og utrente individer (N = 11) (Leong et al., 2012). Senetykkelsen av supraspinatus ble målt på tre ulike steder langs senens forløp og gjennomsnittet av disse var grunnlaget for analysen. Tykkelsesmålingene ble utført av en fysioterapeut, med tre måneders intensiv erfaring med UL. Studien presenterer utmerkede resultater for relativ reliabilitet (ICC = 0,92 med 95% KI: 0,90 – 0,98), og absolutt reliabilitet (Sw = 0,23 mm).

Intra- og inter-tester reliabiliteten av UL målinger for tykkelsen av supraspinatus undersøkt av Temes og kolleger i 2014 (Temes et al., 2014). Tykkelsesmålingen av senen ble gjort i kontrahert og ikke-kontrahert tilstand, på friske deltakere (N = 15). Målingen med UL ble gjennomført av tre fysioterapeuter som på forhånd ble trent i gjennomføringen av målingene med UL av en mer erfaren tester. Det var en uke mellom test og re-test. Resultatene for relativ reliabilitet omtales som meget høy (ICC > 0.9) for både intra-, og inter-tester reliabilitet. Den absolutte reliabiliteten varierte fra 0.07 mm – 0.12 mm.

En nylig publisert studie undersøkte intra-, og inter-tester reliabiliteten av UL målinger av tykkelsen av supraspinatussenen hos både symptomatiske (N = 45) og ikke-symptomatiske (N = 25) deltakere (McCreesh et al., 2015a). I tykkelsesmålingene av supraspinatussenen ble det gjennomført to målinger per sene. I beregningene av reliabilitet ble gjennomsnittet av disse to målingene brukt. Målingene ble gjennomført av to testere. Det blir ikke utdypet grad av erfaring med UL og klinisk bakgrunn av testerene. Studiens resultater for relativ reliabilitet indikerer utmerket intra-, og inter-tester reliabilitet for syke og friske deltakere (ICC = 0,93 – 0,97). Studiens absolutte reliabilitet varierte fra 0,2 mm – 0,5 mm.



## **2. Hensikt og problemstilling**

### **2.1 Hensikt**

Hensikten med denne studien er å undersøke inter-tester og intra-tester reliabiliteten i måling av senetykkelsen av supraspinatus og biceps brachii caput longum på samme ultralydbildet hos pasienter med skuldertendinopati.

### **2.2 Problemstilling**

Hvordan er inter-tester, og intra-tester reliabiliteten ved måling av senetykkelsen til m. biceps brachii caput longum og m. supraspinatus med ultralydsavbildning hos pasienter med skuldertendinopati?

## **3. Metode**

### **3.1 Valg av forskningsdesign**

For å svare på problemstillingen ble det gjennomført en kvantitativ ikke-eksperimentell reliabilitetsstudie. Kvantitative metoder er forskning der måling, kontroll og reproduserbarhet står i fokus (Carter et al., 2011). Med ikke-eksperimentelt design menes et forskningsdesign der studiens avhengige variabler ikke blir manipulert (Drageset and Ellingsen, 2009; Carter et al., 2011). I reliabilitetsstudier er formålet blant annet å dokumentere og forbedre reliabiliteten av kliniske måleinstrumenter og tester (Walter et al., 1998; Carter et al., 2011).

Denne studien er en del av et større doktorgradsprosjekt som foregår i Rogaland (Haslerud, 2012). Dette doktorgradsprosjektet utføres som en randomisert kontrollert studie som undersøker effekten av en kunnskapsbasert fysioterapipakke gitt til pasienter med vedvarende rotatorcufftendinopati. Intervensjonene som ble gitt var forskjellige styrkeøvelser, manuelle mobiliseringer, tøyning av kapsel og laserterapi.

Videre beskrivelse av metoden i denne studien vil være basert på forskningsprotokollen til dette doktorgradsarbeidet.

## 3.2 Utvalg

Utvalget til denne studien er deltakere i doktorgradsstudien, og består av kvinnelige og mannlige pasienter med vedvarende skuldersmerter. Deltakerne blir rekruttert fra ventelistene til kirurgisk avdeling ved Stavanger Universitetssykehus fra september 2014 til ut juni 2016. Videre måtte potensielle deltakere oppfylle forutsatte inklusjon- og eksklusjonskriterier for å kunne delta.

### 3.2.1 Inklusjonskriterier

For å kunne inkluderes i studien måtte pasientene:

1. Ha smerter lokalisert til øvre del av C5 dermatom, samt ha en historie om problemer med arbeid som foregår med armene løftet over hodet.
2. Ha en aktuell smertevarighet vedvarende i minimum tre måneder.
3. Ha tre eller flere reproduerbare tegn til subacromialt impingementsyndrom eller rotatorcufftendinopati
  - Positiv isometrisk abduksjon og/eller lateral rotasjon (Ombregt et al., 2003)
  - Positiv painful arc i aktiv abduksjon av skulder (Ombregt et al., 2003)
  - Positiv Neers tegn (Neer, 1972; Tennent et al., 2003)
  - Positiv Jobs test (Jobe and Moynes, 1982; Tennent et al., 2003)
  - Positiv Hawkins-Kennedy impingement test (Hawkins and Kennedy, 1980)

### 3.2.2 Eksklusjonskriterier

Deltakere ble ekskludert fra studien dersom de:

1. Har blitt behandlet med ikke-steroid antiinflammatoriske midler (NSAID) eller glukokortikoide injeksjoner de siste fire ukene.
2. Har skuldersmerter relatert til traume, som for eksempel fall.
3. Redusert bevegelsesutslag som ved adhesiv kapsulitt/frossen skulder
4. Har en anamnese og klinisk undersøkelse (apprehension/relocation) som danner mistanke om patologisk instabilitet i skulderen.
5. Har totalruptur av rotatorcuffsenen, definert som positiv drop-arm test og smertefri svakhet av isometrisk abduksjon og utadrotasjon.
6. Har artrose i glenohumeralledet, har os acromion hvor formen gir redusert subacromial plass, tidligere frakturer eller kirurgi i skulderkomplekset på affisert side.

7. Har annen komorbiditet som revmatiske eller nevrologiske sykdommer, fibromyalgi eller symptomer fra cervicalcolumna.
8. Har alvorlige psykiatriske lidelser.
9. Ikke forstår norsk eller engelsk.
10. Har færre enn tre positive, reproduserbare tegn til rotatorcuffpatologi eller SIS.

Alle deltakerne ble vurdert med tanke på inklusjons- og eksklusjonskriterier av en erfaren fysioterapeut som er medvirkende i doktorgradsarbeidet, ved et legesenter i Rogaland.

### 3.3 Etiske hensyn

I doktorgradsarbeidet, og min studie, ble de etiske retningslinjene nedfelt i Helsinkideklarasjonen fulgt (World Medical, 2013). Doktorgradsarbeidet ble godkjent i regional komité for medisinsk forskningsetikk (REK) nord den 22.02.13 med referanse 2012/2134 *Effekten av kunnskapsbasert fysioterapibehandling for pasienter med rotator cuff tendinopati* (Vedlegg 1). For inklusjon i doktorgradsarbeidet måtte alle deltakere underskrive en informert samtykkeerklæring (Vedlegg 2). Samtlige deltakere ble muntlig og skriftlig informert om deres mulighet for å trekke seg når som helst. Alle deltakere ble anonymisert, og forskningspersonell var underlagt taushetsplikt. All data ble oppbevart innelåst, og var kun tilgjengelig for personer involvert i min studie. Min studie har ingen kommersiell interesse og ble ikke gjort på oppdrag.

### 3.4 Måleinstrumenter

For å bedømme effekt av intervensjonen i doktorgradsarbeidet ble det utført en kombinasjon av pasientutfylte spørreskjemaer og objektive fysiske målinger. Alle målinger ble gjort ved baseline, og målt på nytt etter tre og seks måneder etter intervensjon. Utfallsmålene i doktorgradsarbeidet var blant annet skulderfunksjon målt med Shoulder Pain and Disability Index (SPADI) (Breckenridge and McAuley, 2011), livskvalitet målt med spørreskjemaet EQ-5D (Rabin and de Charro, 2001), smerte målt med numeric rating scale, samt målinger av senetykkelse med UL (Haslerud, 2012).

Da studien min fokuserer på inter-, og intra-tester reliabiliteten av UL målinger av senetykkelsen til supraspinatus og biceps brachi caput longum ble allerede eksisterende data fra doktorgradsarbeidet brukt. Dette ble gjort av tidsmessige og praktiske årsaker. All UL data ble

samlet inn ved bruk av et ultralydapparat fra General Electric, modell Logiq e (Figur 1) (General Electric, 2016). Det ble brukt en lineær probe på 10 millihertz (MHz).



Figur 1: Ultralydapparat fra General Electric, modell Logiq e

## 3.5 Datainnsamling

### 3.5.1 Datainnsamling til doktorgradsarbeid

Innsamlingen av data med UL blir gjort av en erfaren fysioterapeut som er medvirkende til doktorgradsarbeidet. Denne fysioterapeuten har lang erfaring med UL som verktøy i forskning og i klinisk sammenheng. Dataen som ble brukt i denne studien ble innhentet fra 6.7.14 til 18.5.16 på et legesenter i Rogaland. En standardisert undersøkelsesprotokoll for innsamling av UL bildene ble brukt i doktorgradsarbeidet. Denne protokollen er reliabilitetstestet og hentet fra tidligere studier (Bjordal et al., 2003; Joensen et al., 2009).

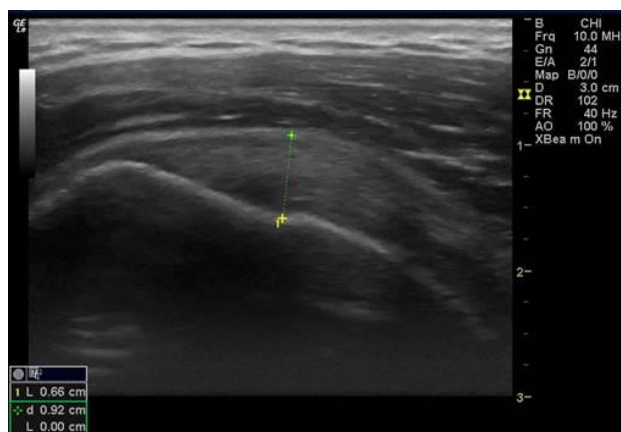
For supraspinatus ble pasienten plassert på en stol med armen bak på ryggen for å få frem senen til m. supraspinatus. Senetykkelsen til m. supraspinatus ble målt med UL apparatets innebygde kalibrerte kalipere på forhåndsbestemte benete landemerker for å sikre plasseringen av sonden longitudinelt på senen. Dernest ble en transvers glidning av sonden utført for å avgjøre den posisjonen der senen blir bedømt som tykkest. Tykkelsen ble målt langs en imaginær linje som løper 90 grader på hudens overflate og dratt til de overfladiske og dype dele av paratendon. Det benete landemerket som brukes i målingen av senen til supraspinatus er knekkpunktet mellom den humerale leddflate og basis av tuberculum majus. Målingene ble utført bilateralt.

Målingene av m. biceps brachii caput longum ble gjennomført i henhold til europeiske retningslinjer for UL (Beggs et al., 2010). Pasienten blir plassert sittende med armen hvilende i fanget med skulderen i lett innadrotasjon. Albuen flekteres til 90 grader og supineres fullt. Proben påføres transversalt på den lange senen til m. biceps brachii, mellom tuberculum majus og minus på os humerus. Målingene ble gjennomført bilateralt.

### 3.5.2 Datainnsamling til mastergradsoppgave

De innsamlede UL bildene fra doktorgradsarbeidet ble brukt i min studie. Inter-tester reliabilitetsmålingene ble gjennomført av to testere. Intra-tester reliabilitetsmålingene ble gjennomført av hovedforfatteren. Tykkelsesmålingene ble gjort på samme apparat som ved scanningene i forbindelse med doktorgradsarbeidet (Figur 1) (General Electric, 2016).

Før målingene startet ble parameterne for målingene av supraspinatus og biceps brachii caput longum forhåndsbestemt mellom de to raterne. Senetykkelsen av supraspinatus ble definert som den korteste avstanden fra “nebbet” på scapula (beinlinjens knekkpunkt mellom den humerale leddflate og basis av tuberculum majus) til og med distale del av bursa subacromialis (Figur 2). Senetykkelsen av biceps caput longum ble definert som bunnen av sulcus intertubercularis til de kraftige hvite strekene som tenkes å være slutten av epitendon (Figur 3). Samtlige inkluderte deltakere hadde skuldertendinopati, men måling av senetykkelsen ble gjennomført bilateralt, uavhengig av sykdomsside, der dette var tilgjengelig. Dette ble besluttet å gjøre for å øke antall målte sener, og fordi studiens fokus er reliabilitet av målingene med UL, og ikke måleforskjeller på friske og syke sener.



Figur 2: Tykkelsesmåling supraspinatus



Figur 3: Tykkelsesmåling biceps brachii caput longum

En praktisk øvelsesrunde for utførelsen av tykkelsesmålingene med UL ble foretatt før selve gjennomføringen av datainnsamlingen til studien. Denne øvelsesrunden involverte begge testerne og varte omtrent en time.

Under selve gjennomføringen av tykkelsesmålingene målte tester 1 datasettene to ganger. Disse målingene vil bli omtalt som  $A^1$  og  $A^2$ .  $A^1$  representerer tester 1 sin første måling, og  $A^2$  representerer tester 1 sin andre måling. Tester 2 målte datasettene én gang og vil bli omtalt som  $B$ . Analysen for inter-tester reliabiliteten ble målt mellom målesettene  $A^1$  og  $B$ . Analysen for intra-tester reliabiliteten ble målt mellom målesettene  $A^1$  og  $A^2$ . Tykkelsesmålingene ble gjennomført over to dager i den siste uken i mai 2016 ved Høgskolen i Bergen institutt for ergoterapi, fysioterapi og radiografi.

En mal for datainnsamlingen ble laget, og brukt, av begge testerne (Vedlegg 3). Malen inneholdt pasientnummer samt dato og klokkeslett for scanningen. Disse ble registrert og ble brukt for å identifisere riktig bilde for tykkelsesmåling. Ved enkelte scanninger var tykkelsesmål av senen synlige fra tidligere målinger. For å sikre blinding av tidligere målinger, ble dette området av skjermen tildekket under tykkelsesmålingene.

## 3.6 Analyse

### 3.6.1 Blinding

For å foreta en blindet analyse ble datasettene sendt til en uavhengig måler som kodet målingene. Etter den statistiske analysen var gjennomført ble kodenøkkelen åpnet opp av samme uavhengige måler.

### 3.6.2 Statistisk metode

Analysen vil gjøres på tidligere målinger av senetykkelse i forbindelse med doktorgradsarbeidet og nye målinger, gjort av en annen person, på det samme ultralydbildet. Graden av enighet vil bestemmes utifra disse målingene og vil i den forstand være en sekundæranalyse. Data vil bli analysert ved bruk av Statistical Package for the Social Sciences 23 (SPSS) for Windows (IBM Corp, 2013).

I denne studien vil ICC med one-way random ( $ICC_{1,1}$ ) og ICC med two-way mixed og absolute agreement ( $ICC_{3,1}$ ) være det primære utfallsmålet for å vurdere relativ reliabilitet. Single measures ble brukt for begge modellene av ICC. Resultatet av analysen for relativ reliabilitet resulterte i en ICC mellom 0,00 – 1,00. I min studie vil Koch sin beskrivelse for kvantifiseringen av ICC bli brukt (Tabell 1) (Koch, 1982). Pearsons R ble gjennomført som sekundær utfallsmål. For Pearsons R ble en sterk sammenheng definert som 0,50 – 1,00 (Cohen, 1988). Innen-subjekt standard deviasjon (Sw),  $Sw \times 1,96$  og  $Sw \times 2,77$  ble anvendt for å vurdere absolutt reliabilitet. Alle resultater ble presentert med signifikansnivå, 95% konfidensintervall og standardavvik der dette er tilgjengelig og relevant.

ICC	Grad av reliabilitet
<0	Ingen
0.00-0.20	Svak
0.21-0.40	Brukbart
0.41-0.60	Moderat
0.61-0.80	Betydelig
0.81-1.00	Tilnærmet perfekt

Tabell 1: Beskrivelse av grad av enighet ved ulike ICC

Datamaterialet ble analysert for normalfordeling ved bruk av Shapiro-Wilk. Denne metoden ble valgt ovenfor Kolmogorov-Smirnov, da sistnevnte har blitt kritisert for å være uhensiktsmessig ved små utvalg (Wilcox, 1997).

Demografiske data (alder, høyde og vekt) er beskrevet som deskriptiv statistikk.

## 4. Resultater

### 4.1 Demografiske data

Målingene med UL ble gjennomført på totalt N = 27 deltakere fra doktorgradsarbeidet. 33% var menn (n = 9) og 67% var kvinner (n = 18). I forbindelse med doktorgradsarbeidet ble noen av deltakerne scannet flere ganger og antall datasett var derfor høyere enn 27. Analysen av biceps brachii caput longum venstre side ble basert på 36 målinger med UL. Analysen av biceps brachii caput longum høyre side ble gjennomført basert på 37 målinger. Det ble brukt 37 målinger i analysen av supraspinatus venstre side. Antall målinger til analysen av supraspinatus høyre side var 38.

Gjennomsnittsalderen til deltakerne var 52 år (95% konfidensintervall (KI): 48 år – 56 år). Den eldste deltakeren var 74 år ved gjennomført scanning. Den yngste deltakeren var 29 år ved gjennomført scanning. Gjennomsnittsvekten til deltakerne var 75,37 kg (95% KI: 69,54 kg – 81,20 kg). Den tyngste deltakeren veide 120 kg. Den letteste deltakeren veide 55 kg.

Gjennomsnittshøyden til deltakerne var 169,52 cm (95% KI: 166,35 cm – 172, 69 cm).

Demografiske data

	Minimum	Maksimum	Gjennomsnitt	Standardavvik
<b>Alder</b>	29 år	74 år	52 år	10,3 år
<b>Vekt</b>	55 kg	120 kg	75 kg	14,7 kg
<b>Høyde</b>	156 cm	183 cm	169 cm	8,0 cm

Tabell 2: Oppsummering av demografiske data av inkluderte deltakere (N=27) med alder, vekt og høyde

### 4.2 Normalfordeling av data

Normalfordelingen av datamaterialet ble analysert med Shapiro-Wilk. Hovedanalysene for supraspinatus og biceps brachii caput longum gjøres på datamaterialet med høyre og venstre side slått sammen (HØVE). Det er derfor valgt å gå videre med parametriske tester siden datamaterialet er tilfredsstillende normalfordelt. Det er også valgt å gjøre subgruppeanalyser med høyre (HØ) og venstre (VE) side separat. Ved analyse med Shapiro-Wilk slo variabelen A<sup>2</sup> biceps høyre side ut som signifikant ikke-normalfordelt (p = 0,030). Etter en samlet vurdering ble det, på tross av dette, valgt å gjøre parametrisk testing. Dette gjøres da den parametriske testingen baseres på ANOVA som er en robust test som tåler data som ikke er normalfordelt (Norman, 2010).



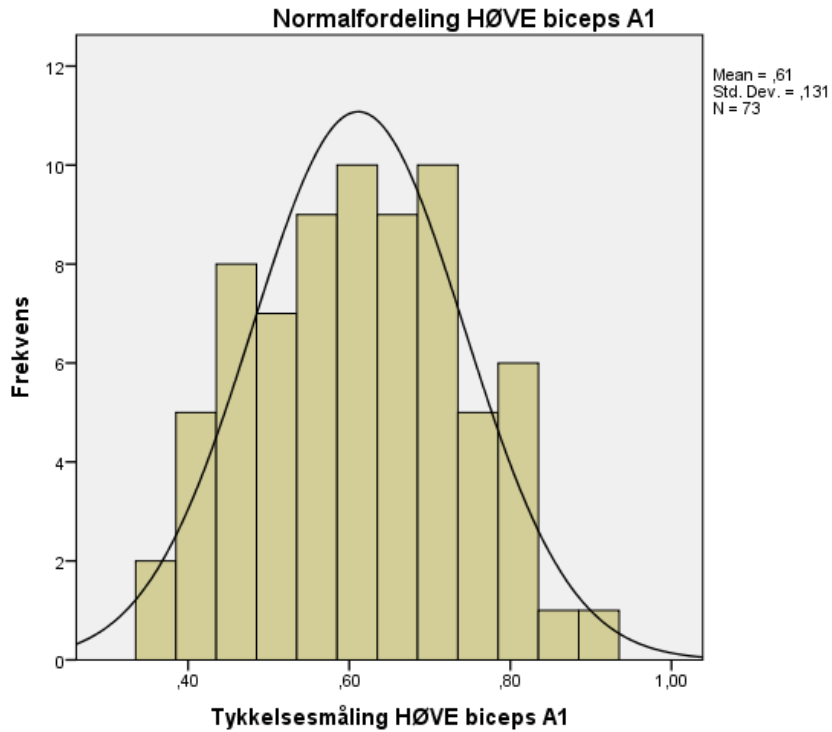
Videre viste analysen med Shapiro-Wilk for normalfordeling av data for biceps brachii caput longum, der høyre og venstre side ble slått sammen, at datasett A<sup>1</sup> hadde en gjennomsnittsmåling på 6,1 mm ( $\pm$  SD = 1,31 mm). Gjennomsnittsmålingen til datasett A<sup>2</sup> for høyre og venstre side var 6,1 mm ( $\pm$  SD = 1,23 mm). Analysen av datasett B gav et gjennomsnitt for høyre og venstre side på 5,8 mm ( $\pm$  SD = 1,35 mm). Resultatene for testing av normalfordeling av biceps brachii caput longum er oppsummert i tabell 3.

Analysen med Shapiro-Wilk for normalfordeling av data for supraspinatus, der høyre og venstre side ble slått sammen, viste at datasett A<sup>1</sup> hadde en gjennomsnittsmåling på 8,0 mm ( $\pm$  SD = 1,20 mm). Videre viste analysen av samme anatomiske struktur at datasett A<sup>2</sup> hadde en gjennomsnittsmåling på 8,3 mm ( $\pm$  SD = 1,13 mm). Resultatene for datasett B for supraspinatus der høyre og venstre side ble slått sammen gav en gjennomsnittsmåling på 7,5 mm ( $\pm$  SD = 0,93 mm). Resultatene for normalfordeling av supraspinatus er summert opp i tabell 4.

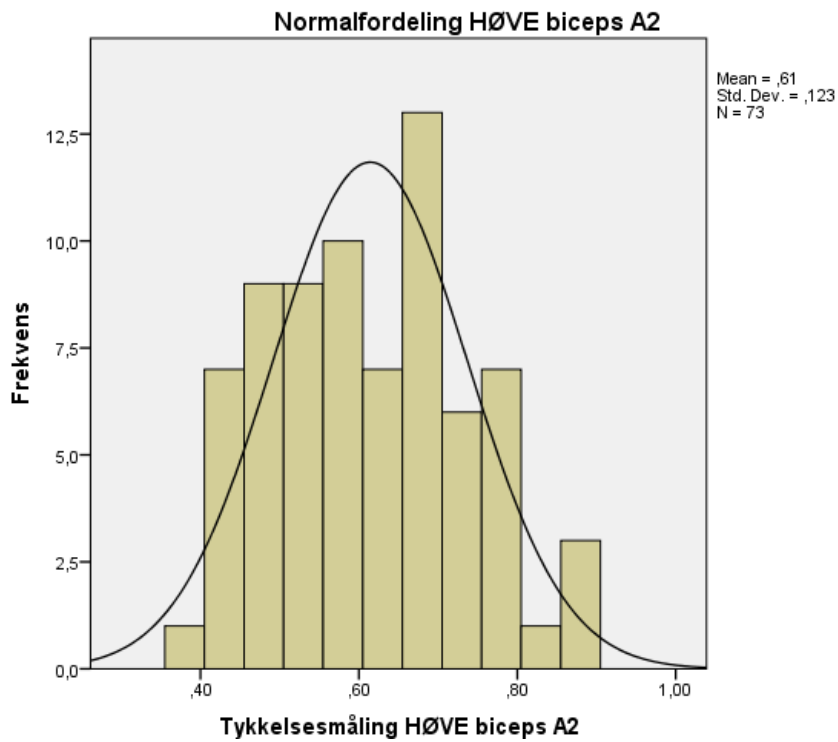
Normalfordeling av data for biceps brachii caput longum

	Shapiro-Wilk		
	Statistic	Frihetsgrader	Signifikansnivå
VE Biceps A <sup>1</sup>	,964	36	,280
VE Biceps B	,964	36	,285
VE Biceps A <sup>2</sup>	,971	36	,461
HØ Biceps A <sup>1</sup>	,974	37	,535
HØ Biceps B	,957	36	,168
HØ Biceps A <sup>2</sup>	,934	37	,030
HØVE Biceps A <sup>1</sup>	,979	73	,253
HØVE Biceps B	,975	72	,169
HØVE Biceps A <sup>2</sup>	,975	73	,151

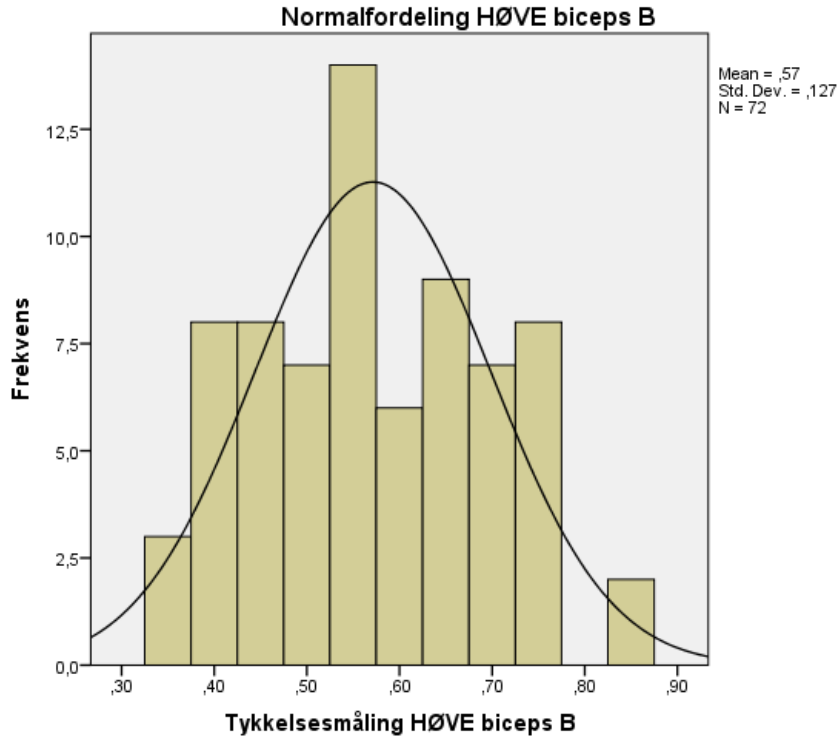
Tabell 3: Oppsummering av test for normalfordeling av data med Shapiro-Wilk for biceps brachii caput longum



Figur 4: Normalfordeling av datasett A<sup>1</sup> for biceps brachii caput longum der høyre og venstre side er slått sammen, gjennomsnitt, antall malinger og standardavvik.



Figur 5: Normalfordeling av datasett A<sup>2</sup> for biceps brachii caput longum der høyre og venstre side er slått sammen, gjennomsnitt, antall malinger og standardavvik.

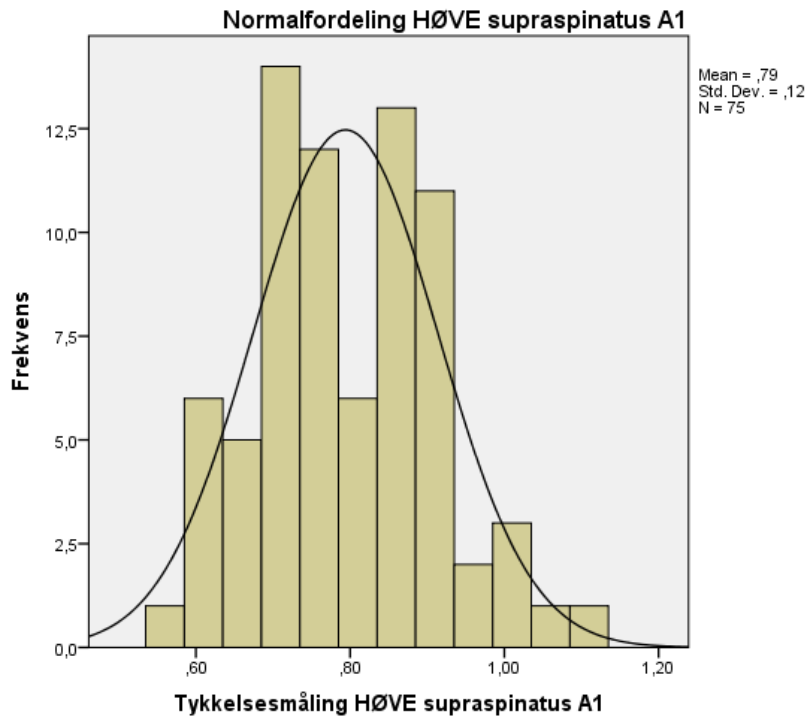


Figur 6: Normalfordeling av datasett B for biceps brachii caput longum der høyre og venstre side er slått sammen

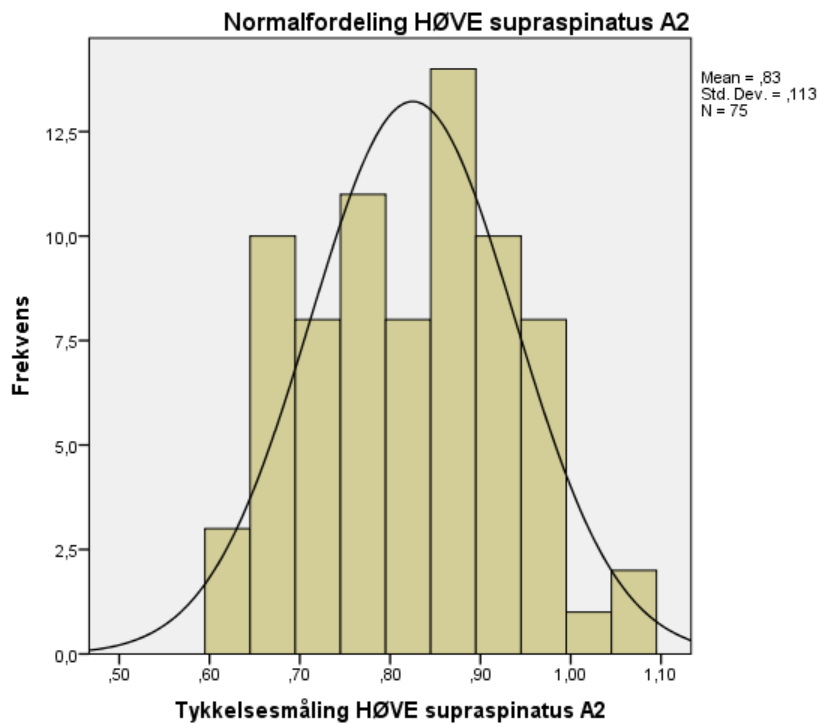
**Normalfordeling av data for supraspinatus**

	Shapiro-Wilk		
	Statistic	Frihetsgrader	Signifikansnivå
VE Supraspinatus A <sup>1</sup>	,971	37	,438
VE Supraspinatus B	,959	36	,197
VE Supraspinatus A <sup>2</sup>	,972	37	,477
HØ Supraspinatus A <sup>1</sup>	,981	38	,758
HØ Supraspinatus B	,983	37	,846
HØ Supraspinatus A <sup>2</sup>	,949	38	,081
HØVE Supraspinatus A <sup>1</sup>	,979	75	,252
HØVE Supraspinatus B	,987	73	,644
HØVE Supraspinatus A <sup>2</sup>	,972	75	,094

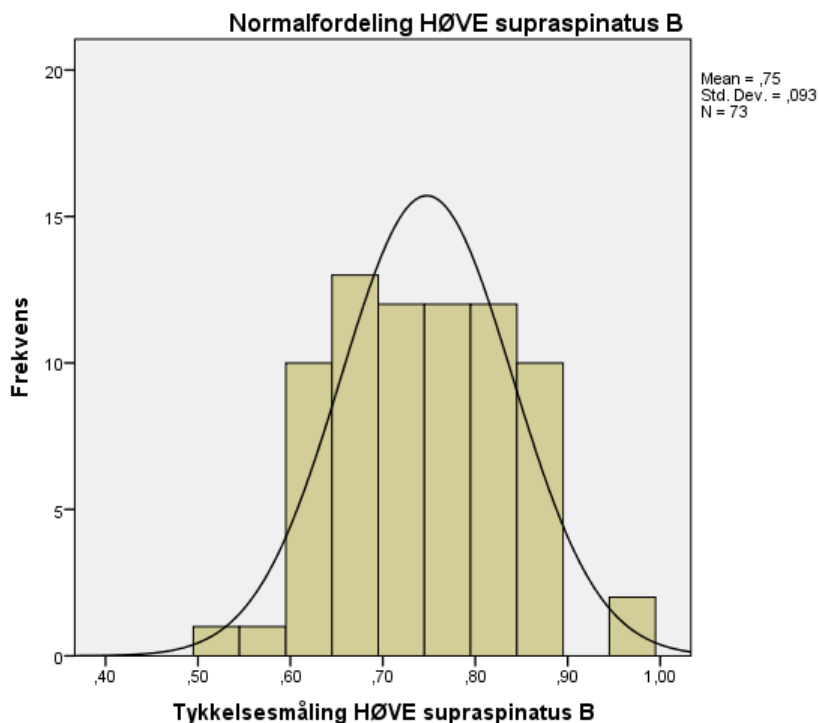
Tabell 4: Oppsummering av test for normalfordeling av data med Shapiro-Wilk for supraspinatus



Figur 7: Normalfordeling av datasett A<sup>1</sup> for supraspinatus der høyre og venstre side er slått sammen, gjennomsnitt, antall malinger og standardavvik.



Figur 8: Normalfordeling av datasett A<sup>2</sup> for supraspinatus der høyre og venstre side er slått sammen, gjennomsnitt, antall malinger og standardavvik.



Figur 9: Normalfordeling av datasett B for supraspinatus der høyre og venstre side er slått sammen, gjennomsnitt, antall malinger og standardavvik.

## 4.3 Intra-tester reliabilitet

### 4.3.1 Relativ reliabilitet biceps brachii caput longum

#### 4.3.1.1 Biceps brachii caput longum høyre og venstre side slått sammen

Resultatene for analysen av intra-tester reliabilitet med  $ICC_{1,1}$  for biceps brachii caput longum der høyre og venstre side ble slått sammen var “tilnærmet perfekt” (= 0,89) (Koch, 1982). Et tilsvarende resultat avdekket analysen av  $ICC_{3,1}$  for samme anatomiske struktur (= 0,89).

Resultatene for analysen er summert opp i tabell 5 nedenfor.

#### 4.3.1.2 Biceps brachii caput longum venstre side

Resultatene for analysen av intra-tester reliabilitet med  $ICC_{1,1}$  for biceps brachii caput longum venstre side var “tilnærmet perfekt” (= 0,94) (Koch, 1982). Videre gav analysen av intra-tester reliabiliteten med  $ICC_{3,1}$  av biceps brachii caput longum venstre side tilsvarende resultat (= 0,94). Resultatene for intra-tester reliabiliteten av biceps brachii caput longum venstre side er oppsummert i tabell 5.

#### 4.3.1.3 Biceps brachii caput longum høyre side

Analyseresultatene for intra-tester reliabiliteten målt med  $ICC_{1,1}$  for biceps brachii caput longum høyre side var også “tilnærmet perfekt” (= 0,85) (Koch, 1982). Ytterligere analyse av intra-tester reliabiliteten med  $ICC_{3,1}$  for biceps brachii caput longum høyre side viste et likt resultat (= 0,85). Resultatene for intra-tester reliabiliteten for biceps brachii caput longum er oppsummert i tabell 5.

Resultater for av analysen av intra-tester relativ reliabilitet for biceps

Biceps brachii caput longum	$ICC_{1,1}$	$ICC_{3,1}$
$A^1 - A^2$ høyre og venstre side slått sammen	0,89 (95% KI: 0,82 – 0,93)	0,89 (95% KI: 0,82 – 0,93)
$A^1 - A^2$ venstre side	0,94 (95 % KI: 0,88 – 0,97)	0,94 (95% KI: 0,88 – 0,97)
$A^1 - A^2$ høyre side	0,85 (95% KI: 0,72 – 0,92)	0,85 (95% KI: 0,72 – 0,92)

Tabell 5: Resultater for analysen av intra-tester relativ reliabilitet med  $ICC_{1,1}$  og  $ICC_{3,1}$  for biceps brachii caput longum venstre side, høyre side og høyre og venstre side slått sammen. ICC: Intraklasse korrelasjonskoeffisient, KI: Konfidensintervall

#### 4.3.2 Relativ reliabilitet supraspinatus

##### 4.3.2.1 Supraspinatus høyre og venstre side slått sammen

Resultatet for analysen av intra-tester reliabiliteten med  $ICC_{1,1}$  av supraspinatus der høyre og venstre side ble slått sammen var “betydelig” (= 0,78) (Koch, 1982). Videre analyse med  $ICC_{3,1}$  for supraspinatus der høyre og venstre side slått sammen viste et tilsvarende resultat (= 0,79). Oppsummerte resultater for intra-tester reliabilitet av supraspinatus er presentert i tabell 6.

##### 4.3.2.2 Supraspinatus venstre side

For supraspinatus venstre side viste resultatet av analysen med  $ICC_{1,1}$  en “betydelig” intra-tester reliabilitet (= 0,66). En “betydelig” intra-tester reliabilitet ble også funnet ved ytterligere analyse med  $ICC_{3,1}$  av supraspinatus på venstre side (= 0,66). Resultatene for intra-tester reliabilitet for supraspinatus venstre side er oppsummert i tabell 6.

##### 4.3.2.3 Supraspinatus høyre side

En “tilnærmet perfekt” intra-tester reliabilitet ble avdekket av analysen med  $ICC_{1,1}$  på supraspinatus høyre side (= 0,90). Et identisk resultat ble funnet ved analysen med  $ICC_{3,1}$  av høyre supraspinatus (= 0,90). En oppsummering av resultatene for intra-tester reliabiliteten for supraspinatus høyre side er presentert i tabell 6.

Resultater for analysen av intra-tester relativ reliabilitet for supraspinatus

Supraspinatus	ICC <sub>1,1</sub>	ICC <sub>3,1</sub>
A <sup>1</sup> – A <sup>2</sup> høyre og venstre side slått sammen	0,78 (95% KI: 0,68 – 0,86)	0,79 (95% KI: 0,64 – 0,87)
A <sup>1</sup> – A <sup>2</sup> venstre side	0,66 (95% KI: 0,44 – 0,81)	0,67 (95% KI: 0,44 – 0,82)
A <sup>1</sup> – A <sup>2</sup> høyre side	0,90 (95% KI: 0,81 – 0,94)	0,90 (95% KI: 0,70 – 0,96)

Tabell 6: Resultater for analysen av intra-tester relativ reliabilitet med ICC<sub>1,1</sub> og ICC<sub>3,1</sub> for supraspinatus venstre side, høyre side og høyre og venstre side slått sammen. ICC: Intraklasse korrelasjonskoeffisient, KI: Konfidensintervall

### 4.3.3 Absolutt reliabilitet

#### 4.3.3.1 Biceps brachii caput longum

Analysen av innen-subjekt standard deviasjon (Sw) mellom A<sup>1</sup> – A<sup>2</sup> for biceps brachii caput longum der høyre og venstre biceps ble slått sammen, viste et gjennomsnitt av senetykkelse på 6,1 mm (± SD = 1,3 mm). Analysen viste ytterligere at Sw = 0,4 mm, Sw \* 1,96 = 0,8 mm og Sw \* 2,77 = 1,2 mm.

Ved analyse av Sw mellom A<sup>1</sup> – A<sup>2</sup> for biceps brachii caput longum der venstre og høyre side ble delt opp hver for seg var resultatene noe høyere. Gjennomsnittsmålingene av senen for venstre og høyre side separat var identiske (= 6,0 mm), men med et noe lavere standardavvik på venstre side (= 1,2 mm). Resultatene for analysen av innen-subjekt standard deviasjon for biceps brachii caput longum er oppsummert i tabell 7.

Resultater for analysen av intra-tester absolutt reliabilitet for biceps

Biceps brachii caput longum	Gjennomsnittstykkelse	Standardavvik	Sw	Sw * 1,96	Sw * 2,77
A <sup>1</sup> – A <sup>2</sup> høyre og venstre side sammen	6,1 mm	1,3 mm	0,4 mm	0,8 mm	1,2 mm
A <sup>1</sup> – A <sup>2</sup> venstre side	6,0 mm	1,2 mm	0,5 mm	0,9 mm	1,3 mm
A <sup>1</sup> – A <sup>2</sup> høyre side	6,0 mm	1,3 mm	0,5 mm	1,0 mm	1,4 mm

Tabell 7: Oppsummering av analysen av intra-tester absolutt reliabilitet for biceps brachii caput longum, høyre og venstre side slått sammen, venstre side og høyre side. Mm = millimeter, Sw = innen-subjekt standard deviasjon.

#### 4.3.3.2 Supraspinatus

For analysen av Sw for supraspinatus mellom A<sup>1</sup> – A<sup>2</sup> der høyre og venstre side ble slått sammen viste resultatene en gjennomsnittstykkelse på 8,1 mm (± SD = 1,2 mm). Videre viste analysen at Sw = 0,6 mm, Sw \* 1,96 = 1,1 mm og Sw \* 2,77 = 1,5 mm.

Videre analyse for supraspinatus der venstre og høyre side ble analysert individuelt gav et tilsvarende resultat for høyre side. Et lavere resultat ble observert for analysen av venstre side, med en gjennomsnittstykkelse på 7,8 mm, men med en generelt høyere Sw (= 0,7 mm.) Resultatene for analysen av innen-subjekt standard deviasjon for supraspinatus er summert opp i tabell 8.

Resultater for analysen av intra-tester absolutt reliabilitet for supraspinatus

Supraspinatus	Gjennomsnittstykkelse	Standardavvik	Sw	Sw * 1,96	Sw * 2,77
<b>A<sup>1</sup> – A<sup>2</sup> høyre og venstre side sammen</b>	8,1 mm	1,2 mm	0,6 mm	1,1 mm	1,5 mm
<b>A<sup>1</sup> – A<sup>2</sup> venstre side</b>	7,8 mm	1,1 mm	0,7 mm	1,4 mm	2,0 mm
<b>A<sup>1</sup> – A<sup>2</sup> høyre side</b>	8,0 mm	1,1 mm	0,6 mm	1,1 mm	1,6 mm

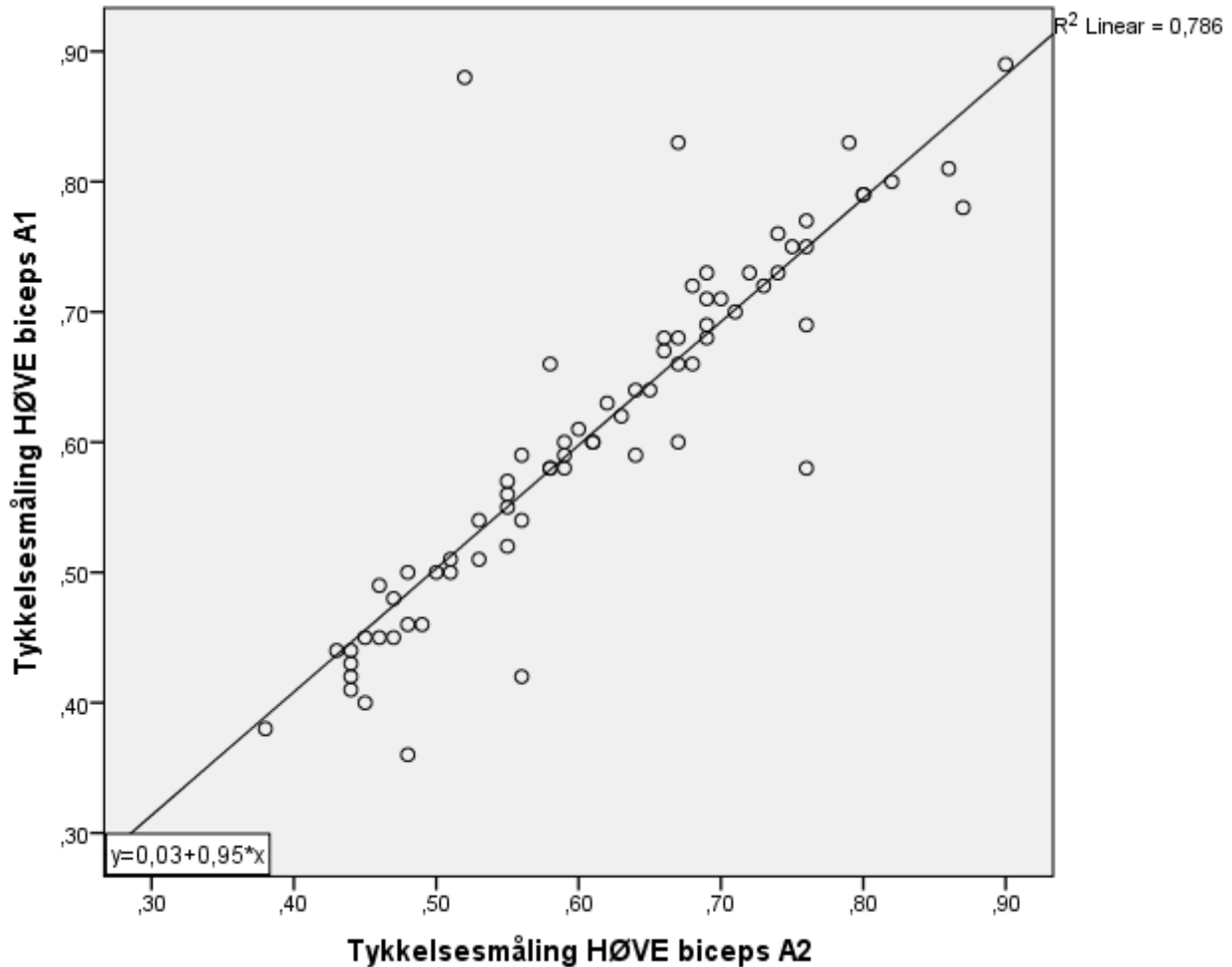
Tabell 8: Oppsummering av analysen av intra-tester absolutt reliabilitet for supraspinatus høyre og venstre side slått sammen, venstre side og høyre side. Mm = millimeter, Sw = Innen-subjekt standard deviasjon.

#### 4.3.4 Spredningsplot og Pearsons korrelasjonskoeffisient

##### 4.3.4.1 Biceps brachii caput longum

Det var en sterk positiv korrelasjon mellom A<sup>1</sup> – A<sup>2</sup> for biceps caput longum der høyre og venstre side slått sammen hvor  $r = 0,886$  ( $p < 0,005$ ). Videre var det også en sterk positiv korrelasjon mellom A<sup>1</sup> – A<sup>2</sup> for biceps brachii caput longum venstre side hvor  $r = 0,946$  ( $p < 0,005$ ). Det ble også identifisert en sterk positiv korrelasjon mellom A<sup>1</sup> – A<sup>2</sup> av biceps brachii caput longum høyre side hvor  $r = 0,843$  ( $p < 0,005$ ). Spredningsplot for datasett A<sup>1</sup> og A<sup>2</sup> for biceps brachii caput longum der høyre og venstre side er slått sammen er presentert i Figur 9.

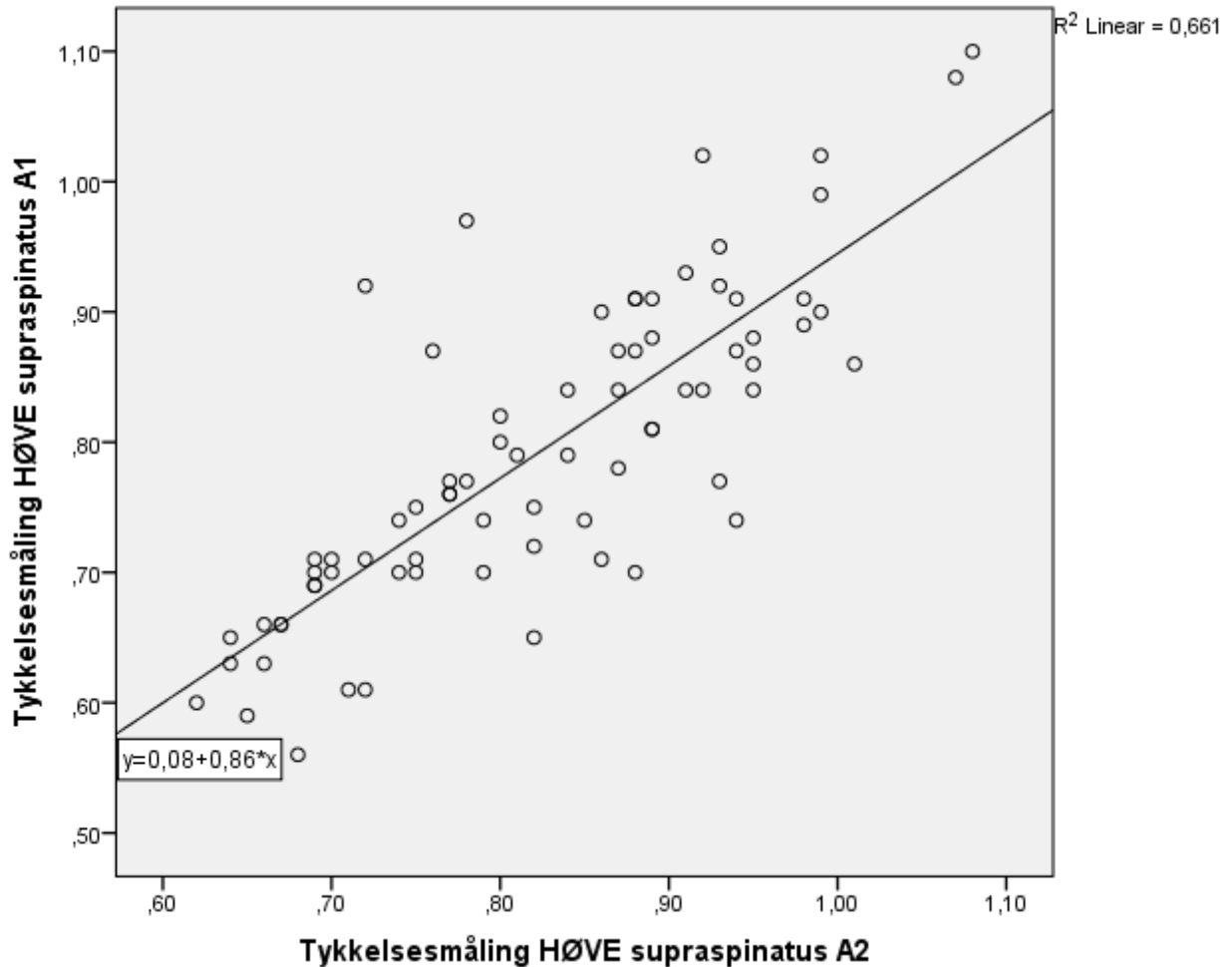




Figur 10: Spredningsplot for datasett A<sup>1</sup> og A<sup>2</sup> for biceps brachii caput longum der høyre og venstre side er slått sammen.

#### 4.3.4.2 Supraspinatus

Det ble funnet en sterk positiv korrelasjon mellom A<sup>1</sup> – A<sup>2</sup> for supraspinatus der høyre og venstre side ble slått sammen hvor  $r = 0,813$  ( $p < 0,005$ ). Et liknende resultat ble funnet for analysen mellom A<sup>1</sup> – A<sup>2</sup> av supraspinatus på venstre side hvor  $r = 0,691$  ( $p < 0,005$ ). En sterk positiv korrelasjon ble også funnet mellom A<sup>1</sup> – A<sup>2</sup> for supraspinatus på høyre side hvor  $r = 0,925$  ( $p < 0,005$ ). Spredningsplot for datasett A<sup>1</sup> og A<sup>2</sup> for supraspinatus der høyre og venstre side er slått sammen er presentert i Figur 10.



Figur 11: Spredningsplot for datasett A<sup>1</sup> og A<sup>2</sup> for supraspinatus der høyre og venstre side er slått sammen

## 4.4 Inter-tester reliabilitet

### 4.4.1 Relativ reliabilitet biceps brachii caput longum

#### 4.4.1.1 Biceps brachii caput longum høyre og venstre side slått sammen

Resultatene av analysen med ICC<sub>1,1</sub> for biceps brachii caput longum høyre og venstre side slått sammen gav en “betydelig” inter-tester reliabilitet (= 0,64). Et tilsvarende resultat ble avdekket av videre analyse med ICC<sub>3,1</sub> (= 0,64). Resultatene for analysen av inter-tester reliabilitet er presentert i tabell 7.

#### 4.4.1.2 Biceps brachii caput longum gjennomsnitt av A<sup>1</sup> og A<sup>2</sup> – B

En “betydelig” inter-tester reliabilitet fremkom av analysen med ICC<sub>1,1</sub> mellom gjennomsnittet av A<sup>1</sup> og A<sup>2</sup> – B for biceps brachii caput longum der høyre og venstre side ble slått sammen (= 0,64). Et liknende resultat ble avdekket av analysen med ICC<sub>3,1</sub> (= 0,65). Resultatene for inter-tester reliabilitet for biceps brachii caput longum er summert opp i tabell 9.

#### 4.4.1.3 Biceps brachii caput longum venstre side

En “tilnærmet perfekt” inter-tester reliabilitet ble avdekket ved analyse av  $ICC_{1,1}$  for biceps brachii caput longum venstre side (= 0,81). Et tilsvarende resultat ble avdekket av videre analyse med  $ICC_{3,1}$  (= 0,82). Resultatene for inter-tester reliabilitet for biceps brachii caput longum venstre side er summert opp i tabell 9.

#### 4.4.1.4 Biceps brachii caput longum høyre side

Resultatene av analysen ved bruk av  $ICC_{1,1}$  gav en “tilnærmet perfekt” inter-tester reliabilitet (= 0,93). Tilsvarende resultat ble resultatet av analysen med  $ICC_{3,1}$  mellom  $A^1 - B$  av biceps brachii caput longum høyre side (= 0,93). En oppsummering av resultatene for inter-tester reliabilitet for biceps brachii caput longum høyre side er presentert i tabell 9.

Resultater for analysen av inter-tester relativ reliabilitet for biceps

Biceps brachii caput longum	$ICC_{1,1}$	$ICC_{3,1}$
$A^1 - B$ høyre og venstre side slått sammen	0,64 (95% KI: 0,47 – 0,76)	0,64 (95% KI: 0,47 – 0,77)
Gjennomsnitt av $A^1$ og $A^2 - B$ høyre og venstre side slått sammen	0,64 (95% KI: 0,48 – 0,76)	0,65 (95% KI: 0,47 – 0,77)
$A^1 - B$ venstre side	0,81 (95% KI: 0,66 – 0,90)	0,82 (95% KI: 0,39 – 0,93)
$A^1 - B$ høyre side	0,93 (95% KI: 0,86 – 0,96)	0,93 (95% KI: 0,80 – 0,97)

Tabell 9: Resultater for analysen av inter-tester relativ reliabilitet med  $ICC_{1,1}$  og  $ICC_{3,1}$  for biceps brachii caput longum venstre side, høyre side, høyre og venstre side slått sammen og gjennomsnittet av  $A^1$  og  $A^2$  der høyre og venstre side ble slått sammen. ICC: Intraklasse korrelasjonskoeffisient, KI: Konfidensintervall

#### 4.4.2 Relativ reliabilitet supraspinatus

##### 4.4.2.1 Supraspinatus høyre og venstre side slått sammen

Analysen for inter-tester reliabilitet med  $ICC_{1,1}$  mellom  $A^1 - B$  for supraspinatus der høyre og venstre side var slått sammen viste et “moderat” resultat (= 0,52) (Koch, 1982). Videre analyse med  $ICC_{3,1}$  for samme anatomiske struktur og variabler resulterte i et sammenliknbart resultat (= 0,54). Resultatene av inter-tester reliabilitet for supraspinatus der høyre og venstre side ble slått sammen er oppsummert i tabell 10.

##### 4.4.2.2 Supraspinatus $A^1$ og $A^2 - B$ , gjennomsnitt av målinger

En “moderat” inter-tester reliabilitet ble avdekket av analysen med  $ICC_{1,1}$  av supraspinatus mellom gjennomsnittet av  $A^1$  og  $A^2 - B$  der høyre og venstre side ble slått sammen (= 0,54). Et tilsvarende resultat viste analysen av inter-tester reliabilitet med  $ICC_{3,1}$  (= 0,54). I tabell 10 foreligger en oppsummering av analyseresultatene for inter-tester reliabilitet av supraspinatus.

#### 4.4.2.3 Supraspinatus venstre side

En “betydelig” inter-tester reliabilitet ble avdekket av analysen med  $ICC_{1,1}$  mellom  $A^1 - B$  ved supraspinatus venstre side (= 0,63). Videre analyse av inter-tester reliabilitet med  $ICC_{3,1}$  avdekket et liknende resultat (= 0,66). Resultatene for inter-tester reliabilitet for supraspinatus venstre side er oppsummert i tabell 10.

#### 4.4.2.4 Supraspinatus høyre side

Analysen av inter-tester reliabilitet med  $ICC_{1,1}$  av supraspinatus høyre side gav en “betydelig” reliabilitet (= 0,74). Et liknende resultat ble avdekket ved analyse med  $ICC_{3,1}$  for supraspinatus høyre side (= 0,75). Resultatene for inter-tester reliabilitet for supraspinatus høyre side er summert opp i tabell 10.

Resultater for analysen av inter-tester relativ reliabilitet for supraspinatus

Supraspinatus	$ICC_{1,1}$	$ICC_{3,1}$
$A^1 - B$ høyre og venstre side slått sammen	0,52 (95% KI: 0,33 – 0,67)	0,54 (95% KI: 0,30 – 0,71)
Gjennomsnitt av $A^1$ og $A^2 - B$ høyre og venstre side slått sammen	0,54 (95% KI: 0,16 – 0,50)	0,54 (95% KI: 0,16 – 0,74)
$A^1 - B$ venstre side	0,63 (95% KI: 0,39 – 0,79)	0,66 (95% KI: 0,22 – 0,84)
$A^1 - B$ høyre side	0,74 (95% KI: 0,55 – 0,86)	0,75 (95% KI: 0,47 – 0,88)

Tabell 10: Resultater for analysen av inter-tester relativ reliabilitet med  $ICC_{1,1}$  og  $ICC_{3,1}$  for supraspinatus venstre side, høyre side, høyre og venstre side slått sammen og gjennomsnitt av  $A^1$  og  $A^2$  der høyre og venstre side ble slått sammen. ICC: Intraklasse korrelasjonskoeffisient, KI: Konfidensintervall

#### 4.4.3 Absolutt reliabilitet

##### 4.4.3.1 Biceps brachii caput longum

Resultatene for analysen av innen-subjekt standard deviasjon (Sw) mellom  $A^1 - B$  for biceps brachii caput longum der høyre og venstre side ble slått sammen gav en gjennomsnittsmåling på 5,9 mm ( $\pm$  SD = 1,3 mm). Analyseresultatene viste videre at  $Sw = 0,5$  mm,  $Sw * 1,96 = 0,9$  mm og  $Sw * 2,77 = 1,3$  mm.

Det ble også gjennomført analyse av Sw for biceps brachii caput longum på høyre og venstre side individuelt. Disse analysene avdekket et generelt lavere resultat for høyre side ( $Sw = 0,4$ ), og et høyere resultat for venstre side ( $Sw = 0,6$ ). En oppsummering av resultatene for innen-subjekt standard deviasjon er presentert i tabell 11.

Resultater for analysen av inter-tester absolutt reliabilitet for biceps

Biceps brachii caput longum	Gjennomsnitt	Standardavvik	Sw	Sw * 1,96	Sw * 2,77
A <sup>1</sup> – B høyre og venstre side sammen	5,9 mm	1,3 mm	0,5 mm	0,9 mm	1,3 mm
A <sup>1</sup> – B venstre side	6,0 mm	1,2 mm	0,6 mm	1,1 mm	1,5 mm
A <sup>1</sup> – B høyre side	6,0 mm	1,4 mm	0,4 mm	0,7 mm	1,0 mm

Tabell 11: Oppsummering for analysen av inter-tester absolutt reliabilitet for biceps brachii caput longum høyre og venstre side slått sammen, venstre side og høyre side. Mm = millimeter, Sw = innen-subjekt standard deviasjon.

#### 4.4.3.2 Supraspinatus

Analysen av innen-subjekt standard deviasjon mellom A<sup>1</sup> – B for supraspinatus der høyre og venstre side ble slått sammen, gav en gjennomsnittstykkelse på 7,7 mm ( $\pm$  SD = 1,1 mm).

Resultatene av analysen viste ytterligere at Sw = 0,6 mm, Sw \* 1,96 = 1,2 mm og Sw \* 2,77 = 1,7 mm.

Subgruppeanalyser av Sw for supraspinatus på høyre og venstre side separat ble gjennomført.

Resultatene for disse analysene avdekket tilsvarende resultater for høyre side (Sw = 0,6 mm) og noe høyere resultater for venstre side (Sw = 0,7). En oppsummering av resultatene for innen-subjekt standard deviasjon mellom A<sup>1</sup> – B for supraspinatus er presentert i tabell 12.

Resultater for analysen av inter-tester absolutt reliabilitet for supraspinatus

Supraspinatus	Gjennomsnitt	Standardavvik	Sw	Sw * 1,96	Sw * 2,77
A <sup>1</sup> – B høyre og venstre side sammen	7,7 mm	1,1 mm	0,6 mm	1,2 mm	1,7 mm
A <sup>1</sup> – B venstre side	7,6 mm	1,1 mm	0,7 mm	1,3 mm	1,9 mm
A <sup>1</sup> – B høyre side	7,8 mm	1,1 mm	0,6 mm	1,1 mm	1,6 mm

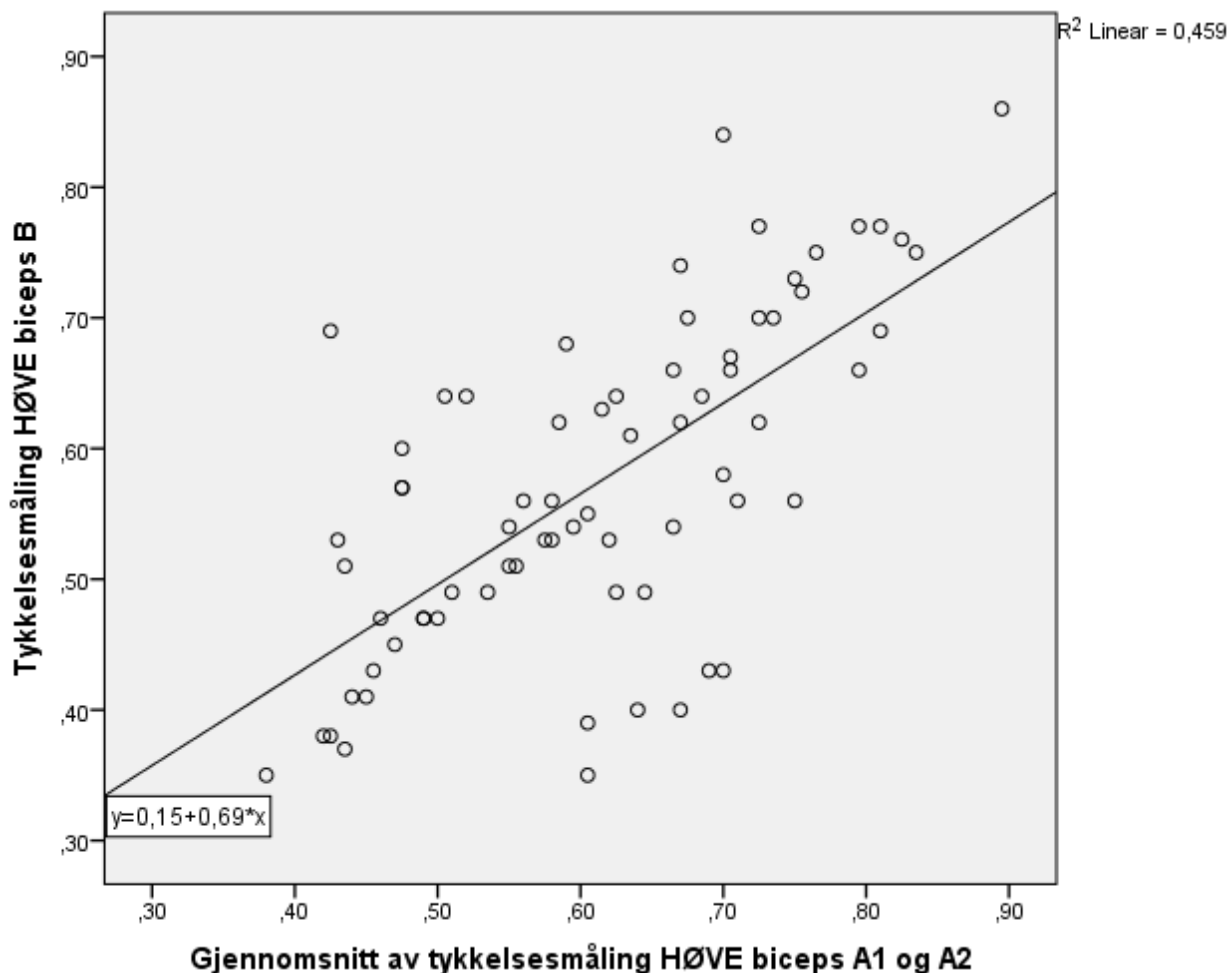
Tabell 12: Oppsummering for analysen av inter-tester absolutt reliabilitet for supraspinatus høyre og venstre side slått sammen, venstre side og høyre side. Mm = millimeter, Sw = innen-subjekt standard deviasjon.

#### 4.4.4 Spredningsplot og Pearsons korrelasjonskoeffisient

##### 4.4.4.1 Biceps brachii caput longum

En sterk positiv korrelasjon ble funnet mellom A<sup>1</sup> – B av biceps brachii caput longum på høyre og venstre side slått sammen hvor  $r = 0,664$  ( $p < 0,005$ ). Ytterligere ble det avdekket en sterk positiv korrelasjon mellom gjennomsnittet av A<sup>1</sup> og A<sup>2</sup> – B for biceps brachii caput longum der høyre og venstre side var slått sammen hvor  $r = 0,678$  ( $p < 0,005$ ). Mellom A<sup>1</sup> – B for biceps brachii caput longum på høyre, og venstre side, ble det også avdekket en sterk positiv korrelasjon hvor  $r = 0,946$  ( $p < 0,005$ ) og  $r = 0,883$  ( $p < 0,005$ ), respektivt. Spredningsplot for gjennomsnittet av

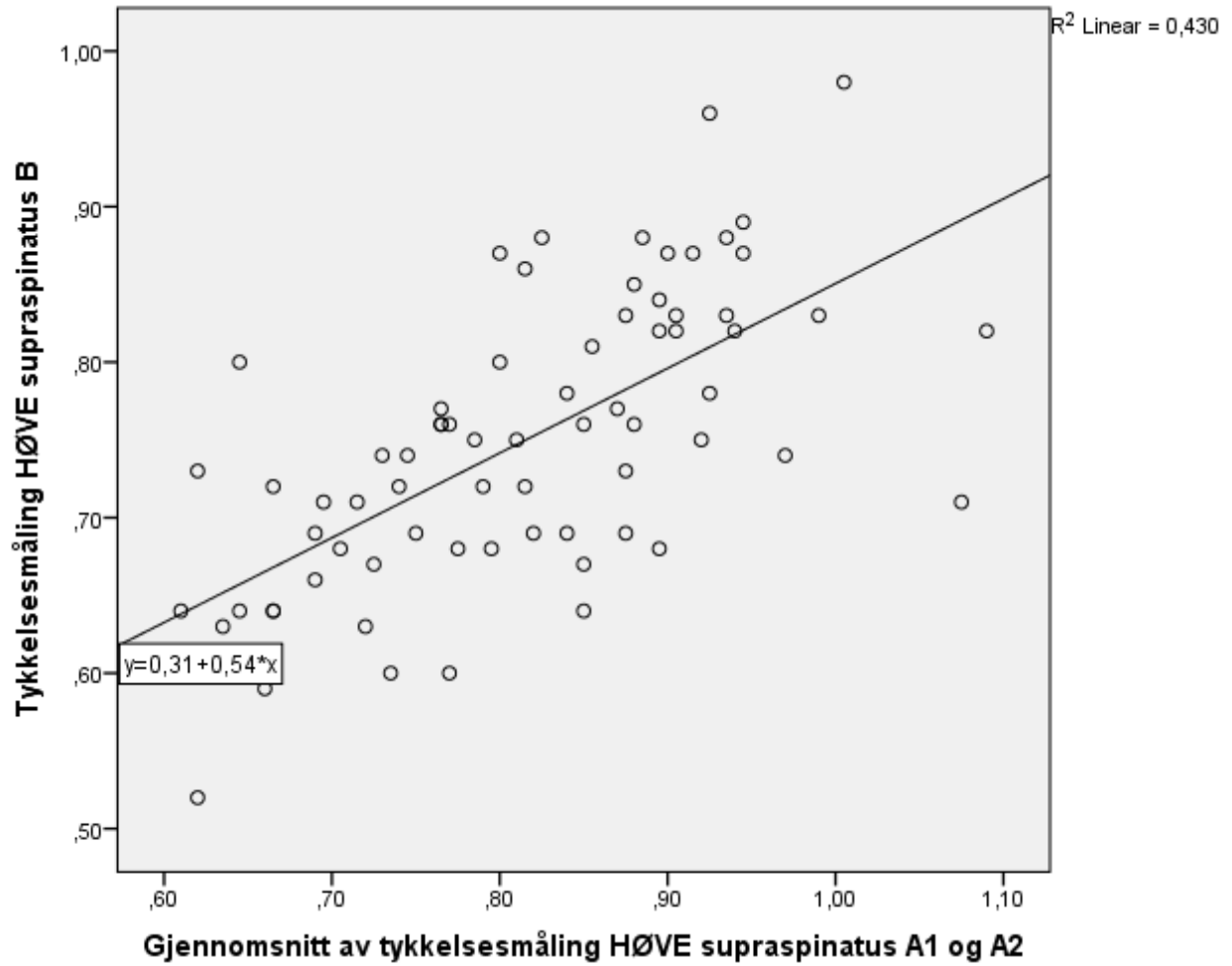
datasett A<sup>1</sup> og A<sup>2</sup> for biceps brachii caput longum der høyre og venstre side er slått sammen og datasett B er presentert i Figur 11.



Figur 12: Spredningsplot for datasett B, og gjennomsnittet av A<sup>1</sup> og A<sup>2</sup> for biceps brachii caput longum der høyre og venstre side er slått sammen

#### 4.4.4.2 Supraspinatus

Mellom A<sup>1</sup> - B for supraspinatus, der høyre og venstre side ble slått sammen, viste analysen en sterk positiv korrelasjon hvor  $r = 0,618$  ( $p < 0,005$ ). En sterk positiv korrelasjon ble også identifisert mellom gjennomsnittet av A<sup>1</sup> og A<sup>2</sup> - B for supraspinatus der høyre og venstre side ble slått sammen hvor  $r = 0,656$  ( $p < 0,005$ ). Analysen avdekket videre en sterk positiv korrelasjon mellom A<sup>1</sup> - B ved supraspinatus høyre side, og venstre side hvor  $r = 0,821$  ( $p < 0,005$ ) og  $r = 0,779$  ( $p < 0,005$ ), respektivt. Spredningsplot for datasett B, og gjennomsnittet av A<sup>1</sup> og A<sup>2</sup> for supraspinatus der høyre og venstre side er slått sammen er presentert i Figur 12.



Figur 13: Spredningsplot for datasett B og gjennomsnittet av  $A^1$  og  $A^2$  for supraspinatus der høyre og venstre side er slått sammen

## 5. Diskusjon

### 5.1 Metode

I denne metodediskusjonen vil det bli diskutert hvilke metodiske forhold som kan tenkes å ha påvirket resultatene i min studie. Diskusjonen vil være relatert til valg av forskningsdesign, utvalget og ekstern validitet, målemetoden, datainnsamling og standardisering, variabilitet og testerne.

#### 5.1.1 Valg av forskningsdesign

Hensikten med denne studien var å undersøke reliabiliteten av UL-målinger av senetykkelse i en populasjon med skuldertendinopati. For å svare på problemstillingen ble det i denne studien valgt å foreta en reliabilitetsstudie med et ikke-eksperimentelt design. Reliabilitet av UL vil være et premiss for å si noe om validiteten (i hvilken grad resultatene er gyldige), selv om det ikke nødvendigvis er en direkte sammenheng mellom de (Carter et al., 2011).

I klinisk praksis brukes UL stadig hyppigere av manuellterapeuter hos pasienter med skulderplager. I en klinisk hverdag skjer det til stadighet at man foretar flere målinger av samme pasient for å se eventuell fremgang i forbindelse med behandling. Det skjer også til stadighet at man spør en kollega om tolkning av et UL bilde for å etterprøve sine egne resultater. For å gjenspeile klinisk praksis ble det, i denne studien, ansett som hensiktsmessig å undersøke både intra-, og inter-tester reliabiliteten av UL.

For å vurdere designet og gjennomføringen av denne studien, kan det være hensiktsmessig å sammenlikne med tidligere reliabilitetsstudier. Det er gjort relativt lite tidligere forskning på intra-, og inter-tester reliabiliteten på målinger med UL av senetykkelse i skulderen.

I 2003 undersøkte Bjordal og kolleger blant annet intra-tester reliabiliteten ved måling med UL av senetykkelsen til supraspinatus (Bjordal et al., 2003). En potensiell svakhet med overnevnte studie er at kun absolutt reliabilitet ble brukt i analysen av intra-tester reliabilitet. Litteratur som omhandler reliabilitetsstudier anbefaler at man presenterer både absolutt og relativ reliabilitet (Carter et al., 2011, s. 240; Kottner et al., 2011). Dette er anbefalt av flere grunner. Først fordi de to analysene gir utfyllende informasjon i forhold til hverandre. Videre fordi det ikke finnes en universal enighet om graderingen av resultatene, eller hvilken statistisk analyse, man skal bruke



ved bedømmingen av relativ reliabilitet (Bartko and Carpenter, 1976; Hart, 1989; Carter et al., 2011). På en annen side vil resultatene av den absolutte reliabiliteten i større grad være overførbare til klinisk praksis på bakgrunn av at resultatene blir oppgitt i den måleenheten målemetoden har (Carter et al., 2011, s. 240). I motsetning til min studie var alle deltakerne i studien til Bjordal og kolleger friske. Dette kan tenkes å være en svakhet da det anbefales å måle reliabiliteten av et klinisk måleverktøy på den populasjonen som den ville blitt anvendt på i klinisk praksis (Carter et al., 2011). Det kan problematiseres om studien til Bjordal burde ha inkludert en deltakergruppe med skulderplager. En ytterligere svakhet kan være at ble det brukt en transduserfrekvens på 7,5 MHz ved målingene med UL (Bjordal et al., 2003). I min studie ble målingene foretatt med en transduserfrekvens på 10 MHz. Dette fremstår mer i tråd med hva som anbefales i målingen med UL av skuldresener (Jacobson, 2003).

I 2012 undersøkte Leong et al. blant annet intra-tester reliabiliten på tykkelsesmålinger med UL på supraspinatussenen, hos symptomfrie og utrente individer (N = 11) (Leong et al., 2012). I likhet med min studie ble senetykkelsen målt bilateralt. I motsetning til min studie indikerer studien å bruke biceps brachii caput longum som referansepunkt for tykkelsesmålingene til supraspinatus. Dette anser jeg som en potensiell svakhet med studien, da fravær av ossøse referansepunkt kan gi økt risiko for målefeil (Duerr, 2010). I studien til Leong et al. ble senetykkelsen av supraspinatus målt 10 mm, 20 mm og 30 mm fra referansepunktet, og gjennomsnittet av disse dannet grunnlaget for analysen (Leong et al., 2012). Dette er en potensiell grunn til målefeil ved min studie, da jeg kun gjennomført én måling. Målefeil tilskrevet denne faktoren ble forsøkt minimert i min studie ved å bruke et ossøst landemerke som referansepunkt til tykkelsesmålingene.

Studien til Leong og kolleger oppgir utmerkede resultater for både relativ reliabilitet (ICC = 0,92) og absolutt reliabilitet (Sw = 0,23 mm). Jeg anser det som en styrke ved denne, og min studie, at begge typer reliabilitet blir presentert, da dette anbefales ved reliabilitetsstudier (Rankin and Stokes, 1998; Kottner et al., 2011). En mulig svakhet med studien til Leong et al. er at de statistiske analysene for relativ reliabilitet kun ble gjort med ICC<sub>3,1</sub>. Som nevnt, tar denne analysen ikke høyde for systematiske feil, og anses derfor som en mulig kilde til målefeil. For å ta høyde for dette i min studie, ble det derfor besluttet å gjøre både ICC<sub>1,1</sub> og ICC<sub>3,1</sub>.

I likhet med min studie ble tykkelsesmålingene utført av en fysioterapeut. Fysioterapeuten i studien til Leong et al. gjennomgikk UL kurs og intensiv trening i tre måneder på gjennomføringen av målingene med en erfaren UL tester. For å øke min erfaring med bruken av UL, kunne treningsperioden min før gjennomføringen av målingene vært lenger. For denne studien utførte jeg ikke selve scanningen av UL bildene, men målte kun senetykkelsen på eksisterende bilder. Resultatene fra min studie indikerer et mindre krav til erfaring ved denne begrensede oppgaven innen bruk av UL.

I 2014 ble intra-, og inter-tester reliabiliteten av UL målinger for tykkelsen av supraspinatus undersøkt av Temes og kolleger (Temes et al., 2014). Målingene av supraspinatus ble gjort i kontrahert og ikke-kontrahert tilstand. I likhet med min studie, presenterer studien både relativ og absolutt reliabilitet. Resultatene for både intra-, og inter-tester reliabiliteten omtales som ICC > 0.9 med tilhørende innen-subjekt standardavvik fra 0.07 – 0.12 mm.

En mulig svakhet i studien var at de inkluderte kun friske deltakere (N = 15) (Temes et al., 2014). Sammenliknet med min studie var oppholdet mellom testingen for intra-tester reliabilitet lenger. I studien til Temes gikk det en uke mellom målingene, mens det i min studie kun gikk én dag. Det kan tenkes at det ville vært hensiktsmessig med et lengre testintervall i min studie. Dette øker sjansen for at man ikke husker resultatene fra første måling, som kan påvirke bedømmingen av andre måling. I min studie var dette ikke mulig grunnet tidsmessige årsaker. Videre anses det som usannsynlig at man husker nøyaktige individuelle mål av omtrentlig 150 sener som ble målt i min studie.

En nylig publisert studie undersøkte intra-, og inter-tester reliabiliteten av UL målinger av tykkelsen av supraspinatussen hos både symptomatiske (N = 45) og ikke-symptomatiske (N = 25) deltakere (McCreesh et al., 2015a). En styrke med studien, i likhet med min studie, er at både relativ og absolutt relativ reliabilitet blir presentert. Ytterligere ble ICC<sub>2,1</sub> anvendt som statistisk analyse. Denne analysen tar både hensyn til systematiske og tilfeldige feil, noe som anbefales ved reliabilitetsstudier (Weir, 2005; Carter et al., 2011).

Videre kan det tenkes å være en styrke at studien til McCreesh et al. undersøkte både friske og syke deltakere, som, i større grad, gjenspeiler klinisk praksis. I min studie hadde alle inkluderte deltakere skuldertendinopati, men det ble gjennomført målinger på både friske og syke

skuldereener, noe jeg anser som en styrke. I tykkelsesmålingene av supraspinatussenen ble det i studien til McCreesh et al. gjennomført to målinger per sene. I beregningene av reliabilitet ble gjennomsnittet av disse to målingene brukt. Dette er en mulig styrke ved studien, som kunne vært gjort i min studie, da senen har en varierende tykkelse i sitt forløp (Figur 2) (McCreesh et al., 2015a). I min studie ble denne potensielle feilkilden forsøkt minimert ved å utarbeide en standardisert protokoll for tykkelsesmålingene.

I målingene av tykkelsen til supraspinatus blir det beskrevet at biceps caput longum ble brukt som landemerke i studien til McCreesh (McCreesh et al., 2015a). Dersom man ikke forankrer målingene mot et ossøst landemerke, kan dette tenkes å være en mulig kilde til målefeil av senetykkelsen (Duerr, 2010). I min studie ble knekkpunktet mellom den humerale leddflate og basis av tuberculum majus brukt som ossøst landemerke (Figur 2). Dette er tilsvarende hva som har blitt gjort i tidligere studier (Bjordal et al., 2003).

For å forsøke å standardisere målingene i min studie, ble det opprettet en målingsprotokoll som ble gjennomgått av begge målere på forhånd. Dette ble også utført i studien til McCreesh og kolleger (McCreesh et al., 2015a). En ytterligere potensiell svakhet med studien er, som problematisert tidligere, en varierende oppgitt transduserfrekvens, som blir oppgitt til 7-12 MHz, men ikke utdypet videre.

### **5.1.2 Utvalg og ekstern validitet**

Utvalget av studiens deltakere, og hvordan de blir utvalgt, vil i stor grad påvirke studiens validitet. Videre vil det inkluderte utvalget påvirke den eksterne validiteten av resultatene (Carter et al., 2011, s. 92). Med ekstern validitet menes i hvilken grad resultatene kan generaliseres. Når man utfører reliabilitetsstudier er det hensiktsmessig at man inkluderer en populasjon som måleinstrumentet er tiltenkt brukt på i klinikken (Carter et al., 2011).

I min studie var målingene med UL basert på deltakere fra et pågående doktorgradsprosjekt i Rogaland. Deltakerne var menn og kvinner med vedvarende skulderplager, som ble rekrutert fra ventelistene til kirurgisk avdeling ved Stavanger Universitetssykehus. Som nevnt tidligere måtte deltakerne oppfylle inklusjon, og eksklusjonskriterier for deltakelse (punkt 3.2.1 og 3.2.2). Klare inklusjon, og eksklusjonskriterier er beskrevet som en nødvendighet for høy ekstern validitet ved kliniske forsøk (Carter et al., 2011). Deltakerne i min studie regnes som representative

demografisk, til den tiltenkte pasientgruppen med skuldertendinopati der UL blir brukt i klinikken.

For å estimere en relativt presis reliabilitet har det blitt beskrevet at det er ønskelig med omkring 50 deltakere, og minst tre målinger (Hopkins, 2000). På en annen side, indikerer tidligere forskning at to målinger er mest anvendt (Bjordal et al., 2003; Temes et al., 2014; McCreesh et al., 2015a). Videre vil konfidensintervallet for resultatene av ICC gi informasjon om spredningen i den inkluderte populasjonen. Dersom det er inkludert for få deltakere i en studie, kan dette føre til tilfeldige utslag og svake ICC resultater med stor spredning i konfidensintervallet (Batterham and George, 2003). Dersom jeg hadde kunnet inkludere ytterligere deltakere og målinger, kunne jeg økt robustheten av resultatene mine. Dette lot seg ikke gjøre av tidsmessige årsaker.

I min studie ble målingene basert på UL bildene til 27 deltakere fra doktorgradsarbeidet, og det ble foretatt to målinger. Noen av disse deltakerne var blitt scannet ved flere anledninger, slik at totalt antall bilder var mer enn 27. Dette kan være en potensiell svakhet ved studien. Det kan problematiseres at man kan kjenne igjen tidligere målte sener, som videre vil påvirke målingen. Det ble valgt og inkludere bildene fra deltakere som ble scannet flere ganger da hver sene ble bedømt individuelt og derfor vil være høyst usannsynlig å identifisere samme sene. Der det ble foretatt flere målinger av samme deltaker, kan det problematiseres at mulig sykdomsprogresjon vil kunne påvirke resultatene. Dette anses som lite trolig, da målingene ble gjennomført på både symptomatisk og asymptomatisk side, og jeg kun målte tykkelsen på senen uten kunnskap om seneaffeksjon.

Alle inkluderte deltakere ble scannet på både venstre og høyre side, uavhengig av hvilken side som var symptomatisk. I denne sammenhengen har jeg ikke skilt mellom frisk og syk side, men tidligere studier har vist en økt tykkelse på symptomatisk side (Joensen et al., 2009). Mine resultater har en relativt lik spredning over hele skalaen, og derved relativt lik spredning også med økende verdier av tykkelsen på senene (figur 10-13). Dette støtter validiteten til mitt valg om å inkludere både syke og friske sener. Det kunne vært interessant å undersøke i hvilken grad mine resultater kan generaliseres til sener på andre anatomiske lokasjoner. På samme måte svekkes generaliserbarheten i min studie av et lavere antall deltakere enn det som er vanlig i reliabilitetsstudier (Hopkins, 2000).

### 5.1.3 Testere

Det var to målere i min studie. En måler hadde minimal erfaring i bruk av UL (tester 1), og en har nesten 20 års erfaring på området (tester 2). En mulig svakhet med studien min var at intra-tester reliabiliteten ble analysert basert på målingene gjort av tester 1. Tidligere studier har vist større grad av reliabilitet på UL målinger dersom målingene blir gjennomført av en erfaren tester (Hebert et al., 2009; Ok et al., 2013). Dersom tester 2 hadde gjennomført målingene i forbindelse med intra-tester reliabiliteten, kan det tenkes at dette kunne styrket resultatene mine. På tross av dette ble det observert en “tilnærmet perfekt” intra-tester reliabilitet i min studie (Koch, 1982). Basert på tester 1 sin minimale erfaring vil man kunne forvente en læringskurve mellom test og re-test (Moosmayer and Smith, 2005). På tross av dette indikerer observasjon av datamaterialet mitt at den første målingen til tester 1 var likest målingen til tester 2. Det foreligger ingen åpenbar forklaring på dette. En potensiell faktor som kan hatt påvirkning var at måling 2 ble utført under et noe større tidspress enn måling 1. Dette kan ha påvirket presisjonen i negativ retning. Med bakgrunn i dette ble det derfor valgt å gjennomføre analyser av inter-tester reliabiliteten på datasettene A<sup>1</sup> – B.

### 5.1.4 Variabilitet, standardisering og målemetoder

Med variabilitet menes i hvilken grad det er spredning i dataene (Carter et al., 2011, s. 261). I min studie vil potensielle kilder til variabilitet stamme fra feil ved ultralydsapparatet, individuelle biologiske variasjoner hos deltakerne, eller feil hos testerne.

Det er lite trolig at det har forekommet feil ved ultralydsapparatet. Apparatet ble kalibrert av en erfaren bruker av UL før scanningene ble utført, uten å finne feil. Transduserfrekvensen brukt i scanningen kan være en potensiell kilde til variabilitet. Det ble brukt en lineær probe med transduserfrekvens på 10 MHz, som er anbefalt i litteraturen ved scanning av de superfisielle senene i skulderen (Jacobson, 2003).

En annen kilde til variabilitet ved bruk av UL, er biologiske variasjoner hos enkeltpersoner. I min studie ble det ikke fokusert på forholdet mellom reliabiliteten og kroppsmasseindeks (KMI). Dette kunne vært gjort, og er en potensiell svakhet, i min studie. Særlig på bakgrunn i tidligere litteratur som indikerer at det er essensielt å vite noe om avstanden fra hudoverflate til sene for å estimere energitap, slik at man kan justere transduserfrekvensen deretter (Bureau and Roederer, 1998; Stolik et al., 2000). Videre har det blitt påpekt at avstanden fra hudoverflaten til senen til

supraspinatus er doblet hos personer som er moderat overvektige (27-30 KMI) (Bjordal et al., 2003). På tross av dette ble det observert en relativt lik spredning av resultatene fra målingene på tross av økt senetykkelse (figur 10-13). Dersom deltakeren som skal scannes har høy KMI, kan det forsvares å bruke en transduserfrekvens på 7,5 MHz for å fremstille senen bedre (Jacobson, 2003). Ultralydbølgen er omvendt proporsjonal med vevspenetrasjonen, og frekvensen er direkte proporsjonal med den aksiale oppløsningen (Jacobson, 2012). Med andre ord betyr dette at man bruker en lavere transduserfrekvens på dypere strukturer, og en høyere transduserfrekvens på mer superfisielle strukturer (Moosmayer et al., 2004).

Tidsintervallet mellom intra-testermålingene vil kunne være en kilde til variabilitet. Det har blitt anbefalt at tidsintervallet mellom målinger skal være kort nok til å minimere sjansen for en læringseffekt, men lang nok til at måleren ikke husker de opprinnelige målene (Blacker and Endicott, 2002). Mellom test og retest i min studie ble gjennomført med én dags mellomrom. Denne tidsperioden anses som tilstrekkelig, og vil derfor være en usannsynlig kilde til stor variabilitet.

For å identifisere hvert individuelle bilde av senene ble pasient-ID og klokkeslettet for scanning brukt. Dette er en potensiell svakhet med studien, da det kan tenkes at man husker tidligere målinger. Det ville trolig styrket studiens metodologi ved å blinde testerene som utførte målingene for intra-tester reliabilitet for denne informasjonen. Dette kunne blitt gjort ved å bruke en ekstern person til å hente frem bildene som skulle måles, slik at testerene ble blindet for pasient-ID og klokkeslett. Grunnet ressurs-, og tidsproblemer ble dette ikke gjennomført. På en annen side anses det som vanskeligere for testerene å huske et nummer, sammenliknet med dersom bildene ble identifisert med navn. Videre er det usannsynlig at man er i stand til å huske ett bilde når det totalt ble målt omtrent 150 bilder.

Testerens erfaring med bruk av det verktøyet som man undersøker reliabiliteten av vil kunne påvirke variabiliteten (O'Connor et al., 2005; Carter et al., 2011). Testernes evne til å måle senetykkelsen på ultralydapparatet vil trolig være en av de største årsakene til variabilitet i min studie. For å redusere variabiliteten fra testerene kan det være hensiktsmessig å etablere intra-tester reliabiliteten for hver rater før man sammenlikner inter-tester reliabiliteten (Portney and Watkins, 2000). Personlig har jeg begrenset erfaring med bruk av UL. Dette kan være en stor svakhet med min studie, da litteratur indikerer at økt erfaring med bruk av UL, vil øke

reliabiliteten av resultatene (O'Connor et al., 2005; Hebert et al., 2009). Indikasjonene fra forutnevnte studier henviser til bruken av UL der man må identifisere de ulike anatomiske strukturene og håndtere transduseren for å fremstille et tydelig bilde. I min studie har jeg ikke gjennomført UL scanningen selv, men gjort målinger av senetykkelse på bilder som er tatt av en erfaren bruker av UL. Basert på gode resultater fra intra-tester reliabiliteten i min studie, kan det tenkes at det ikke stilles like store krav til erfaring på denne avgrensede bruken av UL som måling av senetykkelse innebærer. For å minske sjansen for at min korte erfaring påvirket resultatene ble det utført flere prøvemålinger av senetykkelsen med tester 2, før studiens start. Dette anser jeg som en styrke ved studien.

UL bildene som ble brukt i bedømmelsen av senetykkelsen ble innhentet i forbindelse med det tidligere nevnte doktorgradsarbeidet (Haslerud, 2012). Bildene ble tatt av samme tester (tester 2 i min studie). Denne testeren har nesten 20 års erfaring med bruk av UL og har utgitt flere studier omhandlende UL. I doktorgradsarbeidet ble fulgt en høyt standardisert protokoll for scanningene med UL (punkt 3.4.1). Denne protokollen var tilnærmet identisk som de som har blitt brukt i tidligere studier (Bjordal et al., 2003; Joensen et al., 2009). Disse faktorene vil trolig redusere sjansen for at variabilitet i min studie.

I min studie ble det opprettet en standardisert protokoll for målingene av senetykkelse på de eksisterende bildene. Denne protokollen ble gjennomgått og diskutert av begge testerne til en felles enighet var oppnådd. På tross av en standardisert protokoll vil målingene av senetykkelse kunne være en kilde til variabilitet, da målingen i seg selv setter store krav til presisjon. På tross av en høy grad av standardisering for protokollen, var min personlige erfaring at enkelte bilder var vanskeligere å bedømme. Dette var ofte grunnet uklare bilder der senens avgrensning var diffus og dermed vanskelig å fastsette. Dette kan være en mulig kilde til variabilitet. Dette kunne muligens vært forbedret ved å endre transduserfrekvens, flytte eller foreta en hel-tå manøver av transduseren i de opprinnelige målingene for å få tydeligere bilder.

## 5.2 Valg av statistiske metoder

Innen reliabilitetsstudier finnes det ingen klar konsensus angående hvilke statistiske analyser man bør benytte seg av (Rankin and Stokes, 1998). I litteraturen blir flere mulige statistiske metoder foreslått for å beskrive reliabiliteten av et måleinstrument (Atkinson and Nevill, 1998). I min studie ble ICC<sub>1,1</sub> og ICC<sub>3,1</sub> brukt for å undersøke den relative reliabiliteten. Den absolutte

reliabiliteten ble undersøkt med innen-subjekt standard deviasjon ( $S_w$ ),  $S_w \times 1,96$  og  $S_w \times 2,77$ . I avsnittene under vil valgene av statistiske metoder bli diskutert.

### 5.2.1 Normalfordeling av data

Ved analyse av normalfordelingen med Shapiro-Wilk slo variabelen  $A^2$  for biceps høyre side ut, som eneste variabel, som signifikant ikke-normalfordelt ( $p = ,030$ ). Det foreligger ingen klar årsak til hvorfor denne variabelen var signifikant ikke-normalfordelt. Det kan tenkes å ha vært en tastefeil eller, mindre sannsynlig, en systematisk feil. Det anses som mindre trolig at det foreligger en systematisk feil, da kun den ene variabelen slo ut som signifikant ikke-normalfordelt (Bjørndal and Hofoss, 2012, s. 36). I tilfeller der data er signifikant ikke-normalfordelt reduseres grunnlaget for statistiske analyser ved parametriske tester. I min studie ble relativ reliabilitet analysert med ICC, som baseres på variansanalyse (ANOVA) (Rankin and Stokes, 1998). På tross av om datamaterialet ikke er normalfordelt, legges normalitet ved fordeling av gjennomsnittet, og ikke datamaterialet, som en forutsetning for ANOVA (Norman, 2010). Videre viser sentralgrenseteoremet at for gruppestørrelser større enn 5 eller 10 per gruppe vil gjennomsnittet, uavhengig av den opprinnelige fordelingen, bli omtrent normalfordelt (ibid). Tidligere studier har vist at ANOVA er en robust analysemetode, selv på signifikant ikke-normalfordelte data (Pearson, 1931). Dette fører til at bruk av parametriske metoder, selv på ikke-normalfordelte data, vil gi tilnærmet korrekte svar (Norman, 2010). Ytterligere var ikke variabelen som var signifikant ikke-normalfordelt en del av hovedresultatene, men en subgruppeanalyse. På bakgrunn av dette godtas analysemetoden.

### 5.2.2 Relativ reliabilitet

I analysen av relativ reliabilitet ble  $ICC_{1,1}$  og  $ICC_{3,1}$  brukt. Disse to ble valgt på bakgrunn av at  $ICC_{1,1}$  tar høyde for både tilfeldige og systematiske feil, mens  $ICC_{3,1}$  kun tar høyde for tilfeldige feil (Weir, 2005). Dersom resultatene av analysen av  $ICC_{3,1}$  er høyere enn  $ICC_{1,1}$  kan dette indikere at det foreligger en systematisk feil. Dersom de systematiske feilene er betydelige vil valg av analytisk modell påvirke utfallet av ICC hvor modell 3 > modell 2 > modell 1 (Portney and Watkins, 2000; Weir, 2005). I min studie var grad av systematiske målefeil mindre, og tilfeldige målefeil større. Dette fører til at valg av analytisk modell vil ha mindre betydning for størrelsen av ICC, og dermed føre til likere resultater av ulike analyser med ICC. Det ville ytterligere begrenset generaliserbarheten av resultatene mine dersom kun  $ICC_{3,1}$  hadde blitt



anvendt som analysemetode. I analysene er det tatt utgangspunkt i *single measures*, i motsetning til *average measures*, i SPSS. Single measures gir et generelt lavere resultat enn average measures (Shrout and Fleiss, 1979).

Ved å øke variabiliteten mellom studiens deltakere og redusere antall tilfeldige og systematiske målefeil, kan studiens reliabilitet forbedres (Streiner et al., 2014). I min studie kunne jeg endret kriteriene for deltakelse, målt senetykkelse på økt antall bilder og målinger, og potensielt økt antall testere. Ytterligere, kunne det blitt gjennomført flere runder med øving før målingene startet. Dette ville ha kunnet redusere antall målefeil i min studie.

### 5.2.3 Absolutt reliabilitet

I analysen av absolutt reliabilitet ble innen-subjekt standard deviasjon (Sw) brukt. Dette er en analysemetode som har blitt rapportert som en hyppig brukt analysemetode for absolutt reliabilitet i reliabilitetsstudier (Weir, 2005). Da jeg ikke har hatt tilgang til statistiker, ble det ikke gjennomført Bland-Altman plot (Bland and Altman, 1986). Dette er en potensiell svakhet med studien. På en annen side kan det problematiseres at et Bland-Altman plot, i all hovedsak, kun tilføyer en mer visuell fremstilling av de analysene som har blitt gjort i min studie (Kottner et al., 2011). Videre indikerer litteratur at man står fritt til å velge mellom Sw eller Bland-Altman plot ved rapportering av absolutt reliabilitet (Rankin and Stokes, 1998). Ved inspeksjon av spredningsplot for mine resultater (figur 10-13), gir det liten grunn til å mistenke at et Bland-Altman plot ville vist en vesentlig forskjell fra mine resultater.

I min studie kunne *limits of agreement* (LoA) vært brukt til å beskrive graden av variasjon på en repetert måling (Bland and Altman, 1986). Dette ble ikke gjort da det har blitt hevdet at å presentere både Sw og LoA er u hensiktsmessig da disse er for nært forbundet med hverandre (Hopkins, 2000). I motsetning til relativ reliabilitet, finnes det ingen klar enighet om hvilke verdier av absolutt reliabilitet som anses som tilfredstillende ved tverrsnittsmålinger av skuldereener (Carter et al., 2011).

## 5.3 Resultater

I min studie ble intra-, og inter-tester reliabiliteten av tykkelsesmåling med UL for senene til supraspinatus og biceps brachii caput longum hos pasienter med skuldertendinopati undersøkt. Resultatene fra studien min indikerer en generelt høy intra-tester reliabilitet, og generelt lavere resultater for inter-tester reliabilitet ved begge senene som ble målt. Disse resultatene samsvarer med det som er forventet å finne ved reliabilitetsstudier (Pijls et al., 2010; Carter et al., 2011, s. 240). Resultatene for biceps brachii caput longum var generelt bedre enn resultatene for supraspinatus. I avsnittene under vil resultatene fra studien min med tanke på intra-, og inter-tester reliabilitet bli diskutert.

### 5.3.1 Intra-tester reliabilitet

Resultatene for intra-tester reliabiliteten indikerer i hvilken grad tester 1 er konsistent ved to separate tykkelsesmålinger (datasett A<sup>1</sup> og A<sup>2</sup>) på det samme ultralydbildet av senene til supraspinatus og biceps brachii caput longum.

I min studie var resultatene av relativ og absolutt reliabilitet for biceps brachii caput longum generelt bedre enn for supraspinatus. Disse resultatet kan ha blitt påvirket av flere faktorer. For det første er det en større variasjon i utformingen av senen til supraspinatus (McCreesh et al., 2015a). Dette fører til at tykkelsen av supraspinatussenen vil variere i større grad gjennom forløpet, enn senen til biceps. Ved tykkelsesmåling av supraspinatussenen opplevde jeg det som vanskeligere å avgrense senens superfisielle endepunkt, sammenliknet med bicepssenen.

En ytterligere variabel som kan ha påvirket resultatene var vanskeligheten ved å avgjøre supraspinatussenens tykkeste punkt under selve målingene. Dette ble målt til 90 grader fra det ossøse referansepunktet i beinlinjens knekkpunkt mellom den humerale leddflate og basis av tuberculum majus. Sist kan det tenkes at tidsaspektet av målingene påvirket resultatene.

Tykkelsesmålingene til samtlige supraspinatussener ble gjennomført etter målingene av samtlige bicepssener. Da hver testrunde varte omtrent fire timer, er det ikke utenkelig at konsentrasjonsnivået til testeren kan ha påvirket de noe dårligere resultatene for supraspinatus.

Resultatene for supraspinatus på venstre side avdekket et dårligere resultat, ved relativ reliabilitet, enn høyre side ( ICC<sub>1,1</sub> = 0,66 og 0,90, respektivt). Forskjellen ved absolutt reliabilitet underbygger til en viss grad dette (Sw = 0,7 mm og 0,6 mm, respektivt). Da bildene

av høyre og venstre side ble presentert likt, og inneholdt både friske og syke skuldre, er dette noe overraskende. Denne forskjellen er vanskelig å forklare, men det kan problematiseres om tidsaspektet og tilfeldigheter kan ha hatt en påvirkning.

Resultatene av  $ICC_{1,1}$  var tilnærmet lik, eller lik, resultatene av  $ICC_{3,1}$  for både supraspinatussenen og senen av biceps brachii caput longum. Dette indikerer at det foreligger liten grad av systematiske målefeil (Moe-Nilssen, 1998).

### 5.3.2 Inter-tester reliabilitet

Resultatene for intra-tester reliabiliteten indikerer i hvilken grad tester 1 (datasett A<sup>1</sup>) og tester 2 (datasett B) er konsistente ved tykkelsesmålinger på det samme ultralydbildet av senene til supraspinatus og biceps brachii caput longum. I likhet med intra-tester reliabiliteten, viste resultatene for inter-tester reliabilitet en generelt sterkere absolutt og relativ reliabilitet for biceps, i forhold til supraspinatus. Mulige årsaker til dette har blitt diskutert i avsnitt 5.3.1. Videre var resultatene for  $ICC_{1,1}$  lik eller tilnærmet like resultatene for  $ICC_{3,1}$  ved begge sener. Dette indikerer at det foreligger liten grad av systematiske feil (Moe-Nilssen, 1998).

Basert på den manglende erfaringen til tester 1, er det forventet å se en læringseffekt ved testrunde to, sammenliknet med testrunde en. Med dette menes at etterhvert som tester 1 opparbeider seg mer erfaring, vil målingene av senene komme nærmere “det sanne svaret”. I studien min er det tenkt at tester 2, gitt testerens lange erfaring med bruk av UL, vil ligge nærmere dette “sanne svaret”. Ved analyse av resultatene fra datasett A<sup>1</sup> og A<sup>2</sup> for tester 1, sammenliknet mot resultatene for tester 2, viser det seg at målingene fra datasett A<sup>1</sup> ligger nærmere resultatene til tester 2. Med andre ord var resultatene av datasett A<sup>2</sup> lenger fra det tiltenkte “sanne svaret”. Denne “omvendte” læringseffekten ble observert ved resultatene til begge senene. Dette fenomenet ble ytterligere observert ved subgruppeanalyse av normalfordeling av data. Der slo testrunde to av bicepssenen på høyre side ut som signifikant ikke-normalfordelt ( $p = ,030$ ). Dette var den eneste subgruppen som ikke viste normalfordelt data, så det kan problematiseres om dette var en tilfeldig feil. En annen mulig årsak til den “omvendte” læringskurven, kan trolig tilskrives en grad av tidsnød ved gjennomføringen av testrunde to. Det kan tenkes at en mer forventet læringseffekt hadde blitt observert, dersom tidsnøden ved testrunde to ble eliminert. På bakgrunn av dette, ble det valgt å ta utgangspunkt i resultatene fra målingene A<sup>1</sup> – B i analysen av inter-tester reliabilitet.

Det ble observert svakere resultater av absolutt og relativ reliabilitet for målingene av begge senene på venstre side sammenliknet med høyre side (Tabell 9-12). Videre indikerer resultatene for supraspinatussenen og bicepsen en lavere relativ reliabilitet ved sammenlikning av gjennomsnittet for målingene  $A^1$  og  $A^2 - B$  ( $ICC_{1,1} = 0,54$  og  $0,64$ , respektivt). Det er forventet å se et bedre resultat ved å bruke gjennomsnittet av de to målingene til tester 1 sammenliknet med målingen til tester 2. Etter gjentakende analyse, og diskusjon med flere veiledere, ble det ikke kommet frem til en god forklaring på dette resultatet.

Resultatene for målingene  $A^1 - B$  der høyre og venstre side ble slått sammen, indikerer en gjennomsnittstykkelse av senen til supraspinatus på 7,7 mm. Dette er 1,2 mm mer enn hva Bjordal og kolleger fant i sin studie (Bjordal et al., 2003). I motsetning til min studie inkluderte studien fra 2003 kun friske deltakere, med en grenseverdi på 0,8 mm for syke sener. Da studien min inkluderte både friske og tendinopatiske skuldersener, er det naturlig å tenke at siden jeg hadde flere sener med patologi var senetykkelsen over denne grenseverdien. Det er dermed forventet at supraspinatussenenes gjennomsnittstykkelse er større i min studie, enn i studien med friske sener av Bjordal og kolleger (Bjordal et al., 2003).

#### 5.4 Kliniske implikasjoner

Som manuellterapeut er man opptatt av å finne reliable og anvendbare målemetoder til bruk i klinisk praksis. Ved å undersøke reliabiliteten av, blant annet, tykkelsesmålinger av skuldersener med UL, vil man i større grad kunne vektlegge resultatene av slike målinger i klinikken. Denne kartleggingen vil være viktig, sett i sammenheng med oppblomstringen av bruk av UL i primærhelsetjenesten og den hyppige forekomsten av skulderplager (Luime et al., 2004). I forlengelsen av dette vil det være samfunnsøkonomisk gunstig å bruke UL sammenliknet med, for eksempel, MR. Litteratur indikerer at senetykkelse har et potensiale for å gi en relativt objektiv bedømming av senepatologi (Allen and Wilson, 2001; Joensen et al., 2009; Hodgson et al., 2014; Khan et al., 2014). Resultatene fra min studie indikerer at det er mulig å måle senetykkelse til supraspinatus og biceps med relativt stor nøyaktighet, selv med begrenset erfaring. Dette vil kunne føre til en mer nøyaktig diagnostiseringsprosess hos pasienter med skuldersmerter, som potensielt kan resultere i en bedre behandling. Disse faktorene medfører at UL fremstår som et meget relevant klinisk verktøy for helsepersonell i primærhelsetjenesten.

## 5.5 Videre forskning

For og bedre forstå reliabiliteten av UL bør det gjennomføres ytterligere forskning på området. Kartlegging av reliabilitet er viktig for å kunne videre utforske validiteten av en målemetode (Portney and Watkins, 2000). Disse studiene bør inkludere flere deltakere og, i tillegg til skulder, undersøke andre anatomiske områder der UL blir brukt i klinisk praksis. For å styrke metodologien av videre studier ville det vært hensiktsmessig å øke grad av blinding av testerne under målingene. Det ville vært interessant å undersøke i hvilken grad mine resultater kan generaliseres til sener ved andre anatomiske lokasjoner. Ytterligere ville det vært av interesse å gjennomføre prospektive studier der utfallet av skuldertendinopati ble sammenliknet med tykkelsesmålinger ved UL av skuldere sener. I hvilken grad tykkelsesmåling av sener med UL har potensiale til å avdekke senepatologi er et annet aspekt som bør undersøkes videre.

Med tanke på klinisk overførbarhet, kan det være hensiktsmessig om studiene blir gjennomført i primærhelsetjenesten. I forlengelsen av dette bør videre forskning undersøke og presentere både absolutt og relativ reliabilitet. Begge disse kan med fordel innarbeides som en standard ved reliabilitetsstudier. Utformingen av retningslinjer for bedømmingen av absolutt reliabilitet ved tykkelsesmålinger med UL på skuldere sener ville vært hensiktsmessig for en felles forståelse og klinisk overførbarhet av resultater.

Andre potensielle interesseområder for videre forskning på temaet er grad av diagnostisk presisjon og reliabilitet av UL ved mindre prevalente smertelokasjoner, epidemiologisk forskning av skulderplager i primær-, og sekundærhelsetjenesten, og studier med fokus på den samfunnsøkonomiske innvirkningen av UL som diagnostisk verktøy hos pasienter med skulderplager.

## 6. Konklusjon

Hensikten med denne oppgaven var å undersøke inter-tester og intra-tester reliabiliteten i måling av senetykkelse av m. supraspinatus og m. biceps brachii caput longum med UL hos pasienter med skuldertendinopatier. Det ble totalt gjennomført tilnærmet 150 individuelle målinger per tester.

Resultatene av relativ reliabilitet for *biceps brachii caput longum* er basert på syv analyser der fem var “tilnærmet perfekt” (range ICC: 0,81 – 0,94) og to var “betydelig” (range ICC: 0,64 – 0,65). ICC<sub>1,1</sub> oppgitt som gjennomsnitt for alle analyseresultatene av relativ reliabilitet for biceps var 0,81. Resultatene indikerer en “tilnærmet perfekt” relativ reliabilitet for biceps brachii caput longum. For *supraspinatus* viste to av syv analyser av relativ reliabilitet en “moderat” reliabilitet (range ICC: 0,52 – 0,54), fire indikerer en “betydelig” reliabilitet (range ICC: 0,63 – 0,79) og én avdekket en “tilnærmet perfekt” reliabilitet (ICC: 0,90). Resultatene for relativ reliabilitet av ICC<sub>1,1</sub> presentert som gjennomsnitt av de syv målingene var 0,68. Resultatene indikerer en “betydelig” relativ reliabilitet for supraspinatus.

## 7. Referanser

- Abate M, Gravare-Silbernagel K, Siljeholm C, Di Iorio A, De Amicis D, Salini V, Werner S, Paganelli R. Pathogenesis of tendinopathies: inflammation or degeneration? *Arthritis Research & Therapy* 2009; 11 (3): 235-235.
- Allen G, Wilson D. Ultrasound of the shoulder. *European journal of ultrasound* 2001; 14 (1): 3-9.
- Altman D. *Practical statistics for medical research*. 2nd ed. London: Chapman and Hall, 2000.
- American Educational Research Association. *Standards for educational and psychological testing*: Amer Educational Research Assn, 1999.
- Atkinson G, Nevill AM. Statistical Methods For Assessing Measurement Error (Reliability) in Variables Relevant to Sports Medicine. *Sports medicine [journal article]*1998; 26 (4): 217-238.
- Bartko JJ, Carpenter WT. On the methods and theory of reliability. *The Journal of nervous and mental disease* 1976; 163 (5): 307-317.
- Batterham AM, George KP. Reliability in evidence-based clinical practice: a primer for allied health professionals☆. *Physical therapy in Sport* 2003; 4 (3): 122-128.
- Beggs I, Bianchi S, Bueno A, Cohen M, Court-Payen M, Grainger A, Kainberger F, Klauser A, Martinoli C, McNally E, O'Connor P, Peetrons P, Reijnierse M, Remplik P, Silvestri E. Musculoskeletal ultrasound technical guidelines I. Shoulder. 2010.
- Bjordal JM. Behandling av tendinopati. *Tidsskrift for den Norske legeforening* 2010; 130 (23): 2337.
- Bjordal JM, Demmink JH, Ljunggren AE. Tendon thickness and depth from skin for supraspinatus, common wrist and finger extensors, patellar and Achilles tendons: ultrasonography study of healthy subjects. *Physiotherapy* 2003; 89 (6): 375-383.
- Bjørndal A, Hofoss D. *Statistikk for helse- og sosialfagene*  
In: (Ed.: Gyldendal Norsk Forlag AS, 2012.
- Blacker D, Endicott J. Psychometric properties: concepts of reliability and validity. *Handbook of psychiatric measures* Washington: American Psychiatric Association 2002: 7-14.
- Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* 1986; 327 (8476): 307-310.
- Bland JM, Altman DG. Statistics notes: measurement error. *BMJ* 1996; 313 (7059): 744.

- Breckenridge JD, McAuley JH. Shoulder Pain and Disability Index (SPADI). *J Physiother* 2011; 57 (3): 197.
- Bureau N, Roederer G. Sonography of Achilles tendon xanthomas in patients with heterozygous familial hypercholesterolemia. *AJR Am J Roentgenol* 1998; 171 (3): 745-749.
- Carter R, Lubinsky J, Domholdt E. *Rehabilitation Research Pageburst on Kno Retail Access Code: Principles and Applications*: Elsevier Science Health Science Division, 2011.
- Chen C-H, Hsu K-Y, Chen W-J, Shih C-H. Incidence and Severity of Biceps Long-Head Tendon Lesion in Patients with Complete Rotator Cuff Tears. *Journal of Trauma and Acute Care Surgery* 2005; 58 (6): 1189-1193.
- Clark NC. Functional performance testing following knee ligament injury. *Physical therapy in Sport* 2001; 2 (2): 91-105.
- Cohen J. *Statistical power analysis for the behaviour sciences* (2nd). Hillsdale, NJ: erlbaum; 1988.
- de Jesus JO, Parker L, Frangos AJ, Nazarian LN. Accuracy of MRI, MR arthrography, and ultrasound in the diagnosis of rotator cuff tears: a meta-analysis. *AJR Am J Roentgenol* 2009; 192 (6): 1701-1707.
- De Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine: a practical guide*: Cambridge University Press, 2011.
- Drageset S, Ellingsen S. Forståelse av kvantitativ helseforskning-en introduksjon og oversikt. *Nordisk tidsskrift for helseforskning* 2009; 5 (2): 100-113.
- Duerr M. Reliability and accuracy of distance measurements between shoulder bony landmarks evaluated by ultrasound in asymptomatic subjects. Auckland University of Technology; 2010.
- Ferrari FS, Governi S, Burrese F, Vigni F, Stefani P. Supraspinatus tendon tears: comparison of US and MR arthrography with surgical correlation. *Eur Radiol* 2002; 12 (5): 1211-1217.
- Fredberg U, Stengaard-Pedersen K. Chronic tendinopathy tissue pathology, pain mechanisms, and etiology with a special focus on inflammation. *Scandinavian journal of medicine & science in sports* 2008; 18 (1): 3-15.
- General Electric. LOGIQ e Ultrasound – BT12. 2016; Available from: [http://www3.gehealthcare.com/en/products/categories/ultrasound/point\\_of\\_care\\_ultrasound/logiq\\_e](http://www3.gehealthcare.com/en/products/categories/ultrasound/point_of_care_ultrasound/logiq_e) [Accessed 24.02.2016].
- Hart DL. Invited Commentary. *Phys Ther* 1989; (69): 102-103.



- Harvie P, Ostlere S, Teh J, McNally E, Clipsham K, Burston B, Pollard T, Carr A. Genetic influences in the aetiology of tears of the rotator cuff SIBLING RISK OF A FULL-THICKNESS TEAR. *Journal of Bone & Joint Surgery, British Volume* 2004; 86 (5): 696-700.
- Haslerud S. The effect of a standardized physiotherapy treatment regime for patients with rotator cuff tendinopathy. A randomized controlled trial.; 2012.
- Hawkins R, Kennedy J. Impingement syndrome in athletes. *Am J Sports Med* 1980; 8 (3): 151-158.
- Hebert JJ, Koppenhaver SL, Parent EC, Fritz JM. A Systematic Review of the Reliability of Rehabilitative Ultrasound Imaging for the Quantitative Assessment of the Abdominal and Lumbar Trunk Muscles. *Spine* 2009; 34 (23): E848-E856.
- Hodgson R, O'Connor P, Grainger A. Tendon and ligament imaging. *The British journal of radiology* 2014.
- Hopkins C, Fu S-C, Chua E, Hu X, Rolf C, Mattila VM, Qin L, Yung PS-H, Chan K-M. Critical review on the socio-economic impact of tendinopathy. *Asia-Pacific Journal of Sports Medicine, Arthroscopy, Rehabilitation and Technology* 2016; 4: 9-20.
- Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med* 2000; 30 (1): 1-15.
- IBM Corp. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.; 2013.
- Jacobson JA. Sonography of the shoulder. *Applied Radiology* 2003; 32: 19-25.
- Jacobson JA. *Fundamentals of Musculoskeletal Ultrasound: Expert Consult-Online*: Elsevier Health Sciences, 2012.
- Jobe FW, Moynes DR. Delineation of diagnostic criteria and a rehabilitation program for rotator cuff injuries. *Am J Sports Med* 1982; 10 (6): 336-339.
- Joensen J, Coupe C, Bjordal JM. Increased palpation tenderness and muscle strength deficit in the prediction of tendon hypertrophy in symptomatic unilateral shoulder tendinopathy: an ultrasonographic study. *Physiotherapy* 2009; 95 (2): 83-93.
- Kambouris M, Ntalouka F, Ziogas G, Maffulli N. Predictive genomics DNA profiling for athletic performance. *Recent patents on DNA & gene sequences* 2012; 6 (3): 229-239.
- Khan K, Fields KB, Grayzel J. Overview of overuse (chronic) tendinopathy. UpToDate, Waltham, MA(Accessed on May 1, 2015) 2014.

- Koch GG. Intraclass correlation coefficient. Encyclopedia of statistical sciences 1982.
- Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011; 64 (1): 96-106.
- Leong H-T, Tsui S, Ying M, Leung VY-f, Fu SN. Ultrasound measurements on acromio-humeral distance and supraspinatus tendon thickness: Test-retest reliability and correlations with shoulder rotational strengths. *Journal of Science and Medicine in Sport* 2012; 15 (4): 284-291.
- Levine BD, Motamedi K, Seeger LL. Imaging of the Shoulder: A Comparison of MRI and Ultrasound. *Current Sports Medicine Reports* 2012; 11 (5): 239-243.
- Lewis JS. Rotator cuff tendinopathy: a model for the continuum of pathology and related management. *Br J Sports Med* 2010; 44 (13): 918-923.
- Luime JJ, Koes BW, Hendriksen IJ, Burdorf A, Verhagen AP, Miedema HS, Verhaar JA. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol* 2004; 33 (2): 73-81.
- Lundgreen K, Lian ØB, Engebretsen L, Scott A. Tenocyte apoptosis in the torn rotator cuff: a primary or secondary pathological event? *Br J Sports Med* 2011: bjsports83188.
- Maffulli N, Longo UG, Denaro V. Novel approaches for the management of tendinopathy. *J Bone Joint Surg Am* 2010; 92 (15): 2604-2613.
- Maganaris CN, Narici MV, Almekinders LC, Maffulli N. Biomechanics and pathophysiology of overuse tendon injuries. *Sports medicine* 2004; 34 (14): 1005-1017.
- Magra M, Maffulli N. Genetic aspects of tendinopathy. *Journal of Science and Medicine in Sport* 2008; 11 (3): 243-247.
- Martinoli C, Bianchi S, Prato N, Pugliese F, Zamorani MP, Valle M, Derchi LE. US of the Shoulder: Non-Rotator Cuff Disorders 1. *Radiographics* 2003; 23 (2): 381-401.
- McCreesh KM, Anjum S, Crotty JM, Lewis JS. Ultrasound measures of supraspinatus tendon thickness and acromiohumeral distance in rotator cuff tendinopathy are reliable. *Journal of clinical ultrasound* 2015a.
- McCreesh KM, Crotty JM, Lewis JS. Acromiohumeral distance measurement in rotator cuff tendinopathy: is there a reliable, clinically applicable method? A systematic review. *Br J Sports Med* 2015b; 49 (5): 298-305.

- McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological methods* 1996; 1 (1): 30.
- Moe-Nilssen R. Test-retest reliability of trunk accelerometry during standing and walking. *Arch Phys Med Rehabil* 1998; 79 (11): 1377-1385.
- Moosmayer S, Heir S, Aaser P, Smith H-J. Ultralyddiagnostikk av skulderbuen—en metodebeskrivelse 177–80. 2004.
- Moosmayer S, Smith H-j. Diagnostic ultrasound of the shoulder—a method for experts only? *Acta Orthop* 2005; 76 (4): 503-508.
- Murtaugh B, Ihm J. Eccentric Training for the Treatment of Tendinopathies. *Current Sports Medicine Reports* 2013; 12 (3): 175-182.
- Neer CS, 2nd. Anterior acromioplasty for the chronic impingement syndrome in the shoulder: a preliminary report. *J Bone Joint Surg Am* 1972; 54 (1): 41-50.
- Nicholas J. Clinical observations on sports-induced soft tissue injuries. *Sports-Induced Inflammation* Park Ridge, IL: American Academy of Orthopaedic Surgeons 1990: 129-148.
- Norman G. Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education [journal article]* 2010; 15 (5): 625-632.
- O'Connor PJ, Rankine J, Gibbon WW, Richardson A, Winter F, Miller JH. Interobserver variation in sonography of the painful shoulder. *J Clin Ultrasound* 2005; 33 (2): 53-56.
- Ok J-H, Kim Y-S, Kim J-M, Yoo T-W. Learning curve of office-based ultrasonography for rotator cuff tendons tears. *Knee Surgery, Sports Traumatology, Arthroscopy [journal article]* 2013; 21 (7): 1593-1597.
- Oliva F, Misiti S, Maffulli N. Metabolic diseases and tendinopathies: the missing link. *MLTJ Muscles, Ligaments and Tendons Journal* 2014; 4 (3): 273-274.
- Oliva F, Piccirilli E, Berardi AC, Frizziero A, Tarantino U, Maffulli N. Hormones and tendinopathies: the current evidence. *Br Med Bull* 2016.
- Ombregt L, Bisschop P, Veer HJt. *A System of orthopaedic medicine: with accompanying CD-ROM*. London: Churchill Livingstone, 2003.
- Ottenheijm RP, Jansen MJ, Staal JB, van den Bruel A, Weijers RE, de Bie RA, Dinant GJ. Accuracy of diagnostic ultrasound in patients with suspected subacromial disorders: a systematic review and meta-analysis. *Arch Phys Med Rehabil* 2010; 91 (10): 1616-1625.

- Pearson ES. The analysis of variance in cases of non-normal variation. *Biometrika* 1931; 114-133.
- Pijls BG, Kok FP, Penning LI, Guldmond NA, Arens HJ. Reliability study of the sonographic measurement of the acromiohumeral distance in symptomatic patients. *J Clin Ultrasound* 2010; 38 (3): 128-134.
- Portney LG, Watkins MP. Statistical measures of reliability. *Foundations of clinical research: applications to practice* 2000; 2: 557-586.
- Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 2001; 33 (5): 337-343.
- Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil* 1998; 12 (3): 187-199.
- Rees JD, Maffulli N, Cook J. Management of tendinopathy. *Am J Sports Med* 2009; 37 (9): 1855-1867.
- Riggin CN, Morris TR, Soslowsky LJ. Tendinopathy II: Etiology, Pathology, and Healing of Tendon Injury and Tendon Regeneration: Understanding Tissue Physiology and Development to Engineer Functional Substitutes 2015: 149.
- Rio E, Moseley L, Purdam C, Samiric T, Kidgell D, Pearce AJ, Jaberzadeh S, Cook J. The Pain of Tendinopathy: Physiological or Pathophysiological? *Sports medicine [journal article]* 2014; 44 (1): 9-23.
- Rutten MJ, Spaargaren GJ, van Loon T, de Waal Malefijt MC, Kiemeney LA, Jager GJ. Detection of rotator cuff tears: the value of MRI following ultrasound. *Eur Radiol* 2010; 20 (2): 450-457.
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86 (2): 420-428.
- Smith TO, Back T, Toms AP, Hing CB. Diagnostic accuracy of ultrasound for rotator cuff tears in adults: a systematic review and meta-analysis. *Clin Radiol* 2011; 66 (11): 1036-1048.
- Stolik S, Delgado J, Perez A, Anasagasti L. Measurement of the penetration depths of red and near infrared light in human "ex vivo" tissues. *Journal of Photochemistry and Photobiology B: Biology* 2000; 57 (2): 90-93.
- Stratford PW, Goldsmith CH. Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Phys Ther* 1997; 77 (7): 745-750.

- Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use*: Oxford University Press, USA, 2014.
- Sunde P, Tariq R. [Ultrasound--a good alternative to MR in shoulder problems]. *Tidsskrift for den Norske laegeforening: tidsskrift for praktisk medicin, ny raeke* 2008; 128 (7): 842-842.
- Tallon C, Maffulli N, Ewen SW. Ruptured Achilles tendons are significantly more degenerated than tendinopathic tendons. *Medicine and Science in Sports and Exercise* 2001; 33 (12): 1983-1990.
- Temes WC, Temes Clifton A, Hilton V, Girard L, Strait N, Karduna A. Reliability and validity of thickness measurements of the supraspinatus muscle of the shoulder: an ultrasonography study. *J Sport Rehabil* 2014: Mar 12; Epub ahead of print.
- Tennent TD, Beach WR, Meyers JF. A review of the special tests associated with shoulder examination. Part I: the rotator cuff tests. *Am J Sports Med* 2003; 31 (1): 154-160.
- Thain LM, Adler RS. Sonography of the rotator cuff and biceps tendon: technique, normal anatomy, and pathology. *Journal of clinical ultrasound* 1999; 27 (8): 446-458.
- Van der Windt D, Koes BW, de Jong BA, Bouter LM. Shoulder disorders in general practice: incidence, patient characteristics, and management. *Annals of the Rheumatic Diseases* 1995; 54 (12): 959-964.
- Walter S, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Statistics in medicine* 1998; 17 (1): 101-110.
- Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength & Conditioning Research* 2005; 19 (1): 231-240.
- Wilcox RR. Some practical reasons for reconsidering the Kolmogorov-Smirnov test. *British Journal of Mathematical and Statistical Psychology* 1997; 50 (1): 9-20.
- World Medical A. World medical association declaration of helsinki: Ethical principles for medical research involving human subjects. *JAMA* 2013; 310 (20): 2191-2194.

## Vedlegg 1: Godkjenning REK



<b>Region:</b>	<b>Saksbehandler:</b>	<b>Telefon:</b>	<b>Vår dato:</b>	<b>Vår referanse:</b>
REK nord	Monika Rydland Gaare	77620756	22.02.2013	2012/2134/REK nord
			<b>Dens dato:</b>	<b>Dens referanse:</b>
			19.02.2013	
			<b>Vår referanse må oppgis ved alle henvendelser</b>	

Sturla Haslerud

### **2012/2134 Effekten av kunnskapsbasert fysioterapibehandling for pasienter med rotator cuff tendinopati**

**Forskningsansvarlig:** Universitetet i Bergen  
**Prosjektleder:** Sturla Haslerud

Vi viser til søknad om prosjektendring av 19.02.2013 vedrørende endring av prosjektleder til Professor Jan Magnus Bjordal og til tidligere tilbakemelding vedrørende rekruttering og revidert forespørselskriv. Både tilbakemelding vedrørende rekruttering, revidert forespørselskriv og endring av prosjektleder er i tråd med komiteens merknader i utsettelsesvedtak av 29.01.2013.

Etter fullmakt er det fattet slikt

#### **vedtak**

*Med hjemmel i helseforskningsloven § 10 og forskningsetikkloven § 4 godkjennes prosjektet.*

#### **Sluttmelding og søknad om prosjektendring**

Prosjektleder skal sende sluttmelding til REK nord på eget skjema senest (et halvt år etter prosjektslutt), jf. hfl. § 12. Prosjektleder skal sende søknad om prosjektendring til REK nord dersom det skal gjøres vesentlige endringer i forhold til de opplysninger som er gitt i søknaden, jf. hfl. § 11.

#### **Klageadgang**

Prosjektleder kan klage på komiteens vedtak, jf. forvaltningslovens § 28 flg. Klagen sendes til REK nord. Klagefristen er tre uker fra mottak av dette brevet. Dersom vedtaket opprettholdes av REK nord, sendes klagen videre til Den nasjonale forskningsetiske komité for medisin og helsefag for endelig vurdering.

Med vennlig hilsen

May Britt Rossvoll  
sekretariatsleder

Monika Rydland Gaare  
seniorkonsulent

**Kopi til:** [roiv.lie@smis.uib.no](mailto:roiv.lie@smis.uib.no); [jmb@hib.no](mailto:jmb@hib.no); [postmottak@uib.no](mailto:postmottak@uib.no)

## Vedlegg 2: Forespørsel om deltakelse

# Forespørsel om deltakelse i forskningsprosjektet

*Effekten av kunnskapsbasert fysioterapibehandling for pasienter med rotatorcuff tendinopati*

***I forbindelse med et doktorgradsstudium i fysioterapi ved Institutt for samfunnsmedisinske fag, Universitetet i Bergen, og Institutt for ergoterapi, fysioterapi og radiografi, Høgskolen i Bergen, skal det iverksettes en forskningsstudie for pasienter med smerter i skulderen.***

### **Bakgrunn og hensikt**

Dette er et spørsmål til deg om å delta i en forskningsstudie for å finne ut om en spesifikk øvelsesstrategi, manuell mobilisering, tøyning og laser behandling av muskel/sene apparatet i den vonde skulderen din, vil føre til bedre funksjon og mindre smerter. Vi ønsker også å undersøke hvorvidt denne behandlingen kan føre til at du ikke lenger vil ha behov for å operere skulderen din. Tilbudet er til deg som er henvist til kirurgisk behandling (Subacromial dekompresjon) av skulder, og står på venteliste ved Stavanger Universitets Sykehus. Du må ha vært plaget med skuldersmerter i over 3 måneder.

Ansvarlig for prosjektet er fysioterapeut og doktorgradsstipendiat Sturla Haslerud ved Institutt for samfunnsmedisinske fag, Universitetet i Bergen og Institutt for ergoterapi, fysioterapi og radiografi, Høgskolen i Bergen.

### **Hva innebærer studien?**

Ved den første konsultasjonen vil det bli foretatt en grundig undersøkelse bestående av samtale om skulderplagene dine, skuldertester og ultralydundersøkelse av skuldervenene. Dersom du oppfyller kriteriene for å bli med i studien, vil det bli foretatt videre målinger av skulderfunksjon og smerteintensitet. Du vil også bli bedt om å svare på et spørreskjema om skulderplagene og hvordan disse påvirker/hemmer deg i hverdagen. Totalt vil selve konsultasjonen vare cirka 1 klokke.

Deretter fordeles du til en av to grupper ved loddtrekning. Den ene gruppen vil få den nye behandlingen som er beskrevet over, mens den andre gruppen vil få vanlig oppfølging som opprinnelig planlagt i venteperioden. Behandlingen vil foregå hos fysioterapeut på fysikalsk institutt, 3 ganger ukentlig i 3 uker. Hver behandling varer 45-60 minutter. I de påfølgende 9 ukene vil du få behandling hos fysioterapeut 1 gang per uke. I tillegg vil du få noen enkle hjemmeøvelser. Du vil også bli bedt om å føre treningsdagbok.

### **Mulige fordeler og ulemper**

Fordelen med studien er at du vil få en grundig undersøkelse og kartlegging av skulderplagene dine. Du vil motta informasjon og råd angående dine skulderplager.

Blir du tildelt gruppen som får den nye behandlingen kan det gjøre skulderen din frisk, slik at du slipper operasjon. Deltakelse i studien vil ikke medføre noen merkostnader. Det vil heller ikke ha noen konsekvenser for din mulighet til å bli operert, eller påvirke din plassering på ventelisten til kirurgi ved Stavanger Universitetssykehus.

Ulempen for deg vil være at du må bruke tid til å gå til undersøkelse og behandlinger over en periode på 12 uker. I tillegg vil det være oppfølging med ny undersøkelse og målinger 3 måneder etter behandling er avsluttet. Det er ingen kjente bivirkninger knyttet til behandlingen, men noe øvelsesrelatert ømhet er vanlig de første ukene.

### **Hva skjer med informasjonen om deg?**

Informasjonen som registreres om deg skal kun brukes slik som beskrevet i hensikten med studien. Alle opplysningene vil bli behandlet uten navn og fødselsnummer eller andre direkte gjenkjennende opplysninger. En kode knytter deg til dine opplysninger gjennom en navneliste. Det er kun autorisert personell knyttet til prosjektet som har adgang til navnelisten og som kan finne tilbake til deg. Informasjon som er samlet inn vil bli anonymisert innen prosjektslutt 01.07.2016. Det vil ikke være mulig å identifisere deg i resultatene av studien dersom disse publiseres.

### **Frivillig deltakelse**

Det er frivillig å delta i studien. Du kan når som helst og uten å oppgi noen grunn trekke ditt samtykke til å delta i studien. Dette vil ikke få konsekvenser for din videre behandling eller mulighet for å bli operert. Dersom du ønsker å delta, undertegner du samtykkeerklæringen på siste side. Om du nå sier ja til å delta, kan du senere trekke tilbake ditt samtykke uten at det påvirker din øvrige behandling.

Dersom du senere ønsker å trekke deg eller har spørsmål til studien, kan du kontakte Sturla Haslerud på telefon 934 24 347 eller epost: [shas@hib.no](mailto:shas@hib.no)

Tidspunkt for første konsultasjon avtales på telefon med Sturla Haslerud (prosjektleder), og vil bli utført på Hillevåg Legesenter.

Ytterligere informasjon om studien finnes i kapittel A

## **Kapittel A- utdypende forklaring av hva studien innebærer**

Du vil i forbindelse med undersøkelse og oppfølging bli bedt om å svare på noen spørreskjema om skulderfunksjon og hvordan du håndterer smertene. Vi vil også måle tykkelsen av supraspinatussenen og øvrige rotator cuff sener i skulderen med diagnostisk ultralyd. Ultralyd undersøkelse er smertefritt og gir avbildning på samme måte som av foster under graviditet, og har ingen bivirkninger. Vi vil også undersøke deg med kliniske skuldertester. Skulderen kan bli litt forbigående øm etter disse testene, men ikke noe mer enn ved en vanlig skulderundersøkelse



Den vonde skulderen din kan være forårsaket av flere forhold som opptrer samtidig. Dette kan være redusert samspill mellom skulderblad og skulderledd, sykdom i sener, muskulær svakhet eller ubalanse, og enkelte bevegelsesinnskrenkinger. Ved å kombinere ulike fysikalske behandlingsformer vil vi undersøke hvorvidt det er mulig å korrigere disse forholdene og reversere sykdomsprosessen i skulderen. Behandlingen er basert på en systematisk gjennomgang av forskningslitteratur innenfor området og klinisk erfaring.

Deltakere fordelt til behandlingsgruppen vil kunne gå vederlagsfritt for å gjennomføre en potensielt effektiv behandling, som forventes å gi redusert smerte og økt skulder funksjon. En nylig publisert forskningsstudie har vist at denne type behandling er virkningsfull og reduserte behovet for kirurgi. Behandlingen vil også kunne fungere som pre-operativ trening, og på den måten bidra til et godt utgangspunkt før en eventuell operasjon. Du vil også bli grundig instruert i et spesifikt hjemmeøvelsesprogram tilpasset din skulder. Dette øvelsesprogrammet vil ikke ta lang tid å utføre. Det forventes imidlertid at du utfører programmet i henhold til behandlingsprotokoll, noe som medfører at du etter tre uker vil bli bedt om å styrke øvelser to ganger ukentlig og skuldertøyning daglig.

Det er ingen kjente bivirkninger knyttet til behandlingen, men noe øvelsesrelatert ømhet er vanlig de første ukene. Dersom du opplever ubehag utover dette er det viktig at du underretter fysioterapeuten, slik at behandlingen kan justeres. Du kan også kontakte prosjektleder angående spørsmål om studiet og behandlingen (se kontakt informasjon side 2)

Blir du etter undersøkelsen fordelt til gruppen som har vanlig oppfølging, får du ikke tilbud om den nye behandlingen. Du vil da ha den oppfølgingen som du i samråd med fastlegen din opprinnelig hadde planlagt i venteperioden. Du må da beregne tid til ny undersøkelse etter 12 uker . I tillegg vil du bli bedt om å svare på spørreskjema via e-post etter 4 uker og 6 måneder.

Ansvarlig for prosjektet er fysioterapeut og doktorgradsstipendiat Sturla Haslerud ved Institutt for samfunnsmedisinske fag, Universitetet i Bergen og Institutt for ergoterapi, fysioterapi og radiografi, Høgskolen i Bergen.

Oppstart av studien er planlagt april 2014. Dersom du samtykker til å delta i studien vil vi så fort det lar seg gjøre forsøke å finne en tid som passer for din første undersøkelse. For å bli inkludert i studien og bli fordelt til behandling- eller kontrollgruppen må du ha hatt symptomer i 3 måneder eller mer. Du må også ha en sykehistorie og kliniske funn som er forenlig med en subacromial impingement/rotator cuff tendinopati diagnose. Det er også en forutsetning for deltakelse at du kan forstå og gjøre deg forstått på engelsk eller norsk.

# Samtykke til deltakelse i studien

Jeg er villig til å delta i studien

-----

(Signert av prosjektdeltaker, dato)

Jeg bekrefter å ha gitt informasjon om studien

-----

(Signert, prosjektleder , dato)

