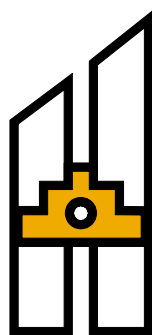


# **En reliabilitetsstudie av ultralydsmåling for å vurdere skuldersener**

Er ultralydsavbildning en reliabel metode å bruke ved måling av tverrsnittet på supraspinatus og biceps tendon longus i skulderen hos personer med og uten skuldersmerter?



**HØGSKOLEN I BERGEN**

**Silje Ingebrigtsen**

Klinisk master i fysioterapi med fordypning i ortopedi og revmatologi.

Institutt for ergoterapi, fysioterapi og radiografi

Masteroppgave

Veileder: Jon Joensen  
Innleveringsdato: 10.10.2012  
Antall ord: 16254

## Avtale om elektronisk publisering av mastergradsoppgave

Denne avtalen om elektronisk publisering av masteroppgave er inngått mellom Høgskolen i Bergen (HiB), Postboks 7030, 5020 Bergen og

Silje Ingebrigtsen

Forfatteren  
(heretter kalt studenten)

På de vilkår som er nevnt nedenfor, gir studenten Høgskolen i Bergen vederlagsfri adgang til å publisere følgende masteroppgave:

En reliabilitetsstudie av ultralydsmåling for å vurdere skuldersener.  
Er ultralydsavbildning en reliabel metode å bruke ved måling av tverrsnittet på supraspinatus og biceps tendon longus i skulderen hos personer med og uten skuldersmerter?

tittel

1. Studenten skal levere oppgaven elektronisk via Its learning.
2. **Etter** godkjenning gjøres oppgaven tilgjengelig gjennom BORA-HiB, høgskolens institusjonelle arkiv for fulltekstpublisering av faglige dokument. Høgskolen i Bergen plikter å publisere oppgaven slik den foreligger ved innlevering, med tekst, tabeller, grafikk, bilder m.m.
3. Høgskolen i Bergen har ikke adgang til kommersiell utnytting av oppgaven.
4. Denne avtalen begrenser ikke studentens rettigheter etter Lov om opphavsrett til åndsverk, og er dermed ikke til hinder for at oppgaven senere publiseres, uendret eller bearbeidet, i elektronisk eller annen form.
5. Oppgaven inneholder sensitive opplysninger og bør behandles konfidensielt.

Dato: 08.10.2012

Sted: Bergen

Navn: Silje Ingebrigtsen

## Forord

Min motivasjon for å gjøre ett klinisk masterprosjekt var i utgangspunktet å lære om forskningsmetode. Gjennom selv å ha gjennomført datainnsamling og databehandling i tillegg til skriveprosessen sitter jeg nå igjen med nye erfaringer og mye nyttig kunnskap. Jeg håper og tror at disse erfaringene og denne kunnskapen er noe jeg kan ta med meg videre og benytte i framtidig jobbsammenheng.

Det har vært en krevende, men spennende og lærerik prosess å jobbe med denne oppgaven. Til tider har det vært vanskelig å balansere tiden mellom oppgaveskriving, jobb og familie. Det har vært mange utfordringer underveis men med god støtte har jeg kommet i mål.

Jeg vil med dette rette en stor takk til:

Min sønn, Christopher, som alltid er en solstråle. Du er det beste mamma har!

Takk til familien som alltid støtter meg i det jeg gjør og til Ivar for korrekturlesning, tålmodighet og oppmuntrende ord.

Jeg er også dypt takknemlig ovenfor alle de 20 personene som takket ja til å delta i masterprosjektet mitt.

Til slutt en takk til veileder Jon Joensen for konstruktive tilbakemeldinger, gode diskusjoner og nyttige tips.

*Silje Ingebrigtsen*

Bergen, oktober 2012

## Abstract

Is ultrasound imaging a reliable method to use when measuring cross section tendon thickness of the supraspinatus and biceps tendon longus in persons with and without shoulder pain ?

**Background:** Rotatorcuff pathology is the most common type of shoulder disorder causing pain and loss of function. The use of diagnostic ultrasonography examination in shoulder problems are increasing, but the diagnostic accuracy of US for rotatorcuff tendinopathy is yet unclear. Acceptable reliability is required to include measuring tools as part of an examination. The purpose of this study was to evaluate the interrater- and the test-retest reliability of ultrasound imaging when measuring tendon thickness of the supraspinatus and the biceps.

**Methods and subjects:** A total of 20 participants with a mean age of 36.7 years (range 18-56) were included in the study. Participants (n=10) who had experienced shoulder pain lasting more than 3 months were included in the pain group (6 women/4 men). Participants (n=10) who were healthy were included in the healthy group (5 women/5 men). Two physiotherapists did a total of 1280 measurements of tendon thickness in the supraspinatus and the biceps. The analyzes of the interraterreliability was made on the basis of both raters measurements and re-measurements on the same ultrasound image from two repeated picture takings. An investigation of test-retest reliability was made for each rater by measurements on their own pictures. Outcome measures used are intraclasscorrelationcoefficient ICC (1.1) and  $\pm 1.96$  times the standard deviation of the difference ( $\pm 1.96SD_{diff}$ ). The relative reliability was graded into three categories: good reliability = 0.76 to 1.00, moderate reliability = 0.50 to 0.75, poor reliabiliet=0.00 to 0.49.

**Results:** The relative reliability was good for biceps longitudinal in 15 investigations (range:0.78-0.99), and poor in 1 investigation (0.47). For biceps transverse relative reliability was good in 9 investigations (range:0.76-0.97) and moderate in 7 investigations (range:0.54-0.73). For supraspinatus relative reliability was good in 5 investigations (range:0.77-0.87), moderate in 8 investigations (range:0.59-0.75) and poor in two investigations (0.46 and 0.48). For rotator cuff relative reliability was



good in 3 investigations (range:0.92-0.96), moderate in 11 investigations (range:0.51-0.73) and poor in two investigations (0.27 and 0.32).

**Conclusion:** The reliability of ultrasound imaging was good in the measurement of the cross section of the biceps tendon longus. In the cross section measurements of the supraspinatus tendon the reliability was moderate. The variability was highest in the cross section measurements of the supraspinatus in the transverse plane (rotator cuff). The results of this study shows that the raters, who are both physiotherapists, after a short period of practical training were able to make accurate cross section measurements on ultrasound images of the biceps tendon longus.

## Sammendrag

Er ultralydsavbildning en reliabel metode å bruke ved måling av tverrsnittet til supraspinatus og biceps lange sene i skulderen hos personer med og uten skuldersmerter?

**Bakgrunn:** Rotatorcuffpatologi er den vanligste årsaken til skulderproblematikk som forårsaker smerte og redusert funksjon. Bruken av diagnostisk ultralyd ved undersøkelse av skulderproblematikk er økende, men den diagnostiske nøyaktigheten av UL ved rotatorcuff tendinopati er foreløpig uklar. Tilfredsstillende reliabilitet er nødvendig for å kunne inkludere måleinstrumenter som del av en undersøkelse. Hensikten med denne studien var å evaluere inter-rater- og test-retest reliabilitet ved måling av senetykkelse på biceps og supraspinatus med ultralydsavbildning som metode.

**Metode og Deltakere:** Total 20 deltakere med gjennomsnittsalder på 36.7 år (range 18-56) ble inkludert i studien. Deltakere (n=10) som hadde hatt skuldersmerter med varighet over 3 måneder ble inkludert i smertegruppe (6 kvinner/4 menn). Deltakere (n=10) som var friske ble inkludert i frisk gruppe (5 kvinner/5 menn). To fysioterapeuter gjorde tilsammen 1280 tverrsnittsmålinger av skuldresenene supraspinatus og biceps lange sene. Analysene av inter-raterreliabiliten ble gjort på grunnlag av begge raterene sine målinger og remålinger på samme ultralydbilde fra to gjentatte billedtakinger. Det ble i tillegg gjort en test-retest undersøkelse for hver rater ved målinger på sine egne bilder. Utfallsmålene som er brukt er intraklassekorrelasjonskoeffisient ICC (1.1) for relativ reliabilitet og  $\pm 1.96$  ganger standardavviket til differansen ( $\pm 1.96 * SD_{diff}$ ) for absolutt reliabilitet. Den relative reliabiliteten uttrykt i ICC ble gradert i 3 kategorier: God reliabilitet = 0.76-1.00, moderat reliabilitet = 0.50-0.75, dårlig reliabilitet = 0.00-0.49.

**Resultat:** Den relative reliabiliteten var for biceps longitudinal god ved 15 undersøkelser (range:0.78-0.99), og dårlig ved 1 undersøkelse (0.47). For biceps tverrsnitt var relativ reliabilitet god ved 9 undersøkelser (range:0.76-0.97) og moderat ved 7 undersøkelser (range:0.54-0.73). For supraspinatus var relativ reliabilitet god ved 5 undersøkelser (range:0.77-0.87), moderat ved 8 undersøkelser (range:0.59-0.75)

og dårlig ved 2 undersøkelser (0.46 og 0.48). For rotatorcuff var relativ reliabilitet god ved 3 undersøkelser (range:0.92-0.96), moderat ved 11 undersøkelser (range:0.51-0.73) og dårlig ved 2 undersøkelser (0.27 og 0.32).

**Konklusjon:** Reliabiliteten ved ultralydsmåling av tverrsnittet til biceps lange sene var god ved bilder i longitudinalt og transversalt plan. Ved tverrsnittsmålinger av supraspinatus var reliabiliteten moderat. Ved tverrsnittsmålinger av supraspinatus i transversalt plan (rotatorcuff) var variabiliteten størst. Resultatene fra undersøkelsen viser at raterene i studien, som begge er fysioterapeuter, med kun grunnleggende opplæring og uten tidligere erfaring *kan* foreta presise ultralydsmålinger av tverrsnittet til biceps lange sene.

## Innholdsfortegnelse

<b>1.0 Innledning</b> .....	<b>1</b>
1.1 Introduksjon til temaet og bakgrunn for valg av oppgave .....	1
1.2 Hensikt med prosjektet.....	4
1.3 Oppbygning av oppgaven .....	5
1.4 Problemstillinger .....	5
1.5 Oppgavens avgrensning .....	6
<b>2.0 Teori</b> .....	<b>7</b>
2.1 Teoretisk forankring.....	7
2.2. Måleinstrumenter .....	8
2.2.1 Generelt om reliabilitet .....	9
2.2.2 Målefeil/variasjon .....	10
2.2.3 Interrater- og test-retest reliabilitet .....	10
2.2.4 Relativ reliabilitet.....	11
2.2.5 Absolutt reliabilitet .....	14
2.3 Variabler og standardisering .....	15
2.4 Validitet.....	16
2.4.1 Utvalg og ekstern validitet .....	17
2.5 Smerteregistrering.....	18
2.6 Skulderens kompleksitet .....	18
2.6.1 Rotatorcuff tendinopati .....	19
2.7 Ultralydsavbildning av skuldereener.....	20
2.7.1 Måling av senetykkelse .....	22
<b>3.0 Metode</b> .....	<b>23</b>
3.1 Utvalget.....	23
3.2 Ratere .....	24
3.3 Etikk og godkjenning .....	24
3.4 Gjennomføring av undersøkelsene.....	25
3.4.1 Utstyr.....	25
3.4.2 Protokoll for undersøkelse .....	26
3.5 Protokoll for måling av senetykkelse.....	26
3.5.1 Biceps lange sene .....	27

3.5.2 Rotatorcuff tverrsnitt og supraspinatus longitudinal .....	28
3.6 Pilottesting og rating av bilder .....	28
3.7 Statistiske metoder .....	29
<b>4.0 Resultat .....</b>	<b>31</b>
4.1 Demografiske data .....	31
4.2 Grafisk framstilling .....	31
4.3 Normalfordeling .....	31
4.4 Interraterreliabilitet .....	31
4.4.1 Gradering av relativ reliabilitet .....	32
4.4.2 Absolutt reliabilitet .....	33
4.5 Test-retest reliabilitet .....	34
4.5.1 Gradering av relativ reliabilitet .....	35
4.5.2 Absolutt reliabilitet .....	36
<b>5.0 Diskusjon.....</b>	<b>39</b>
5.1 Metodediskusjon .....	39
5.1.1 Målemetoden.....	39
5.1.2 Variabilitet og standardisering .....	41
5.1.3 Utvalg og ekstern validitet .....	44
5.1.4 Datainnsamling og prosedyre for undersøkelse .....	45
5.2 Valg av statistiske metoder .....	46
5.2.1 Relativ reliabilitet.....	47
5.2.2 Absolutt reliabilitet.....	47
5.3 Resultatdiskusjon .....	48
5.3.1 Interraterreliabilitet .....	48
5.3.2 Test-retest reliabilitet .....	50
5.4 Helhetlig tilnærming og helsereelatert livskvalitet.....	51
5.5 Etske hensyn .....	53
5.6 Kliniske implikasjoner .....	53
5.7 Videre forskning .....	54
<b>6.0 Konklusjon.....</b>	<b>55</b>

## Tabellindeks

Tabell 1: Eksempel på beregning av ICC .....	14
Tabell 2: Inklusjons- og eksklusjonskriterier for prosjektdeltakelse .....	23
Tabell 3: Demografiske data .....	24
Tabell 4: IRR. Relativ og absolutt reliabilitet, frisk gruppe.....	33
Tabell 5: IRR. Relativ og absolutt reliabilitet, smerte gruppe .....	34
Tabell 6: TRR. Relativ og absolutt reliabilitet frisk gruppe rater 1 og rater 2.....	37
Tabell 7: TRR. Relativ og absolutt reliabilitet smerte gruppe rater 1 og rater 2 .....	37

## Figurindeks

Figur 1: Eks. på ulike måleinstrument for undersøkelse av skulderplager ihht ICF .....	8
Figur 2: De ulike komponentene i et observert måleresultat .....	9
Figur 3: Ultrasonography GE Logic E.....	25
Figur 4: Transversalt bildesnitt biceps lange sene .....	27
Figur 5: Longitudinalt bildesnitt biceps lange sene .....	27
Figur 6: Transversalt bildesnitt rotatorcuff .....	28
Figur 7: Longitudinalt bildesnitt supraspinatus .....	28
Figur 8: Oversikt over de ulike reliabilitetsundersøkelsene i oppgaven.....	30

## Vedlegg

**Vedlegg 1:** Informasjonsskriv og samtykkeskjema

**Vedlegg 2:** Kopi av godkjenning prosjektprotokoll, Regional Etisk Komitè

**Vedlegg 3:** Normalfordelingskurver av data i studien

**Vedlegg 4:** Bland-Altman Plot for interraterreliabilitet

**Vedlegg 5:** Bland-Altman Plot for test-retest reliabilitet

## Forkortelser

UL	- Ultralyd
ICF	- International Classification of Functioning and Health
SPADI	- Shoulder Pain Disability Index
SF-36	- Short Form 36
(Quick)DASH	- Disabilities of the Arm, Shoulder and Hand
WORC	- Western Ontario Rotator Cuff Index
MRI	- Magnetresonanstomografi
CT	- Computertomografi
IQ	- Intelligence Quotient
IRR	- Interraterreliabilitet
TRR	- Test-retest reliabilitet
SEM	- Standard error of measurement
LOA	- Limits of agreement
ICC	- Intraclass Correlation Coefficient
SD	- Standardavvik
KI	- Konfidensintervall
VAS	- Visual Analogue Scale
VAS	- Numeric Rating Scale
KMI	- Kroppsmasseindeks
SPSS	- Statistical Package of Social Science

## 1.0 Innledning

### 1.1 Introduksjon til temaet og bakgrunn for oppgaven

Muskelskjelettplager er svært vanlig i befolkningen og innebærer stor medisinsk, sosial og økonomisk belastning både for den enkelte og for samfunnet (Andersen et al, 2009, Brage et al, 2010). Statistikk fra Arbeids- og velferdsdirektoratet (NAV) viser at det i 1. kvartal 2012 var 163495 legemeldte sykefraværstilfeller grunnet muskel- og skjelettlidelser i Norge, hvorav i overkant av 37000 av disse skyldtes plager i nakke/skulder/arm. De samlede kostnadene for muskel- og skjelettlidelser i Norge er anslått til å ligge på mellom 37 og 44 milliarder kroner årlig (Ihlæbek og Lærum, 2004). Skuldresmerter er rapportert å ha en prevalens på mellom 7-34% generelt i befolkningen, og 21% hos de over 70 år (Badcock et al, 2002; Pope et al, 1997; Luime et al, 2004; Makela et al, 1999; Svebak et al, 2006; Chakvarty et al, 1990 og Walker-Bone et al, 2004). Gjennomsnittlig varighet for en episode med skuldresmerter er 10 måneder, og mellom 20-51% har fremdeles skuldresmerter etter ett år (Pope et al, 1997 og Svebak et al, 2006).

Skuldresmerter er ofte assosiert med redusert funksjon og begrensninger i aktivitet og deltakelse (Van Der Heijden, 1999). Daglige aktiviteter som av-/og påkledning, å spise, sove og fritidsaktiviteter eller å gjøre aktiviteter med armene over hodehøyde og å være i arbeid vanskeliggjøres. Sykefravær kan medføre at mennesker mister en viktig arena for fellesskap og mestring. Videre er personer med skuldresmerter vist å skåre vesentlig lavere enn normale verdier for friske personer på SF-36, et generisk måleinstrument konstruert for å gjøre en rask vurdering av helse relatert livskvalitet (Hill et al, 2010).

Patologi av rotatorcuff er antatt å være den hyppigste grunnen til skuldresmerter - rundt 60 prosent av skuldresmerter skyldes problematikk tilknyttet senene (Lewis, 2009; Murrell og Walton, 2001). Normal funksjon i skulderen forutsetter ett komplekst samspill av mobilitet og stabilitet. I tillegg til skulderens kompleksitet er pasientgruppen med skuldresmerter en heterogen gruppe. Dette gjør det utfordrende å diagnostisere og klassifisere skulderplager, og dermed også å finne en optimal behandling for disse plagene (Awerbuch, 2008). Diagnostiske kriterier for ulike



diagnoser knyttet til skulderen har vist seg å ikke være gjensidig utelukkende, og flere studier har vist dårlig reproduserbarhet av blant annet Cyriax's diagnostiske klassifikasjonssystem (Buchbinder et al, 1996; de Winter et al, 1999; Hayes og Petersen, 2003 ). De fleste kliniske testene av enkelte anatomiske strukturer er i tillegg upålitelige (Beaudreuil et al, 2009) og kan heller ikke nøyaktig diagnostisere rotatorcuff patologi (Hughes et al, 2008). Degenerative forandringer av senefiberstruktur forekommer uten å gi symptomer og det er store forskjeller i senetykkelse fra individ til individ (Bjordal et al, 2003). Ved skulderproblematikk kan det dermed enkelte ganger være vanskelig å skille mellom naturlige individuelle variasjoner og viktige avvik.

Over halvparten av pasientene som blir diagnostisert med skuldersmerter av allmennleger, blir henvist til fysioterapi (van der Windt et al, 1995). Målet med kliniske undersøkelser er å samle informasjon som kan bli brukt som veiledning i vurdering av prognose og behandling. Klinisk undersøkelse av skulder kan med fordel kombineres med ultralydsavbildning. Ultralydsavbildning er regnet for å være en god metode for kartlegging av anatomiske strukturer og for å vurdere patologiske endringer i muskel- og skjelettapparatet (Jacobson og van Holsbeck, 1998 og Bolvig et al, 2011). Det er i litteraturen beskrevet høy sensitivitet og spesifisitet for påvisning av rotatorcuffrupturer og patologiske forhold i den lange bicepssenen med ultralyd (Teefey et al, 2000 og Bouffard et al, 2000) og MR og ultralyd er dokumentert likeverdige i vurderingen av skuldresener (Rutten et al, 2010). Ultralydsavbildning er en enkel, sikker og kostnadseffektiv metode. Metoden er strålefri, ikke-invasiv og er lett portabelt sammenlignet med eksempelvis magnetresonanstomografi (MRI) og computertomografi (CT). Det finnes ingen kjente bivirkninger med ultralydsavbildning og det har fordeler som blant annet muligheten til gjennomføre sanntids avbildning og bildediagnostikk i flere plan (Bolvig et al, 2011). Ultralydsavbildning gjennomføres i direkte samhandling med pasienten, noe som gjør at undersøkelsen kan spisses nøyaktig til det symptomatiske området. Den direkte samhandlingen mellom kliniske symptomer og bildefunn er uvurderlig, spesielt med tanke på at moderne bildediagnostikk ofte avslører asymptomatisk patologi (Bolvig et al, 2011). Fysioterapeuter kan på en annen måte enn radiologen sette sammen resultatene fra anamnesen, den kliniske undersøkelsen og sanntids ultralydsundersøkelse til ett helhetsbilde.

Diagnostisk ultralyd ble introdusert på 70-tallet, men det har skjedd en betydelig utvikling av metoden siden den gang (Bolvig et al, 2011). Det er først de siste 10-15 årene at teknologien er blitt god nok til å kunne brukes ved undersøkelse av muskel- og skjelettområdet. Moderne ultralydutstyr med lineære transdusere med høy oppløsning (fra 10 MHz til > 20MHz), kombinert med tekniske nyvinninger som økt kontrastoppløsning og støyreduksjon har ført til en bildekvalitet som for få år siden ikke var mulig (Moosmayer et al, 2004; Bolvig et al, 2011; Bjordal et al, 2003; Le Corroller et al, 2008; Ohrndorf et al, 2010 og Allen, 2008).

Moosmayer et al (2004) og Bolvig et al (2011) hevder at metoden er både operatør og utstyrs-avhengig og at den har en lang læringskurve. For å beherske metoden vil det være nødvendig med standardisert teknikk, utstyr og god opplæring. De sier videre at den største "faren" ved ultralydsavbildning er feildiagnostisering grunnet tekniske eller operatør betingede begrensninger (Bolvig et al, 2007 og Jacobson, 2011). Andre studier viser derimot til at det er viet en overdreven respekt til læringskurven (Read og Perko, 1998; Rutten et al, 2010; Coroller et al, 2008; Ohrndorf et al, 2010 og Bjordal et al, 2003). Til tross for at operatør avhengighet hyppig er nevnt som en begrensning ved ultralydsavbildning og også er den mest sannsynlige grunnen til variasjon i realibilitet, er det relativt få studier som har undersøkt intertaterreliabilitet ved ultralydsavbildning (O'Connor et al, 2005). Etersom alle studier på området er gjennomført i sekundærhelsetjenesten, er den kliniske nytten og den prognostiske validiteten av ultralydsavbildning foreløpig ikke etablert i primærhelsetjenesten (Ottenheim et al, 2010). Pasienter med mistanke om skulderpatologi, som eksempelvis rotatorcuffruptur, blir per i dag henvist videre fra primærhelsetjenesten til spesialisthelsetjenesten for videre diagnostisering og utredning.

Både i forskning og i klinisk sammenheng kreves det reliable og valide måleinstrumenter for å kunne gjøre sikker diagnostikk og evaluere om en intervensjon har hatt effekt. Evidens bør innarbeides i alle aspekter av fysioterapi som undersøkelse, evaluering, diagnose, prognose og intervensjon. Som fagutøvere må vi kritisk vurdere om resultatene fra måleinstrumentene er anvendbare og overførbare til vår fysioterapi praksis (Jamtvedt et al, 2007). Måleinstrumenter skal fungere både i klinikken så vel som i forskningsøyemed og gi pasienten og behandlere mulighet til en felles forståelse og målsetning. For arbeidsgivere og trykdevesen kan et godt

dokumentert og klinisk brukbart måleinstrument bidra til tidlig identifisering og igangsetting av tiltak for pasienter med risiko for å utvikle langvarige plager og dermed kan reliable måleinstrumenter ha store økonomiske konsekvenser. Viktige beslutninger om pasienter gjøres ofte med bakgrunn i resultater fra måleinstrumenter. Hvis et måleinstrument har for stor variasjon kan vi egentlig ikke stole på det. Før det kan implementeres, må det dermed testes for både reliabilitet og validitet. Reliabilitet refererer til konsistensen av en test eller et måleinstrument og sier noe om i hvilken grad en test er fri for systematiske og tilfeldige målefeil (Weir, 2005 og Solberg, 2002) Validitet sier noe om i hvilken grad resultatene fra en studie er gyldige, hensiktsmessige og meningsfulle i forhold til det vi ønsker å undersøke (Domholdt, 2005).

## 1.2 Hensikt med prosjektet

Vi har i klinikken behov for å gi hver enkelt pasient optimal behandling og ivareta dem på alle funksjonsnivå. I tillegg til objektivt å kunne måle sener med tendinopati, vil kvantitative tiltak med ultralydsavbildning kunne legge til rette for ny forskning for å identifisere risikofaktorer og for å forebygge progresjon av lidelsen. En økt presisjon i forhold til lokalisering av problemområdet vil kunne føre til en mer målrettet og presis behandling noe som igjen vil kunne gi en samfunnsøkonomisk gevinst i tillegg til økt livskvalitet hos pasienten.

Formålet med studien er å undersøke hvorvidt to fysioterapeuter uten tidligere erfaring og med kun grunnleggende opplæring innen diagnostisk ultralyd, kan oppnå god intrarater-, interrater- og test-retest reliabilitet i observasjonen av senetykkelse på suprapinatus- og biceps tendon longus. Denne mastergradsoppgaven er en del av ett prosjekt som inneholder to oppgaver. Datainnsamling har blitt gjennomført av forfatter av denne oppgaven i samarbeid med en medstudent. Det er i studien valgt å gjøre to sett med målinger på hver enkelt sene. Ved test-retest er det i de to masteroppgavene benyttet ulike sett med målinger for å vurdere grad av reliabilitet. Medstudent vil bruke innsamlede data for å beskrive grad av test-retest (1. måling) - og intraraterreliabilitet (alle målinger). I denne oppgaven vil jeg beskrive grad av interraterreliabilitet (alle målinger) og test- retest reliabilitet (2. måling). Kapittel om metode vil inneholde en nærmere beskrivelse av dette.

### 1.3 Oppbygning av oppgaven

Oppgaven er bygd opp som en forskningsstudie. Videre i dette kapittelet vil problemstillingen presenteres og det redegjøres for oppgavens avgrensninger. I neste kapittel vil det først redegjøres for den teoretiske forankringen til studien. Videre vil aktuell teori omkring reliabilitet, validitet, skulderleddet og ultralydsavbildning bli redegjort for. Metodekapittelet vil inneholde beskrivelse av de inkluderte deltakerne, testprotokoll for undersøkelse, testprotokoll for måling av senetykkelse og statistisk analyse. I resultatkapittelet vil det foreligge beskrivelse av resultater og fremstilling av dette ved hjelp av tabeller. I diskusjonskapittelet vil undersøkelsesmetode og resultater bli drøftet opp i mot aktuell litteratur og den definerte problemstillingen. Styrker og svakheter ved studien, kliniske implikasjoner og videre forskning vil også bli belyst. Til slutt vil konklusjonen inneholde en oppsummering av de viktigste funnene og om ultralydsdiagnostikk på skuldereener har en tilfredsstillende reliabilitet til å kunne implementeres som en del av den fysioterapeutisk skulderundersøkelse i primærhelsetjenesten. Vedlagt er godkjenning fra REK-Vest, informasjonsskriv og samtykkeskjema, normalfordelingskruver av data i studien samt Bland-Altman plot for de ulike undersøkelsene av absolutt reliabilitet i studien.

### 1.4 Problemstillinger:

I denne oppgaven vil grad av intertaterreliabilitet og test-retest reliabilitet til ultralydsavbildning ved tverrsnittsmål på skuldereener hos 20 deltakere bli vurdert. Med bakgrunn i dette har følgende problemstillinger blitt definert:

1. "Å estimere intertaterreliabilitet hos to testere ved tverrsnittsmål av skuldereenerne supraspinatus og biceps tendon longus på ultralydsbilder"
2. "Å estimere test-restest reliabilitet hos to testere ved tverrsnittsmål av skuldereenerne supraspinatus og biceps tendon longus på ultralydsbilder".

## 1.5 Oppgavens avgrensning

Ultralydsavbildning er ett måleinstrument som kan kategoriseres under kroppsstruktur- og funksjonsnivå i henhold til World Health Organization's (WHO) International Classification of Functioning, Disability and Health (ICF). Det er ikke valgt å inkludere måleinstrumenter som kartlegger deltakerenes aktivitets- og deltakelsesnivå grunnet omfang av oppgave og begrensede tidsressurser. Hovedfokus i oppgaven er på reliabilitet, og dette vil dermed vektlegges i større grad enn validitet i drøftingsdelen av oppgaven.

## 2.0 Teori

### 2.1 Teoretisk forankring

Som fysioterapeut befinner man seg hele tiden i skjæringsfeltet mellom humaniora, samfunnsvitenskap og naturvitenskap (Thornquist, 2003). Å gjennomføre forskningsbaserte studier kan på mange måter være utfordringsfullt. En forskers grunnleggende posisjon kommer til uttrykk i hva han eller hun velger å undersøke, og i hvordan den innsamlede informasjonen bearbeides, tolkes og settes inn i en større sammenheng. Det er viktig å være bevisst det komplekse ved praksis. Forskning bør med andre ord være gjenstandsspesifikk og fenomener må kontekstualiseres (Thornquist, 2003).

Den medisinske kulturen har i liten grad vært opptatt av vitenskapsteori og grunnlagsspørsmål, noe som gjenspeiles i faglitteratur og forskning. Dagens biomedisin er i hovedsak forankret i en empiristisk tenkning, hvor objektiv og tallfestet informasjon fremskaffet via teknologisk utstyr har høyest prestisje (Thornquist, 2003). Det medisinske anliggende har på mange måter blitt et spørsmål om presisjon og nøyaktighet. Den empiristiske retningen hevdet at det bare er den fysiske, observerbare og målbare verden vi kan utvikle vitenskapelig kunnskap om (Thornquist, 2003). Empirismen bidro dermed til å skape en kropp-/sjel dualisme. Fenomenologien som senere oppstod, hadde en mer erfaringsbasert retning hvor personen og hvordan verden erfares for denne står i fokus (Thornquist, 2003). Fenomenologi, som tar sterk avstand til kropp-/sjel dualismen som empirismen representerer, har utviklet seg til å bli en viktig grunntanke i helsefag. Forskning som har et biopsykososialt fokus vil være forankret i en fenomenologisk forståelse, hvor utgangspunktet er at mennesker lever i en menings- og betydningsfull verden, og at kropp og sjel ikke kan skilles (Thornquist, 2003).

I denne studien er det inkludert et måleinstrument hvor resultat i form av objektive tallverdier vies mye oppmerksomhet. Selv om det ikke finnes noen prinsipiell motsetning mellom å foreta målinger ved hjelp av ultralydsavbildning og å dokumentere meningsfulle fenomener, synes det viktig å understreke at slike bilder og målinger kun er en del av den fysioterapeutiske undersøkelsen for å vurdere pasienter

med skuldersmerter. *Hva* som gjøres til gjenstand for kvantifisering og om det er *samsvar* mellom fenomeners vesen og forskningsmessig tilnærming, er avgjørende. Det er med andre ord viktig at det å vurdere reliabilitet ikke går på bekostning av relevans og validitet.

I 2001 utviklet World Health Organization (WHO) en biopsykososial modell kalt "International classification of functioning, disability and Health" (ICF) (Domholdt 2005). Det begrepsmessige grunnlaget for klassifikasjonen omfatter to hovedområder: funksjon/funksjonshemming og kontekst. Hensikten er å klassifisere menneskers helse slik den kommer til uttrykk i funksjon og funksjonshemming, både i forhold til kroppsfunksjoner og kroppsstrukturer på den ene side, og til dagliglivets aktiviteter og deltakelse i sosial sammenheng på den andre side. I tillegg klassifiserer ICF miljøfaktorer som kan påvirke funksjon og funksjonshemming. En helhetlig skulderundersøkelse bør inneholde måleinstrumenter innenfor alle de 3 nivåene i ICF- modellen slik at pasientens situasjon vurderes i sin helhet. Både subjektive og objektive mål bør vektlegges (Figur 1).

Figur 1: Eksempel på ulike måleinstrument for undersøkelse av skulderplager inndelt ihht ICF

Kroppsfunksjon og -strukturer	Aktivitet	Deltakelse
UL- avbildning	Simple shoulder test	WORC
ROM	SPADI	QuickDASH
Smerte	SF-36	SF-36

## 2.2 Måleinstrumenter

Både i forskning og i tilnærming til pasienter er det svært viktig at de målemetoder man benytter seg av er reliable og valide (Atkinson og Nevill, 1998). For å vite om et måleinstrument er anvendelig som evalueringsverktøy må datakvaliteten til instrumentet vurderes. Dette krever en systematisk testing av måleegenskapene til instrumentet, og kan ses på som en kontinuerlig flertrinnsprosess. Måleegenskapene analyseres ved å vurdere instrumentets reliabilitet og instrumentets validitet. Man begynner med å teste ut reliabiliteten, en metode som ikke er reliabel kan heller ikke være valid (McDowell og Newell, 1996). Tester som benyttes i klinisk praksis må være reliable innen testere (intratester), mellom testere (intertester) og mellom

forskjellige observasjoner (test-retest). Jo høyere reliabiliteten er, jo lettere er det å påvise reelle forskjeller og sammenhenger mellom uavhengige målinger.

### 2.2.1 Generelt om reliabilitet

American Educational Research Association (1999) definerer reliabilitet som *"The degree to which test scores are free from errors of measurement"* (referert i Domholt, 2005, s. 255). På norsk oversettes reliabilitet ofte til pålitelighet. Krav til reliable skåringer er repeterbarhet og stabilitet. Med repeterbarhet menes at målingene skal gi de samme skårene ved gjentatte undersøkelser med samme metode på samme subjekt under identiske forhold. Målingen må også være stabil. Med det menes at ulike ratere skal kunne skåre det samme uavhengig av tidspunkt og i hvilket miljø undersøkelsen foretas (Streiner og Norman, 2008). Reliabiliteten til et måleinstrument omhandler altså dets evne til å produsere like resultater når repeterte målinger foretas under identiske forhold (Jamtvedt et al, 2007).

Reliabilitet blir ofte vurdert i forhold til to ulike teorier: den klassiske teorien og generaliserbarhetsteorien (Domholdt, 2005). Førstnevnte går ut i fra at enhver måling har en sann måling i tillegg til en feilkomponent. Dette er illustrert i figur 2.

Figur 2. De ulike komponentene i et observert måleresultat



Den sanne skåren for en måling kan man aldri vite med sikkerhet og sammenhengen mellom repeterte målinger må derfor brukes for å estimere graden av målefeil.

Målefeil er beskrevet som variasjon mellom flere målinger. Variabiliteten mellom ulike målinger av en person vil alene skyldes feilkomponenten.

Generaliserbarhetsteorien sier på den annen side at det er ulike kilder for variabilitet for alle målinger. Her deles målefeil inn i grupper av variabilitet, eller ulike fasetter.

Det vil si at en måling kan bli påvirket av flere faktorer, som for eksempel



individuelle biologiske variasjoner hos testpersonen, ferdigheten til undersøkeren, eller nøyaktigheten til apparatet som brukes i målingen (Domholdt, 2005).

### 2.2.2 Målefeil/variasjon

Alle kvantitative måleresultater består av to komponenter; en som reflekterer det som blir målt og en som representerer feilkilder (Domholdt, 2005). Hensikten med reliabilitetsteori er å anslå målefeilene og foreslå metoder for å minimere disse. Dette vil si at når man undersøker reliabilitet undersøker man i hvilken grad måleresultatene er fri for systematiske og tilfeldige målefeil. Systematiske feil er ett uttrykk for at måleverdiene viser en generell endring i en retning (Atkinson og Nevill, 1998). En systematisk feil er en feil det er mulig å identifisere og forutsi.

Den andre komponenten av variasjon er tilfeldige målefeil som refererer til ”støy” i målingene (Streiner og Norman, 2008). Tilfeldige målefeil er utforutsigbare og kan påvirke resultatet forskjellig fra gang til gang. Slike tilfeldige feil kan skyldes biologisk variabilitet hos testpersonene, mekanisk variabilitet i målemetoden og/eller ikke tilfredsstillende testprotokoll (Atkinson og Nevill, 1998). De fleste målinger er utsatt for en kombinasjon av instrumentfeil, testers feil og subjektets variasjon (Domholdt, 2005). I tillegg er komponenten av tilfeldige målefeil vanligvis større enn systematiske feil (Atkinson og Nevill, 1998).

### 2.2.3 Interrater- og test-retest reliabilitet

Det finnes ulike typer reliabilitet. I denne oppgaven anvendes interraterreliabilitet (IRR) og test-retest reliabilitet (TRR). Interraterreliabilitet er et mål på samsvar mellom to eller flere ratere som evaluerer samme subjekt ved samme tidspunkt og som bruker samme informasjon, som i denne studien er samme ultralydsbilde. Interraterreliabilitet måler altså instrumentets stabilitet ”på tvers” av ulike ratere (Domholdt, 2005). Om raterene kommer frem til samme skåre, tyder dette på høy eller god interraterreliabilitet. Test-retest vil i denne studien være et mål på samsvar hvor det kreves at samme rater undersøker samme deltaker med instrumentet på to forskjellige tidspunkt. Fordelen med denne tilnærmingen er at den gir et reelt bilde av raterens stabilitet over tid. Den betinger imidlertid at deltakerens skuldere som

målingene gjøres på ikke endrer seg nevneverdig. I tillegg må tidsintervallet være lite nok til at det er minimal, eller ingen sjanse for at læringseffekt hos rateren finner sted, men stort nok til at rateren ikke husker hva som ble målt ved den første målingen (Blacker og Endicott, 2003). Fokus for reliabiliteten i denne studien er å undersøke målefeil relatert til rateren og/eller testprosedyren.

#### 2.2.4 Relativ reliabilitet

Reliabilitet kan kvantifiseres på to måter; relativ reliabilitet og absolutt reliabilitet (Domholdt, 2005). Tanken bak relativ reliabilitet er at dersom et måleinstrument er reliabelt, vil individuelle målinger innenfor en gruppe beholde samme posisjon i gruppen ved gjentatte målinger (Domholdt, 2005). Et eksempel kan være at dersom en deltaker skårer nær toppen av fordelingen ved første måling vil denne deltakeren være forventet å holde seg nær toppen av fordelingen, til tross for at skåren ved måling endres fra tid til annen.

Relativ reliabilitet kan beskrives med en korrelasjonskoeffisient kalt Intraclass Correlation Coefficient (ICC) (Fisher, 1925). ICC finnes i flere former, og brukes ved måling av kontinuerlige variabler. Koeffisienten benyttes når variablene som skal sammenlignes befinner seg i samme klasse, altså der variablene og variansen har den samme metriske målestokk. Hvis variablene ikke er i samme klasse, vil Pearsons  $r$ , som er en interklassekoeffisient, være den foretrukne korrelasjonskoeffisient (McGraw og Wong, 1996). Pearsons  $r$  benyttes dersom man for eksempel ønsker å måle korrelasjonen mellom høyde og vekt. De metriske verdiene (centimeter og kilogram) er ikke de samme, og en kan heller ikke anta at variansen mellom de to datasettene er sammenfallende. Pearsons  $r$  er dermed heller ikke egnet til å avdekke systematiske målefeil (Domholdt, 2005). Hvis en ønsker å måle korrelasjonen mellom resultatene til to ratere ved måling av senetykkelse på ultralydsbilder, vil ICC være den foretrukne korrelasjonskoeffisient, da både de metriske variablene (mm) og variansen er de samme.

Definisjonen av 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*” (Moe-Nilssen, 1998). ICC beskrives med andre ord

som forholdet mellom den sanne variansen og den totale variansen som skyldes målefeil, og indikerer graden av assosiasjon mellom gjentatte målinger av variablene vi er interessert i. Hvis ICC er 1.0 er forbindelsen mellom gjentatte målinger ideell - hvis ICC er 0.0 finnes det ingen sammenheng mellom variablene (Domholdt, 2005). Nivået for nedre aksepterte grense av ICC kan variere avhengig av hva en ser på. Hva som defineres som et "akseptabelt" nivå av reliabilitet må sees i sammenheng med faktorer som valg av ICC modell, nøyaktigheten til måleinstrumentet, subjektene vi måler og hva vi skal bruke resultatene til. Ulike ICC verdier for hva som regnes som høy og lav reliabilitet kan dermed variere og akseptabel reliabilitet må bestemmes av klinikerne som utfører testene for at de skal få klinisk relevans (Van Genderen et al, 2003). Domholdt (2005) understreker at en vurdering om reliabiliteten til ett måleinstrument er akseptabel eller ikke hovedsakelig bør være klinisk fundert. En allmenngyldig enighet kan dermed ikke være basert på ICC alene, noe som vil beskrives nærmere etterhvert.

Ulike forfattere viser til ulike verdier for å vurdere om reliabiliteten kan beskrives som høy eller lav. En generell retningslinje er å betrakte ICC verdier under 0,50 som dårlig, verdier fra 0.50-0.75 som moderat og ICC verdier over 0.75 som god reliabilitet (Portney og Watkins, 2000). Valgt nivå for gruppering av ICC i denne oppgaven vil følge den generelle retningslinjen. Denne graderingen av ICC verdier vil være hensiktsmessig å bruke til kliniske problemstillinger som rotatorcuffpatologi, der resultatene skal benyttes til å vurdere indikasjoner på tendinopati. Ved undersøkelse og behandling av alvorlige sykdommer som kreft, ville nedre grense for akseptabel og god ICC antakeligvis vært høyere enn 0.75. Reliabiliteten til målemetoder som skal brukes til diagnostikk og råd for videre behandling på individnivå krever som regel høyere reliabilitet enn for tester som skal brukes til et beskrivende formål alene (Streiner og Norman, 2008).

En skiller mellom enveis- og toveis ICC-modeller, avhengig av om det er en eller to kilder til systematisk varians (Shrout og Fleiss, 1979). ICC (1.1) er basert på antagelsen at variasjonen i målingene til den enkelte deltager alene skyldes målefeil. Denne modellen kalles "one way random" i SPSS. En toveis ICC-modell gir mulighet til å samtidig undersøke to eller flere uavhengige faktorer for en og samme avhengige faktor. I denne studien vil deltakerne være en av de uavhengige faktorene. Ved test-

retest vil de to ulike undersøkelsesrundene være den andre faktoren, mens ved undersøkelse av interraterreliabilitet vil raterene være den andre faktoren. ICC (3.1) antar at variabiliteten i tillegg kan skyldes læringseffekt eller tilvenningseffekt uten å anta at systematiske feil er del av målefeilen. Dette betyr at  $ICC(1.1) = ICC(3.1)$  når det ikke forekommer noen læring eller tilvenning (Moe-Nilssen, 1998). ICC (3.1) kalles "two way mixed" i SPSS og indikerer at raterene er sett på som avhengige faktorer mens deltakerne er sett på som uavhengige faktorer. Med andre ord er ett bestemt sett med ratere de eneste en er interessert i å vurdere, noe som er fornuftig hvis man kun vil ha ICC verdien til enkelte ratere i et studie og ikke ønsker å generalisere utover disse. Raterene vil da teste ett bestemt sett med deltakere som er trukket fra et potensielt ubegrenset sett med slike deltakere. På denne måten vil ICC (3.1) kunne vurdere reliabiliteten av kun disse raterene, som om de ville ha testet alle slike deltakere. I denne studien vil ICC (3.1) benyttes ved undersøkelse av test-retet reliabilitet, i tillegg til ICC (1.1).

ICC (2.1) er utviklet for å vurdere grad av enighet mellom ratere, og om raterene er "interchangeable", altså utskiftbare. Denne kalles "two way random" i SPSS (Shrout og Fleiss 1979; McGrow og Wong, 1996). ICC (2.1) er basert på en "two-way random effects model" der både raterene og deltakerne er sett på som uavhengige faktorer. Med dette menes at ett bestemt sett med ratere er trukket fra en større og ubegrenset populasjon med potensielle ratere. Disse bestemte raterene vil teste ett bestemt sett med deltakere trukket fra en potensielt ubegrenset mengde av slike deltakere. På denne måten vil ICC (2.1) kunne gjelde alle slike ratere som rater alle slike type deltakere. I denne studien vil ICC (2.1) benyttes ved undersøkelse av interraterreliabilitet, i tillegg til ICC (1.1).

ICC estimatet kan oppgis som single measure reliabilitet (ICC 1.1), basert på enkeltmålinger, eller som average measures reliabilitet (ICC 1.k) som er gjennomsnittet av  $k$  målinger (McGraw og Wong, 1996). Graden av sammenheng mellom målinger kan uttrykkes som konsistens eller som absolutt samsvar. Ved konsistens er systematiske målefeil mellom raterene ikke tatt med i beregningen, mens ved absolutt samsvar er det tatt hensyn til disse variasjonene.

Følgende eksempel (tabell 1) illustrerer dette: Tre ratere gir to deltakere følgende skåringer.

Tabell 1. Eksempel på beregning av ICC

	Rater 1	Rater 2	Rater 3
Pasient 1	4	6	8
Pasient 2	6	8	10

Ved å estimere ICC (2.1) ut fra et ønske om å måle konsistens, oppnår man en ICC-verdi på 1.0. Rater 1, 2 og 3 er perfekt konsistente i sine vurderinger. Ved å estimere ICC (2.1) ut fra et ønske om å måle absolutt samsvar, oppnår man en ICC verdi på 0.67. Rater 1, 2 og 3 har dermed ikke absolutt samsvar i sine målinger.

ICC vil ikke beskrive nøyaktigheten av målingene uttrykt i en gitt målenhet (Domholdt, 2005). Med tanke på at ICC også i stor grad påvirkes av heterogeniteten til utvalget, er det ikke anbefalt å bruke ICC alene som eneste mål på reliabilitet. Det bør i tillegg benyttes et mål på absolutt reliabilitet, som kan ha en større praktisk nytteverdi (Domholdt, 2005).

### 2.2.5 Absolutt reliabilitet

For å styrke reliabilitetsinformasjonen er det viktig å beskrive omfanget av variasjon på en repetert måling. Dette kan måles med blant annet *limits of agreement* (LoA) (Bland og Altman, 1986). Absolutt reliabilitet undersøker variabiliteten mellom skåringene fra måling til måling, altså i hvilken grad repeterte målinger varierer for ett individ. Det er viktig å vite den absolutte reliabiliteten til et måleinstrument blant annet for å kunne evaluere om variasjonen i målingene skyldes målefeil. Bland og Altman introduserte metoden "*limits of agreement*" (LoA) som gir 95 % sikkerhet for at den sanne scoren finnes innen  $\pm 1,96$  ganger standardavviket til differansen av målingene ( $SD_{diff}$ ) (Bland og Altman, 1986). Dette er definisjonen på en repeterbarhetskoeffisient vedtatt av British Standards Institution (BSI, 1975). Metoden baserer seg på å analysere forskjellene mellom de parvise skårene enten mellom ratere (interrater) for hvert individ eller fra test-retest for hvert individ. Disse forskjellene blir plottet mot de respektive individuelle gjennomsnittene i et

Bland-Altman plot. En grov indikasjon på systematiske målefeil blir gitt ved å undersøke retningen og størrelsen av spredningen rundt null-linjen. Standardavviket ( $SD_{diff}$ ) til differansen mellom målingene brukes for å danne øvre og nedre grense (LoA) for hvor de to målingene samsvarer (Bland og Altman, 1986). Repeterbarhet er viktig med tanke på å måle grad av samsvar mellom to ratere, da raterenes evne til å repetere sine egne målinger vil begrense graden av samsvar som er mulig. Hvis repeterbarheten er dårlig for én rater, til eksempel hvis det er betydelig variasjon i repetererte målinger på samme sene, så vil grad av enighet mellom de to raterene også bli dårlig (Bland og Altman, 1986).

Absolutt reliabilitet er variasjon angitt på den skala målingene er tatt på. Klinisk sett vil dette være nyttig, da man kan anvende disse verdiene ( $X \pm SD$ ) direkte i praksis (Domholdt 2005 s. 258). Jo mindre SD er, jo bedre vil reliabiliteten til målingene være (Atkinson og Nevill, 1998). Små variasjoner innen hvert individ legger til rette for å oppdage små, men likevel potensielt klinisk meningsfulle forskjeller i variabelen vi måler (Batterham og George, 2003). Det er ikke alltid slik at en høy korrelasjonskoeffisient er assosiert med ett lite standardavvik (SD) eller en lav korrelasjonskoeffisient med et stort SD. Det er derfor viktig å få med informasjon om begge disse i en presentasjon av resultater (Altman, 2000; Domholdt, 2005). Relativ og absolutt reliabilitet utfyller hverandre og anbefalingene er at man i en reliabilitetsanalyse vurderer variabiliteten i forhold til begge (Domholdt, 2005). Hvilken statistisk metode man skal benytte ved analysering av reliabilitet blir blant annet bestemt av om dataene er *heteroscedastiske* eller *homoscedastiske*. Et datamateriale er heteroscedastisk dersom endringer i nivå av målte verdier systematisk fører til høyere eller lavere nivå på den målte variabiliteten. Dersom det ikke fins noen sammenheng mellom størrelsen på måleverdiene og målt variabilitet er datamaterialet homoscedastisk. Ett normalfordelt, homoscedastisk datamateriale kan analyseres ved parametrisk analyse (Atkinson og Nevill, 1998).

### 2.3 Variabler og standardisering

Ved planlegging og gjennomføring av en reliabilitetsstudie er det viktig å ta høyde for de mange faktorer som kan spille inn og påvirke reliabiliteten. Variasjonene kan reduseres ved å standardisere tester og trene raterene til å utføres tester mest mulig likt

(Strenger et al, 1997; Domholdt, 2005). Nøyaktigheten og stabiliteten til raterene avhenger blant annet av kjennskap og erfaring til målemetodene. Som for forsøkspersonene som blir målt i en funksjon eller en aktivitet, vil det også kunne forekomme en læringseffekt blant raterene når ett instrument anvendes. Det vil derfor være viktig å inkludere en form for opplæring eller pilottesting i design av studien. Videre vil nøyaktighet i beskrivelse av protokoll for undersøkelse og utførelse av målinger være viktig for reproduserbarheten og for at målemetodene skal kunne implementeres i klinisk praksis (Van Genderen et al, 2003). Reproduserbarheten til ett ultralydsapparat er et uttrykk for at like resultater skal kunne oppnås med samme metode på "identiske" subjekter, men i ulike laboratorier med ulike ratere som bruker ulike ultralydsapparater (BSI, 1975).

Det er viktig at forskeren identifiserer flest mulig variabler som vil være viktige å ta hensyn til i en reliabilitetsstudie (Domholdt, 2005). Variablene kan settes i sammenheng med instrumentet, raterene eller deltakerne. Variabler som kan virke inn på reliabiliteten til ultralydsapparatet kan være en ikke-optimal frekvens, artefakter eller instrumentfeil. Raterene kan på sin side ha varierende dagsform og motivasjon, ulik plassering av transduseren på de enkelte senestrukturene, eller det kan være unøyaktighet i testposisjon til pasient. Deltakerne kan på sin side ha varierende smertenivå. Når de ulike variablene er identifisert, er det viktig å bestemme grad av standardisering i testprotokollen (Domholdt, 2005). Graden av standardisering er hvor mange av de ulike variablene innenfor en reliabilitetsundersøkelse som kontrolleres. Dette deles som oftest inn i; ikke standardisert, delvis standardisert og høyt standardisert. Høyt standardiserte testprotokoller kontrollerer mange variabler som vil påvirke graden av reliabilitet det er mulig å oppnå i en klinisk setting. (Domholdt, 2005). Det har i denne studien vært et mål at standardiseringen av protokoll for undersøkelse og protokoll for måling av senetykkelse er reproduserbar og dermed kan anvendes i klinikken.

## 2.4 Validitet

Høy reliabilitet er nødvendig, men ikke tilstrekkelig for høy validitet. Domholdt definerer validitet som: *"The appropriateness, meaningfulness, and usefulness of the specific inferences made from test score"* (Domholdt, 2005 s. 259) og som *"The*

*meaningfulness of test scores as they are used for specific purposes.*” (Domholdt, 2005 s. 561). Et måleinstrument som ikke er reliabelt har heller ikke god validitet siden målinger med mange feil gir liten mening og nytte. Et reliabelt måleinstrument er kun valid hvis det i tillegg til å være repeterbart gir meningsfull informasjon (Domholdt, 2005). Det finnes to distinkte former av validitet: intern validitet og ekstern validitet (Domholdt, 2005). En studie har intern validitet om målingene og metodene er reliable, om målingene måler det de skal måle og om feilen mellom raterene og innenfor subjektene er små. Ekstern validitet sier noe om resultatene fra en klinisk kontrollert studie kan appliseres på en større populasjon, såkalt generaliserbarhet. I kliniske forsøk kan ekstern validitet vurderes ved at man blant annet har klare inklusjons- og eksklusjonskriterier (Peat et al, 2002). Inklusjons- og eksklusjonskriteriene i denne studien vil bli beskrevet under kapittel om metode.

#### 2.4.1 Utvalg og ekstern validitet

Den eksterne validiteten til en studie vil være avhengig av utvalget av deltakere inkludert i studien (Domholdt, 2005). I prinsippet kan man aldri generalisere sikkert til noen andre enn til den populasjonen utvalget er trukket fra. Reliabiliteten av ett instrument bør derfor vurderes ved å bruke ett utvalg deltakere med samme karakteristika som instrumentet er ment å brukes på i klinikken (Domholdt, 2005). Videre bør deltakerne som undersøkes ha variasjonsbredde i forhold til anvendt skala. Ved lite variasjon i resultatene, som for eksempel ved testing av friske individer, vil dette gi lav reliabilitets koeffisient selv om resultatene kan være nærmest like mellom første og andre måling. På den andre siden vil det å ha en ekstremt heterogen gruppe, som for eksempel ved å blande deltakere med smertefulle og ikke-smertefulle skuldre, kunne føre til at reliabiliteten generelt blir overestimert for bruk i praksis. Det mest ideelle vil dermed være å ha to ulike grupper, en hvor det er deltakere med smertefulle skuldre og en hvor det er deltakere med ikke-smertefulle skuldre, i tillegg aldersspredning og kjønnsfordeling for å oppnå en god reliabilitet for måleinstrumentene som skal vurderes. En inndeling av deltakerne i frisk/syk gruppe er aktuell også for å finne ut om metoden UL differensierer bra. Forskeren bør nøye vurdere hvilke deltakerne som skal inkluderes i en studie slik at de er mest mulig lik pasientene som målemetodene er ment brukt til i klinikken (Domholdt, 2005). For å



kunne generalisere til populasjonen er størrelsen på utvalget av betydning. Små utvalg har større sjanse for ikke å representere populasjonen og kan føre til utvalgsskjevhet.

## 2.5 Smertereistrering

The international Association for the Study of Pain definerer smerte som "*en ubehagelig sensorisk og emosjonell opplevelse som opptrer i sammenheng med vevsskade, eller blir beskrevet som om den skyldes vevsskade*" (IASP). Smerte er en subjektiv opplevelse, noe som vanskeliggjør objektiv evaluering av smerte (Danneskiold-Samsøe et al, 2011). I en klinisk hverdag og i forskning er det allikevel ønskelig å kunne evaluere smerte kvantitativt for å kunne evaluere smerteutvikling. Ulike metoder for å registrere smerte er utviklet, blant annet smerteskalaer. Smerte kan oppleves svært ulikt for to personer til tross for samme plager, noe som skyldes at smerte er en biopsykososial tilstand. Smerte kan ha ulik karakter og beskrives på ulike måter, og kvantitativ måling av smerte som standardiseres i en skala kan derfor bare sammenlignes innen samme individ.

Smerte kan måles med blant annet verbal rating scale (VRS), numeric rating scale (NRS) og visual analogue scale (VAS). Disse metodene er alle vist å være reliable og valide måter å måle smerte på (ICC range: 0.67-0.83) (Li et al, 2007). I denne oppgaven graderte deltakerne smerte før og etter testingen ved hjelp av visual analogue scale (VAS), som er en av smerteskalaene som ofte benyttes i klinikken (Li et al 2007).

## 2.6 Skulderens kompleksitet

Skulderen er utfordrende å undersøke på grunn av et stort antall strukturer, store bevegelsesutslag og de mange skadene som kan oppstå inne i eller rundt leddet (Magee, 2002). Glenohumeralleddet er et multiaksialt synovialledd som hovedsakelig stabiliseres av ligamenter og muskulatur og i mindre grad av benstrukturer.

Glenohumeralleddet er det mest mobile leddet i kroppen (Kapandji, 1982) og skulderen består av en rekke ledd som krever et nøye samspill for å kunne fungere optimalt (Solberg, 2002). Ved undersøkelse av skulder og skulderbue kan det være utfordrende å tolke funnene nøyaktig da det er flere strukturer som jobber samtidig.

Skuldersmerter og dysfunksjon kan ha ulike etiologi og patologi. Rotatorcuff tendinopati er den mest vanlige skulderlidelsen som forårsaker smerte og nedsatt funksjon, og supraspinatus er som regel involvert (Lundgreen et al, 2011). Papatheodorou et al (2006) viser i tillegg til at skulderproblematikk som ikke er knyttet til rotatorcuff`en også bør adresseres ved uspesifikke skuldersmerter, inkludert ultralydsavbildning av biceps lange sene. Problematikk knyttet til biceps lange sene kan klinisk arte seg likt rotatorcuff problematikk, og kan bli oversett ved undersøkelse (Martinoli et al, 2003). Ved skuldersmerter vil det med andre ord være viktig å undersøke både rotatorcuffsener i tillegg til biceps lange sene ved hjelp av ultralydsavbildning.

### 2.6.1 Rotatorcuff tendinopati

Tendinopati er en generell terminologi uten etiologiske, biokjemiske eller histologiske implikasjoner som blir brukt for å beskrive patologi og smerter fra sener (Maffuli et al, 2003; Carr og Harvie, 2005). Dette betyr at diagnosen tendinopati blir stilt ut fra kliniske undersøkelser, og ikke utfra histopatologiske funn (Carcia et al, 2010). Hovedsymptomet på tendinopati er smerte, som også er det første innsettende varselet om tilstanden. Klinisk sett karakteriseres tendinopati av smerte mot ytre motstand og ved palpasjon samt hevelse og redusert fysisk funksjon (Woo et al, 2007; Maffuli et al, 2004 og Paavola et al, 2005).

Akutt og kronisk overbelastning av senene kan bidra til utviklingen av rotatorcuff tendinopati (Woo et al, 2007; Maffuli et al, 2003). Tendinopatiske forandringer i sener er påvist hos rotter med ultralyd etter vedvarende repetitiv overbelastning. Disse forandringene er forenlige med endringer man observerer hos mennesker med rotatorcuff tendinopati (Soslowky et al, 2002). Overbelastning av rotatorcuff sener kan bidra til økt volum i det trange subakromiale rom, som muligens kan ha en katabolsk effekt på den intratendinøse og bursale homeostase og kan trigge en kaskade av cytokiner, neuropeptider og andre kjemikalier som er blitt identifisert i senevevet og i bursae (Lewis, 2008). Nøyaktig etiologi og patogenese ved tendinopati er imidlertid fortsatt uklar, og det trengs mer forskning for å få en fullstendig forståelse av årsakssammenhengen (Lewis et al, 2008 og Woo et al, 2007). Sannsynligvis er

patogenesen ved rotator cuff tendinopati multifaktoriell og resultat av en kombinasjon av indre (intrinsiske), ytre (ekstrinsiske) og miljømessige faktorer (Lewis, 2008; Abate et al, 2009; Fu et al, 2010).

Patologi i rotatorcuff'en inkluderer ett bredt spekter av tilstander, inkludert reversibel innflammasjon av sener, irreversibel degenerasjon av sener, delvise cuffrupturer, reversibel kalsifisert tendinopati, totale cuffrupturer og degenerativ glenohumeral artritt (Carr og Harvie, 2005). Det er en glidende overgang fra tidlige stadier av tendinopati til sluttstadiet hvor det foreligger en degenerasjon av rotatormansjett og eventuell ruptur (Aune, 2010).

## 2.7 Ultralydsavbildning av skuldersener

Ultralydsavbildning tillater en dynamisk evaluering av rotatorcuff'en med resultater i sanntid i tillegg til den viktige fordelene med bilateral undersøkelse (Carr og Harvie, 2005). Diagnostisk ultralyd er ikke en kontinuerlig lyd, men ekstremt korte impulser. Definisjonen på ultralyd er lyd med en frekvens som ligger over det mennesket kan høre (Van Hoolsbreck og Intercaso 2001 s 1). Mens ett øre kan oppfatte mellom 15000 til 20000 svigninger per sekund, har ultralyd en frekvens som ligger over dette. Valg av frekvens er avgjørende for bildekvaliteten og frekvensen blir bestemt utifra hvilke strukturer en skal undersøke. Frekvensen er direkte proporsjonal med den aksiale oppløsningen, og omvendt proporsjonal med vevspenetrasjonen til ultralydsbølgen (Jacobson, 2007). Dette betyr at man bruker høyere frekvenser til undersøkelse av overflatiske strukturer og lavere frekvenser til dype strukturer (Moosmayer, 2004). Undersøkelse av sener krever relativt høy oppløsning, da sener er overflatiske strukturer. Dette oppnås ved å bruke transdusere med høy frekvens (10-20 MHz) (Jacobson, 2007).

Ved ultralydsavbildning blir ultralydsbølger sendt inn i vevet fra en transduser. Bildet blir dannet på bakgrunn av en pulsekkomåling, som betyr at pulsen som blir sendt ut fra transduseren og igjennom mediet blir reflektert tilbake til transduseren hvor det så gir grunnlag for å dannes et bilde (Jacobson, 2007). Anatomi slik den vises ved ultralyd kalles sonoanatomi. Grad av ekko viser seg som ulike gråtoner på bildet. Forskjellige vevstyper gir ulik grad av ekko og en kan dermed skille de fra hverandre.

Veskefylte rom i vevet framstår nesten som svart. Muskulatur, som er relativt hypoekkoisk, vil ha en mørk gråtone. Beinvev vil gi sterkest ekko (hyperekkkoisk) og framstår dermed med en lys gråtone. Friske sener vil ved ultralydsavbildning fremstå med en homogen og hyperekkkoisk struktur med typisk mønster av parallelle kollagene fibre (Van Holsbeeck et al, 1994). En patologisk sene blir ofte beskrevet med et delvis eller totalt oppbrudd av det intratendinøse mønsteret, eller som en fortykkelse av senen (Paavola et al, 1998, Nehrer et al, 1997). Studier har vist at hyperekkkoiske områder i sener på ultralydsbilder korrelerer bra med patologi sett ved kirurgi (Movin et al, 1998 og Kader et al, 2002).

Ultralydbølger blir reflektert hver gang de treffer overganger mellom vev. Faktorer som virker inn på refleksjonen er innfallsvinkel på lydbølgene og akustisk motstand. Det oppnås best ekko med vinkelrett lydbølge på det reflekterende vevet. Akustisk motstand er ett uttrykk for forskjeller i vevet sin massetetthet og forskjeller i lydshastighet (Van Holsbeeck og Introcaso, 2001). Bolvig et al (2011) understreker at det ved ultralydsavbildning er obligatorisk å skanne vinkelrett på senen for å unngå artefakter i form av anisotropier. Begrepet artefakt er brukt for å beskrive hvilken som helst uønsket bildeinformasjon som blir generert i det bildet dannes (Bolvig et al, 2011). Å forstå artefakter er viktig for å unngå feiltolkning av bilder. Enkelte ganger kan artefakter bidra til å gi viktig diagnostisk informasjon, som eksempelvis ossifikasjoner eller fremmedlegemer i muskulatur (Van Holsbeeck og Introcaso, 2001) Artefakter i form av anisotropier er derimot ønskelig å unngå. Slike artefakter er sterke reflektorer som vil oppføre seg som speil hvor retningen av den reflekterte ultralyden er lik innfallsvinkelen. En konsekvens av dette er at eksempelvis sener, kan opptre både hypo-og hyperekkkoisk, avhengig av vinkelen på innsonajonen (Bolvig et al, 2011).

Bolvig et al (2011 s. 7-8) og Jacobson (2007) beskriver "gullstandarden" for undersøkelse med ultralyd på følgende måte:

- Inngående kjennskap til den kliniske settingen og til de spesifikke spørsmål som skal besvares, som i denne studien vil være tendinopati og måling av senetykkelse.

- Inngående kunnskap om anatomi
- Høy kvalitet på ultralydsutstyret
- Kjennskap til begrensningene og fallgruvene ved ultralydsavbildning

### 2.7.1 Måling av senetykkelse

Ultralydsdiagnostikk kan brukes til å måle hevelse og intratendinøs strømming, noe som også er kriteriene for å kunne stille diagnosen rotatorcufftendinopati ved ultralydsavbildning (Hoksrud et al, 2008; Maffuli et al, 1987). Individuer med kronisk tendinopati er vist å ha større tverrsnittsmål av sener i enkelte områder sammenlignet med en asymptomatisk kontrollgruppe (Leung og Griffith, 2008). Forandringer i en sene trenger imidlertid ikke alltid å være patologisk, og funn ved ultralydsdiagnostikk er observert også i asymptotiske sener (Khan et al, 2003) Forskning indikerer at friske sener etter trening vil vise endringer i intratendinøs vaskularitet, noe som kan demonstreres ved ultralydsavbildning med Doppler (Boesen et al, 2006). Det er dog ikke klart enda på hvilken måte denne endrede intratendinøse vaskulariteten kan relateres til endring i senetykkelse. I vårt studie var det ingen av deltakerne som hadde deltatt i fysisk aktivitet rett før undersøkelse med ultralyd.

Studier har vist at gjennomsnittlig tykkelse av supraspinatussenen er 6.5 mm hos friske individer. Mellom individer varierer verdien med  $\pm 0.8$  mm standardavvik (SD). Forskjellen mellom høyre og venstre side innen ett individ er imidlertid liten. Gjennomsnittlig differanse i senetykkelse mellom sidene hos ett individ kan være kun 0.1 mm, og er svært sjelden mer enn 0.8 mm i de mest vanlige områdene for tendinopati (Bjordal et al, 2003). En heterogen sene med en økning i mer enn 0.8 mm i senetykkelse er vurdert til å indikere tendinopati (Thain et al, 1999 og Uri, 1997) Videre har studier vist at senetykkelsen reduseres med økende alder (Katayose og Magee, 2001). De fleste sener varierer i tykkelse i sin egen lengderetning (Middleton, 1989). Med bakgrunn i dette bør det benyttes en standardisert protokoll for måling av senetykkelse for å være sikker på at målinger settes på samme punkt.

Relativt få forsøk har blitt gjort på å relatere kvantitative målinger av senen som den kommer til uttrykk på ett ultralydsbilde til å klinisk dokumentere smerte eller patologi.

### 3.0 Metode

#### 3.1 Utvalget

Totalt 20 deltakere ble inkludert i studien, hvorav 10 deltakere var friske (uten skuldersmerter) mens 10 deltakere hadde skuldersmerter.

For rekruttering av deltakere ble det tatt kontakt med primærhelsetjenesten med informasjon om studien. På grunn av liten respons og et begrenset tidsperspektiv med tanke på innleveringsdato for oppgaven, ble imidlertid deltakerne rekruttert via ett bekvemmelighetsutvalg bestående av medstudenter, venner og bekjente. Enkelte meldte seg uoppfordret til deltakelse, mens andre ble kontakt med forespørsel om å delta.

Deltakerne som er inkludert i denne studien har spredning i kjønn og alder. Valg om å hente inn informasjon om alder er gjort på bakgrunn av at studier viser at senetykkelsen reduseres med økende alder (Katayose og Magee, 2001) og med tanke på at tendinopati forekommer hos voksne personer i alle aldre. Kroppsmasseindeks kalkuleres etter innhenting av vekt og høyde, da det er vist at avstanden fra hudoverflaten og inn til senen kan virke inn på bildet (Bjordal et al, 2003) Informasjon om smerte er hentet inn for å kunne vurdere hvorvidt det kan være en påvirkende faktor i vurderingen av reliabiliteten til ultralyd som måleinstrument blant raterene. Det er i denne studien anslått ut i fra lignende, tidligere studier at 20 deltakere vil være tilstrekkelig for å kunne gi en vurdering av intertaterreliabilitet (Skovdal et al, 2011; Ohrndorf et al, 2010 og O`Connor et al, 2005)

Tabell 2. Inklusjons- og eksklusjonskriterier for prosjektdeltakelse

Inklusjonskriterier	Eksklusjonskriterier
<ul style="list-style-type: none"> <li>Skuldersmerter med varighet over 3 måneder</li> </ul>	<ul style="list-style-type: none"> <li>Tidligere kirurgiske inngrep i skulder</li> <li>Brudd under tilheling i skulder</li> <li>Nevrologiske sykdommer</li> <li>Revmatiske sykdommer</li> </ul>

Tabell 3. Demografiske data

Deltakere	Alder (år)		KMI		VAS før		VAS etter	
	Mean	Range	Mean	Range	Mean	Range	Mean	Range
<b>u/skuldersmerter</b>								
Kvinner (N=5)	40.8	25-54	23.3	20.5-27.2	0	-	0	-
Menn (N=5)	35.6	24-56	24.4	23.0-25.2	0	-	0	-
Alle (N=10)	38.2	24-56	23.8	20.5-27.2	0	-	0	-
<b>m/skuldersmerter</b>								
Kvinner (N=6)	36.2	27-56	22.3	20.7-25.7	1.6	0.0-5.9	2.0	0.0-5.9
Menn (N=4)	33.6	18-48	24.2	22.8-25.9	1.8	0.0-5.8	1.8	0.0-6.2
Alle (N=10)	35.1	18-56	23.1	20.7-25.9	1.7	0.0-5.9	1.9	0.0-6.2
<b>Totalt (N=20)</b>	<b>36.7</b>	<b>18-56</b>	<b>23.5</b>	<b>20.5-27.2</b>	<b>0.9</b>	<b>0.0-5.9</b>	<b>1.0</b>	<b>0.0-6.2</b>

### 3.2 Rater

To fysioterapeuter (Rater 1 og Rater 2) var ansvarlig for undersøkelsen av deltakerne. Ingen av fysioterapeutene hadde tidligere erfaring med bruk av ultralyd som undersøkelsesmetode. Før oppstart av datainnsamling hadde vi halvannen time med veiledning i bruk av ultralyd som undersøkelsesmetode. I tillegg hadde vi seks timers praktisk øving i bruk av instrumentet.

### 3.3 Etikk og godkjenning

Prosjektprotokollen "Reliabilitetsundersøkelse av diagnostisk ultralyd på skuldere" har blitt godkjent av Regional Etisk Komite Vest, dato 18/8-2011 med ref nr 2011/1251 (vedlegg 1). Jmfør retningslinjer gitt i Helsinki-deklarasjonen fulgte vi prinsippet om frivillig deltakelse, innhenting av informert skriftlig samtykke (vedlegg 2) og at forsøkspersonene hadde rett til å trekke seg fra studien når som helst og uten nærmere årsaksforklaring.

I forkant av undersøkelsen ble telefonnummer registrert for kontakt og timeavtale. Telefonnummer ble slettet så snart undersøkelsen var gjennomført. Opplysningene var kun tilgjengelige for forskerne tilknyttet prosjektet. Forskningsgruppen er databehandlingsansvarlig. Personopplysninger vil være indirekte personidentifiserbare, og aidentifiseres ved hjelp av en koblingsnøkkel som ivaretas av prosjektleder. Bakgrunnsinformasjon ble oppbevart adskilt fra annen informasjon.

Etter prosjektslutt vil all identifiserbar rådata bli slettet da det ikke er planlagt noen etterkontroll eller senere bruk av data.

### 3.4 Gjennomføring av undersøkelsene

Data ble samlet inn i perioden fra medio november 2011 til medio mars 2012. Begge testerne tok ultralydbilder av begge skuldre til alle deltakerne i to omganger. Hver enkelt omgang/undersøkelse varte mellom 10-20 minutter. Dette ga tilsammen fire bilder av hver enkelt sene og totalt 320 bilder. Bildene ble "frost" og deretter lagret. Tverrsnittsmål på senene ble satt på bildene i etterkant av all datainnsamling ved hjelp av den innebygde, kalibrerte kalkulatoren i apparatet. Ved undersøkelse ble informasjon om alder, vekt, høyde og smerte innhentet. Smerte ble registrert ved hjelp av Visual Analogue Scale (VAS) før oppstart og etter gjennomført testprosedyre, på en skyvbar linjal. Målelinjen er 10 cm, hvor 0 cm er ingen smerte og 10 cm er verst tenkelige smerte.

#### 3.4.1 Utstyr

Ultralydsapparatet brukt i studien er av typen "Ultrasonography GE Logic E" med en 10-Mhz lineær transduser (Gråtone, B-mode).

Figur 3 . Ultrasonography GE Logic E





### 3.4.2 Protokoll for undersøkelse

Det ble først tatt bilde av biceps lange sene. Deltaker satt da på en stol med skulderen i 0-posisjon med hånden hvilende på låret i supinert stilling. I denne posisjonen ligger bicepssenen antero-medialt på skulderen (Allen, 2008). Det tas nå ett tverrsnittsbilde av senen (figur 4) samt ett bilde i longitudinal retning (figur 5).

Deretter ble det tatt bilde av rotatorcuff/supraspinatussenen. Deltakeren sitter på en stol med hånden på ryggen (håndbak inntil rygg). I denne posisjonen er humerus innadrotert, og supraspinatussenen ligger antero-lateralt på skulderen (Allen, 2008). Det tas nå ett tverrsnittsbilde av rotatorcuff som inkluderer supraspinatus (figur 6) samt ett longitudinalt bilde av supraspinatussenen (figur 7)

Ovennevnte prosedyre med bildetaking av biceps lange sene og supraspinatussenen ble gjennomført bilateralt på hver deltaker to ganger av hver fysioterapeut etter følgende prosedyre:

- Deltakerne tas imot av begge terapeutene, og får informasjon om gjennomføringen av forestående undersøkelse. Deretter forlater fysioterapeut B lokalet. Fysioterapeut A tar bilde av skuldre bilateralt, for så å forlate rommet.
- Fysioterapeut B kommer deretter inn i rommet, tar bilder av skuldre bilateralt og forlater deretter rommet.
- Fysioterapeut A kommer tilbake til rommet, og tar på ny bilder av skuldre bilateralt for så å igjen forlate rommet.
- Fysioterapeut B kommer inn i rommet, og tar på ny bilder av skuldre bilateralt.

### 3.5 Protokoll for måling av senetykkelse

Tverrsnittet på biceps lange sene og supraspinatussenen ble målt i antero-posterior retning (figur 4, 5, 6 og 7). Når målepunktet ble satt på øverste synlige linje, søkte en

å sette målepunktet direkte på linjen, verken ovenfor eller nedenfor, men midt på linjen.

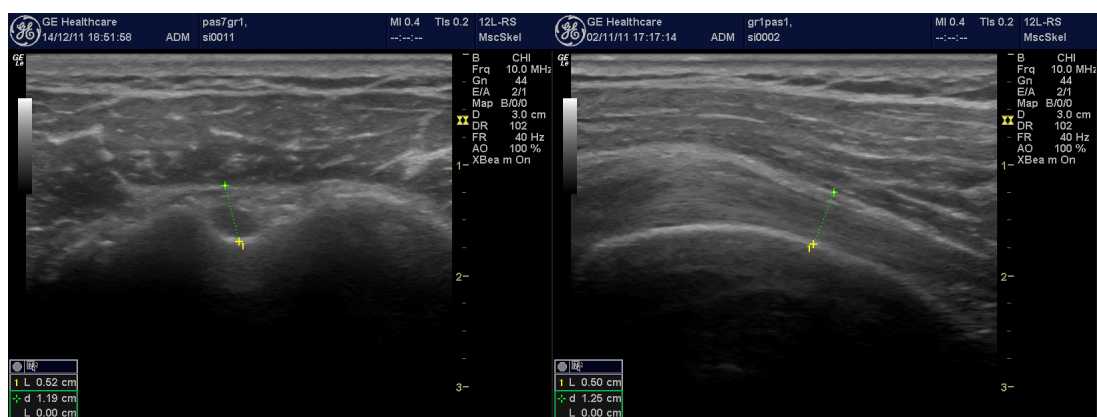
### 3.5.1 Biceps lange sene

Transversalt bildesnitt med skulderen i nøytral posisjon (figur 4): Første målepunkt (1) settes i midten av den bicipitale sulcusen mellom tuberculum majus og minor. På grunn av at humerus på bildet kan ha en forholdsvis bred linje, velges det å sette målepunkt 1 i midten av beinlinjen slik den fremkommer på bildet. Andre målepunkt (2) settes på øverste synlige linje av ligamentum transversum, så direkte som mulig på øverste senedrag. Ved uklarhet av linjen, settes målepunkt på tenkt fortsettelse av ligamentum transversum. Gjennom retningen av målelinjen søker en å dele senen i to mest mulig like halvdel.

Longitudinalt bildesnitt med skulderen i nøytral posisjon (figur 5): Første målepunkt (A) settes i midten av nedre beinlinje, i første område med parallelle linjer etter at senen har passert forhøyning. Andre målepunkt (B) settes på utsiden av synovialskjeden som omslutter m.biceps longus. Det er videre besluttet at andre målepunkt (B) settes på utsiden av den subacromiale-subdeltoid bursaen, som ligger mellom biceps og deltoideus.

Figur 4. Transversalt bildesnitt, biceps lange sene

Figur 5. Longitudinalt bildesnitt, biceps lange sene

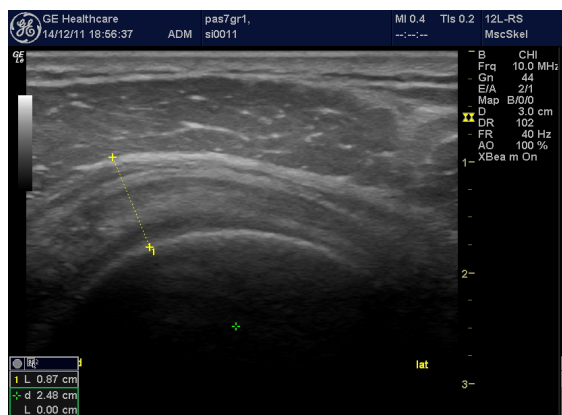


### 3.5.2 Rotatorcuff tverrsnitt og supraspinatus longitudinal

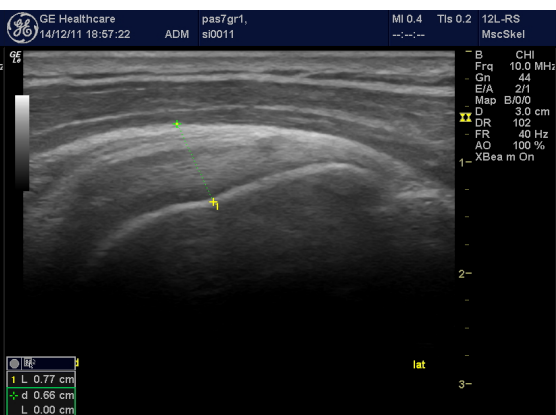
Transversalt bildesnitt med skulderen i innadrotasjon (figur 6): Lokaliserer mediale og laterale del av bildet. Dersom tverrsnittet av senen til m. Biceps longus er tilstede, beregnes senen til m. Supraspinatus å ligge omtrent to cm lateralt for denne. Videre lateralt ligger senen til m. Infraspinatus. Det velges et område med relativt parallelle linjer innenfor gitte antatte beliggenhet av senen til m. Supraspinatus. Målepunkt (1) settes i midten av beinlinjen til caput humeri. Målepunkt (2) settes i skillet mellom supraspinatus og deltoideus, på øverste linje av synlig senedrag. Dersom den subacromiale bursaen dekker supraspinatus, settes målepunkt på øverste linje av det peribursielle vevet mellom rotatorcuffen og deltoideus.

Longitudinalt bildesnitt med skulderen i innadrotasjon (figur 7): Første målepunkt (A) plasseres i collum anatomicum (midten av beinlinjens brytningspunkt mellom tuberculum majus og caput humeri). Målelinjen dras deretter vinkelrett i forhold til senefibrene, til den superfisielle kanten av senen. Andre målepunkt (B) plasseres på øverste linje av senedraget, i skillet mellom supraspinatus og deltoideus.

Figur 6. Transversalt bildesnitt rotatorcuff



Figur 7. Longitudinalt bildesnitt supraspinatus



### 3.6 Pilottesting og rating av bilder

Praktisk trening i å sette mål på ultralydsbildene i følge protokoll ble gjort i fellesskap med medstudent på datasett tatt i forberedelsen til studien. Varighet for den praktiske treningen var ca en time. Skåring av bilder inkludert i studien ble gjennomført i løpet av de to siste ukene i mars 2012. Alle bildene/senene ble skåret 2 ganger av hver

tester, hvor det gikk minimum 7 dager mellom 1. og 2. måling av samme bildesett. Tilsammen ga dette 640 målinger per tester, totalt 1280 målinger. Det ble gjennomført en randomiseringsprosedyre for hvilke bilder som skulle rates når - Pasientene ble nummerert fra 1-10 i henholdsvis smerte gruppe og frisk gruppe. Lappene ble så lagt i en konvolutt, hvor man trakk en lapp for deretter å rate aktuelt nummer med dens tilhørende bilder/sener.

### 3.7 Statistiske metoder

De statistiske analysene er blitt gjennomført i dataprogrammet "Statistical Package for Social Sciences" (SPSS) Version 20.0 og Excel Version 2008 for Macintosh.

For interraterreliabilitet ble relativ reliabilitet vurdert ved hjelp av intraklasse korrelasjons koeffisient one-way random (ICC 1.1) med konfidensintervall på 95% samt ved intraklasse korrelasjonskoeffisient two-way random (ICC 2.1), absolute agreement, for å vurdere grad av samsvar mellom raterene.

Relativ reliabilitet for test-retest ble undersøkt ved hjelp av intraklasse korrelasjons koeffisient, one-way random (ICC 1.1) med konfidensintervall på 95% samt med two-way mixed (ICC 3.1), absolute agreement. ICC (3.1) ble sammenlignet med ICC (1.1) for å kontrollere mulig læringseffekt.

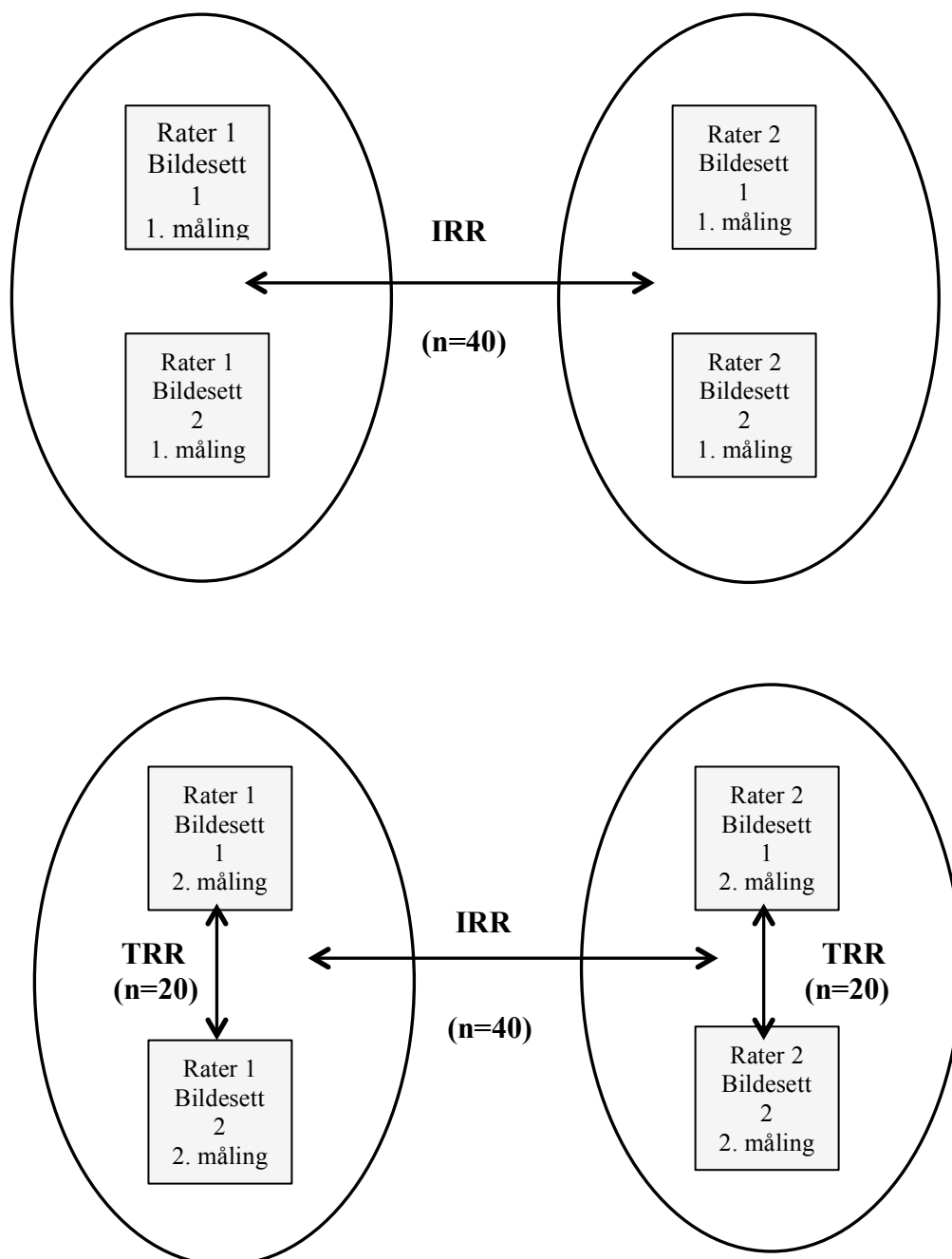
For både interrater- og test-retest ble absolutt reliabilitet beregnet ved å estimere standardavviket til differansen av de parvise målingene. Forskjellene mellom målingene ble plottet mot gjennomsnittet av de innhentede målingene gjort av begge ratere. Systematiske målefeil mellom de to ratere og limits of agreement (gjennomsnittlig forskjell  $\pm 1.96SD$ ) ble beregnet som beskrevet av Bland og Altman (1986).

I alle beregninger utført for å avgjøre interraterreliabilitet ble en måleverdi per rater brukt. Alle bilder har som tidligere nevnt blitt skåret 2 ganger av hver tester, og det er i data-analysen valgt å skille *mellom* 1. og 2. måling, men *med begge bildesett* 1 og 2 samlet. Dette betyr at interraterreliabilitet har blitt vurdert for 1. måling frisk gruppe,

1. måling smerte gruppe, 2. måling frisk gruppe og 2. måling smerte gruppe. Test-retest beregnes for 2. måling av raterenes egne bilder mellom bildesett 1 og 2, for henholdsvis smerte gruppe og frisk gruppe (figur 6).

Demografiske data (smerte, KMI og alder) er beskrevet som deskriptiv statistikk.

Figur 8. Oversikt over reliabilitetsundersøkelsene i denne oppgaven. De ulike undersøkelsene ble gjentatt for henholdsvis frisk gruppe og smerte gruppe. Her vises undersøkelsene for 1 gruppe



## 4.0 Resultat

### 4.1 Demografiske data (tabell 3)

I studien var det seks deltakere som rapporterte om smerter før oppstart av undersøkelse med ultralyd. En deltaker rapporterte ingen endring i smertenivå fra start til slutt, mens to deltakere rapporterte en viss bedring i smertenivået. Tre deltakere rapporterte om økt smerte fra oppstart av undersøkelse til etter fullført undersøkelse med ultralyd.

4 deltakere (2 i frisk gruppe, 2 i smerte gruppe) hadde en KMI på  $> 25 \text{ kg/m}^2$ , noe som i følge WHO's klassifikasjon betegner dem som overvektige (WHO).

### 4.2 Grafisk framstilling

Interrater- og test-retest reliabilitet er presentert som relativ og absolutt reliabilitet oppgitt i tabell 4-7. En fremstilling av systematiske- og tilfeldige målefeil, samt øvre og nedre LoA er presentert i Bland-Altman plott for variablene for å belyse de individuelle variasjonene til målingene visuelt (vedlegg 4 og 5). Verdier til Bland-Altman plottene kan leses direkte ut i fra reliabilitetstabellene hvor  $X_d$  representerer systematiske målefeil. Ved visuell inspeksjon ble dataene ikke funnet å være heteroscedastiske.

### 4.3 Normalfordeling.

Datamaterialet viste seg å være normalfordelt ved tverrsnittsmål på alle senene/bildene hos begge raterene (vedlegg 3).

### 4.4 Interraterreliabilitet

Tabell 4 og 5 viser en oversikt over variabilitetsmål som uttrykk for interraterreliabiliteten ved måling av senetykkelse på ultralydsbildene. Ved relativ reliabilitet varierte ICC (1.1) fra 0.51-0.93 i frisk gruppe og fra 0.27-0.95 i smerte gruppe. Bortsett fra 2. måling venstre rotatorcuff i smerte gruppe er ICC (2.1) lik eller

tilnærmet lik ICC (1.1) ved alle undersøkelser, noe som tyder på at det ikke foreligger betydelige systematiske målefeil. Ved absolutt reliabilitet varierte de systematiske målefeilene ( $x_d$ ) fra 0-0.06 cm i frisk gruppe, og fra 0-0.05 cm i smerte gruppe.

I denne oppgaven er det valgt å gradere den relative reliabiliteten (ICC) i følgende 3 kategorier:

0.00-0.49: Dårlig reliabilitet

0.50-0.75: Moderat reliabilitet

0.76-1.00: God reliabilitet

#### 4.4.1 Gradering av relativ reliabilitet

*God reliabilitet (>0.75):* I frisk gruppe oppnådde 2. måling av venstre biceps longitudinal høyest korrelasjon, med en ICC (1.1) verdi på 0,93, noe som tilsier god reliabilitet. Også 2. måling av supraspinatus bilateral, høyre biceps tverrsnitt og begge målinger av høyre biceps longitudinal oppnådde god reliabilitet. I smerte gruppe viste 2. måling av høyre biceps longitudinal høyest korrelasjon, med en ICC (1.1) verdi på 0.95. Videre viste også 2. måling av biceps tverrsnitt bilateral, 1. måling av høyre biceps longitudinal og begge målinger av venstre biceps longitudinal god reliabilitet.

*Moderat reliabilitet (0.50-0.75):* I frisk gruppe viste begge målinger av venstre biceps tverrsnitt samt rotatorcuff bilateral moderat reliabilitet. I tillegg viste også 1. måling av venstre supraspinatus, høyre biceps tverrsnitt og høyre supraspinatus kun moderat reliabilitet. I smerte gruppe viste begge målinger av venstre rotatorcuff og supraspinatus moderat reliabilitet. Også 1. måling av høyre biceps tverrsnitt, supraspinatus og venstre biceps tverrsnitt viste moderat reliabilitet.

*Dårlig reliabilitet frisk gruppe (0.00-0.49):* I frisk gruppe viste 1. måling av venstre biceps longitudinal dårlig reliabilitet. I smerte gruppe ga også begge målinger på venstre rotatorcuff og 2. måling på høyre supraspinatus longitudinal dårlig reliabilitet.

#### 4.4.2 Absolutt reliabilitet

Absolutt reliabilitet målt ved standardavviket til differansen viste høyest reliabel verdi ved 2. måling biceps longitudinal bilateral i frisk gruppe (0.04mm) og ved 2. måling biceps longitudinal bilateral i smerte gruppe (0.04mm/0.02mm). Den sanne verdien vil forventes å være 0.08mm for 2. måling biceps longitudinal bilateral i frisk gruppe, og 0.08mm/0.04mm for 2. måling biceps longitudinal bilateral i smerte gruppe ved  $1.96 \cdot SD_{diff}$  for 95% av målingene.

Dårligst reliable verdi ble vist ved 1. måling høyre rotatorcuff i frisk gruppe (0.12mm) og i smerte gruppe ved 1. måling venstre biceps tverrsnitt (0.10mm). Den sanne verdien vil forventes å være 0.24mm for 1. måling høyre rotatorcuff i frisk gruppe og 0.20mm for 1. måling venstre biceps tverrsnitt i smerte gruppe ved  $1.96 \cdot SD_{diff}$  for 95% av målingene.

Analyser for absolutt reliabilitet ( $SD_{diff}$ ) viste totalt sett noe lavere reliabilitet for frisk gruppe enn for smerte gruppe (tabell 4 og 5). Videre viser alle verdier for 2. måling i frisk gruppe høyere reliabilitet enn 1. måling i samme gruppe. Også i smerte gruppe viser tilnærmet alle verdiene ved 2. måling høyere reliabilitet enn ved 1. måling.

Tabell 4. IRR: Relativ og absolutt reliabilitet. Frisk gruppe, (n= 40)

	Måling	Relativ reliabilitet				Absolutt reliabilitet			
		ICC 1.1	CI 95%	ICC 2.1	CI 95%	$X_d$	SD	$SD \cdot 1.96$	LOA
<b>Venstre skulder</b>									
Biceps tverrsnitt	1.	0.54	0.28-0.73	0.54	0.28-0.73	0.00	0.09	0.18	-0.18+0.18
	2.	0.70	0.50-0.83	0.70	0.49-0.83	0.00	0.08	0.16	-0.16+0.16
Biceps longitudinal	1.	0.47	0.18-0.68	0.48	0.20-0.69	0.03	0.08	0.16	-0.13+0.19
	2.	0.93	0.87-0.96	0.93	0.87-0.96	0.01	0.04	0.08	-0.07+0.09
Rotatorcuff tverrsnitt	1.	0.62	0.39-0.78	0.63	0.39-0.78	0.02	0.09	0.18	-0.16+0.20
	2.	0.70	0.49-0.83	0.71	0.31-0.87	0.06	0.08	0.16	-0.10+0.22
Supraspinatus longitudinal	1.	0.75	0.58-0.86	0.75	0.58-0.86	0.00	0.08	0.16	-0.16+0.16
	2.	0.86	0.75-0.92	0.86	0.74-0.93	0.02	0.06	0.12	-0.10+0.14
<b>Høyre skulder</b>									
Biceps tverrsnitt	1.	0.66	0.44-0.80	0.66	0.44-0.80	0.01	0.08	0.16	-0.15+0.17
	2.	0.76	0.59-0.86	0.76	0.59-0.86	0.02	0.07	0.14	-0.12+0.16
Biceps longitudinal	1.	0.78	0.62-0.88	0.79	0.60-0.89	0.02	0.05	0.10	-0.08+0.12
	2.	0.86	0.74-0.92	0.86	0.71-0.93	0.02	0.04	0.08	-0.06+0.10



Rotatorcuff tverrsnitt	1.	0.51	0.23-0.70	0.51	0.23-0.71	0.02	0.12	0.24	-0.22+0.26
	2.	0.73	0.55-0.85	0.74	0.51-0.86	0.04	0.08	0.16	-0.12+0.20
Supraspinatus longitudinal	1.	0.65	0.43-0.80	0.65	0.42-0.80	0.00	0.10	0.20	-0.20+0.20
	2.	0.87	0.76-0.93	0.87	0.74-0.93	0.02	0.05	0.10	-0.08+0.12

CI= konfidensintervall,  $X_d$ = mean difference (systematiske målefeil) SD=standard deviation of the differences,  $SD*1.96=95\%CI$  av målefeilen. LOA=limits of agreement ( $X_d\pm(SD*1.96)$ )

Tabell 5. IRR: Relativ og absolutt reliabilitet. Smerte gruppe, (n=40)

Måling		Relativ reliabilitet				Absolutt reliabilitet			
		ICC 1.1	CI 95%	ICC 2.1	CI 95%	$X_d$	SD	$SD*1.96$	LOA
<b>Venstre skulder</b>									
Biceps tverrsnitt	1.	0.68	0.47-0.82	0.68	0.46-0.82	0.00	0.10	0.20	-0.20+0.20
	2.	0.81	0.67-0.90	0.81	0.66-0.90	0.02	0.08	0.16	-0.14+0.18
Biceps longitudinalt	1.	0.87	0.76-0.93	0.87	0.76-0.93	0.00	0.05	0.10	-0.10+0.10
	2.	0.89	0.79-0.94	0.89	0.79-0.94	0.01	0.04	0.08	-0.07+0.09
Rotatorcuff tverrsnitt	1.	0.72	0.53-0.84	0.73	0.50-0.86	0.03	0.06	0.12	-0.09+0.15
	2.	0.70	0.50-0.83	0.71	0.38-0.86	0.04	0.07	0.14	-0.10+0.18
Supraspinatus longitudinalt	1.	0.64	0.42-0.79	0.64	0.42-0.79	0.01	0.08	0.16	-0.15+0.17
	2.	0.59	0.35-0.76	0.61	0.29-0.79	0.04	0.06	0.12	-0.08+0.16
<b>Høyre skulder</b>									
Biceps tverrsnitt	1.	0.73	0.54-0.85	0.73	0.55-0.85	0.02	0.08	0.16	-0.14+0.18
	2.	0.89	0.80-0.94	0.89	0.70-0.95	0.03	0.05	0.10	-0.07+0.13
Biceps longitudinalt	1.	0.84	0.71-0.91	0.84	0.71-0.91	0.00	0.05	0.10	-0.10+0.10
	2.	0.95	0.90-0.97	0.95	0.86-0.98	0.01	0.02	0.04	-0.03+0.05
Rotatorcuff tverrsnitt	1.	0.32	0.00-0.57	0.34	0.05-0.59	0.03	0.08	0.06	-0.03+0.09
	2.	0.27	-0.39-0.54	0.34	0.02-0.60	0.05	0.08	0.15	-0.10+0.20
Supraspinatus longitudinalt	1.	0.59	0.34-0.76	0.59	0.34-0.76	0.02	0.08	0.16	-0.14+0.18
	2.	0.46	0.18-0.67	0.47	0.20-0.68	0.03	0.09	0.18	-0.15+0.21

CI= konfidensintervall,  $X_d$ = mean difference (systematiske målefeil) SD=standard deviation of the differences,  $SD*1.96=95\%CI$  av målefeilen. LOA=limits of agreement ( $X_d\pm(SD*1.96)$ )

#### 4.5 Test-retest reliabilitet

Test-retest reliabiliteten til måling av senetykkelse på ultralydsbildene er presentert i tabell 6 og 7. Ved relativ reliabilitet varierte ICC (1.1) fra 0.53-0.85 i frisk gruppe og fra 0.60-0.93 i smerte gruppe for rater 1. For rater 2 varierte ICC (1.1) fra 0.61-0.99 i frisk gruppe og fra 0.72-0.99 i smerte gruppe. Ved absolutt reliabilitet varierte de systematiske målefeilene ( $x_d$ ) fra -0.11-0.09 cm i frisk gruppe og fra -0.05-0.02 cm i smerte gruppe hos rater 1. Hos rater 2 varierte de systematiske målefeilene ( $x_d$ ) fra -

0.01-0.02 i frisk gruppe og fra -0.02-0.03 i smerte gruppe. Bortsett fra 1 målepunkt hos rater 1 i frisk gruppe (høyre supraspinatus), er ICC (3.1) lik eller tilnærmet lik ICC (1.1) for begge raterene. Det foreligger med andre ord lite systematiske målefeil i datamaterialet.

#### 4.5.1 Gradering av relativ reliabilitet

*God reliabilitet (>0.75):* I frisk gruppe var det for rater 1 målinger på venstre biceps longitudinal som ga høyest reliabilitet med en ICC (1.1) verdi på 0.85. Rater 1 oppnådde også god reliabilitet ved målinger på høyre biceps longitudinal. Rater 2 oppnådde på lik linje med rater 1 høyest reliabilitet ved målinger på venstre biceps longitudinal, med en ICC (1.1) verdi på 0.99. Dette tilsier nærmest perfekt korrelasjon mellom målingene. Videre oppnådde rater 2 også god reliabilitet ved målinger på venstre biceps tverrsnitt og rotatorcuff samt høyre biceps tverrsnitt/longitudinal og rotatorcuff. I smerte gruppe oppnådde rater 1 høyest reliabilitet ved målinger på venstre biceps tverrsnitt med en ICC (1.1) verdi på 0.93. Rater 1 oppnådde også god reliabilitet ved målinger på venstre biceps longitudinal og supraspinatus samt høyre biceps tverrsnitt/longitudinal. I smertegruppe oppnådde rater 2 høyest ICC (1.1) verdi på høyre biceps longitudinal (0.99). Rater 2 oppnådde også god reliabilitet på målinger av venstre biceps tverrsnitt/longitudinal, supraspinatus og høyre biceps tverrsnitt, rotatorcuff og supraspinatus.

*Moderat reliabilitet (0.50-0.75):* For rater 1 ble det i frisk gruppe oppnådd moderat reliabilitet ved målinger på venstre biceps tverrsnitt, rotatorcuff, supraspinatus og høyre biceps tverrsnitt og rotatorcuff. Rater 2 oppnådde moderate reliable målinger i frisk gruppe på både høyre og venstre supraspinatus longitudinal. I smerte gruppe oppnådde rater 1 moderate reliable målinger på høyre og venstre rotatorcuff samt høyre supraspinatus. Rater 2 hadde i smertegruppe kun moderat reliabel skåring på venstre rotatorcuff.

*Dårlig reliabilitet (0.00-0.49):* Rater 1 fikk dårlig reliabilitet ved test-retest målinger på høyre supraspinatus i frisk gruppe.

#### 4.5.2 Absolutt reliabilitet

Absolutt reliabilitet målt ved standardavviket til differansen viste for rater 1 høyest reliabilitet ved biceps longitudinal bilateralt (0.06 mm) og ved høyre supraspinatus (0.06mm) i frisk gruppe. I smerte gruppe hadde rater 1 høyest absolutt reliabilitet ved biceps longitudinal bilateralt (0.05mm/0.04 mm) og ved rotatorcuff høyre skulder (0.05mm). Den sanne verdien vil forventes å være 0.12mm for biceps longitudinal og for høyre supraspinatus i frisk gruppe, og 0.10mm/0.08mm for biceps longitudinal bilateral og 0.10 for høyre rotatorcuff i smerte gruppe ved  $1.96 * SD_{diff}$  for 95% av målingene.

Dårligst absolutt reliabilitet ble for rater 1 vist ved venstre supraspinatus (0.10mm), høyre biceps tverrsnitt (0.10mm) og høyre rotatorcuff (0.10mm) i frisk gruppe, og ved venstre biceps tverrsnitt (0.07mm) og venstre supraspinatus i smerte gruppe (0.07mm). Den sanne verdien vil forventes å være 0.20mm for venstre supraspinatus, høyre biceps tverrsnitt og høyre rotatorcuff i frisk gruppe, samt 0.14mm for venstre biceps tverrsnitt og supraspinatus i smerte gruppe ved  $1.96 * SD_{diff}$  for 95% av målingene.

Rater 2 oppnådde høyest absolutt reliabilitet i frisk gruppe for biceps longitudinal venstre (0.02mm) og biceps tverrsnitt høyre (0.02mm) og i smerte gruppe ved høyre biceps longitudinal (0.01mm) og høyre rotatorcuff tverrsnitt (0.02mm). Den sanne verdien vil forventes å være 0.04mm for venstre biceps longitudinal og høyre biceps tverrsnitt i frisk gruppe samt 0.02mm for høyre biceps longitudinal og 0.04mm for høyre rotatorcuff tverrsnitt i smerte gruppe ved  $1.96 * SD_{diff}$  for 95% av målingene.

Dårligst absolutt reliabilitet ble for rater 2 vist ved supraspinatus bilateral i frisk gruppe (0.08mm/0.08mm) og i smerte gruppe ved høyre biceps tverrsnitt og supraspinatus (0.06mm/0.06mm). Den sanne verdien vil forventes å være 0.16mm for supraspinatus bilateral i frisk gruppe og 0.12mm for høyre bicepstverrsnitt og rotatorcuff ved  $1.96 * SD_{diff}$  for 95% av målingene.

Analysen for absolutt reliabilitet ( $SD_{diff}$ ) viste totalt sett lavere reliable målinger for rater 1 enn for rater 2 i begge gruppene (tabell 4 og 5). Begge raterne viser til målinger med høyere absolutt reliabilitet i smerte gruppe enn i frisk gruppe.

Tabell 6. TRR: Relativ og absolutt reliabilitet. Frisk gruppe, begge ratere (n=20)

	Rater	Relativ reliabilitet				Absolutt reliabilitet			
		ICC 1.1	CI 95%	ICC 3.1	CI 95%	$X_d$	SD	SD*1.96	LOA
<b>Venstre skulder</b>									
Biceps tverrsnitt	1	0.67	0.14-0.91	0.67	0.07-0.13	0.01	0.09	0.18	-0.17+0.19
	2	0.90	0.66-0.97	0.90	0.63-0.97	0.00	0.04	0.08	-0.08+0.08
Biceps longitudinalt	1	0.85	0.54-0.96	0.85	0.55-0.96	0.02	0.06	0.12	-0.10+0.14
	2	0.99	0.97-1.00	0.99	0.97-1.00	0.00	0.02	0.04	-0.04+0.04
Rotatorcuff tverrsnitt	1	0.61	-0.35-0.79	0.61	-0.02-0.90	0.09	0.08	0.16	-0.07+0.25
	2	0.96	0.85-0.99	0.96	0.85-0.99	0.00	0.03	0.06	-0.06+0.06
Supraspinatus longitudinalt	1	0.66	0.10-0.91	0.66	0.01-0.91	0.02	0.10	0.20	-0.18+0.22
	2	0.61	0.01-0.90	0.61	-0.03-0.90	0.02	0.08	0.16	-0.14+0.18
<b>Høyre skulder</b>									
Biceps tverrsnitt	1	0.53	-0.07-0.86	0.53	-0.11-0.86	0.03	0.10	0.20	-0.17+0.23
	2	0.97	0.88-0.99	0.97	0.87-0.99	-0.01	0.02	0.04	-0.05+0.03
Biceps longitudinalt	1	0.81	0.43-0.95	0.81	0.36-0.94	0.00	0.06	0.12	-0.12+0.12
	2	0.92	0.73-0.98	0.92	0.73-0.98	0.01	0.04	0.08	-0.07+0.09
Rotatorcuff tverrsnitt	1	0.74	0.27-0.93	0.74	0.26-0.93	0.04	0.10	0.20	-0.16+0.24
	2	0.97	0.86-0.99	0.97	0.85-0.99	0.00	0.03	0.06	-0.06+0.06
Supraspinatus longitudinal	1	0.48	-0.14-0.84	0.56	0.46-0.96	-0.11	0.06	0.12	-0.23+0.01
	2	0.61	0.01-0.90	0.61	-0.03-0.90	0.02	0.08	0.16	-0.14+0.18

CI= konfidensintervall,  $X_d$ = mean difference (systematiske målefeil) SD=standard deviation of the differences, SD\*1.96=95%CI av målefeilen. LOA=limits of agreement ( $X_d \pm (SD*1.96)$ )

Tabell 7. TRR: Relativ og absolutt reliabilitet. Smerte gruppe, begge ratere (n=20)

	Rater	Relativ reliabilitet				Absolutt reliabilitet			
		ICC 1.1	CI 95%	ICC 3.1	CI 95%	$X_d$	SD	SD*1.96	LOA
<b>Venstre skulder</b>									
Biceps tverrsnitt	1	0.93	0.73-0.98	0.93	0.71-0.98	0.01	0.07	0.14	-0.13+0.15
	2	0.97	0.89-0.99	0.97	0.89-0.99	0.01	0.03	0.06	-0.05+0.07
Biceps longitudinalt	1	0.85	0.52-0.96	0.85	0.52-0.96	0.02	0.05	0.10	-0.08+0.12
	2	0.94	0.81-0.99	0.94	0.77-0.99	0.02	0.03	0.06	-0.04+0.08
Rotatorcuff tverrsnitt	1	0.60	-0.01-0.89	0.60	-0.12-0.89	0.00	0.06	0.12	-0.12+0.12
	2	0.72	0.24-0.92	0.73	0.23-0.93	0.03	0.05	0.10	-0.07+0.13

Supraspinatus longitudinalt	1	0.77	0.35-0.94	0.78	0.27-0.94	-0.05	0.07	0.14	-0.19+0.09
	2	0.79	0.39-0.94	0.79	0.39-0.94	-0.02	0.05	0.10	-0.12+0.08
<b>Høyre skulder</b>									
Biceps tverrsnitt	1	0.79	0.35-0.95	0.79	0.31-0.95	-0.01	0.06	0.12	-0.11+0.13
	2	0.86	0.57-0.96	0.86	0.57-0.96	0.02	0.06	0.12	-0.10+0.14
Biceps longitudinalt	1	0.90	0.66-0.97	0.90	0.63-0.97	0.00	0.04	0.08	-0.08+0.08
	2	0.99	0.95-1.00	0.99	0.95-1.00	0.00	0.01	0.02	-0.02+0.02
Rotatorcuff tverrsnitt	1	0.66	0.08-0.91	0.66	-0.01-0.91	0.00	0.05	0.10	-0.10+0.10
	2	0.92	0.73-0.98	0.92	0.72-0.98	0.01	0.02	0.04	-0.03+0.05
Supraspinatus longitudinalt	1	0.74	0.28-0.93	0.74	0.22-0.93	0.00	0.06	0.12	-0.12+0.12
	2	0.79	0.38-0.94	0.79	0.36-0.94	-0.01	0.06	0.12	-0.13+0.11

CI= konfidensintervall,  $\bar{X}_d$ = mean difference (systematiske målefeil) SD=standard deviation of the differences,  
 $SD*1.96=95\%CI$  av målefeilen. LOA=limits of agreement ( $\bar{X}_d \pm (SD*1.96)$ )

## 5.0 Disuksjon

### 5.1 Metodediskusjon

Det vil i dette kapittelet diskuteres de metodiske forhold som kan ha påvirket resultatene i denne studien relatert til målemetode, variabilitet og standardisering, utvalg, ratere, studiedesign, datainnsamling og valg av statistiske analyser.

#### 5.1.1 Målemetoden

Ultralydsavbildning kan være en nyttig metode å bruke ved undersøkelse av skulderen (Allen og Wilson, 2001; Dinnes et al, 2003; Jacobson og Van Holsbeeck, 1998; Middleton et al, 2004; Martinoli et al, 2003 og de Jesus et al, 2009), men den diagnostiske nøyaktigheten av UL ved rotatorcuff tendinopati er fremdeles uklar. Videre er UL som metode foreløpig ikke etablert i primærhelsetjenesten.

I denne oppgaven er målsettingen å undersøke interraterreliabilitet og test-retest reliabilitet av ultralydsmåling på tverrsnittet til skuldereener. Slike målinger kan brukes til å måle hevelse og sammen med vurdering av intratendinøs strømming med doppler, vil diagnosen rotatorcufftendinopati ved ultralydsavbildning kunne stilles eller utelukkes (Hoksrud et al, 2008; Maffuli et al, 1987).

Kun få studier har undersøkt interrater- og test-retest reliabilitet ved ultralydsmålinger på tverrsnittet til skuldereener. I ett studie av Naredo et al (2006) blir blant annet interraterreliabilitet hos eksperter i muskelskjelett UL ved rotatorcuff tendinopati vurdert. Absolutt enighet uttrykt i ICC var gjennomsnittlig 0.50 (moderat enighet) for ekspertene. Den relativt lave interraterreliabiliteten i studien til Naredo et al (2006) kan muligens skyldes at raterene ikke var kjent med utstyret på forhånd og heller ikke fulgte en bestemt standardisert protokoll. Raterene hadde ikke trent sammen på undersøkelse med ultralydsavbildning før studien og var heller ikke kjent med pasientenes sykehistorie. Det fantes på forhånd ikke noen definert enighet i hva rotatorcuff tendinopati er. Disse uklarhetene forårsaket dermed ulike tolkninger av samme patologiske funn. Ultralydsavbildning av smertefull skulder ble ikke

sammenlignet med ikke-smertefull skulder. Sammenligning med motsatt skulder er ett viktig moment ved vurdering av om det er tilstedeværelse av tendinopati. Forskjellen i senetykkelse fra høyre til venstre skulder hos en frisk person vil sjelden være mer enn 0.8 mm (Bjordal et al, 2003). En heterogen sene med økning i mer enn 0.8 mm er imidlertid en indikasjon på tendinopati (Thain et al, 1999 og Uri, 1997).

I artikkelen er det angitt at det er brukt lineære transdusere med frekvens fra 7-14 Mhz i studien av Naredo et al (2006). Det er imidlertid ikke spesifisert i artikkelen hvilken frekvens som ble benyttet ved undersøkelse av skuldereener.

I studien til Naredo et al (2006) er kun relativ reliabilitet vurdert. Domholdt (2005 s. 258) påpeker at det er et problem at mange reliabilitetsstudier kun vurderer relativ reliabilitet. Skal man gjøre en meningsfull klinisk vurdering av resultatene og implementere disse i praksis, bør også absolutt reliabilitet vurderes. I motsetning til en korrelasjonskoeffisient, er absolutt reliabilitet uttrykt i den måleenheten målemetoden har. Videre vil "ekstremverdier/uteliggere" som gjør at en større del av måleskalaen enn normalt forventet for pasientgruppen benyttes, kunne påvirke ICC og gi en kunstig høy verdi av denne. Dette vil heller ikke avdekkes med mindre man i tillegg vurderer absolutt reliabilitet eller oppgir i studien hvilke enkeltmålinger/rådata som er grunnlaget for analyse av ICC.

Bjordal et al (2003) har i sitt studie vurdert blant annet intraraterreliabilitet ved test retest av ultralydsmåling av senetykkelse på supraspinatus. Det ble der vist høy reliabilitet ved hjelp av Bland Altman LoA. En svakhet med studien er at det ikke er gjennomført beregning av relativ reliabilitet. Domholdt (2005) anbefaler som tidligere nevnt å bruke både relativ og absolutt reliabilitet ved undersøkelse av reliabilitet, da disse utfyller hverandre. Samtidig er det en styrke med studien at det faktisk er beregnet absolutt reliabilitet, ettersom dette er klinisk anvendbart og overførbart til praksis. En annen mulig svakhet med studien til Bjordal et al (2003) er at alle inkluderte deltakere var friske. Reliabiliteten av ett instrument bør vurderes ved å bruke ett utvalg deltakere med samme karakteristika som instrumentet er ment å brukes på i klinikken (Domholdt 2005), og studien kunne da med fordel ha inkludert en gruppe med deltakere med smertefulle skuldre. Videre kan også transduserfrekvens

benyttet i studien (7.5 Mhz) være en svakhet, da det i dag er anbefalt høyere frekvenser enn dette ved vurdering av skuldereener (Jacobson, 2003).

Ett mål med ultralydsmåling av tverrsnittet til skuldereener er å differensiere mellom personer som har rotatorcuff tendinopati og personer som er friske. Dette er ett spørsmål om metoden sin sensitivitet og spesifisitet (Domholdt, 2005). Med sensitivitet menes i hvilken grad ultralydmåling av tverrsnittet til skuldereener gir ett riktig positivt svar dersom personen har tendinopati. Spesifisitet er ett uttrykk for målemetodens evne til å gi ett riktig negativt svar dersom personen ikke har rotatorcuff tendinopati. Relativt få studier er tidligere blitt gjort på å undersøke rotatorcuff tendinopati med UL. Til tross for at det ser ut til at UL kan diagnostisere tendinopati nøyaktig, er sensitiviteten i ulike studier ikke entydige (Martin-Hervas et al, 2001 og Naredo et al, 1999). Den lave sensitiviteten (66,7 %) i studien til Martin-Hervas et al (2001) kan muligens forklares med ett forholdsvis lite populasjonsutvalg, en transduser med "utdatert"/for lav frekvens (7.5 Mhz) samt mulig sykdomsprogresjon som vil kunne føre til systematiske målefeil.

### 5.1.2 Variabilitet og standardisering

I følge generaliserbarhetsteorien finnes det ulike kilder for variabilitet for alle målinger som kan settes i sammenheng med instrumentet, testingen eller deltakerne. Variabiliteten i denne studien kan med andre ord skyldes feil ved ultralydsapparatet, feil hos raterene som utfører undersøkelsene og målingene, eller individuelle biologiske variasjoner hos den enkelte deltaker.

Teoretisk sett vil det kunne forekomme feil ved ultralydsapparatet, men dette er allikevel lite sannsynlig. Veileder Jon Joensen har kalibrert instrumentet både før og etter undersøkelsen, uten å finne feil. Imidlertid kan transdusertype- og frekvens man benytter ved billedtaking være en mulig kilde til variabilitet. Høy frekvens transdusere vil gi bilder med høy oppløsning, men til gjengjeld vil penetrasjonsevnen bli redusert. Det er i denne studien valgt å benytte en lineær probe med 10 Mhz frekvens. Jacobson (2003) anbefaler lineære transdusere med frekvens på minimum 10 Mhz ved undersøkelse av skulder med bakgrunn i at skuldereenerene er relativt overflatiske



strukturer, i tillegg til at den høye frekvensen vil gi bilder med høy oppløsning. Med tanke på at det har skjedd en teknologisk utvikling av UL de siste årene finnes det nå kvalitetsmessig bedre og mer moderne ultralydsutstyr. Tidligere studier gjort på vurdering av skuldresener i forhold til senetykkelse, har i stor grad brukt transduserfrekvens på 7.5 Mhz. Dette blir i dag betegnet som en utdatert og for lav frekvens i vurderingen av skuldresener, ettersom man nå har tilgang til transdusere med høyere frekvens.

Ved ultralydsavbildning blir resultatene fremstilt digitalt. I følge Domholdt (2005) er det en fordel å bruke ett mekanisk måleinstrument der måleresultatene fremstilles digitalt. Det blir da lettere å skille mellom instrumentreliabilitet og raterreliabilitet, ettersom måleresultatene ikke gir grunnlag for tolkning.

Raterene sine ferdigheter i å ta gode bilder samt i å sette nøyaktige mål på bildene vil være hovedkilden til variabiliteten i målingene i denne studien. Intrarater reliabilitet for hver rater bør ideelt sett etableres før man sammenligner ratere (Portney og Watkins, 2000). Det ble derfor gjennomført en pilottesting i forkant av studien som innebar praktisk trening i å sette mål på ultralydsbildene i følge protokoll, noe som er en styrke med denne studien. Dette ble gjort i fellesskap med medstudent på datasett tatt i forberedelsen til studien. Vi fulgte en høyt standardisert protokoll for undersøkelse, men det er allikevel vanskelig å sikre at begge raterene opptrår helt likt og tolker den standardiserte protokollen likt. Ingen av raterene hadde noen tidligere erfaring med ultralydsavbildning som undersøkelsesmetode. Erfaringen jeg gjorde meg undervegs var at begge raterene i stor grad oppnådde god kvalitet på bildene av senene. Imidlertid var setting av mål en større kilde til variabilitet, da dette er en prosedyre som krever stor grad av presisjon. Ved ultralydsavbildning er standardisering av testprotokoll en forutsetning for å kunne sette konsistente og presise målinger. Til tross for at protokoll for måling av senetykkelse i denne studien var høyt standardisert, var det erfaringsmessig på enkelte bilder vanskelig å vurdere hvor man faktisk skulle sette målingene. En god kjennskap til de anatomiske forholdene i skulderen er avgjørende for å kunne gjøre konsistente målinger. Muligens hadde raterene ulike ferdigheter i forhold til dette og ulike ferdigheter i å

tolke sonoanatomien på bildene, noe som dermed vil kunne være en kilde til variabilitet.

Individuelle variasjoner hos deltakerne kan være en mulig kilde til variabilitet ved ultralydsavbildning. Bjordal et al (2003) viser i sitt studie til at avstand fra hudoverflate til supraspinatus senen er doblet hos personer som er moderat overvektige (KMI 27-30 kg/m<sup>2</sup>). I denne studien viste de demografiske data at 4 deltakere (2 i frisk gruppe og 2 i smerte gruppe) hadde en KMI på > 25 kg/m<sup>2</sup>, noe som i følge WHO's klassifikasjon betegner dem som overvektige. Flere studier viser til at avstand fra hudoverflate til sene vil være essensielt å vite noe om, for å kunne estimere energitap (Enwemeka, 2001; Bureau og Roederer, 1998; Stolik et al, 2000) og dermed vurdere hvilken transduserfrekvens som kan være aktuell å bruke.

Jacobson (2003) påpeker at det hos personer med høy KMI kan forsvares å bruke en transduserfrekvens på 7.5 Mhz, ettersom frekvensen er direkte proporsjonal med den aksiale oppløsningen, og omvendt proporsjonal med vevspenetrasjonen til ultralydsbølgen (Jacobson, 2007). Dette betyr at man bør bruke høye frekvenser til undersøkelse av overflatiske strukturer, og lavere frekvenser til dype strukturer (Moosmayer, 2004). Ettersom vi i denne studien brukte 10 MhZ transduserfrekvens, vil det faktisk at 4 deltakere var overvektige ha kunnet påvirket bildekvaliteten ved undersøkelse av disse deltakerne. Ved minst ett tilfelle var dette noe jeg erfarte selv. Det var utfordrende å lokalisere rett struktur med transduseren, og bildene var uklare.

Smerteregistrering ved hjelp av Visual Analogue Scale viste at undersøkelsen ga enkelte deltakere en viss økning i smerteverdier. Det har vist seg at reliabiliteten vil være høyere der deltakerne har lave smerteverdier (Nussbaum og Downes, 1998). Utifra angitt smertenivå og range for VAS fra oppstart til slutt av undersøkelse (tabell 4), er det ikke noe som tyder på at smerte har hatt negativ innvirkning på kvaliteten av undersøkelsen og ultralydsavbildningen.

Mulig sykdomsprogresjon er en lite sannsynlig kilde til variabilitet i denne studien, ettersom alle undersøkelsene av en og samme deltaker ble gjennomført i løpet av noen timer.

Tidsintervallet fra test til retest vil kunne påvirke variabiliteten i denne studien. Tidsintervallet må være lite nok til at det er minimal, eller ingen sjanse for at læringseffekt hos rateren finner sted, men stort nok til at rateren ikke husker hva som ble målt i den originale situasjon (Blacker og Endicott, 2003). Mellom test og retest undersøkelsene i denne studien gikk det rundt 30 minutter. Det kan da være diskutabelt om dette er ett stort nok tidsintervall til at rateren ikke husket billedtakingen fra originalsituasjonen. Ettersom målingene av senetykkelsen på bildene først ble gjort i etterkant av at all datainnsamling var ferdig, kan tidsintervallet mellom undersøkelsene likevel forsvares.

### 5.1.3 Utvalg og ekstern validitet

Ekstern validitet handler om generaliserbarhet (Domholdt 2005) Det er viktig at reliabiliteten testes ut på akkurat den typen deltakere som måleinstrumentet er tenkt brukt til i klinikken (Domholdt 2005). Det var i denne studien ett mål at utvalget skulle være representativt i forhold til den allmenne populasjon. Et stort randomisert utvalg er den beste måten å sikre representativitet på. Størrelse på utvalg som er nødvendig for å kunne få statistisk signifikante resultat anbefales beregnet statistisk ved hjelp av styrkeanalyse (Altman 1991; Polit og Beck 2004; Batterham og George 2000). I tillegg vil konfidensintervall for ICC verdiene være viktig for å illustrere spredningen i studiepopulasjonen. Dårlig presisjon av ICC verdier med stor spredning i konfidensintervallet kan forekomme på grunn av for få deltakere (Batterham og George, 2000). Hopkins (2000) mener at 50 deltakere med minst tre forsøk vil være nødvendig for å estimere reliabilitet. I litteraturen sees det imidlertid at 2 målinger er mest anvendt. I denne studien ble 20 deltakere inkludert, med 2 testforsøk. For å få mer robuste data kunne vi ha inkludert flere deltakere og gjort flere testforsøk.

Utvalgsstørrelse ble i denne studien estimert ut i fra tidligere, lignende studier. Det ble foretatt ett stratifisert utvalg hvor deltakerene ble trukket systematisk fra de ulike strata. De inkluderte deltakerne i studien representerer på denne måten i stor grad personer som kan få rotatorcuff tendinopati med hensyn til alder, kjønn og KMI.

Peat et al (2002) understreker at høy ekstern validitet i kliniske forsøk blant annet er avhengig av klare inklusjons- og eksklusjonskriterier. I denne studien ble deltakere med skuldersmerter med varighet over 3 måneder inkludert, i tillegg til at deltakere med tidligere kirurgiske inngrep og brudd under tilheling i skulder, samt neurologiske og revmatiske sykdommer ble ekskludert. Av tabell 4; "demografiske data", kan man se at det i smertegruppe var deltakere som faktisk oppga å ikke ha smerte i skulderen/e nå. Videre har vi i smertegruppe valgt å undersøke skuldre bilateralt. Det var ingen av de inkluderte deltakerne som hadde smerte i begge skuldre ved undersøkelse. I analyse av data har vi heller ikke registrert hvilken skulder som faktisk var smertefull. Dette er en klar svakhet med studien. Dette vil si at vi har en blanding av deltakere med friske og smertefulle skuldre i smertegruppen, og også antageligvis friske deltakere/eventuelt deltakere med andre skulderplager enn tendinopati i smertegruppe. Å ha en ekstremt heterogen gruppe vil kunne føre til at reliabiliteten blir overestimerert for bruk i praksis (Domholdt, 2005). For å unngå dette, kunne vi ha inkludert en grundig anamnese i tillegg til en klinisk undersøkelse for rekruttering av passende deltakere til smertegruppen. I tillegg burde vi skilt mellom smertefull og ikke-smertefull skulder i data-analysen, samt registrert hvilken side deltakerne hadde smerte.

I Domholdt (2005) blir variasjon av de inkluderte deltakerne nevnt som en viktig faktor for å unngå over- eller underestimert av reliabiliteten. For like målinger vil kunne gi lavere korrelasjons koeffisient selv om verdien på to målinger er nokså like, mens for stor variasjon vil kunne føre til det motsatte (Domholdt 2005, s.269). Bland Altman plotene i denne oppgaven viste at det var en jevn spredning av materialet, noe som tyder på at relativt like målinger ikke har hatt negativ innvirkning på reliabiliteten.

#### 5.1.4 Datainnsamling og prosedyre for undersøkelse

Av praktiske og ressursmessige grunner har forfatter av denne oppgaven og medstudent valgt å utføre alle målinger og datainnsamling selv. Dette er noe av kritikken til studien, da forforståelse og forventninger i forhold til resultater i studien kan være svakhet for prosjektets validitet og reliabilitet. Det er derimot en styrke i

studien at det er to ratere som har gjennomført datainnsamling og utføring av målinger, og at disse raterene var blindet for hverandres resultater. Flere ratere vil øke generaliserbarheten til klinikken (Van Genderen et al, 2003) ettersom det er viktig at ultralydsmålinger av senetykkelse i tillegg til god intraraterreliabilitet også har god interraterreliabilitet.

## 5.2 Valg av statistiske metoder

Hvilke statistiske metoder man skal bruke ved reliabilitetstudier har vært mye diskutert i løpet av de siste årene (Streiner og Norman, 2003), og flere statistiske tester har blitt foreslått for å beskrive reliabilitet eller validitet til et måleinstrument (Atkinson og Nevill, 1998). I denne studien er ICC benyttet for å beskrive relativ reliabilitet mens en Bland-Altman analyse med  $\pm 1.96SD_{diff}$  og 95% limits of agreement er benyttet for å beskrive absolutt reliabilitet. I følge Atkinson og Nevill (1998) er  $\pm 1.96SD_{diff}$  og LoA en god metode for å beskrive reliabilitet mellom ratere som bruker samme måleinstrument. En av fordelene med Bland-Altman analyse er at måleforskjellene kan vises skjematisk i plot og at det er mulig å se om resultatene er heteroscedastiske. Heteroscedastisitet er et uttrykk for at variasjon mellom målinger øker proporsjonalt med størrelsen på målingen (Domholdt 2005). Plotene viste ingen klar tendens til heteroscedastisitet og resultatene ble dermed behandlet som homoscedastiske. Ved å beregne  $\pm 1.96SD_{diff}$  og 95 % LoA vil en kunne vurdere en mulig sammenheng mellom målefeilen og den sanne verdien, i motsetning til for eksempel om man kun beregner ICC (Bland og Altman, 1986).

Domholdt (2005) anbefaler at en ved undersøkelser gjennomfører flere målinger og beregner en gjennomsnittsverdi av disse. Det vil da bli en økt sannsynlighet for reliable målinger. I denne studien er det allikevel valgt å bruke enkeltmålinger som grunnlag for å vurdere reliabilitet, da dette vil kunne øke generaliserbarheten til den allmenne populasjon.

### 5.2.1 Relativ reliabilitet

I denne studien er det tatt hensyn til de systematiske forskjellene, det vil si forskjeller i gjennomsnittsverdier for interrater- og test-retest målingene, i beregningen av ICC. Dersom de systematiske målefeilene er betydelige, vil valg av modell påvirke størrelsen på ICC hvor modell 3 > modell 2 > modell 1 (Portney og Watkins, 2000; Weir, 2005). I våre resultater var komponenten av tilfeldige målefeil mye større enn de systematiske. Da vil valg av modell ha mindre betydning for størrelsen på ICC, og resultatene fra ulike ICC'er vil være mer like.

Reliabiliteten kan forbedres ved å øke variabiliteten mellom subjektene og redusere andelen systematiske og tilfeldige målefeil (Streiner og Norman, 2008). I denne studien kunne vi oppsummert sett hatt andre inklusjonskriterier, økt antall runder med undersøkelse/antall bilder, økt antall deltakere og eventuelt økt antall ratere. For å redusere målefeilene kunne vi gjennomført grundigere opplæring og samkjøring av raterene. Imidlertid var målet med denne studien å se hvor reliable målinger ratere uten tidligere erfaring og med kun grunnleggende opplæring ville oppnå ved måling av senetykkelse på ultralydsbilder.

### 5.2.2 Absolutt reliabilitet

Absolutt reliabilitet er det viktigste målet for reliabilitet for forskere og helsepersonell da det viser et eksakt estimat av forandring i en måling (Hopkins, 2000). Det vil ikke være nok å bare estimere relativ reliabilitet uttrykt i ICC, da klinikerer må vite noe konkret om marginene omkring målefeilen for å kunne definere om det er oppstått en endring i en måling (Van Genderen et al, 2003). Klinikerer vil ha nytte av å vite om det er skjedd en reell endring eller om variasjon i målinger kan tilskrives målefeil. For å vite noe om variasjon rundt den sanne verdien av målingene må verdier for absolutt reliabilitet estimeres (Bland og Altman, 1986). Det vil være en fordel å framstille et Bland Altman-plot for å illustrere resultatene visuelt, da dette vil vise en mulig sammenheng mellom målefeilen og den sanne verdien (Bland og Altman, 1986). For relativ reliabilitet finnes det graderinger for hva som regnes som lav eller høy reliabilitet (Domholdt, 2005). Disse er anvendbare i forhold til å estimere reliabilitet

ved UL-måling av senetykkelse. Det finnes derimot ingen enighet for hvilke verdier av absolutt reliabilitet som kan anses som tilfredsstillende ved måling av tverrsnittet til skuldereener.

### 5.3 Resultatdiskusjon

I dette kapittelet vil resultatene i denne studien i forhold til interrater- og test retest reliabilitet bli diskutert.

#### 5.3.1 Interraterreliabilitet

Interraterreliabilitet i denne analysen uttrykker i hvilken grad rater 1 og rater 2 er konsistente i utføringen av målinger på tverrsnittet til supraspinatus senen og biceps tendon longus på det samme ultralydsbildet. Utføringen av måleprosedyren med tolkning av sonoanatomien på bildet er bestemmende for grad av enighet i målingene. Noe av utfordringen i forhold til dette er ved billedtaking å justere transduseren slik at en får optimalt ekko av senestrukturen og å definere målepunktet likt.

Interraterreliabiliteten er en analyse av målingene for rater 1 og rater 2 ved 1. måletidspunkt og ved 2. måletidspunkt.

ICC verdiene for 1. måling er høyest ved tverrsnittsmålinger av biceps longitudinal i begge gruppene. Her er det ett unntak - 1. måling av venstre biceps longitudinal i frisk gruppe, som kun oppnådde dårlig reliabilitet med en ICC verdi på 0.47. Grunnen til den lave ICC-verdien kan man se utifra dens tilhørende Bland-Altman plot (vedlegg 4). Her sees det 3 klare uteliggere/ekstremverdier i den ene enden av måleskalaen. Disse ekstremverdiene i en ende av måleskalaen, vil påvirke ICC verdien i negativ retning. Bortsett fra disse 3 verdiene, viser Bland-Altman plottet til god enighet og god samling rundt gjennomsnittsverdien til differansen mellom måingene ( $X_d$ ).

Også ved 2. måling oppnådde raterene høy relativ reliabilitet ved tverrsnittsmålinger av biceps longitudinal bilateralt i begge gruppene. Dette reflekteres også i verdiene for absolutt reliabilitet, som var høyest for 2. måling av biceps longitudinal bilateralt i begge gruppene. Videre var det ved 2. måling også høy relativ reliabilitet ved

målinger på høyre biceps tverrsnitt i begge gruppene og venstre biceps tverrsnitt i smertegruppe, samt supraspinatus bilateralt i frisk gruppe.

Samlet sett er ICC (1.1) verdiene høyere for 2. måling enn ICC (1.1) verdiene for 1. måling, noe som kan tyde på at det har vært en viss læringseffekt blant raterene i studien. Raterene har nå fått noe mer erfaring og man kan derfor anta at det har skjedd en bedring av ferdighetene. Dette gjenspeiles også i variabiliteten i målingene, uttrykt ved  $\pm 1.96 SD_{diff}$ . Disse er samlet sett større ved nesten alle de 1. målingene enn ved de 2. målingene.

Ved begge måletidspunkt i begge grupper er ICC (2.1) tilnærmet lik eller lik ICC (1.1), noe som tyder på lite systematiske målefeil. Det er imidlertid ett unntak: 2. måling venstre rotatorcuff i smerte gruppe. Her er det en differanse i ICC verdi på 0.07. Tilfeldige feil i ett datamateriale vil ikke påvirke gjennomsnittet av målingene, men kun variasjonen rundt gjennomsnittet. Systematiske feil vil derimot påvirke gjennomsnittet, som er uttrykt som  $X_d$  i denne studien.  $X_d$ , som da er ett uttrykk for systematiske målefeil, er ved rotatorcuff venstre i smertegruppe på 0.05mm. I følge Atkinson og Nevill (1998) vil endringer i egenskaper ved måleinstrumentet kunne føre til systematiske målefeil. Dette er etter min mening en lite sannsynlig forklaring i denne studien. En mer sannsynlig forklaring er at disse undersøkelsene har små utvalg, og at tilfeldig variabilitet fører til slike endringer som viser seg som systematiske målefeil. Muligens gjenspeiler differansen i ICC verdi (0.07) og differansen i  $X_d$  (0.05mm) utfordringene med å ta gode, like bilder av denne strukturen, samt å sette konsistente og presise mål på disse bildene.

Oppsummert er det målinger av biceps longitudinal som mest konsistent viser til høye reliable målinger. Dette er ganske forventet, ettersom det er en relativt enkel struktur å sette mål på (parallele linjer). På den annen side hadde vi i forkant en viss forventning om at supraspinatus longitudinal ville gi målinger med høy reliabilitet, grunnet den karakteristiske sonoanatomien av senen med "hakket" for hvor målepunkt settes (beinlinjens brytningspunkt mellom tuberculum majus og caput humeri). Supraspinatus viste derimot kun til høy reliabilitet i frisk gruppe ved 2. måletidspunkt. Ved 1. måletidspunkt i frisk gruppe og ved begge målinger i smertegruppe er IRR



overraskende lav (moderat reliabilitet) for supraspinatus. Hva dette skyldes er vanskelig å svare på. Målinger av rotatorcuff viser som forventet kun dårlig til moderat reliabilitet ved alle målinger. Rotatorcuff, med supraspinatus inkludert, er både utfordrende å ta bilde av, og å sette konsistente og presise målinger på. Det er ett bildesnitt med flere strukturer inkludert, og det finnes ingen klare, åpenlyse plasser for hvor man skal sette de to målepunktene.

### 5.3.2 Test- retest reliabilitet

Forutsetningen for høy test-retest reliabilitet er god korrelasjon mellom verdiene ved gjentatte målinger (Domholdt, 2005). Test-retest reliabilitet for rater 1 og rater 2 er i denne oppgaven bestemt ut i fra hvor konsistent raterne sine 2. målinger er på egne bilder tatt ved 1. undersøkelse sammenlignet med bilder tatt ved 2. undersøkelse. Test-retest reliabiliteten vil her uttrykke hvor konsistente raterne er i plasseringen av transduseren på de ulike senene hos samme deltaker og hvor nøyaktige målinger raterne gjør.

Begge raterne hadde god relativ reliabilitet ved målinger på biceps longitudinal bilateralt i begge gruppene samt biceps tverrsnitt bilateralt og venstre supraspinatus i smertegruppe. Også ved beregning av absolutt reliabilitet var det for begge raterne biceps longitudinal bilateral i begge gruppene som viste til høyest reliable verdier.

Videre hadde rater 2 totalt tre målepunkt med moderat reliabilitet; høyre og venstre supraspinatus frisk gruppe venstre rotatorcuff i smertegruppe. Rater 1 hadde moderat reliabilitet på tilsammen 8 målepunkt; biceps tverrsnitt bilateralt, rotatorcuff bilateralt og venstre supraspinatus i frisk gruppe samt rotatorcuff bilateralt og høyre supraspinatus i smertegruppe. Rater 1 hadde lav reliabilitet ved målinger på høyre supraspinatus i frisk gruppe.

Rater 2 viser generelt til flere målepunkt med høyere relativ reliabilitet enn rater 1, og hadde god reliabilitet også ved målinger på rotatorcuff bilateralt og biceps tverrsnitt bilateralt i frisk gruppe samt høyre supraspinatus og rotatorcuff i smertegruppe.

Variabiliteten uttrykt ved  $\pm 1.96SD_{diff}$  er samlet sett større ved nesten alle rater 1 sine målinger i forhold til rater 2 sine målinger. Rater 1 bruker altså en større del av måleskalaen og har det man kan kalle en ekstremverdi ved blant annet biceps tverrsnitt venstre. En slik ekstremverdi i kun en ende av måleskalaen vil påvirke ICC i negativ retning. Man kan dermed bli lurt til å tro at rater 2 sin ICC ved målinger på biceps tverrsnitt venstre er vesentlig bedre enn ved rater 1 sine målinger enn om den ene ekstremverdien hadde vært ekskludert. Samtidig kan rater 2 vise til generelt mer jevne og presise målinger enn rater 1, og har også derfor følgelig høyere ICC verdier.

Ved alle rater 2 sine målinger er  $ICC(3.1)=ICC(1.1)$  i begge grupper, noe som tilsier at det sannsynligvis ikke har forekommet læring og at andelen av systematiske feil er liten (Moe-Nilssen, 1998). Rater 1 har to tilfeller der  $ICC(3.1)>ICC(1.1)$ . I frisk gruppe viser målinger ved høyre supraspinatus at det er ett betydelig systematisk skift i datamaterialet på 0.08 ( $ICC(1.1)=0.48$ ,  $ICC(3.1)=0.56$ ). Dette kan muligens være betinget at bilder tatt ved 1. og 2. undersøkelse avviker fra hverandre grunnet at rateren har plassert transduseren ulikt de to gangene, eller det har forekommet læring og utvikling av ferdigheter hos rateren. Videre har rater 1 ved venstre supraspinatus i smerte gruppe et systematisk skift på 0.1 ( $ICC(1.1)=0.77$ ,  $ICC(3.1)=0.78$ ). Dette er et svært lite, og ubetydelig systematisk skift. Ved resten av målingene til rater 1 i begge grupper er  $ICC(1.1)=ICC(3.1)$ . Totalt sett foreligger det med andre ord svært lite systematiske feil ved test-retest hos begge raterene.

Robustheten til datamaterialet er mindre ved test-retest enn ved intraraterreliabilitet grunnet færre antall bilder og målinger. Man kan anta at tilfeldige målefeil ville ha utjevnet seg dersom utvalget hadde vært større (Hopkins, 2000).

#### 5.4 Helhetlig tilnærming og helsereelatert livskvalitet

Dagens biomedisin er i stor grad resultat av en empiristisk filosofi hvor tanken er at alle menneskelige problemer har sin opprinnelse - og løsning - i fysiologien (Thornquist, 2003). Forskere og klinikere har i lang tid vært opptatt av å bruke objektive mål på patologi i form av eksempelvis laboratorieprøver og radiologi for veiledning i behandling (Domholdt, 2005). Forskning som har et biopsykososialt fokus

vil være forankret i en fenomenologisk forståelse, hvor utgangspunktet er at mennesker lever i en verden med mening og betydning. Fenomenologien tar sterk avstand til kropp-sjeldualismen, og mener at kropp og sjel henger sammen. (Thornquist, 2003). Som klinikere bør vi ha en helhetlig tilnærming til pasienter og ikke kun gjøre en isolert undersøkelse av pasientens affiserte kroppsdel. Det finnes ingen prinsipiell motsetning mellom å foreta målinger ved hjelp av UL og dokumentere meningsfulle fenomener. Resultatene fra en ultralydsavbildning av skulder kan brukes som et supplement til anamneseopptak og andre kliniske undersøkelser som i større grad kan evaluere funksjon/funksjonshemming i henhold til ICF. Skulderplager er med på å redusere livskvalitet og mange opplever å ikke kunne delta i aktiviteter de ønsker på grunn av smerter (Hill et al, 2010). Livskvalitet er ett uttrykk som ofte benyttes i forbindelse med helse både på administrativt nivå, i vurdering av behandlingstilbud og i klinisk behandling av individuelle tilfeller (Mooney, 2006). Verdens helseorganisasjon (1995 s. 1045) definerer livskvalitet som *"an individual's perception of their position in life in the context of the culture and value system in which they live and in relation to their goals, expectation and standards and concerns"*. Begrepet omfatter individets fysiske helse, psykologiske tilstand, sosiale forhold, grad av uavhengighet og forhold til ulike miljømessige faktorer (Lorenz og Koller, 2002). Helserelatert livskvalitet er et smalere begrep som inkluderer fysisk- og mental helse og konsekvensene av fysisk og mental helse (Mooney, 2006). Dette er viktige faktorer som bør tas hensyn til ved undersøkelse og behandling av skulderrelaterte plager, og som bør kartlegges individuelt. Eksempel på et egnet måleinstrument for pasienter med skulderplager er SF-36, som er et generisk (generelt) livskvalitetsinstrument konstruert for å gjøre en rask måling av helselatert livskvalitet. Det kan benyttes i generelle populasjonsundersøkelser, i klinisk forskning og i daglig klinisk praksis (Loge og Kaasa, 1998). Slike generelle instrumenter er designet for å samle inn et stort spekter av funksjon, handikap og ubehag som er relevant for livskvalitet (Spilker 1990). I følge Loge og Kaasa (1998) kan slike generelle livskvalitetsskjema blant annet vise hva som er av betydning for pasienten i en større sammenheng. Sammenlignet med to andre generiske måleinstrumenter for helselatert livskvalitet, har SF-36 høyest reliabilitet, validitet og sensitivitet for å vurdere livskvalitet (Dempster og Donnelly, 2000).

### 5.5 Etske hensyn.

I 2007 trådte lov om behandling av etikk og redelighet i forskning i kraft (Laake et al, 2008). Jamfør paragraf 3 i loven var det krav om at forskningsprosjekter i Norge som innebærer forsøk på mennesker skal legges fram for Regionale komiteer for medisinsk og helsefaglig forskningsetikk (REK) til godkjenning. Loven fastsetter ikke hvilke regler REK skal bedømme innsendte forskningsprosjekter etter, og man må derfor anta at de fremdeles vil bli bedømt etter et internasjonalt regelsett som omfatter Helsinki - deklarasjonen. Innholdet i informert samtykke er spesifisert i Helsinki-deklarasjonens artikkel 22 (Laake et al, 2008). Hensikten med et slikt samtykke er å ta hensyn til enkeltindividet og etter utforming av Helsinki-deklarasjonen i 1964 skal dette hensynet alltid komme i første rekke. I denne oppgaven ble informert samtykke utarbeidet i forkant av studien og ble sammen med prosjektbeskrivelsen sendt til REK. Det er viktig at personer som skal delta i et forskningsprosjekt får relevant opplysninger om studiet, informasjon om eventuell risiko og at de når som helst kan trekke seg (Ruyter, 2003). I denne studien er det ingen potensiell risiko, men det kunne tenkes at noen av deltakere ville oppleve økt smerte i etterkant av undersøkelse på grunn av ulike bevegelser og statisk hold av armen i en bestemt posisjon. Dette ble nøye informert om muntlig før oppstart av testingen.

### 5.6 Kliniske implikasjoner

Dersom en får konstantert at ultralydsavbildning i forhold til diagnostisering av tendinopati er en reliabel undersøkelsesmetode for denne pasientgruppen i primærhelsetjenesten, åpner dette for nye muligheter for disse pasientene. Muligheten for å gjøre presise målinger av tverrsnittet til skuldereener vil sammen med bruk av Doppler for å vurdere intratendinøs strøming gi ett nytt verktøy, også for fysioterapeuter, i diagnostiseringen av denne pasientgruppen. Ultralydsavbildning kan da bli implementert i primærhelsetjenesten som en standard undersøkelsesmetode for pasienter med smertefulle skuldre for å få en bedre forståelse av skulderplager.

## 5.7 Videre forskning

Skulderplager har en tendens til å bli langvarige, og det kan være fristende å påstå at pasienter med skulderplager er forhåndsbestemt til å utvikle en kronisk tilstand dersom de ikke mottar adekvat behandling. Dette kan føre til betydelig redusert helse og livskvalitet i tillegg til store sosioøkonomiske kostnader. Tilnærming bør fokusere på å forebygge progresjon av tilstanden og tidlig intervensjon ved slike plager er dermed essensielt. Dette krever at man samler mer informasjon på ett tidlig stadium om kvaliteten og dynamikken av rotatorcuff'en.

For å kunne avkrefte eller konkludere med resultatene i denne studien, bør det gjennomføres mer forskning på området som inkluderer større studier med flere deltakere. Disse studiene kan med fordel gjennomføres i primærhelsetjenesten med tanke på implementering på dette nivå. Det kunne videre vært interessant med studier som hadde ett prospektivt design der resultatene ved UL avbildning- og måling ble sammenlignet med utfallet av skulderplager (utfallsmål). Studiene bør inneholde en vurdering av både relativ reliabilitet og absolutt reliabilitet, for å sikre god overførbarhet og mulighet for implementering i praksis. Absolutt reliabilitet bør inkluderes som en standard i reliabilitetsstudier og det kunne med fordel vært utarbeidet retningslinjer for gradering av absolutt reliabilitet ved tverrsnittsmåling av skuldere med UL.

Framtidig forskning bør videre fokusere på diagnostisk nøyaktighet i settinger med lavere prevalens, oppfølgingsstudier av pasienter med skulderplager i primær og sekunderhelsetjenesten for å evaluere sykdomsutvikling og epidemiologi, og studier rundt kostnadseffektivitet ved tilnærming som inkluderer diagnostisk ultralyd i vurderingen av behandlingstiltak.

## 6.0 Konklusjon

I dette prosjektet er det gjort undersøkelser av intrarater-, interrater og test-retest reliabilitet. Denne oppgaven inneholder 16 ulike undersøkelser av interrater reliabilitet og test-retest reliabilitet. To ratere gjorde tilsammen 1080 målinger på to ulike bildesnitt av skuldansenene supraspinatus og biceps tendon longus.

Ultralydsapparatet brukt i studien er "Ultrasonography GE Logic E" med en 10-Mhz lineær transduser. I studien deltok det totalt 20 personer; 10 friske og 10 med smertefulle skuldre. Hele utvalget inngikk i alle undersøkelsene.

For *biceps longitudinal* fant en at den relative reliabiliteten var god ved 15 undersøkelser (range:0.78-0.99) og dårlig ved 1 undersøkelse (0.47). ICC (1.1) uttrykt som gjennomsnittet av de 16 undersøkelsene er 0.87, noe som indikerer god relativ reliabilitet. For *biceps tverrsnitt* fant en at den relative reliabiliteten var god ved 9 undersøkelser (range:0.76-0.97) og moderat ved 7 undersøkelser (range:0.54-0.73). ICC (1.1) uttrykt som gjennomsnittet av de 16 undersøkelsene er 0.77, noe som indikerer god relativ reliabilitet. For *supraspinatus longitudinal* fant en at den relative reliabiliteten var god ved 5 undersøkelser (range:0.77-0.87), moderat ved 8 undersøkelser (range:0.59-0.75) og dårlig ved 2 undersøkelser (0.46 og 0.48). ICC (1.1) uttrykt som gjennomsnittet av de 16 undersøkelsene er 0.67 (range: 0.46-0.87), noe som indikerer moderat grad av relativ reliabilitet. For *rotatorcuff tverrsnitt* fant en god reliabilitet ved 3 undersøkelser (range:0.92-0.96), moderat reliabilitet ved 11 undersøkelser (range:0.51-0.73) og dårlig reliabilitet ved 2 undersøkelser (0.27 og 0.32). ICC (1.1) uttrykt som gjennomsnittet av de 16 undersøkelsene er 0.67, noe som indikerer moderat grad av relativ reliabilitet.

Resultatene i denne studien viser at det er mulig for fysioterapeuter å oppnå tilfredsstillende relativ og absolutt reliabilitet på tverrsnittsmålinger av enkelte senestrukturer i skulderen selv uten lang opplæringsstid og med kun grunnleggende erfaring i bruk av UL. Allikevel er ikke resultatene i denne studien entydige nok, og det er for stor variasjon (range) i grad av reliabilitet for de enkelte strukturene til å kunne anbefale implementering på bakgrunn av undersøkelsene i denne studien. Det ble imidlertid sett en generell forbedring i ferdigheter allerede fra 1. til 2.

---

måletidspunkt (målinger med høyere reliabilitet), og man kan derfor tenke seg at selv med kun noe lenger opplæringstid og noe mer erfaring ville resultatene ha sett annerledes ut. Det vil være nødvendig med ytterligere forskning med flere inkluderte deltakere for å kunne gi ett konkluderende svar på om ultralyd som metode er reliabelt nok til å innføres som en del av den fysioterapeutiske undersøkelsen av pasienter med mistanke om rotatorcufftendinopati i primærhelsetjenesten.

**Litteraturliste**

Abate, M., Gravare-Silbernagel, K., Siljeholm, C., Di Iorio, A., De Amicis, D., Salini, V., Werner, S. og Paganelli, R. (2009) Pathogenesis of tendinopathies: inflammation or degeneration? *Arthritis Research & Therapy*, 11 (3), s. 235.

Allen, G.M. og Wilson, D.J. (2001) Ultrasound of the shoulder. *Eur J Ultrasound*, 14:3-9.

Allen, G.M. (2008) Shoulder ultrasound imaging-integrating anatomy, biomechanics and disease processes. *European Journal of Radiology* 68, 137–146

Altman, D.G. (2000). *Practical statistics for medical research*. (2nd ed ed.) London ; New York: Chapman and Hall.

Andersen, I., Frydenberg, H. og Mæland, J.G. (2009) Muskel- og skjelettplager og fremtidig sykefravær. *Tidsskr Nor Legeforen*, 129: 1210-3.

Arbeids- og velferdsetaten (NAV) *Sykefraværstilfeller 1 kv 2012*. Nedlastet 25.09.2012 fra: <http://nav.no/321006.cms>

Atkinson, G. og Nevill, A.M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.*, 26, 217- 238

Aune, A.K. (2010) Kroniske skulderlidelser i: *Idrettsskader*. Redigert av Bahr, R. og Mæhlum, S. Bergen, Fagbokforlaget.

Awerbuch, M.S. (2008) The clinical utility of ultrasonography for rotator cuff disease, shoulder impingement syndrome and subacromial bursitis. *Med J Aust* 188:50–53

Badcock, L.J., Lewis, M., Hay, E.M., McCarney, R. og Croft, P.R. (2002) Chronic shoulder pain in the community: a syndrome of disability or distress? *Ann Rheum Dis* 61, 128–31.



Batterham, A.M. og George, K.P. (2003). Reliability in evidence-based clinical practice: A primer for allied health professionals. *Physical Therapy in Sport*, 4, 122-128.

Beaudreuil, J., Nizard, R., Thomas, T., Peyre, M., Liotard, J.P., Boileau, P., Marc, T., Dromard, C., Steyer, E., Bardin, T., Orcel, P. og Walch, G. (2009) Contribution of clinical tests to the diagnosis of rotator cuff disease: A systematic literature review. *Joint Bone Spine.*, 76 s.15-19.

Bjordal, J., Demmink, J. og Ljunggren, A. (2003) Tendon Thickness and Depth from Skin for Supraspinatus, Common Wrist and Finger Extensors, Patellar and Achilles Tendons: Ultrasonography study of healthy subjects. *Physiotherapy*, 89 (6) s. 375–383.

Blacker, D., og Endicott, J. (2003). *Psychometric properties: concepts of reliability and validity*. In American Psychiatric Association. Task Force for the Handbook of Psychiatric Measures (Ed.), *Handbook of psychiatric measures*, Washington, DC: American Psychiatric Association. (pp. 7-14).

Bland, J.M. og Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurements. *Lancet*: 307-310.

Boesen, M.I., Koenig, M.J., Torp-Pedersen, S., Bliddal, H. og Langberg, H. (2006) Tendinopathy and Doppler activity: the vascular response of the Achilles tendon to exercise. *Scand J Med Sci Sports*, 16 s. 463–469.

Bolvig, L., Fredberg, U. og Rasmussen, O.S. (2011) *Textbook on musculoskeletal ultrasound - for beginners and trained*. Danmark, Munksgaard.

Bouffard, J.A., Lee, S.M. og Dhanju, J. (2000) Ultrasonography of the shoulder. *Semin Ultrasound CT MR.*, 21 s.164–191

- Brage, S., Ihlebæk, C., Natvig, B. og Bruusgaard, D. (2010) Muskel- og skjelettlidelser som årsak til sykefravær og uføreytelser. *Tidsskr Nor Legeforen*, 130:2369-70
- British Standards Institution. *Precision of test methods 1: Guide for the determination and reproducibility for a standard test method (BS 597, Part 1)*. London: BSI (1975).
- Bureau, N.J og Roederer, G. (1998). Sonography of Achilles tendon xanthomas in patients with heterozygous familial hypercholesterolaemia, *American Journal of Roentgenology*, 171, 745-749.
- Buchbinder, R., Goel, V., Bombardier, C. og Hogg-Johnson, S. (1996) Classification systems of soft tissue disorders of the neck and upper limb: do they satisfy methodological guidelines? *J Clin Epidemiol*. 49: 141-149.
- Carcia, C.R., Martin, R.L., Houck, J. og Wukich, D.K. (2010) Achilles pain, stiffness, and muscle power deficits: achilles tendinitis. *J Orthop Sports Phys Ther*, 40 (9), s. 21-26.
- Carr, A. og Harvie, P. (2005) Rotator Cuff Tendinopathy i: *Tendon Injuries : Basic Science and Clinical Medicine* Edited by Maffulli, N., Renstrom, P. og Leadbetter, W. B. Pp. 332. London: Springer.
- Chakravarty, K.K. og Webley, M. (1990) Disorders of the shoulder: an often unrecognised cause of disability in elderly people, *Br. Med. J.* 300 p. 848–849
- de Jesus, J.O., Parker, L., Frangos, A.J. og Nazarian, L.N. (2009) Accuracy of MRI, MR arthrography, and ultrasound in the diagnosis of rotator cuff tears: a meta-analysis. *AJR Am J Roentgenol* 2009;192:1701-7.
- Danneskiold-Samsøe, B., Lund, H., og Wæhrens, E.E. (2011) *Klinisk Reumatologi*. København, Munksgaard Danmark.

de Winter, A.F., Jans, M.P, Scholten, R.J.P.M., Devillé, W., van Schaardenburg, D. og Bouter, L.M. (1999) Diagnostic classification of shoulder disorders: interobserver agreement and determinants of disagreement. *Ann Rheum Dis.* 58: 272-277.

Dempster, M. og Donnelly, M. (2000) Measuring the health related quality of life people with ischaemic heart disease. *Heart* 83, s. 641-644.

Dinnes, J., Loveman, E., McIntyre, L. og Waugh, N. (2003) The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review. *Health Technol Assess.* 7, p. 1-166.

Domholdt, E. (2005). *Rehabilitation research: principles and applications*. St. Louis, Elsevier Saunders.

Enwemeka, C.S. (2001). Attenuation and penetration depth of red 632.8 nm and invisible infrared 904 nm light in soft tissues, *Laser Therapy*, 13, 95-101.

Fisher, R.A. (1925) *Statistical methods for research workers*. Biological Monographs and Manuals Series. Published in Edinburgh by Oliver and Boyd.

Fu, S.C., Rolf, C., Cheuk, Y.C., Lui, P.P. og Chan, K.M. (2010) Deciphering the pathogenesis of tendinopathy: a three-stages process. *Sports Med Arthrosc Rehabil Ther Technol*, 2, s. 30.

Hopkins, W.G. (2000). Measures og reliability in sports medicine and science. *Sports Med*(1): 1-15.

Hayes, K.W. og Petersen, C.M. (2003) Reliability of classifications derived from Cyriax's resisted testing in subjects with painful shoulders and knees. *J Orthop Sports Phys Ther.* 33(5):235-46.

Hoksrud, A., Ohberg, L., Alfredson, H. og Bahr, R. (2008) Color Doppler ultrasound findings in patellar tendinopathy (jumper's knee). *Am J Sports Med*, 36 (9), s. 1813-20.

- Hughes, P.C., Taylor, N.T. og Green, R. (2008) Most clinical tests cannot accurately diagnose rotator cuff pathology: a systematic review. *Australian Journal of Physiotherapy*, (54) p. 159-170.
- Hill, C.L., Gill, T.K., Shanahan E.M. og Taylor, A.W. (2010) Prevalence and correlates of shoulder pain and stiffness in a population-based study: the North West Adelaide Health Study. *International Journal of Rheumatic Diseases*, 13: 215–222
- Ihlebak, C. og Lærum, E. (2004) Plager flest - koster mest, muskel - skjelettlidelser i Norge. Oslo. Rapport nr. 1. *Nasjonal ryggnettverk Forskningsenheten/ formidlingsenheten*. s. 1-64.
- International Association for the Study of Pain (1979) *Pain journal*. (6), p. 250
- Jacobson, J.A. (2003) Sonography of the shoulder. *Applied Radiology*, s. 19-25
- Jacobson, J.A., van Holsbeeck M.T. (1998) Musculoskeletal ultrasonography. *Orthop Clin North Am* 29:135-67.
- Jacobson, J.A. (2007) *Fundamentals of Musculoskeletal Ultrasound*. Elsevier Health Sciences.
- Jamtvedt, G., Hagen K.B og Bjørndal, A. (2003) *Kunnskapsbasert fysioterapi. Metoder og arbeidsmåter*. Gyldendal Norsk Forlag AS.
- Kader, D., Saxena, A., Movin, T. og Maffulli, N. Achilles tendinopathy: some aspects of basic science and clinical management. *Br J Sports Med* 2002;36:239–49.
- Kapandji, I.A. (1982). *The Physiology of the joints*. Edinburgh London Melbourne And New York, Churchill Livingstone.
- Katayose, M. og Magee D.J. (2001) The cross-sectional area of supraspinatus as measured by diagnostic ultrasound. *J Bone Joint Surg Br* 83-B (4) s. 565-568.

- Laake, P., Olsen B.R. og Benestad, H.B. (2008) *Forskning i medisin og biofag* Oslo, Gyldendal Norsk Forlag AS.
- Le Corroller, T., Cohen, M., Aswad, R., Pauly, V. og Champsaur, P. (2008) Sonography of the painful shoulder: role of the operator's experience. *Skeletal Radiol*, 37(11):979–986.
- Lewis J.S. (2008) Rotator cuff tendinopathy. *British Journal of Sports Medicine*. 43, s. 236–241
- Li, L., Liu, X. og Herr, K. (2007) Postoperative pain intensity assessment: a comparison of four scales in Chinese adults. *Pain Med*, 8 (3), s. 223-34.
- Loge, J. H. og Kaasa, S. (1998) Short Form 36 (SF-36) health survey: normative data from the general Norwegian population. *Scandinavian Journal of Social Medicine*, 26 (4), s. 250-258.
- Lorenz, W. og Koller, M. (2002) Quality of life: a deconstruction for clinicians. *Journal of the royal society of medicine*, 95, s. 481-488.
- Lundgreen, K., Lian, Ø.B., Engebretsen, L. og Scott, A. (2011). Tenocyte apoptosis in the torn rotator cuff a primary or secondary pathological event?" *Br J Sports Med*. p. 1035-1039
- Luime, J.J., Koes, B.W., Hendriksen, I.J., Burdorf, A., Verhagen A.P., Miedema, H.S. og Verhaar, J.A. (2004) Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol* 33, 73–81.
- Leung, J.L.Y. og Griffith, J.F. (2008) Sonography of Chronic Achilles Tendinopathy: A Case–Control Study. *Journal of clinical ultrasound*. (36) p. 27-32
- Maffulli, N., Regine, R., Angelillo, M., Capasso, G. og Filice, S. (1987) Ultrasound diagnosis of Achilles tendon pathology in runners. *Br J Sports Med*, 21 (4), s. 158-62.

Maffulli, N., Sharma, P. og Luscombe, K.L. (2004) Achilles tendinopathy: aetiology and management. *J R Soc Med.* 97:472–6.

Maffulli, N., Wong, J. og Almekinders, L.C. (2003) Types and epidemiology of tendinopathy. *Clin Sports Med*, 22 (4), s. 675-92.

Magee, D.J. (2002). *Orthopedic Physical Assessment*. Philadelphia, Sanders Elsevier.

Mäkelä, M., Heliövaara, M., Sainio, P., Knekt, P., Impivaara, O. og Aromaa, A. (1999) Shoulder joint impairment among Finns aged 30 years or over: prevalence, risk factors and co-morbidity. *Rheumatology* 38, 656–62.

Martin-Hervas, C., Romero, J., Navas-Acien, A., Reboiras, J.J., Mu nuera, L. (2001) Ultrasonographic and magnetic resonance images of rotator cuff lesions compared with arthroscopy or open surgery findings. *J Shoulder Elbow Surg.* 10:410-5.

Martinoli, C., Bianchi, S., Prato, N., Pugliese, F., Zamorani, M.P., Valle, M. og Derchi, L.E. (2003) US of the shoulder: non-rotator cuff disorders. *Radiographics* 23:381-401; quiz 534.

McDowell, I. og Newell, C. (1996). *Measuring health a guide to rating scales and questionnaires*. (2nd ed ed.) New York: Oxford University Press.

McGraw, K.O. og Wong, S.P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1. 30-46

Middleton, W.D., Teefey, S.A., Yamaguchi, K. (1998) Sonography of the shoulder. *Semin Musculoskelet Radiol.* 2 s. 211–221.

Middleton, W., Teefey, S., Yamaguchi, K. (2004) Sonography of the rotator cuff: analysis of interobserver variability. *AJR Am J Roentgenol* 183:1465-8.

Moe-Nilssen, R. (1998). Test-retest Reliability of Trunk Accelerotometry During Standing and Walking. *Arch Phys Med Rehabilitation* (79): 1377-1385.

Mooney, A. (2006) Quality of life: questionnaires and questions. *Journal of health communication*, 11, s. 327-341.

Moosmayer, S., Heir, S., Aaser, P. og Smith, H. (2004) Ultralyddiagnostikk av skulderbuen - en metodebeskrivelse. *Tidsskr Nor Lægeforen*, 124:177-80

Movin, T., Kristoffersen-Wiberg, M., Rolf, C. og Aspelin, P. (1998) MR imaging in chronic Achilles tendon disorder. *Acta Radiol* 1998;39:126–32.

Murrell, G.A. og Walton, J.R. (2001). Diagnosis of rotator cuff tears. *Lancet*, 357: 769-770.

Murphy, K.R. og Davidshofer, C.O. (2005) *Psychological testing: principles and applications*. Upper Saddle River, NJ : Pearson Education International, Prentice Hall.

Naredo, A.E., Aguado, P., Padrón, M., Bernad, M., Uson, J., Mayordomo, L. og Martin-Mola, E. (1999) A comparative study of ultrasonography with magnetic resonance imaging in patients with painful shoulder. *J Clin Rheumatol*. 5:184-92.

Naredo, E., Möller, I., Moragues, C., de Agustin J.J., Scheel, A. K., Grassi, W., de Miguel, E., Backhaus, M., Balint, P., Bruyn, G. A. W., D'Agostino, M. A., Filippucci, E., Iagnocco, A., Kane, D., Koski, J.M., Mayordomo, L., Schmidt, W.A. Swen, W.A.A., Szkudlarek, M., Terslev, L., Torp-Pedersen, S., Uson, J., Wakefield, R.J. og Werner, C. (2006) Interobserver reliability in musculoskeletal ultrasonography: results from a “Teach the Teachers” rheumatologist course. *Ann Rheum Dis*, 65:14–19

Nehrer, S., Breitenseher, M., Brodner, W., Kainberger, F., Fellingner, E.J., Engel, A. og Imhof, F. (1997). Clinical and sonographic evaluation of the risk of rupture in the Achilles tendon, *Archives of Orthopaedic Trauma and Surgery*, 116, 14-18.

Nussbaum, E.L. og Downes L. (1998). Reliability of clinical pressure-pain algometric measurements obtained on consecutive days. *Phys Ther* 78: 160-169.

O'Connor PJ, Rankine J, Gibbon WW, Richardson A, Winter F, Miller JH.

O'Connor, P.J, Rankine, J., Gibbon, W.W., Richardson, A., Winter, F. og Miller, J.H. (2005). Interobserver variation in sonography of the painful shoulder. *Journal of Clinical Ultrasound* **33**(2): 53-56.

Ohrndorf, S., Naumann, L., Grundey, J., Scheel, T., Scheel, A.K., Werner, C. og Backhaus, M. (2010). Is musculoskeletal ultrasonography an operator-dependent method or a fast and reliably teachable diagnostic tool? Interreader agreements of three ultrasonographers with different training levels. *International journal of rheumatology*, 164518.

Ottenheijm, R.P., Jansen, M.J., Bart Staal, J., van den Bruel, A., Weijers, R.E., de Bie, R.A. og Dinant, G-J. (2010) Accuracy of Diagnostic Ultrasound in Patients With Suspected Subacromial Disorders: A Systematic Review and Meta-Analysis. *Arch Phys Med Rehabil* (9) p. 1616-1625

Paavola, M. og Jarvinen, T.A. (2005) Paratendinopathy. *Foot Ankle Clin.* 10:279–92.

Paavola, M., Paakkala, T., Kannus, P. og Jarvinen, M. (1998). Ultrasonography in the differential diagnosis of Achilles tendon injuries and related disorders: A comparison between pre-operative ultrasonography and surgical findings, *Acta Radiologica*, 39, 612-619

Papatheodorou, A., Ellinas, P., Takis, F., Tsanis, A., Maris, L., Batakis, N. (2006) US of the shoulder: rotator cuff and non-rotator cuff disorders. *Radiographics*, 26:e23

Peat, J. K., Xuan, W., Mellis, C., og Williams, K. (2002). *Health science research a handbook of quantitative methods*. London: Sage.

Pope, D.P, Croft, P.R, Pritchard, C.M og Silman, A.J. (1997) Prevalence of shoulder pain in the community: the influence of case definition. *Annals of the Rheumatic Diseases*, 56 s. 308–312.

Portney, L. G. og Watkins, M. P. (2000). *Foundations of clinical research: applications to practice*. (vols. 2nd ed.) Upper Saddle River, N.J.: Prentice Hall.



Polit, D. F. og Beck, C. T. (2004). *Nursing research principles and methods*. (7th ed ed.) Philadelphia, Pa: Lippincott Williams & Wilkins.

Read, J.W. og Perko, M. (1998) Shoulder ultrasound: diagnostic accuracy for impingement syndrome, rotator cuff tear, and biceps tendon pathology. *J Shoulder Elbow Surg*, 7:264-71

Rothstein, J.M. og Echternach J.L. (1993). *Primer on Measurement: An Introductory Guide to Measurement Issues*, American Physical Therapy Association.

Rutten, M.J., Spaargaren, G.J., van Loon, T., de Waal, M., Kiemeney, L.A. og Jager G.J. (2010). Detection of rotator cuff tears: the value of MRI following ultrasound. *European Radiology* 20 (2) s. 450-457.

Skovdal Rathleff, M., Moelgaard, C. og Lykkegaard Olesen, J. (2011). "Intra- and interobserver reliability of quantitative ultrasound measurement of the plantar fascia." *Journal of Clinical Ultrasound* 39(3): 128-134.

Ruyter, K. W. (2003). *Forskningsetikk. Beskyttelse av enkeltpersoner og samfunn*. Oslo, Gyldendal norsk forlag.

Shrout, P. E. og Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol.Bull.*, 86, 420-428.

Solberg, A.S. (2002). *Klinisk undersøkelse av nakke-skulder*. Kristiansand, Høyskoleforlaget.

Soslowsky, L. J., Thomopoulos, S., Esmail, A., Flanagan, C. L., Iannotti, J. P., Williamson, J. D., 3rd og Carpenter, J. E. (2002) Rotator cuff tendinosis in an animal model: role of extrinsic and overuse factors. *Ann Biomed Eng*, 30 (8), s. 1057-63.

Streiner, D. L. og Norman, G.R. (2008). *Health measurement scales: a practical guide to their development and use*. (vols. 4th ed.) Oxford: Oxford University Press.

- Strender, L.E., Sjoblom, A., Sundell, K., Ludwig, R., og Taube, A. (1997). Interexaminer reliability in physical examination of patients with low back pain. *Spine*, 22, 814-820.
- Stolik, S., Delgado, J.A., Perez, A. og Anasagasti, L. (2000). Measurement of the penetration depths of red and near infrared light in human and ex vivo tissues, *Journal of Photochemistry and Photobiology* , 57, 90-93.
- Svebak, S., Hagen, K., Zwart, J.A. (2006) One-year prevalence of chronic musculoskeletal pain in a large adult Norwegian county population relations with age and gender – the HUNT Study. *J Musculoskelet Pain* 14, 21–8.
- Spilker, B. (1990) *Quality of life assessments in clinical trials*. New York. Raven Press.
- Teefey, S.A., Middleton, W.D., Yamaguchi, K. (1999) Shoulder Sonography. State of the Art. *Radiol Clin North Am.*, 37 s. 767–785.
- Teefey, S.A., Hasan, S.A., Middleton, W.D., Patel, M., Wright, R.W og Yamaguchi, K. (2000). Ultrasonography of the rotator cuff: A comparison of ultrasonographic and arthroscopic findings in one hundred consecutive cases, *Journal of Bone and Joint Surgery (Am)*, 82, 498-504.
- Thain L.F. og Adler, R.S. (1999) Shoulder: rotator cuff and long head of biceps tendons. In: Chhem RK, Cardinal E, editors. Guidelines and gamuts in musculoskeletal ultrasound. New York: Wiley-Liss. p. 39-72
- Thornquist, E. (2003). *Vitenskapsfilosofi og vitenskapsteori for helsefag*. Bergen, Fagbokforlaget.
- Uri, D.S. (1997) MR imaging of shoulder impingement and rotator cuff disease. *Radiol Clin North Am*, 35:77-96.

Van der Heijden G.J. (1999) Shoulder disorders: a state-of-the-art review. *Baillieres Best Pract Res Clin Rheumatol*, 13, s. 287 – 309.

Van der Windt D.A, Koes B.W, de Jong B.A, Bouter L.M. (1995) Shoulder disorders in general practice: incidence, patient characteristics, and management. *Ann Rheum Dis*. 54 s. 959–964.

Van Genderen, F.R., De Bie, R.A., Helders, P.J.M. og van Meeteren, N.L.U. (2003). Reliability research: towards a more clinically relevant approach. *Physical Therapy Reviews* 8: 169-176.

Walker-Bone, K., Palmer, K.T., Reading, I., Coggon, D. og Cooper, C. (2004) Prevalence and impact of musculoskeletal disorders of the upper limb in the general population. *Arthritis Rheum* 51, 642–51.

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength & Conditioning Research (Allen Press Publishing Services Inc.)*, 19, 231-240.

Woo, S., Renstrom, P. og Arnoczky, S. (2007) *Tendinopathy in athletes*. Oxford: Blackwell Publishing

World Health Organization (2001). *International classification of functioning, disability and health: ICF*. Geneva: World Health Organization. Nedlastet 20.08.2012 fra: <http://www.who.int/classifications/icf/en/>

World Health Organization (2001). *Body Mass Index*. Nedlastet 10.09.2012 fra: [http://apps.who.int/bmi/index.jsp?introPage=intro\\_3.ht](http://apps.who.int/bmi/index.jsp?introPage=intro_3.ht)

## **Forespørsel om deltakelse i forskningsprosjektet**

### ***Reliabilitetsstudie av diagnostisk ultralyd på skuldersonene***

#### **Bakgrunn og hensikt**

Skulderdiagnostikk med ultralyd brukes hyppig i mange land, men er mindre utbredt i Norge. Diagnostisk ultralyd kan være et godt alternativ til MR og en supplerings til den kliniske fysioterapiundersøkelsen. I flere studier er det påvist at diagnostisk ultralyd er like velegnet til diagnostisering av smerter fra sener som MRI. Noen av ankepunktene mot diagnostisk ultralyd av muskel-skjelettsystemet har vært en antatt lang læringskurve, og usikkerhet knyttet til påliteligheten ved funn på ultralydsbilder. De senere år har teknologiske fremskritt innen ultralyddiagnostikk gjort apparatene bedre, samt bruken av apparatene enklere. Denne studien søker å undersøke hvorvidt fysioterapeuter uten tidligere erfaring innen bruk av diagnostisk ultralyd og med kort opplæring, oppnår høy pålitelighet ved målinger av tykkelsen på sener i skuldrene.

Dette er en forespørsel til deg om å delta i denne studien.

#### **Hva innebærer studien?**

I studien vil det bli gjort bildediagnostikk ved hjelp av ultralyd på to av senene i skulderen, senen til m. Supraspinatus og m. Biceps longus. I tillegg registreres alder, kroppsmasseindeks (KMI) og grad av smerte i skulderen/skuldrene ved bruk av en tallskala fra 1-10. I etterkant av undersøkelsen vil forskerne ved hjelp av ultralydbildene måle senetykkelsen. Deretter analyseres disse målingene for å komme frem til hvor gode forskerne er til å gjøre slike målinger - eller sagt på en annen måte, hvor pålitelige deres målinger er. Ytterlige informasjon om studien finnes i vedlegg A og B.

#### **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, som oppbevares hos prosjektleder, er eneste informasjon som knytter deg til de opplysningene og ultralydbildene som innsamles. Det er kun autorisert personell knyttet til prosjektet

som har adgang til navnelisten og som kan finne tilbake til deg. Det vil ikke være mulig å identifisere deg i resultatene av studien når 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. 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 Ingvill Fjell Naterstad, telefon 951 66 561, email [ingvill.fjell.naterstad@stud.hib.no](mailto:ingvill.fjell.naterstad@stud.hib.no) eller Silje Ingebrigtsen, telefon 984 88 523, email [silje.ingebriigtsen@stud.hib.no](mailto:silje.ingebriigtsen@stud.hib.no).

**Ytterligere informasjon om studien finnes i kapittel A – utdypende forklaring av hva studien innebærer.**

**Ytterligere informasjon om personvern og forsikring finnes i kapittel B – Personvern, biobank, økonomi og forsikring.**

### **Samtykkeerklæring følger etter kapittel B**

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

##### **Kriterier for deltakelse**

For å kunne delta i studien kan du ikke ha gjennomført noen tidligere skulderoperasjoner. Du kan heller ikke ha et brudd som er under tilheling. Vi ønsker å inkludere to grupper med deltakere i vårt prosjekt. I den ene gruppen som er planlagt inkludert, er det ønskelig med deltakere som ikke har hatt skuldersmerter. I den andre gruppen ønsker vi deltakere som har skuldersmerter med en varighet på over tre måneder. Samlet sett vil vi inkludere 20 deltakere i studien.

##### **Bakgrunnsinformasjon om studien**

Skuldersmerter forekommer hyppig i befolkningen, og de fleste skuldersmerter oppstår uten et klart forutliggende traume. Fysioterapeuter har inngående kunnskap om muskel- og skjelettsystemet, og utfører alltid en selvstendig undersøkelse før de igangsetter behandling. Å tolke billeddiagnostiske ultralydfunn kan være vanskelig,

da en ser strukturelle endringer også hos mennesker uten smerter. Men for flere pasienter kan en klinisk undersøkelse av skulder supplert med ultralyd, gjøre at en ikke trenger MR. I vurderingen av senene rundt skulderen er MR og ultralyd dokumentert likeverdige gjennom forskning (Rutten et al. 2010). Det er også vist at pasientene er mer fornøyd med ultralydundersøkelse versus MR (Middleton 2004).

Rutten MJ, Spaargaren GJ, van Loon T et al. Detection of rotator cuff tears: the value of MRI following ultrasound. *Eur Radiol* 2010; 20: 450–7.

Middleton WD, Payne WT, Teefey SA et al. Sonography and MRI of the shoulder: comparison of patient satisfaction. *AJR Am J Roentgenol* 2004; 183: 1449–52.

### **Praktisk utførelse av studien**

Ved deltakelse i studien vil du bli bedt om å ta av på overkroppen, undertøy kan beholdes på. Deretter vil det bli tatt ultralydbilde av begge skuldrene. Dette er smertefritt. Det er to undersøkere, som begge vil ta to sett med ultralydbilder.

Undersøkelsen vil ta rundt en time, og tidspunkt avtales med deg.

I prosjektet er det ikke økonomiske midler til kompensasjon eller dekning av utgifter, som for eksempel tapt arbeidsfortjeneste eller reiseutgifter.

### **Kapittel B - Personvern, økonomi og forsikring**

#### **Personvern**

Opplysninger som registreres om deg er alder, vekt, og smerter i skuldrene uttrykt gjennom en visuell analog skala (fra 1 til 10). I forkant av undersøkelsen vil også telefonnummer registreres (for kontakt og timeavtale). Telefonnummer slettes så snart undersøkelsen er gjennomført. Disse opplysningene er kun tilgjengelige for forskerne tilknyttet dette prosjektet, og opplysningene vil bli slettet så snart dataanalysene er gjennomført. Forskningsgruppen er databehandlingsansvarlig.

#### **Retten til innsyn og sletting av opplysninger om deg og sletting av prøver**

Hvis du sier ja til å delta i studien, har du rett til å få innsyn i hvilke opplysninger som er registrert om deg. Dersom du trekker deg fra studien blir alle innsamlede opplysninger slettet, med mindre opplysningene allerede er inngått i analyser eller brukt i vitenskapelige publikasjoner.

Studien er finansiert gjennom Høgskolen i Bergen sitt masterstudium i klinisk fysioterapi.

### **Forsikring**

Siden ultralyddiagnostikk er en velkjent metode, og er uten kjente bivirkninger eller smerter, ansees det ikke som nødvendig med egen forsikring i denne studien.

### **Informasjon om utfallet av studien**

Du har som deltaker rett til å få informasjon om resultatet av studien. Ta kontakt med [silje.ingebrihtsen@stud.hib.no](mailto:silje.ingebrihtsen@stud.hib.no) eller [ingvill.fjell.naterstad@stud.hib.no](mailto:ingvill.fjell.naterstad@stud.hib.no).

### **Samtykke til deltakelse i studien**

Jeg er villig til å delta i studien

-----  
(Signert av prosjektdeltaker, dato)

Stedfortredende samtykke når berettiget, enten i tillegg til personen selv eller istedenfor

-----  
(Signert av nærstående, dato)

Jeg bekrefter å ha gitt informasjon om studien

-----  
(Signert, rolle i studien, dato)

REK – Regionale komiteer for medisinsk og helsefaglig forskningsetikk

## **Forskningsprosjekt**

### **Reliabilitetsundersøkelse av diagnostisk ultralyd på skuldersonene**

Vitenskapelig tittel: Intra- og interobservatør reliabilitet ved tverrsnittsmål av Supraspinatus senen og Biceps longus senen ved diagnostisk ultralyd.

Prosjektbeskrivelse: REK Vest anser Høgskolen i Bergen som forskningsansvarlig for prosjektet. Formålet med studien er å undersøke hvorvidt to fysioterapeuter med liten erfaring og kun grunnleggende opplæring innen diagnostisk ultralyd, kan oppnå god reliabilitet i observasjonen av senetykkelse i skulder. Ti pasienter med smerter i skulder vil bli rekruttert fra primærhelsetjenesten i Bergen. I tillegg vil man inkludere ti friske deltakere uten smerter i skulderen.

*(Redigert av REK)*

Ref. nr.: 2011/1251    Prosjektstart dato: 15.08.2011    Prosjektslutt dato:  
31.05.2012

Behandlingsstatus: Godkjent

Forskningsstatus: Avsluttet

Prosjektleder: [jon.joensen@hib.no](mailto:jon.joensen@hib.no)

Forskningsansvarlig(e): [Høgskolen i Bergen](#)

Initiativtaker: Bidragsforskning    Finansieringskilder:

Prosjektet gjennomføres ved lån av Høgskolen i Bergen sitt utstyr, og uten ekstern finansiering.

Utvalg: Pasienter/klienter

Forskningsmetode:: Statistiske (kvantitative) analysemetoder

Antall forskningsdeltakere (Norge): 20

Utdanningsprosjekt/doktorgradsprosjekt: Studium: Master innen klinisk fysioterapi

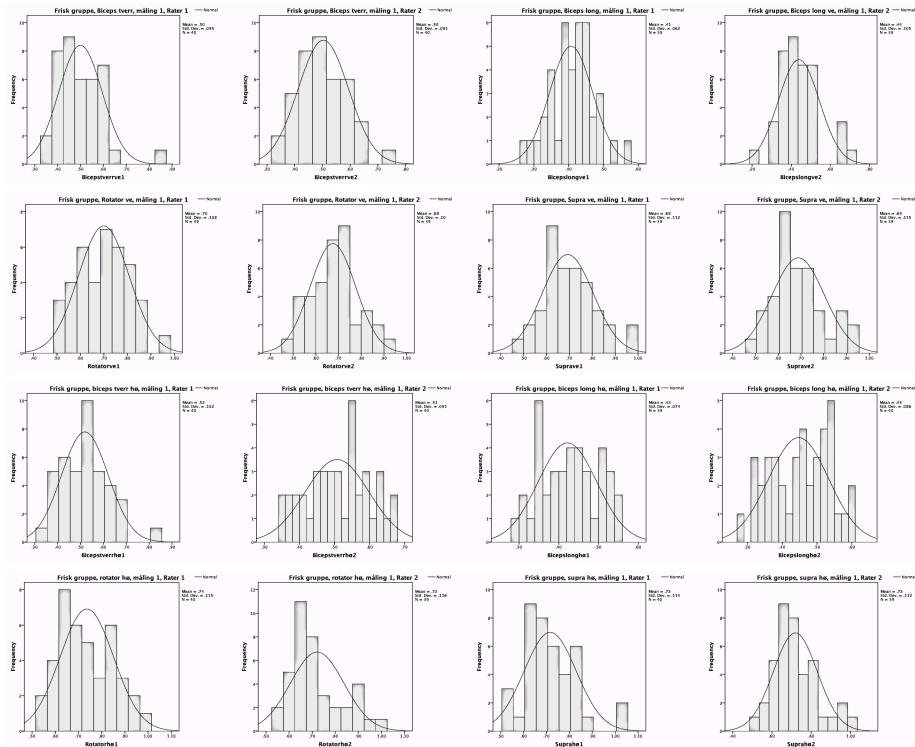
Nivå: Master

**Behandlet i REK Dato 18.08.2011** REK vest

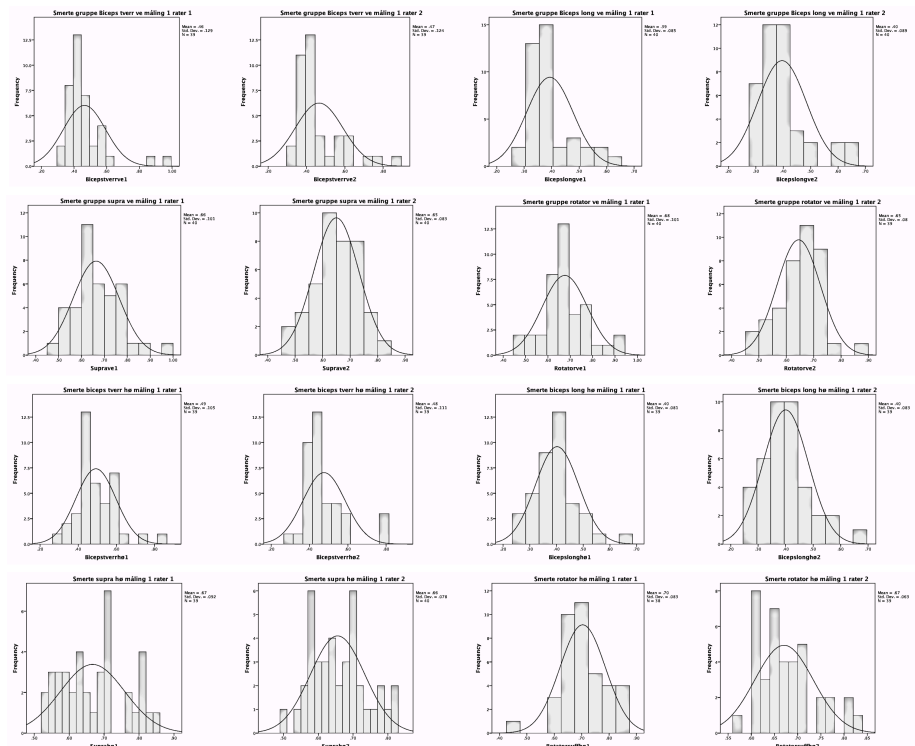


## Vedlegg 3

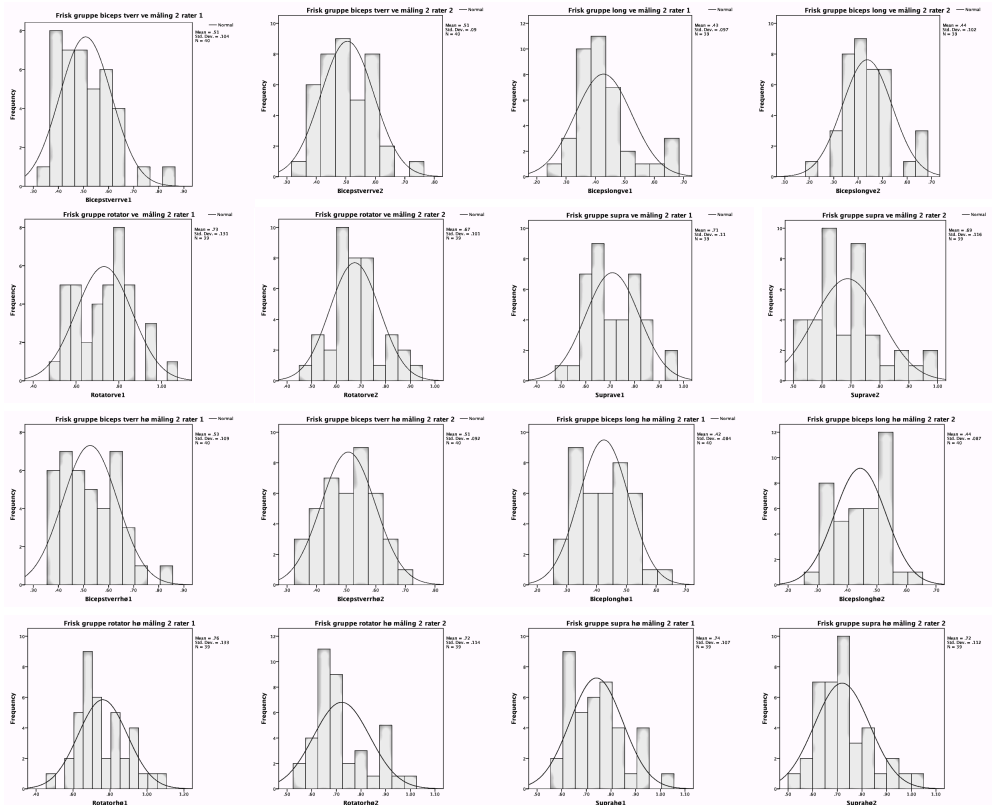
Normalfordelingskurver, frisk gruppe måling 1.



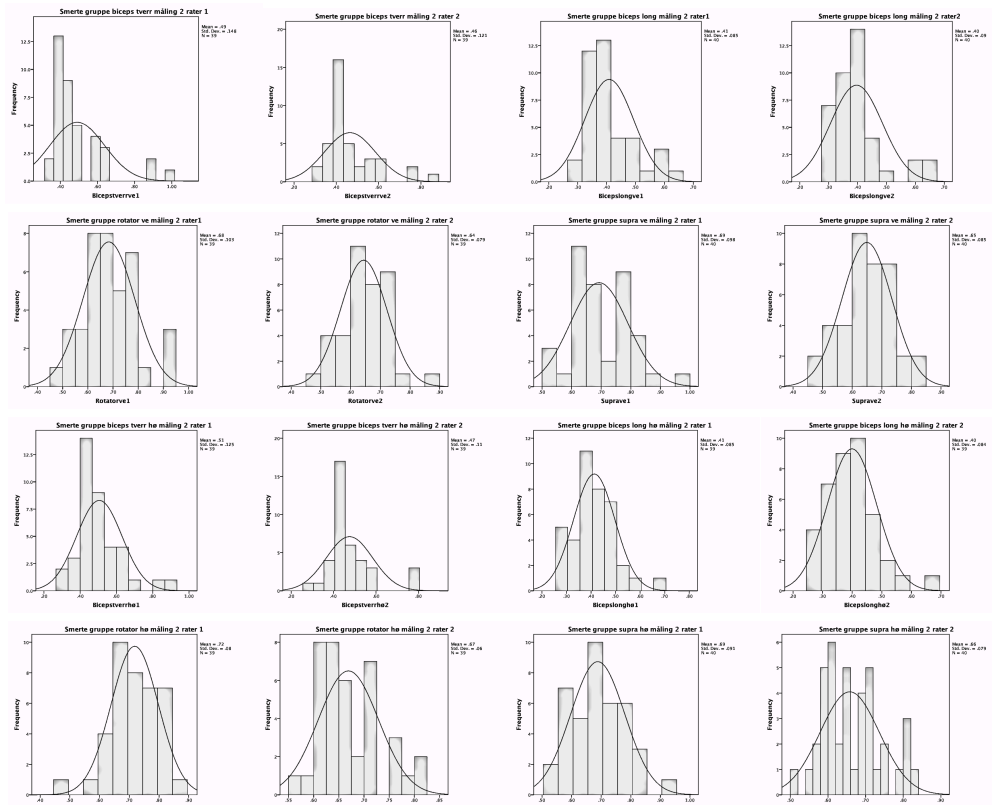
Normalfordelingskurve, smerte gruppe måling 1.



Normalfordelingskurver, frisk gruppe måling 2

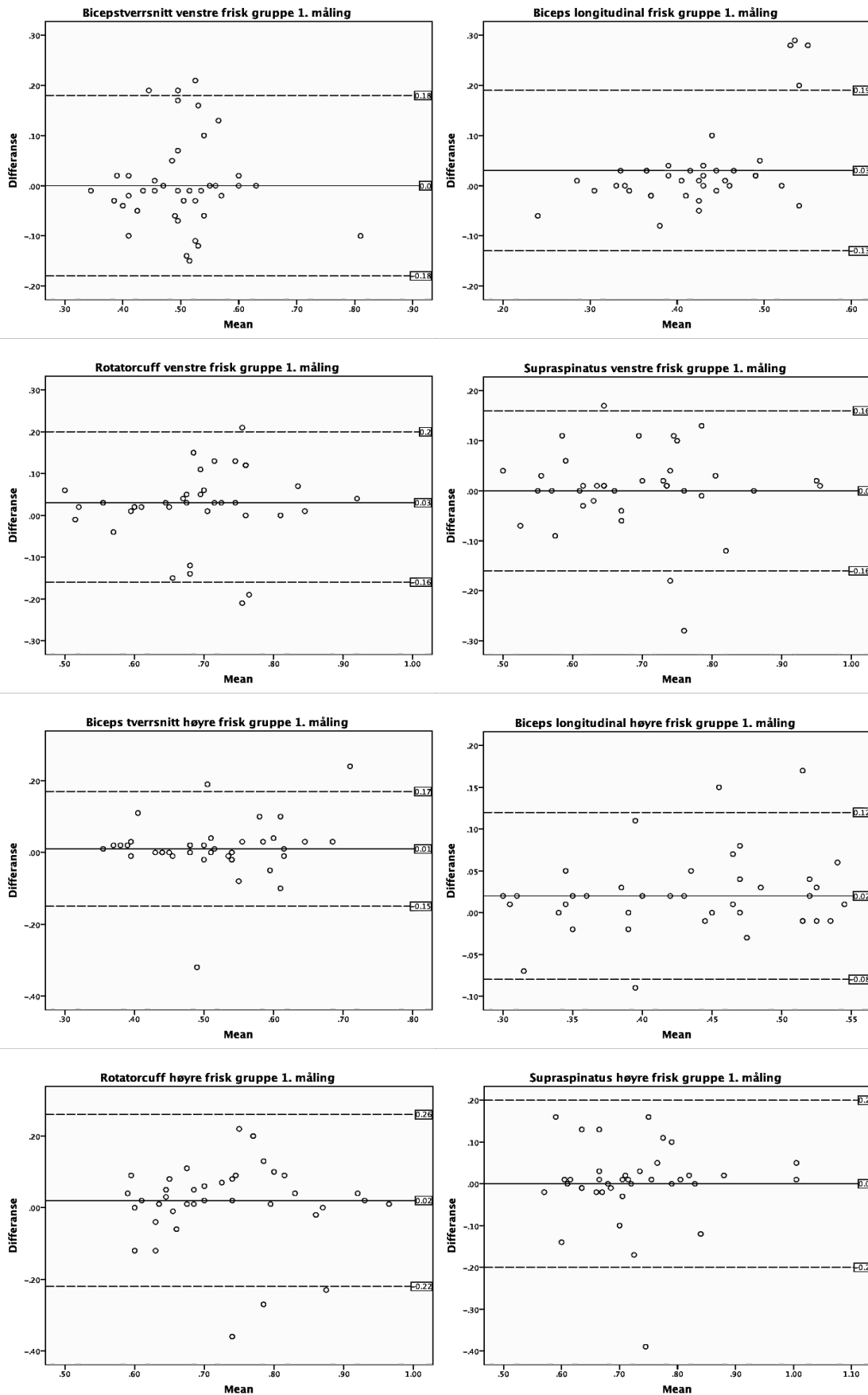


Normalfordelingskurve smerte gruppe måling 2

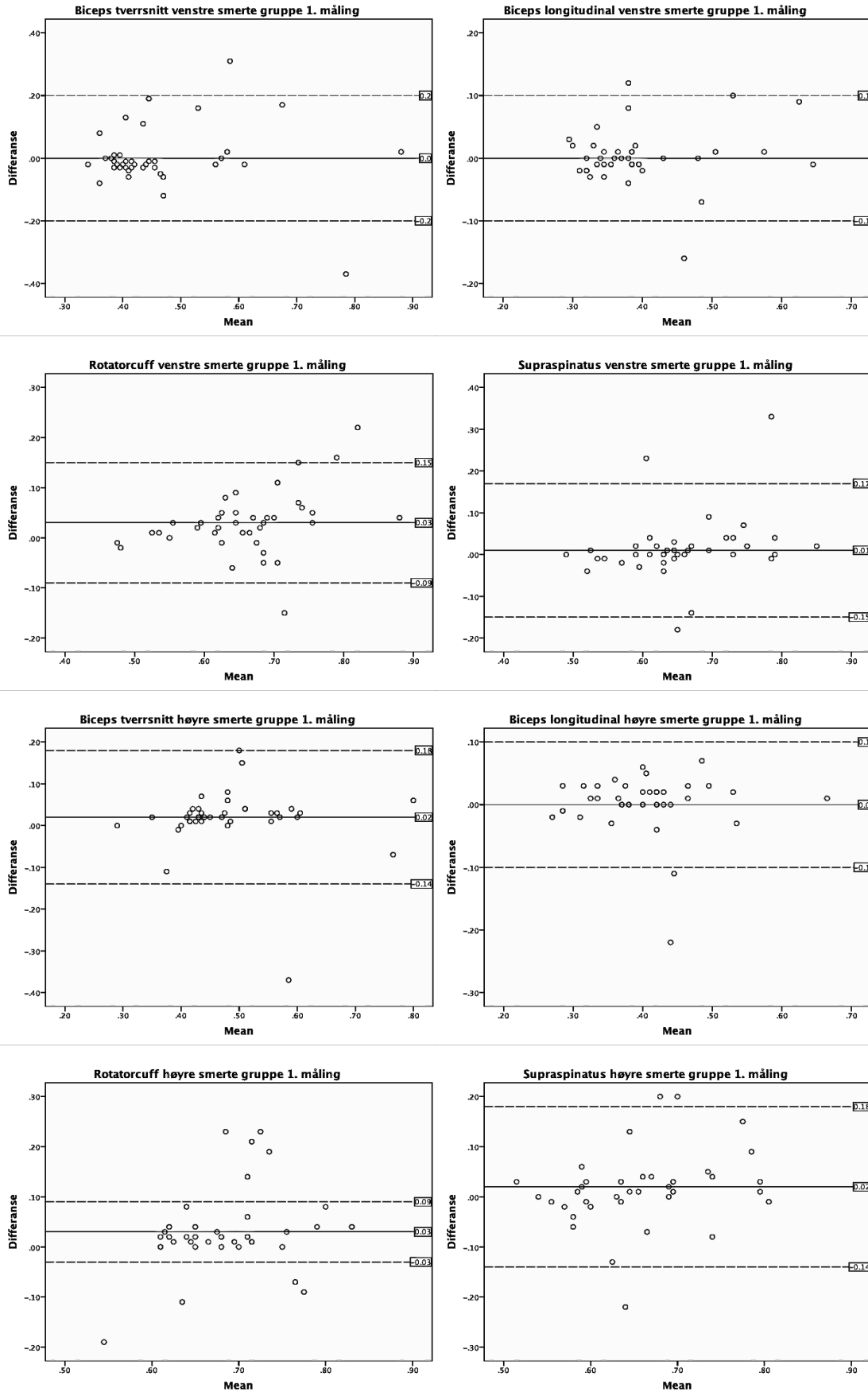


## Vedlegg 4: Interraterreliabilitet

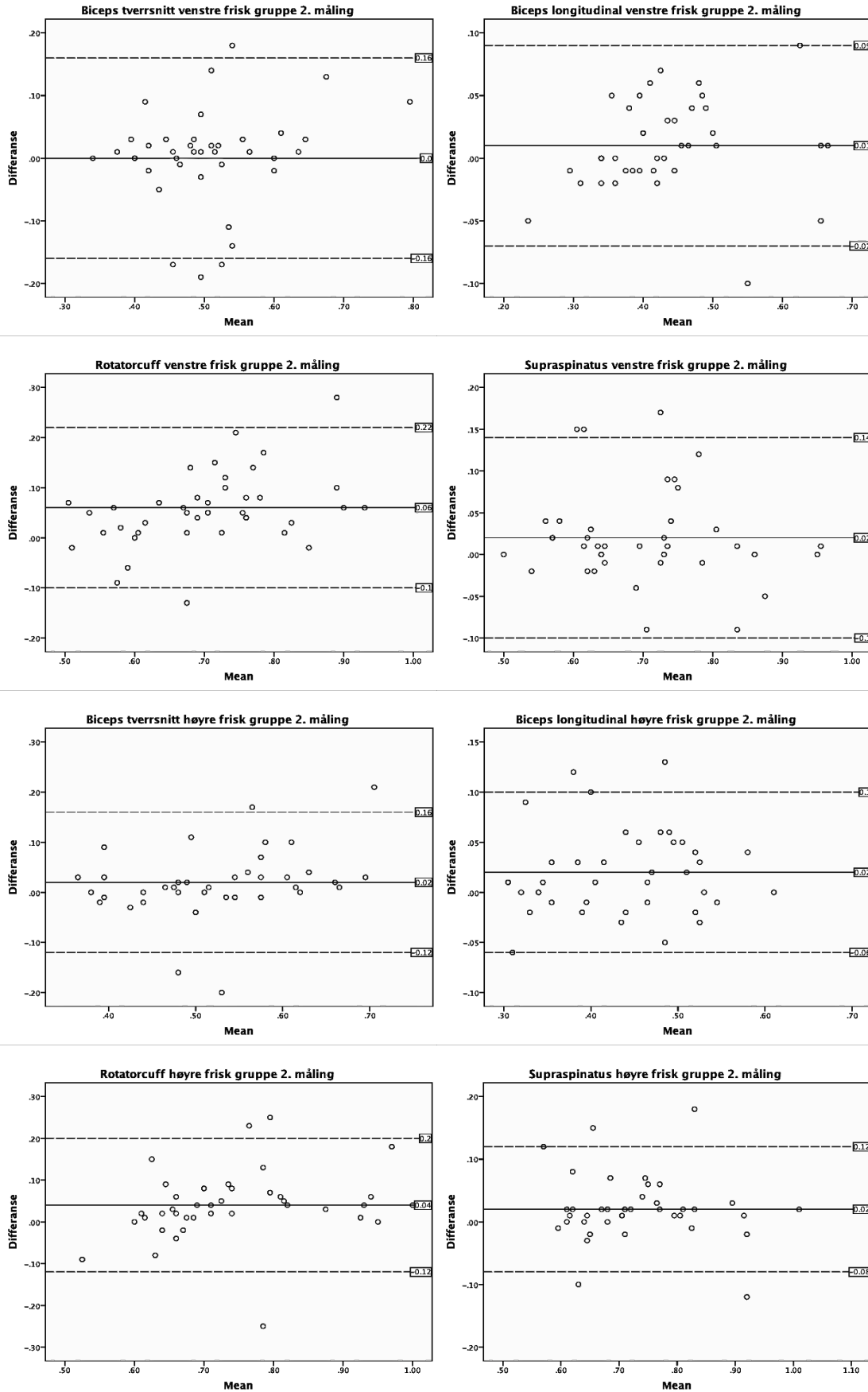
Bland-Altman plot for absolutt reliabilitet, frisk gruppe, 1. måling



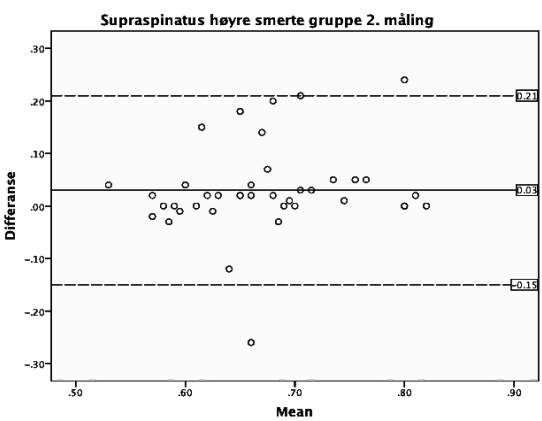
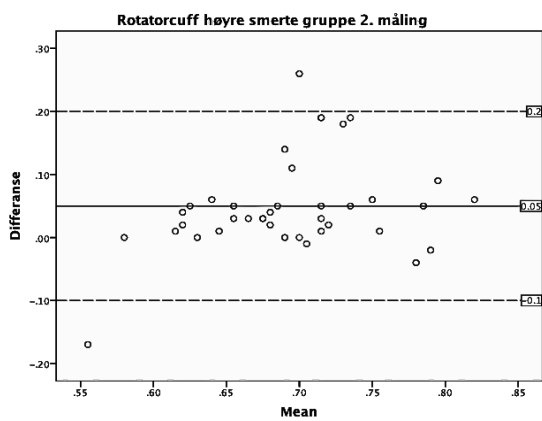
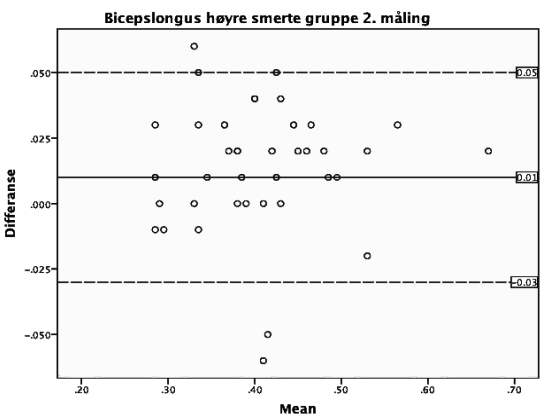
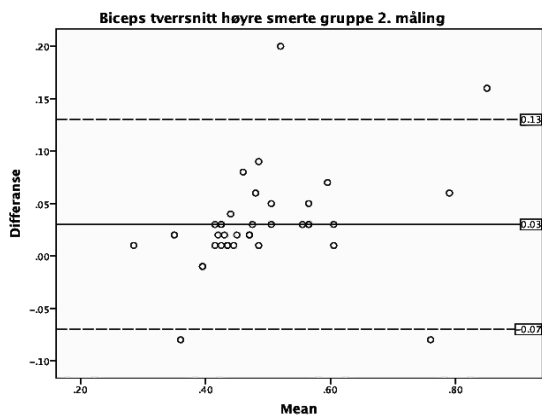
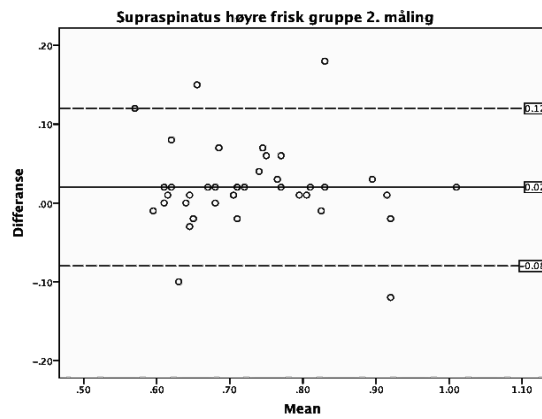
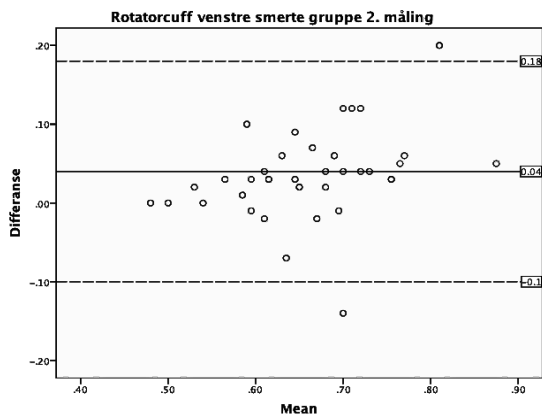
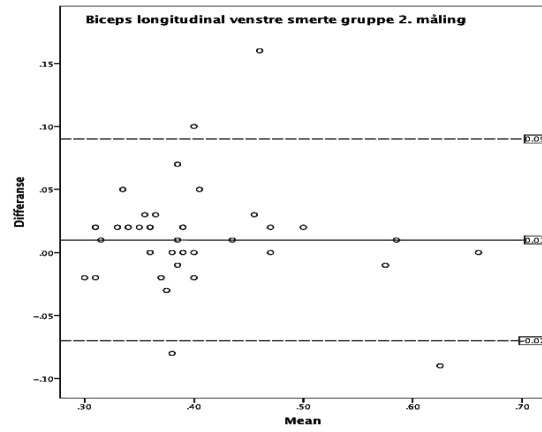
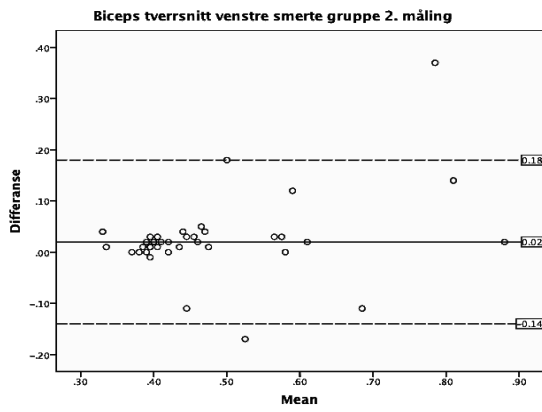
Bland Altman plot for absolutt reliabilitet, smerte gruppe, 1. måling.



Bland-Altman plot for absolutt reliabilitet, frisk gruppe 2. måling

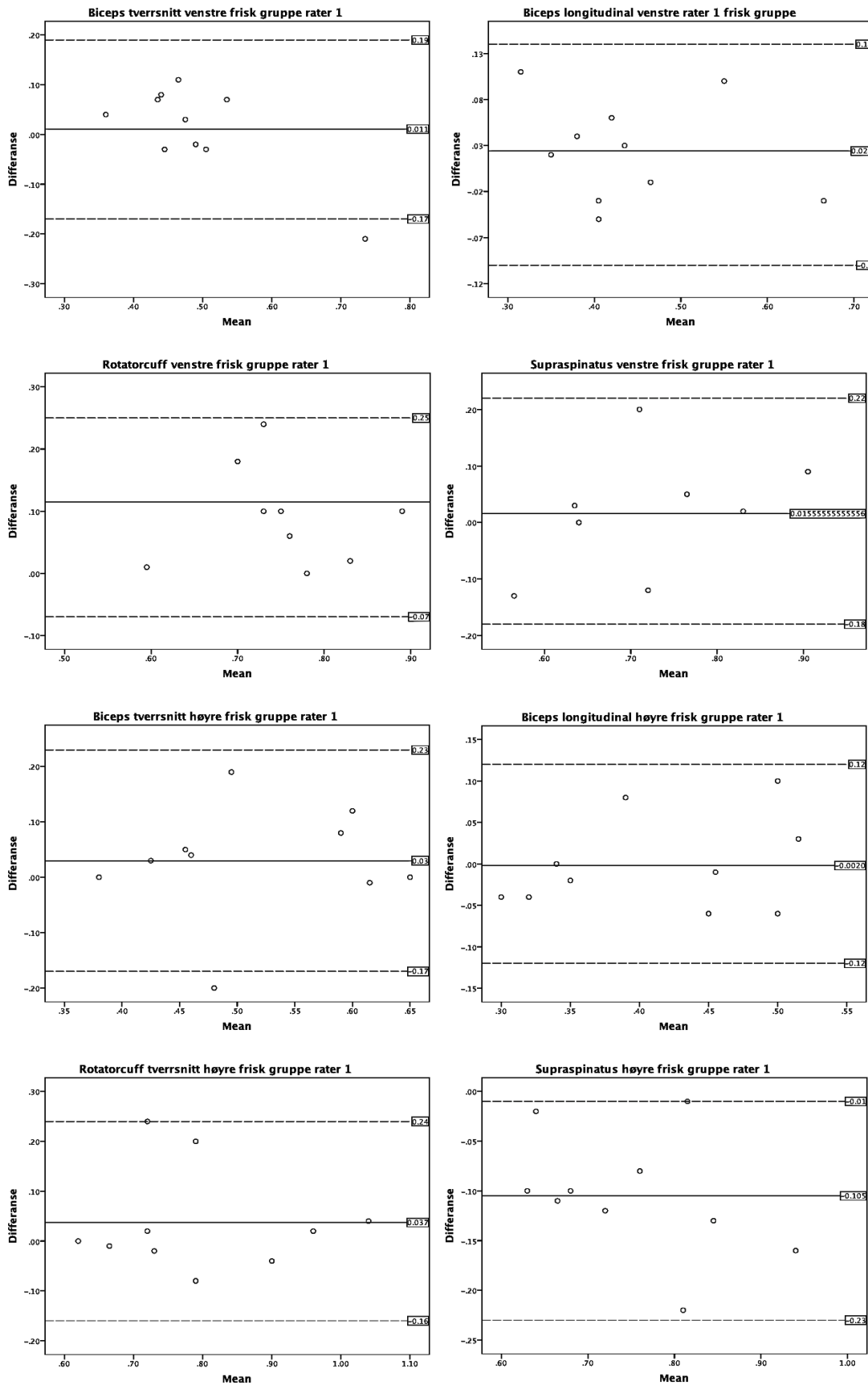


Bland-Altman plot for absolutt reliabilitet, smerte gruppe 2. måling

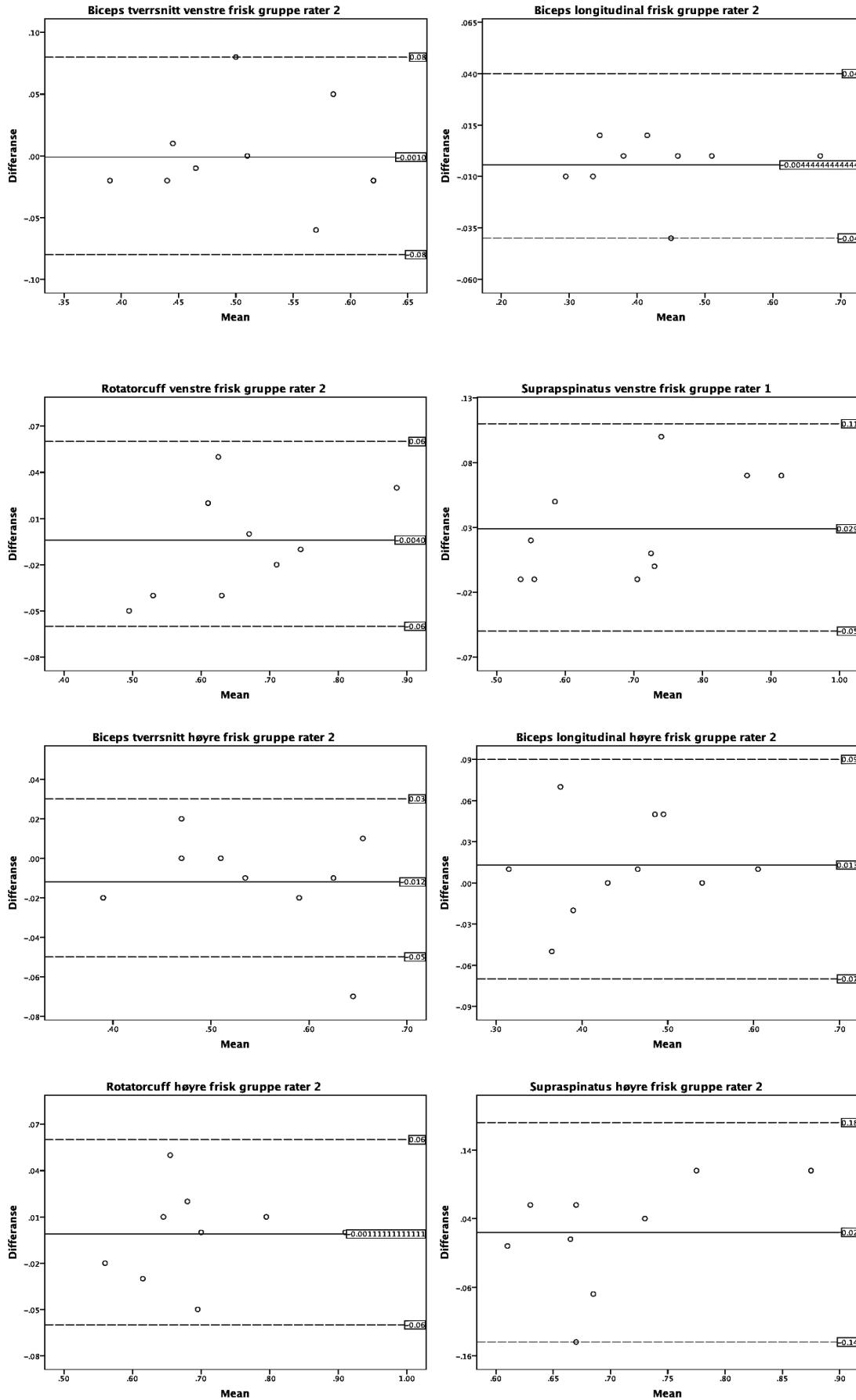


Vedlegg 5: Test retest.

Test-retest frisk gruppe rater 1

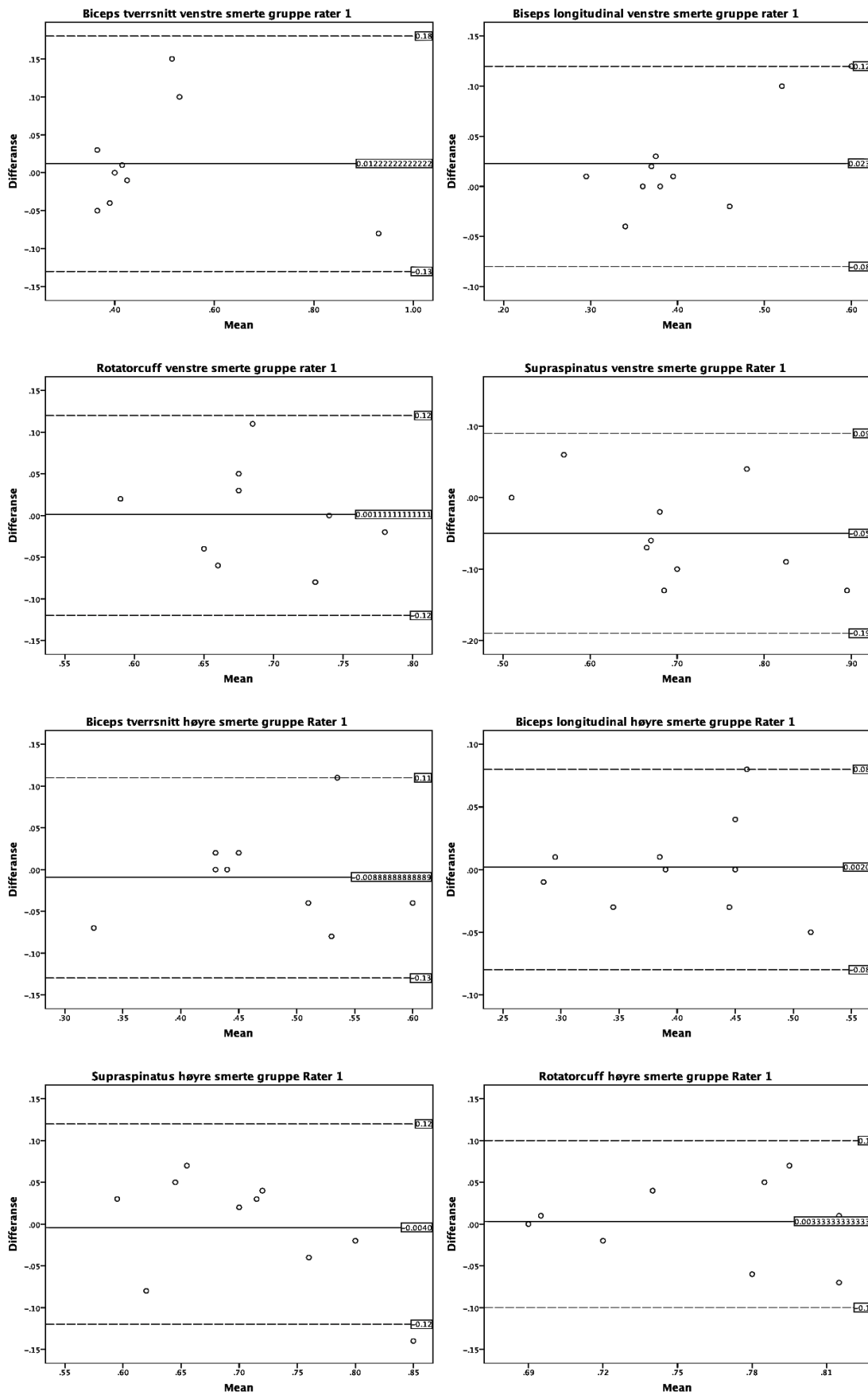


Test-retest frisk gruppe rater 2





# Test-retest smerte gruppe Rater 1



# Test-retest smerte gruppe Rater 2

