

# Test- retest reliabilitet av fire impingementtester for skulder



Manuellterapi teori: MANT395

Masterprogram i helsefag/ Klinisk masterstudium i manuellterapi  
for fysioterapeuter

Institutt for samfunnsmedisinske fag

Kandidatnummer: 209027

Antall ord: 9925

(uten henvisninger og tabeller)

Universitetet i Bergen

Høsten 2012

## Innholdsfortegnelse

<b>SAMMENDRAG</b>	<b>3</b>
<b>ABSTRACT</b>	<b>4</b>
<b>1.0 INTRODUKSJON</b>	<b>5</b>
<b>1.1 BAKGRUNN</b>	<b>5</b>
<b>1.2 TEORI</b>	<b>6</b>
1.2.1 SUBACROMIALT IMPINGEMENT SYNDROM	6
1.2.2 DIAGNOSTISK ULTRALYD	8
1.2.3 DIAGNOSTISKE TESTER VED SKULDER IMPINGEMENT	9
1.2.4 RELIABILITET:	13
1.2.5 RELEVANTE STUDIER	14
<b>2.0 HENSIKT OG PROBLEMSTILLING</b>	<b>16</b>
<b>2.1. HENSIKT</b>	<b>16</b>
<b>2.2. PROBLEMSTILLING</b>	<b>16</b>
<b>3.0 METODE</b>	<b>17</b>
3.1 VALG AV FORSKNINGSDESIGN	17
<b>3.2 UTVALG</b>	<b>17</b>
<b>3.3 DATAINNSAMLING</b>	<b>20</b>
3.3.1 BESKRIVELSE AV TESTENE:	20
<b>3.4 ANALYSE</b>	<b>22</b>
<b>3.5 ETISKE HENSYN</b>	<b>23</b>
<b>4.0 RESULTATER</b>	<b>25</b>
<b>5.0 DISKUSJON</b>	<b>30</b>
<b>5.1 STUDIEN OPP MOT LIGNENDE STUDIER</b>	<b>30</b>
<b>5.2 DISKUSJON AV TESTENE OG UTFØRELSE</b>	<b>32</b>
<b>5.3 VALIDITET</b>	<b>33</b>
5.3.1 INTERN VALIDITET:	33
5.3.2 EKSTERN VALIDITET:	35
<b>5.4 STYRKER OG SVAKHETER VED STUDIEN</b>	<b>37</b>
<b>5.5 VIDERE FORSKNING</b>	<b>38</b>
<b>6.0 KONKLUSJON</b>	<b>39</b>
<b>REFERANSELISTE</b>	<b>40</b>
<b>VEDLEGG 1:</b>	<b>46</b>
<b>VEDLEGG 2:</b>	<b>48</b>
<b>VEDLEGG 3:</b>	<b>50</b>

## **Sammendrag**

### **Test-retest reliabilitet av fire impingementtester for skulder.**

Klinisk masterstudium i manuellterapi for fysioterapeuter, Institutt for samfunnsmedisinsk fag, Universitetet i Bergen, 2012

#### **Bakgrunn**

Mellom 30-50% av befolkningen opplever skuldersmerter hvert år. Subacromialt impingement syndrom (SIS) er den vanligste skulderlidelsen, og representerer 44-65% av registrerte skulderplager i klinisk praksis. Det er betydningsfullt at de diagnostiske testene som brukes er reliable så vel som valide. Målet med denne studien var å undersøke test-retest reliabiliteten til fire kjente impingementtester fra klinisk hverdag. Studien inngår i en større studie hvor også undersøkelse av intertesterreliabilitet og testenes diagnostiske validitet inngår.

#### **Problemstilling**

Hvor god er test-retest reliabiliteten til fire impingementtester for skulder; Hawkins-Kennedy test, Neers tegn, Jobe supraspinatustest og Painful arc?

#### **Metode**

I denne studien ble grad av samsvar i skåringer mellom to terapeuter på fire impingementtester undersøkt ved Kappastatistikk og prosentvis enighet. For å sikre variasjon i skårer besto utvalget av pasienter med verifisert positiv impingement, pasienter med annen skulderpatologi og av skulderfriske personer. Ultralyd ble brukt som diagnostisk verktøy for å verifisere om pasientene hadde impingement eller annen skulderpatologi. Totalt 34 deltakere ble inkludert.

#### **Resultater**

Jobe supraspinatustest og Neers test fikk best resultat med høye kappaverdier på 0,94 og prosentvis enighet på 97%. Hawkins test fikk en kappaverdi på 0,82 og 92% enighet, mens Painful arc oppnådde en kappaverdi på 0,64 og enighet på 82,4 %.

#### **Konklusjon**

Test-retest reliabiliteten av testene i denne studien var alle høyst tilfredsstillende. Neers test, Jobe supraspinatus test og Hawkins-Kennedy test oppnådde alle kappaverdier med nesten perfekt samsvar, mens Painful arc oppnådde bra samsvar fra test til retest.

#### **Nøkkelord**

Test-retest reliabilitet, subacromialt impingement syndrom, kliniske tester.

## **ABSTRACT**

### **Test-retest reliability of four shoulder impingement tests.**

Clinical program in Manual therapy for physical therapists, Institute for Social Medicine, University of Bergen, 2012

#### **Background**

Between 30-50% of the population experience shoulder pain every year. Subacromial impingement syndrome (SIS) is the most common shoulder disorder, and represents 44-65% of shoulder problems encountered in clinical practice. It is important that the diagnostic tests are reliable as well as valid. The aim of this study was to investigate the test-retest reliability of four known impingement tests from clinical practice. The study is part of a larger study in which the investigation of inter tester reliability and diagnostic validity are included.

#### **Research question**

How good is the test-retest reliability of four shoulder impingement tests; Hawkins-Kennedy test, Neer sign, Jobe supraspinatus test and Painful arc?

#### **Methods**

In this study the degree of correlation between the scores from two therapists at four impingement tests were investigated by Kappa statistics and percentage agreement. To ensure variation in scores the samples consisted of patients with verified positive impingement, patients with other shoulder pathology and shoulder healthy subjects. Ultrasound was used as a diagnostic tool to verify whether patients had impingement or other shoulder pathology. A total of 34 participants were included.

#### **Result**

Jobe supraspinatus test and Neers sign got the best results with high kappa values of 0.94 and percentage agreement of 97%. Hawkins test had a kappa value of 0.82 and 92% agreement, while Painful arc achieved a kappa value of 0.64 and 82.4% agreement.

#### **Conclusion**

Test-retest reliability of the tests in this study all come out well. Neers sign, Jobe supraspinatus test and Hawkins-Kennedy test achieved kappa-values with almost perfect agreement, while Painful arc achieved good agreement from test to retest.

#### **Keywords**

Test-retest reliability, subacromial impingement syndrom, clinical tests.

## 1.0 INTRODUKSJON

### 1.1 Bakgrunn

Pasienter med skuldersmerter, og gjerne diffuse skuldersmerter er en diagnosegruppe og problematikk jeg ofte møter i min kliniske hverdag som fysioterapeut. Jeg synes temaet er svært interessant, av stor klinisk betydning, og et tema jeg gjerne vil fordype meg videre innen.

Et sted mellom 30-50 prosent av befolkningen opplever skuldersmerter hvert år (Lewis 2009). Vanligst er plager knyttet opp mot rotatorcuffen og avklemminger i det subacromiale rom i skulderen (Solberg 2002, Lewis 2009, Michener et al. 2009). Subacromial impingement syndrom (SIS) er den vanligste skulderlidelsen i befolkningen, og representerer 44-65% av alle registrerte skulderplager i kliniske praksis (Michener et al. 2009). Lignende tall ser en fra en undersøkelse av Östör et al. (2005) i England som fant tegn til impingement hos 74% av pasientene som oppsøkte allmennlege på grunn av skuldersmerter. En amerikansk studie fra 2005 knytter impingement også til en økonomisk og samfunnsmessig side. De fant at kostnadene til behandling av dysfunksjon i skulder beløp seg til 7 milliarder dollar i 2000 (Meislin et al. 2005). Det angis i studien ikke eksakt hvor mye av denne summen som direkte skyldes impingementproblematikk, men legger man til grunn tallene fra Michener et al. (2009) dreier det seg om 3-4 millioner dollar, hvilket er store kostnader og et klart samfunnsøkonomisk problem.

Det er skrevet og forsket mye innen feltet SIS i skulder. Flere effektstudier er gjort for å utforske den mest effektive behandlingen av ulike skulderdiagnoser. Det foreligger blant annet to systematiske oversikter fra Cochrane library (Greene et al. 2008, Buchbinder et al. 2006) som oppsummerer resultatene fra ulike studier. Det er derimot forsket mindre på de ulike testene som benyttes for å avdekke den aktuelle diagnose, og validering av disse (Hegedus et al. 2008). Skulderplager er en diagnostisk utfordring og den kliniske undersøkelsen utført av fysioterapeuten er av særdeles viktighet. Det er svært sentralt at de diagnostiske testene som brukes for å komme frem til en diagnose er reliable så vel som valide (Park et al. 2005). For å sikre best mulig behandling av pasienter med impingement er det ønskelig med pålitelig diagnostisering (Michener et al. 2009). Det foreligger dog ikke en enhetlig konsensus vedrørende diagnostiske kriterier av impingement (Awerbuch 2008). Dette vanskeliggjør diagnostiseringen, som ikke kan settes på bakgrunn av anamnese alene (ibid). Teknikker som artroskopi og MR (Malhi og Khan 2005, Nørregård et al 2002), eller ultralyd (Kromer et al 2010) blir ofte benyttet i diagnostiseringen. Malhi og Khan (2005) og Park et al

(2005) hevder at artroskopi er gullstandard. Malhi og Khan (2005) har sett grad av samsvar mellom artroskopi, MR og ultralyd. De fant da at det var 95% samsvar mellom funn på MR og ultralyd sammenlignet med artroskopi. Ved skulderpatologi egner ultralyd seg spesielt godt for skulderundersøkelser fordi man enkelt kan sammenlikne med den andre skulderen. Videre kan en ved ultralydundersøkelse bevege skulder/arm i forskjellige posisjoner for å analysere strukturene i forhold til hverandre (Juel et al. 2007). Det er ingen kontraindikasjon for ultralyd og undersøkelsen gir ikke smerte, hvilket begge er store styrker. Ved artroskopi vil derimot en del pasienter bli ekskludert grunnet kontraindikasjoner, og metoden er dyrere (Itoi et al. 2006, Holtby & Razmjou 2004, MacDonald et al. 2000).

I denne studien var vi tre masterstudenter ved Universitetet i Bergen som jobbet sammen og forsket på tre ulike problemstillinger vedrørende måleegenskaper til fire kjente tester for skulderimpingement som brukes i den kliniske hverdagen. Målet med min del av studien var å undersøke test-retest reliabiliteten. De andre studentene undersøkte henholdsvis intertester reliabiliteten og testenenes diagnostiske validitet.

## **1.2 Teori**

God kunnskap om skulderregionens anatomi og biomekanikk er overordnet for gjennomføring av en målrettet funksjonsundersøkelse. Grunnleggende anatomi anses som kjent for leserne av denne oppgaven. I korte trekk består skulderen i hovedsak av tre knokler, tre ekte ledd (glenohumeralledet, akromioklavikulærløddet og sternoklavikulærløddet) med ligamentforsterkninger og mer enn 20 muskler (Juel 2007).

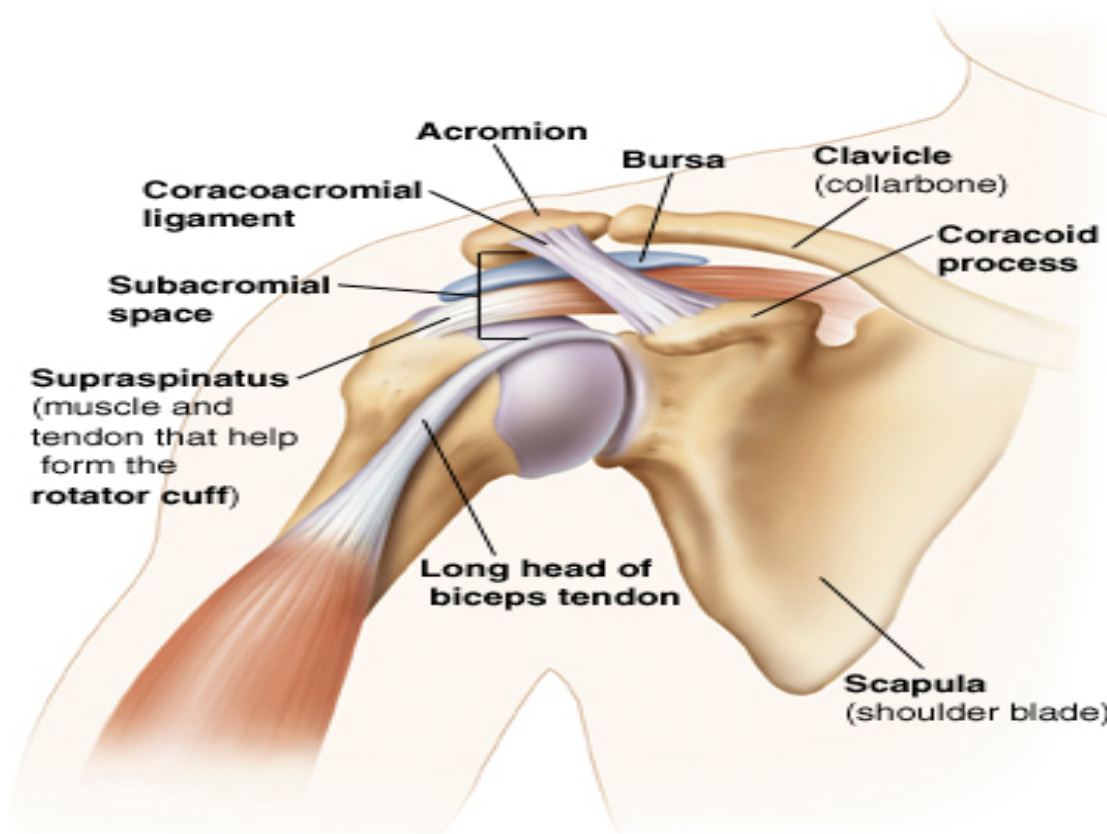
Glenohumeralledet er et kuleledd, men kontaktflaten mellom humerus og scapula er liten sammenlignet med den dypere hofteskålen i hofteløddet. Dette gir skulderen stor bevegelsesfrihet men relativt liten strukturell stabilitet (Dahl og Rinvik 2010, Michener et al. 2009). Det settes dermed store krav til ligamentstrukturene som forsterker stabiliteten, og spesielt til muskulaturen som både skal stabilisere og bevege leddet (Ibid). Små endringer og dysfunksjoner i rekrutteringen av musklene rundt skulder, les rotatorcuffen, vil ofte kunne medføre skulderplager (Juel 2007).

### **1.2.1 Subacromialt impingement syndrom**

Subacromialt impingement syndrom (SIS) er en betegnelse på smertetilstander i skulderen, der smerten sitter subakromialt dvs. under den avsluttende hvelvingen på skulderbladkammen (Hegedus et al. 2008). Tilstanden har sammenheng med anatomisk trange plassforhold og er oftest relatert til rotatormansjettens (rotatorcuffens) muskler, dvs. lange biceps-sene,

supraspinatus-senen, infraspinatus-senen eller til en slimpose, *bursa subacromiale* (Lewis 2009, Michener et al. 2009).

Ved subacromialt impingement skjer det en innklemming av strukturer i det subacromiale rom, rommet mellom undersiden av acromion og tuberculum majus på humerus, se figur 1 under (Solberg 2002, Hegedus et al. 2008). Strukturene som kommer i beknip er rotatorcuffen, særlig supraspinatus-senen og overliggende bursa (slimpose) (Solberg 2002, Lewis 2009). SIS kan skyldes opprykket humerushode, anatomiske forhold som størrelsen på fossa glenoidalis, formen på acromion, instabilitet i glenohumeralledet eller funksjonell instabilitet (Solberg 2002)



Figur 1: Subacromial space: Rommet mellom humerus og acromion der strukturer kan komme i beknip, og gi smertetilstander (Greys Anatomy 2009).

Den hyppigste årsaken til subacromialt impingement er overbelastning av supraspinatus-senen og påfølgende små skader i senen. Skadene gir inflammasjon og degenerasjon av vevet, noe som virker negativt inn på plassforholdene i det subacromiale rommet og medfører

inneklemming (*impingement*) av bløtdelstrukturene under *acromion* (blant annet seneplatene til de nevnte muskler, eller slimposen) (Lewis 2009, Michener et al. 2009).

Brodal (2007) definerer smerte som en ubehagelig sensorisk og emosjonell opplevelse som opptrer i sammenheng med vevsskade eller truende vevsskade. Symptomene ved SIS starter ofte som sviende smerter i skulderen og øvre del av overarm (særlig på yttersiden) og forverres ved ensformig repetitiv belastning, spesielt ved arbeid med armen over skulderens plan. Plagene er initialt anfallsvis for etterhvert å bli mer konstante (Solberg 2002).

Biomekanisk er det interessant å se på hvordan humerus opptrer i forhold til scapula ved armbevegelse. Forholdet mellom bevegelsen i glenohumeralledet og scapula beskrives som den scapulohumerale rytmen (Juel 2007). Nyere forskning viser at det er store individuelle forskjeller i denne rytmen, men litteraturen enes om at bevegelsen av scapula er størst ved abduksjon av skulderen mellom 60 og 130 grader (Juel, 2007, Lewis 2009, Michener et al. 2009). Dette tilsvarer smertebuen (*Painful arc*) ved subacromialt smertesyndrom og tyder på at plassforholdene subacromialt er dårligst i denne delen av bevegelsesbanen (*ibid*). Våre 4 tester i denne studien ble alle utført innenfor dette bevegelsesintervallet. Beaulieu et al. (2006) brukte i en omfattende studie tredimensjonal åpen MRI for å se hvordan plassforholdene subacromialt endrer seg ved Neer og Hawkins-Kennedy impingement test. De fant at begge testene ga betydelig reduksjon av plassforholdene mellom supraspinatus og acromion. De fant også at på 90 % av de som ble testet var det subacromial kontakt mellom supraspinatus og infraspinatus. De konkluderte med at små irritasjoner i aktuelle sener eller bursaer lett vil gi smerte (*ibid*). Smerte gir også ofte en reflektorisk påvirkning av bevegelser i skulderen. Dette kan gi opphav til en dysfunksjonell bevegelse og kan forverre eller vedlikeholde smerten (Juel, 2007).

### **1.2.2 Diagnostisk ultralyd**

Skulderundersøkelse ved hjelp av ultralyd har blitt mer og mer brukt de siste 30 år, og det har i perioden vært stor teknologisk og faglig utvikling (Kelly et al. 2010). Metoden bidrar til en mer presis diagnostikk av ulike bløtvevstrukturer ( Juel, 2007).

Med ultralyd skjer billedannelsen ved hjelp av at lydbølger i området 1 til 15 Mega-Hertz som sendes og mottas av høyfrekvente lydhodet, doppler. Signalene gjøres deretter fortløpende om til bilder, kurver eller lydsignaler (Kelly et al. 2010). Strukturer som hyppig bidrar til subacromiale smerter kan kartlegges ved ultralyd, deriblant tendinopati, kalknedslag, bursareaksjoner og cuffrupturer (Juel, 2007, Kelly et al. 2010).



Det er mange fordeler ved å bruke Ultralyd ved undersøkelse av skulder fremfor annen bildediagnostikk. Ultralydundersøkelsen er kostnadseffektiv, plassbesparende, utføres relativt raskt og regnes som gullstandard for mange tilstander (Juel, 2007, Kelly et al. 2010, Kromer et al. 2010). Undersøkelsen er non-invasiv, uten strålebelastning og gir ingen bivikninger eller kontraindikasjoner. Et annet stort pluss med ultralyd fremfor annen billediagnostikk er at ultralyd kan utføres med aktiv bevegelse i det ønskede leddet (ibid). I mange land anbefales ultralyd som førstevalg bildeundersøkelse av senevev (Juel, 2007, Kelly et al. 2010). Det store minuset med ultralyd er at kvaliteten på billedundersøkelsen sjelden blir god hvis ikke undersøkeren er god. Ultralyddiagnostikk skiller seg fra de fleste andre former for bildedannende diagnostiske undersøkelser ved at kvaliteten er sterkt avhengig av vedkommende som innhenter dataene, undersøkelsen er sterkt operatørvhengig (ibid).

### **1.2.3 Diagnostiske tester ved skulder impingement**

Før en kan vurdere reliabiliteten av impingementtester må en få et overblikk over hvilke tester som faktisk benyttes i diagnostikk av tilstanden, dvs. deres validitet. Søk i PubMed 09.10.11 gav følgende resultat:

Søk 1: Shoulder impingement syndrome ga 1389 treff

Søk 2: Søket kan innsnevres ved å sette diagnose som MESH-term i søket. Ved søkestrengen "Shoulder Impingement Syndrome/diagnosis"[MAJR] ga 280 treff.

Søk 3: Søket ble spisset enda mer ved søkestrengen "Shoulder Impingement Syndrome/diagnosis"[MAJR] AND ("test" OR "tests" OR "physical examination"), dette søket ga 87 treff.

Av totalt 87 treff var det 5 artikler som ble vurdert som særlig relevant, og dette var studiene til Kelly et al. (2010), Michener et al. (2009), Calis et al. (2008), Silva et al. (2008) og Park et al. (2005). Felles for artiklene var at de eksplisitt tok for seg den diagnostiske validiteten ved kliniske impingement-tester. I tillegg ble det funnet studier og 1 systematiske oversikt som undersøker diagnostiske kriterier ved tester som brukes for å kartlegge patologi i strukturer som kan være affisert ved impingement. Disse strukturene, som for eksempel bursa eller rotatorcuffen, kan imidlertid være skadet uten at det foreligger impingement. Dette gjelder studiene til Beaudreuil et al. (2009), MacDonald et al. (2000), Hughes et al. (2008) og Hegedus et al. (2008). De fem mest relevante studiene er satt inn i en tabell 1, som gir oversikt over diagnostiske egenskaper som ble funnet for ulike tester.

**Tabell 1:** Skjematisert oversikt over de mest relevante studiene vi legger til grunn i vår oppgave. Tabellen tar for seg de ulike testene brukt i studiene og skåringene på bakgrunn av: *S* = Sensitivitet. *SP* = Spesifisitet. *PPV* = Positiv prediktiv verdi. *NPV* = Negativ prediktiv verdi. *LR+* = Positiv likelihood-ratio. *LR-* = Negativ likelihood-ratio. \* = regnet ut på bakgrunn av oppgitt sensitivitet (*S*) og spesifisitet (*SP*)

Studie	Test	S	SP	PPV	NP V	Nøy aktig	LR+	LR-
Silva et al 2008  (n= 30)	Neer	68,4	30	65	33,3	55,2	0,98*	1,05*
	Hawkins	73,7	40	70	44,4	62,1	1,23*	0,66*
	Yocum	79	40	71,4	50	65,5	1,32*	0,53*
	Jobe	73,7	30	66,7	37,5	58,6	1,05*	0,88*
	Patte	57,9	60	73,3	42,9	58,6	1,45*	0,70*
	Gerber	68,4	50	72,2	45,5	62,1	1,37*	0,63*
	Passiv abduksjon	73,7	10	60,9	16,7	51,7	0,82*	2,63*
	Abduksjon mot motstand	57,9	20	57,9	20	44,8	0,72*	2,1*
Calis et al. 2000  (n= 125)	Hawkins	92,1	25	75,2	56,2	72,8	1,23*	0,32*
	Neer	88,7	30,5	75,9	52,3	72	1,28*	0,37*
	Horisontal adduksjon	82	27,7	73,7	38,4	66,4	1,13*	0,65*
	Speeds test	68,5	55,5	79,2	41,6	64,8	1,54*	0,57*
	Yergasons test	37	86,1	86,8	35,6	51,2	2,66*	0,73*
	Painful arc	32,5	80,5	80,5	32,5	46,4	1,67*	0,84*
	Drop arm	7,8	97,2	87,5	29,9	33,6	2,78*	0,96*
Kelly et al. 2010  (n= 59)	Neer	62,1	0			54,5	0,62*	∞*
	Hawkins	74,1	50			71	1,48*	0,52*
	Painful arc	29,6	50			31	0,60*	1,41*
	Svakhet i abduksjon	37,9	50			39,4	0,76*	1,24*
	Smerte ved abduksjon	55,2	75			57,6	2,21*	0,59*
	Svakhet i utadrot.	55,2	25			51,5	0,74*	1,79*
	Smerte ved utadrot.	34,5	100			42,4	∞*	0,66*
	Svakhet ved empty can	51,9	66,7			53,3	1,56*	0,72*
	Smerte ved empty can	51,9	33,3			50	0,78*	1,45*
	Svakhet ved full can	44,8	75			48,5	1,79*	0,74*
	Smerte ved full can	34,5	25			33,3	0,46*	2,62*
<b>Park et al. 2005</b>	<b>Neer</b>	68	68,7	80,4	53,2		2,17*	0,46*
	<b>Hawkins</b>	71,5	66,3	79,7	55,7		2,12*	0,43*
	<b>Painful arc</b>	73,5	81,1	88,2	61,5		3,89*	0,33*
	<b>Empty can (jobe)</b>	44,1	89,5	88,4	46,8		4,2*	0,63*
	<b>Speed test</b>	38,3	83,3	80,5	42,9		2,30*	0,74*
<b>(n= 913)</b>	<b>Cross-body adduksjonstest</b>	22,5	82	69,3	36,9		1,25*	0,95*
	<b>Drop arm-test</b>	26,9	88,4	81	39,7		2,32*	0,83*
	<b>Infraspinatus muskeltest</b>	41,6	90,1	90,6	45,8		4,2*	0,65*
Michener et al 2009  (n = 55)	Hawkins	63	62			62	1,63	0,61
	Neer	81	54			68	1,76	0,35
	Painful arc	75	67			71	2,25	0,38
	Empty can (jobe)	50	87			69	3,9	0,57
	Utadrotasjon mot motstand	56	87			72	4,39	0,5

Begrepsforklaringer:

Testens **sensitivitet** (S) og **spesifisitet** (SP) beskriver testens validitet (Benestad & Laake 2004).

**Diagnostisk sensitivitet** kan defineres som testens evne til å identifisere individer med tilstanden (positive) i forhold til antallet med tilstanden identifisert av referansetesten og uttrykkes ved Positiv/Positiv+Falsk Negativ (Benestad & Laake, 2004).

**Diagnostisk spesifisitet** defineres som testens evne til å ekskludere individer uten tilstanden (negativ), det vil si Negativ/Falsk positiv+Negativ. **Sensitivitet** angir dermed testens evne til å avsløre om en person er syk, mens **spesifisitet** angir testens evne til å avsløre om en person er frisk (Hattam & Smeatham 2010). Sensitivitet og spesifisitet regnes ut som ratio, men angis oftest som prosentvise verdier med 95 % konfidensintervall (KI) (Hattam & Smeatham 2010).

**Positiv prediksjonsverdi**, PPV: I screening/diagnostiske tester er dette et mål på nytten av en screeningtest/diagnostisk test. Det er andelen av de som har et positivt testresultat som har sykdommen, og kan tolkes som sannsynligheten for at et positivt testresultat er korrekt.

Verdien beregnes slik:  $PPV = \frac{\text{antall med positivt testresultat som har sykdommen}}{\text{antall med positivt testresultat}}$  (Kunnskapssenteret 2010).

**Negativ prediksjonsverdi**, NPV : I screening/diagnostiske tester: et mål på nytten av en screeningtest/diagnostisk test. Angir andelen av dem som har et negativt testresultat som ikke har sykdommen. Kan tolkes som sannsynligheten for at et negativt testresultat er korrekt.

Beregnes slik:  $NPV = \frac{\text{antall med negativt testresultat uten sykdommen}}{\text{antall med negativt testresultat}}$  (Kunnskapssenteret 2010).

**LR+ og LR -** : Sannsynlighetsandeler (LR). Dette er kanskje den beste måten å vurdere den kliniske nytteverdien av en test. LR gir et forholdstall mellom sensitivitet og spesifisitet. Jo lengre bort fra 1 LR+ befinner seg og jo nærmere 0 LR- ligger, desto mer informativ er testen (Hattam og Smeatham 2010).

Tabell 4: Formlen for sensitivitet (S), spesifisitet (SP), Positiv prediksjonsverdi (PPV), Negativ prediksjonsverdi (NPV), Sannsynlighetsandeler (LR+ og LR-). Konfidensintervall (KI)

<b>Sensitivitet</b>	<b>Positiv/(Positiv+Falsk Negativ)</b>
<b>Spesifisitet</b>	<b>Negativ/ (Falsk Positiv+Negativ)</b>
<b>PPV</b>	<b>Positiv/(Positiv+Falsk Positiv)</b>
<b>NPV</b>	<b>Negativ/(Falsk Positiv+Negativ)</b>
<b>LR+</b>	<b>Sensitivitet/ (1-spesifitet)</b>
<b>LR-</b>	<b>(1- sensitivitet)/spesifisitet</b>
<b>KI</b>	$\frac{a}{n} \pm 1.96 \sqrt{\frac{\frac{a}{n}(1 - \frac{a}{n})}{n}}$
<b>Nøyaktighet</b>	<b>(P+N)/(P+FP+FN+N)</b>

P=Positiv, N=Negativ, FP=Falsk positiv, FN=Falsk negativ, a=antall positive/negative, n=totalt antall

Oppsummering av resultater fra studiene.

I studien til Silva et al. (2008) ble det konkludert med at de fleste tester utført for å avdekke SIS er relativt gode når det gjelder sensitivitet, problemet er bare at de generelt har lav spesifisitet. Studien konkluderer videre med at bildediagnostikk i form av MR bør supplere undersøkelsen for nøyaktig diagnose. Calis et al. (2000) viste at de mest sensitive diagnostiske testene i deres studie var Hawkins test (92,1 %), Neer test (88,7 %) og horisontal adduksjonstest (82,0 %). Testene med høyest spesifisitet var drop arm test (97,2 %), Yergason test (86,1 %) og Painful arc (80,5 %). Det problematiseres at testene med høy sensitivitet virker å ha lav spesifisitet og motsatt. I studien til Kelly et al (2010) ble det sett på diagnostisk nøyaktighet ved SIS. Nøyaktigheten varierte fra 71-62% på de ulike testene med Hawkins-Kennedy test som den mest nøyaktige. Park et al. fra 2005 tok også for seg den diagnostisk nøyaktighet av kliniske tester med tanke på å avdekke SIS. Resultatet i denne omfattende studien viser at Hawkins- test, Neers test, Painful arc og Empty can (jobe) kommer ut som de fire mest treffsikre testene. De har alle høy LR+ og lav LR-. **Resultatene fra denne studien tillegges stor vekt i vår oppgave, også på grunn av et meget stort antall deltakere (n=913).**

Reliabilitet og diagnostisk nøyaktighet av 5 kliniske tester for å avdekke SIS ble undersøkt av Michener et al. (2009). Testene som ble brukt var Neer, Hawkins, Painful arc, Jobe og Utrotasjonstest med motstand. To erfarne terapeuter, en ortoped og en fysioterapeut trente sammen i forkant. Kappa-koeffisienten varierte fra 0,39-0,67 på de fem testene. Neer 0.39 (0.12-0.65) Hawkins 0.40 (0.13-0.67) Empty-can 0.47 (0.22-0.72) Painful arc 0.45 (0.18-0.72) Utadrotasjon mot motstand 0.67 (0.4-0.94). Studien konkluderte blant annet med at alle

testene har akseptabel diagnostisk treffsikkerhet og reliabilitet i forhold til klinisk bruk. Testene painful arc, Jobe og Utrotasjon med motstand kom best ut med tanke på reliabilitet.

Ut fra disse studiene samt har vi valgt å benytte de fire diagnostiske testene Hawkins, Neers, Painful arc og Jobe supraspinatustest (Empty can) i studien. De to førstnevnte beskrives i litteraturen som rene impingementtester (Magee 2008, Solberg 2002, Hattam og Steatham 2010, Bahr og Mæhlum 2003). Likhetsratioene på Jobes test og painful arc er noe mer sprikende. Til tross for dette presenteres både Jobes test (Gross et al. 1996) og Painful arc (Bahr og Mæhlum 2003) i litteratur som impingementtester. Vi har spesielt vektlagt studien til Park et al. (2005) i utvelgelsen av våre fire tester. Dette på grunn av stor utvalgstørrelse, 913. I denne studien kommer våre 4 valgte tester best ut totalt sett, med høy LR + og lav LR.

#### 1.2.4 Reliabilitet:

Reliabilitet defineres av Domholdt (2000) som den grad målinger er fri for målefeil.

Test – retest reliabiliteten sier noe om hvor like måleresultater blir ved repeterte målinger – målinger tatt med samme måleinstrument, på samme person (Hopkins, 2000).

Ringdall 2009 definerer test-retest reliabilitet som; grad av samsvar mellom målinger av samme variabel på ulike tidspunkt. Dette undersøkelsesdesignet med to repeterte målinger kalles test-retest design.

Ved vurdering av kvaliteten til en målemetode, er det å undersøke metodens reliabilitet et sentralt ledd. Repeterte målinger av samme kvalitet på samme person vil vanligvis ikke bli helt like (Altman, 1997). Grunner til dette er naturlig variabilitet deltaker eller tester, variabilitet i målingsprosessen, variabilitet i måleinstrumentet eller til sist en kombinasjon av disse. En vil derfor se at repeterte målinger, av samme person, vil variere rundt den sanne verdi på grunn av målefeil (ibid).

I teorien finner vi to måter å vurdere reliabilitet på:

- Klassisk testteori bygger på det grunnlaget at en måling består av en sann score pluss en feilkomponent i målemetoden. Forholdet mellom repeterte målinger brukes til å estimere feilkomponenten. I denne teorien anses all variabilitet mellom målinger av en person, å skyldes feilkomponenten (Domholdt, 2000).
- Et utvidet syn på årsaker til variabilitet i målinger ses ved generaliserbarhetsteori hvor grunnene til variabilitet ansees å ligge i instrumentet (*instrumentets reliabilitet*),

mellom ratere eller hos rater (*inter- og intratester reliabilitet*), samt hos personen (*intrasubjektvariabilitet*) Domholdt, 2000).

Til *instrumentreliabilitet* hører begrepene *presisjon* (accuracy) og *nøyaktighet* (trueness) til (Altman, 1997). *Presisjon* sier hvor likt samme måleresultat kan oppnås igjen og igjen.

*Nøyaktighet* er et uttrykk for hvor likt måleresultatet er den sanne verdien (ibid).

Ved *raterreliabilitet* beskrives to sider, *interraterreliabilitet* og *intraraterreliabilitet*.

*Interraterreliabilitet* representerer grad av enighet som to eller flere testere gjør på et sett av observasjoner/tester. *Intraraterreliabilitet* gir uttrykk for grad av enighet hos en tester ved repeterte målinger på et sett av observasjoner (Domholdt, 2000).

*Intrasubjektvariabilitet* er variabiliteten i personens resultatoppnåelse ved repeterte målinger (Hopkins, 2000, Domholdt, 2000). Biologiske forhold hos et individ, som fysiske og psykiske endringer, er hovedkilden til variabilitet (ibid). Videre påvirkes testpersonen av hvordan informasjon gis i forbindelse med utførelse av en test, hvor god den gitte informasjon er samt selve opplevelsen av å gjennomføre testen (ibid).

### 1.2.5 Relevante studier

Søk i PubMed 09.10.11 hadde som hensikt å kartlegge studier der test-retest-reliabiliteten for de mest aktuelle impingementtestene var undersøkt. Søkeordene var ("shoulder" OR "impingement" OR "shoulder impingement" OR "shoulder girdle") AND ("test-retest" OR "intra-examiner" OR "reliability"). Fra dette søket ble det funnet kun 2 artikler og en systematisk oversikt som på bakgrunn av abstract ble vurdert som særlig relevante: Hegedus m.fl. 2008, De Wilde et al. (2003), Johansson og Ivarson (2009). Det kan problematiseres at det finnes mange studier hvor det er blitt forsket på intertester reliabiliteten alene (Michener et al. 2009, Nomden et al. 2009, Nørregaard et al. 2002, Ostor et al. 2004, Cadogan et al. 2001, MacDonald et al. 2000), men få studier som har forsket på test retest-reliabiliteten av aktuelle skuldertester.

I studien til Hegedus et al. 2008 sammenfattes studier knyttet til kliniske tester av skulder i en systematiske oversikt. For vår oppgave er funnene knyttet opp mot Neer- test og Hawkins-test for impingement mest interessant. Studien konkludere med at de to impingement- testene er forholdsvis reliable, men at de har begrenset diagnostisk verdi. Det presiseres videre at det er stort behov for metodisk sterke studier i videre forskning på området.

I studien til De Wilde et al. fra 2003 ble det rapportert høy intra-og intertesterreliabilitet på

henholdsvis (ICC 0,97 til 0,93) ved utførelse av en modifisert Hawkins-Kennedy impingement test. Fem sensorer utført tre målinger hver på fem skuldre hos friske individer og brukt en liggende stilling. Denne studien ble altså ikke gjort på målgruppen som er pasienter med skulderplager og er noe mindre relevant.

Intra- og intertestreliabilitet av fire manuelle skuldertester brukt for å kunne identifisere subacromiale smerter ble undersøkt av Johansson og Ivarson (2009). 33 deltakere med skulderplager ble testet. De tok for seg testene Neer impingement sign, Hawkins- Kennedy impingement test, Patte maneuver og Jobe supraspinatustest. De to testerne fant at alle de fire testene hadde tilnærmet perfekt enighet både med tanke på inter og intratesterreliabilitet, med en Kappa-koeffisient på 0,91-1. Alle de fire testene hadde perfekt enighet ved intratesterreliabilitet. Forfatterne konkluderte med at testene er reproduserbare og reliable til klinisk bruk for å kunne identifisere SIS. Verdt å merke seg ved denne studien er at testerne trente mye sammen i forkant av studien.

## **2.0 HENSIKT OG PROBLEMSTILLING**

### **2.1. Hensikt**

I en studie der kliniske tester skal anvendes i diagnostiseringen av en tilstand, forutsettes det at de er reliable. Denne studien hadde som formål å undersøke måleegenskaper til de kliniske testene. Min studie var del av en større tredelt studie der test-retest og inter-tester reliabilitet og diagnostisk nøyaktighet ble undersøket mht fire impingement tester i skulderen.

Ultralydundersøkelse ble brukt som en gullstandard for å diagnostisere impingement og annen skulderpatologi.

I min del av studien undersøkte jeg test-retest reliabiliteten.

### **2.2. Problemstilling**

Hvor god er test-retest reliabiliteten til fire subacromiale impingementtester i skulder; Hawkins-Kennedy test, Neers test, Jobe supraspinatustest og Painful arc.



## 3.0 METODE

### 3.1 Valg av forskningsdesign

Denne oppgaven kan beskrives som en reliabilitetsstudie, der vi har undersøkt reliabiliteten til fire tester på to tidspunkt i et test-retest design. Som forskningsdesign passer dette best inn under en tverrsnittstudie med en referansetest (Ringdal 2009), en test – retest analyse. Studien er kvantitativ, som forutsetter at dataene er målbare og kan gis til uttrykk i tall eller mengdedeterminologi (Halvorsen 1996). I denne studien så vi på samvariasjon mellom de funnene to terapeuter får ved utførelse av fire tester, på det samme utvalget to ganger med ca en ukes mellomrom.

Vi valgte ut tester med relativt høy sensitivitet og samtidig relativt høy spesifisitet.

- Hawkins test presenteres som en impingementtest (Hattam og Smeatham 2010, Gross et al 96, Solberg 2002, Bahr og Mæhlum 2003).
- Neers tegn presenteres som en impingementtest (Solberg 2002, Hattam og Smeatham 2010, Bahr og Mæhlum 2003).
- Jobes test er primært supraspinatustest, sekundært en SIS test. Er også ofte positiv ved cuffruptur, artrogene forhold eller luksasjon i skulder (Solberg 2002).
- Painful arc: Smerter ved deler av abduksjonsbanen i skulder ses ofte ved SIS (Solberg 2002:123, Bahr og Mæhlum 2003).

### 3.2 Utvalg

I studien brukte vi tre kliniske utvalg, ett utvalg bestående av pasienter med diagnostisert impingement, ett utvalg med pasienter som hadde annen skulderproblematikk, og ett utvalg friske uten skulderpatologi. Hensikten med tre utvalg var å styrke oppgaven metodisk. De to testpersonene var blindet overfor hvilken gruppe de ulike pasientene tilhørte. Tre ulike utvalg gjorde det vanskeligere for de blindede testpersonene å vite hvilken gruppe deltakerne tilhørte. Utvalgene åpnet også muligheten for å undersøke de ulike testenes diagnostiske nøyaktighet.

I tillegg til å bestemme utvalgets sammensetning, måtte en definere utvalgets størrelse. I reliabilitetsstudier med dikotome utfallsvariabler benyttes kappastatistikk og  $\kappa$ -verdier gir uttrykk for testenes reliabilitet.  $\kappa$ -verdier er et punkttestimat, og alle punkttestimater er beheftet med usikkerhet. Usikkerheten kommer til uttrykk gjennom konfidensintervallene (Jamtvedt et al. 2003). En ønsker derfor at utvalget skal være av en viss størrelse. Til dette

kan tre grunner trekkes frem:

- Det rådes å ha et stort nok antall i utvalget for å kunne generalisere resultatene til den større kliniske populasjonen som utvalget er trukket fra (Carter et al 2011).
- Enkelte statistiske tester behøver minimum 30 personer i utvalget (ibid).
- Presisjonen ved deskriptiv statistikk vil avhenge av utvalgsstørrelse (ibid)

Det kan dermed sies at jo større utvalg, jo lavere usikkerhet vil være forbundet med punkttestimatene og en vil følgelig også få smalere konfidensintervall (Straus et al 2005). Det ble av disse grunnene regnet som hensiktsmessig med minimum 30 personer i studien. Terwee et al. 2007 anbefaler imidlertid 50 personer. Skulle det i innenfor rammene av prosjektet, la seg gjøre å rekruttere flere enn 30 deltakere til studien ville det bli gjort.

Rekrutteringen til utvalgene med skuldersmerter ble gjort av en erfaren lege som driver en stor skulderklinikk. Hans hovedvirke er ultralydundersøkelse av skulder. Han rekrutterte innenfor tidsrammene 14 pasienter med positiv SIS til vår studie og 4 pasienter med annen skulderproblematikk. En kollega og medforfatter som ikke utførte var involvert i testingen rekrutterte 16 friske deltakere. Til sammen hadde vi da 18 pasienter og 16 friske deltakere. For å bli invitert til å delta i studien og kunne inngå i utvalget måtte pasientene passe inn i henhold til definerte inklusjons- og eksklusjonskriterier:

<b>Gruppe 1: Impingement</b>	
Inklusjonskriterier:	Eksklusjonskriterier:
<ul style="list-style-type: none"> <li>• Diagnostisert med impingement ved ultralyd</li> <li>• Skuldersmerter med varighet mer enn én måned</li> <li>• Skuldersmertene må være pasientens aktuelle hovedproblem</li> <li>• Personer i alderen 18-60 år</li> <li>• Må kunne gi et informert samtykke</li> </ul>	<ul style="list-style-type: none"> <li>• Annen kjent patologi i skulder</li> <li>• Tidligere traume mot skulder</li> <li>• Tidligere skulderoperert</li> <li>• Inflammatoriske leddlidelser</li> <li>• Smerte ved bevegelse av nakken</li> <li>• Nevropatisk smerte i nakke eller overekstremitet</li> <li>• ”Røde flagg” som kan mistenke annen alvorlig patologi</li> </ul>

<b>Gruppe 2: Annen skulderpatologi</b>	
Inklusjonskriterier:	Eksklusjonskriterier:
<ul style="list-style-type: none"> <li>• Diagnostisert med annen skulderpatologi ved ultralyd</li> <li>• Skuldersmerter med varighet mer enn én måned</li> <li>• Skuldersmertene må være pasientens aktuelle hovedproblem</li> <li>• Personer i alderen 18-60 år</li> <li>• Må kunne gi et informert samtykke</li> </ul>	<ul style="list-style-type: none"> <li>• Skulderimpingement</li> <li>• Tidligere traume mot skulder</li> <li>• Tidligere skulderoperert</li> <li>• Inflammatoriske leddlidelser</li> <li>• Smerte ved bevegelse av nakken</li> <li>• Nevropatisk smerte i nakke eller overekstremitet</li> <li>• ”Røde flagg” som kan mistenke annen alvorlig patologi</li> </ul>

<b>Gruppe 3: Friske</b>	
Inklusjonskriterier:	Eksklusjonkriterer:
<ul style="list-style-type: none"> <li>• Ingen smerter i skulderområdet siste 3 måneder</li> <li>• Personer i alderen 18-60 år</li> <li>• Må kunne gi et informert samtykke</li> </ul>	<ul style="list-style-type: none"> <li>• Smerter i skulderområdet</li> <li>• Tidligere traume mot skulder</li> <li>• Tidligere skulderoperert</li> <li>• Inflammatoriske leddlidelser</li> <li>• Smerte ved bevegelse av nakken</li> <li>• Nevropatisk smerte i nakke eller overekstremitet</li> <li>• ”Røde flagg” som kan mistenke annen alvorlig patologi</li> </ul>

Selve invitasjonen til å bli med i studien ble foretatt av andre enn de som utførte testene, med andre ord ble rekrutteringsprosessen administrert av en uavhengig person. Gjennom kriteriene søker en i utgangspunktet et mest mulig heterogent utvalg. Grad av heterogenitet kan i etterkant analyseres langs dimensjoner som kjønn og alder og skårer for deltakerne i utvalget.

### 3.3 Datainnsamling

Datainnsamlingen foregikk på klinikken der forskerne arbeidet. Det ble gjort målinger med hver enkelt av de fire impingementtestene i skulderen, og det ble registrert om den gitte testen fremprovoserte smerte i skulderen eller ikke. Data innsamlingen ble gjort vår/høst 2012, med bearbeidelse høsten 2012. Tidsfristen for ferdig prosjekt var november 2012.

De kliniske testene som vi gjennomførte var Neers test, Hawkins-Kennedy test, Jobe/empty can-test og painful arc. Moen et al. (2010) og Solberg (2002) har gjort rede for hvordan disse 4 testene korrekt skal utføres i henhold til hvordan de opprinnelig ble beskrevet, og den sekundære fremstillingen til Moen et al. (2010) ble lagt til grunn da det ikke lot seg gjøre å fremskaffe de primære artiklene gjennom universitetsbiblioteket i Bergen.

#### 3.3.1 Beskrivelse av testene:

##### *Neers test*

Pasienten plasseres i sittende stilling, terapeuten står bak. Terapeuten fikserer scapula ipsilateralt for å hindre protraksjon av skulderen. Pasientens arm flekteres passivt. Testen er positiv hvis den framprovoserer smerter. Tenkt mekanisme ved testen er inneklemping av supraspinatus-senen og den subacromiale bursaen mellom tuberculum majus og acromion (Moen et al 2010).



##### *Hawkins test*

Pasienten i sittende, med overarmen 90 grader flektert. Scapula på samme side fikseres for å hindre rotasjon av scapula. Overarmen roteres passivt innover til end-feel eller til eventuelle smerter oppstår (Solberg 2002).



*Jobes test*

Pasienten i sittende, skulderen 90 grader abduisert, 30 grader horisontal addusert og maksimalt innadrottert. Terapeuten plasserer hånd på oversiden av pasientens overarm. Pasient skal holde overarmen i ro mens terapeut appliserer et trykk nedover mot pasientens arm. Positiv test hvis smerter (Moen et al. 2010).

*Painful arc*

Pasienten i sittende. Park et al. (2005) instruerer pasienten til å føre armen i fleksjon i scapulas plan/scaption. Testen er positiv hvis pasienten opplever smerter mellom 60 og 120 grader abduksjon (Kessel og Watson 1977, Bahr og Mæhlum 2003, Park et al. 2005, Michener et al. 2009, Kelly et al. 2010).



Terapeutene er begge i gang med en klinisk masterstudie i manuellterapi, har relativt lik yrkes-bakgrunn og ca. 5 år klinisk erfaring innenfor muskel/skjelettproblematikk. Testene ble skåret med utgangspunkt i dikotomien positiv eller negativ, og testene ble skåret som positiv hvis de primært gav smerter lokalt i skulderområdet. Pasienten måtte oppleve relativt lokale og distinkte/kjente smerter for at testen skulle bli registrert som positiv. Andre opplevelser som strekk og diffuse ubehag ble registrert som negativ test.

**Faktisk utførelse:**

De fire testene ble utført av terapeut nr. 1 med 2 minutters mellomrom mellom hver test.

Deretter ble pasienten testet på nytt av terapeut nr. 2 etter tre minutter, med samme prosedyre som terapeut nr. 1.

Valg av 2 minutters pause mellom testene ble valgt for å gi den aktuelle skulderen hviletid for å ta hensyn til en eventuell irritasjon ved den foregående testen.

De to testpersonene var blindet for hvilken gruppe pasientene tilhørte. Pasientene ble videre oppfordret til ikke å si noe om egen helsetilstand eller eventuelle plager før siste retest var

utført. Hver pasient i gruppene med skulderpatologi fikk etter at alle testene var utført utdelt et øvelsesprogram og gitt generelle ergonomiske tips og øvelser til bruk i hverdagen.

I forkant av studien gjennomgikk terapeutene testprosedyrene for testene som beskrevet av Moen et al. (2010). Testene ble også utført på to pasienter som øving før oppstart av testingen. Det var enighet mellom terapeutene at annen standardisering av testutførelse ikke skulle gjennomføres, dette for i større grad prøve å bevare den økologiske validiteten. Med økologisk validitet menes her en form for generalisering hvor en ønsker at de eksperimentelle testresultatene også skal være gjeldende i det virkelige liv (Kruuse, 2007). Vi tenker også at standardiseringen kan bli mer omfattende enn hva som er mulig å få til i klinisk praksis hvis terapeuter har en omfattende trening sammen i forkant av testingen. Dette har vi eksempel på i studien til Johansson m.fl. (2009) der omfattende trening mellom testerne i forkant kan ha ført til et kunstig godt samsvar mellom terapeutene, med resultat mellom 0,91 til perfekt korrelasjon på Kappa-koeffisienten (Johansson et. al. 2009). Test og utførelse hentet fra (Solberg 2002 og Moen 2010)

### 3.4 Analyse

Deltakerne i studien ble beskrevet med en deskriptiv statistikk. Etter at alle testene var utført og data samlet inn ble resultatene analysert. Testene ble skåret etter dikotomien positiv eller negativ, som befinner seg på nominalskala-nivå. Grad av samsvar i skåringene mellom test og retest ble undersøkt med kappa statistikk, ved hjelp av Statistical Package for the Social Sciences (SPSS, versjon 19 for mac). En finner med utgangspunkt i data fra denne programvaren k-verdier og 95% konfidensintervall. Inter- og intratester-reliabilitetsstudier med dikotome variabler gir generelt en kappa-verdi som uttrykk for testenets reliabilitet.

Cohens kappa (k) avspeiler grad av samsvar i resultat mellom eksempelvis en test og en retest utover det samsvar som følger av tilfeldighet (Altman 1997). Kappa kan ha verdier fra -1 til +1, hvor -1 er perfekt uenighet og +1 som gjenspeiler perfekt samsvar (ibid). Middelveien 0 betyr at det ikke er noe grad av samsvar utover det som er tilfeldig. Kappaverdier brukes for å justere for henholdsvis inter- og intratester enighet som oppstår tilfeldig (Landis og Koch 1977, Domholdt 2000). Kappa måler altså enighet ut over det som forventes helt tilfeldig, og dess nærmere skåringene er +1 dess bedre. Kappa inndeles gjerne i fem nivåer:

Kappa  $<0,2$ : Neglisjerbart samsvar.

Kappa  $\geq 0,2 < 0,4$  Minimalt samsvar.

Kappa  $\geq 0,4 < 0,6$  Moderat samsvar.

Kappa  $\geq 0,6 < 0,8$  Bra samsvar.

Kappa  $\geq 0,8$  Nesten perfekt samsvar.

Kappaverdier over eller lik 0,6 anses som bra (Landis og Koch 1977, Altman 1997).

### 3.5 Etiske hensyn

Det var en forutsetning for studien at etiske hensyn ble fulgt. Pasientene ble grundig informert og skrev under på en samtykke erklæring før de ble inkludert i forskningsprosjektet.

Generelle etiske overveielser fulgte reglene og anbefalingene som er utarbeidet av Helsinkideklarasjonen (Ruyter et al, 2000). I deklarasjonen legges det vekt på at hensynet til forsøkspersonenes helse og integritet må gå foran hensynet til forskning og samfunn (Dalland, 2007).

Det er i arbeid med pasienter overordnet at man overholder taushetsplikten og at alle pasientdata blir anonymisert i studien. Alle pasientene ble informert om at kunne trekke seg fra studien når som helst uten at det ville gi noen konsekvenser for eventuell videre oppfølging.

I hele studien var det tre testsituasjoner og to ulike testere som utførte testene. Normalt i klinisk praksis gis det flere intervensjoner, og deltakerne var godt informert om at det i denne studien bare ble utført fire tester og ingen intervensjon for øvrig.

Hvorvidt man bør drive forskning på en utvalgt gruppe er ifølge Birkler (2005) et grunnleggende etisk spørsmål. Det bør nøye overveies om prosjektet kan styrke klinisk praksis og hvilken nytteverdi det har. I denne studien var det ingen samfunnsmessige omkostninger forbundet med arbeidsprosessen.

Som skissert i bakgrunn er skulderplager og SIS en stor diagnosegruppe. Derfor virker det nærliggende at det bør gjøres reliabilitetsstudier som tilstreber å validere de kliniske testene som brukes for å avdekke disse lidelsene. Sett i et slik perspektiv kan prosjekt ha en verdi for den enkelte pasient med skuldresmerter, fysioterapeut/manuellterapeut og samfunn.

Det tilstrebes å presentere studienes resultater så objektivt og korrekt som mulig og dermed unngå å fordreie eller ignorere uønskede resultater. Dette gjøres i respekt for vitenskapens interne etikk, som sier at man ikke bør lyve eller jukse og dermed bedrive vitenskapelig uredelighet (Birkler, 2005). Dette prosjekt har på bakgrunn av disse overveielser og de anvendte metoder, forsøkt å bidra ytterligere til å evidensbasere reliabiliteten av impingement-tester.

**Egen rolle som forsker**

Det er viktig at man som forsker holder seg så nøytral som mulig og at man tilstreber å være så objektiv som mulig. Som forsker har man alltid med seg en viss forforståelse av emnet som det skal forskes innenfor. Malterud (2003) beskriver dette som en ryggsekk som vi bærer med oss inn prosjektet og som hele veien vil påvirke måten vi samler inn og leser våre data på.

Kritisk vurdering av de ”brillene” man ser med er viktig samt prøve å se situasjonen fra en annens forsker ”briller” (Malterud, 2003). Som manuellterapeut er kliniske tester en viktig del av undersøkelsesprotokollen, som tilbys pasientene. Dette betyr at man som manuellterapeut både må tro på testbatteriet som skal utføres og samtidig ha gode kunnskaper og evner slik at de utføres riktig og likt hver gang. Hvis ikke kan dette påvirke resultatene.

For å sikre mest mulig grad av nøytralitet brukte vi en kollega til å registrere og behandle data.

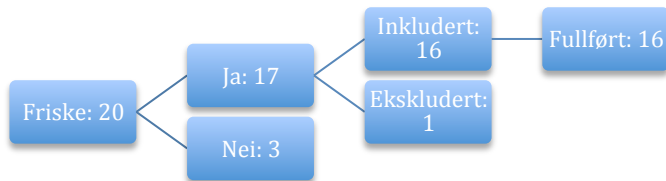


## 4.0 RESULTATER

I denne studien ble totalt 34 pasienter inkludert, fordelt i tre ulike pasientkategorier.

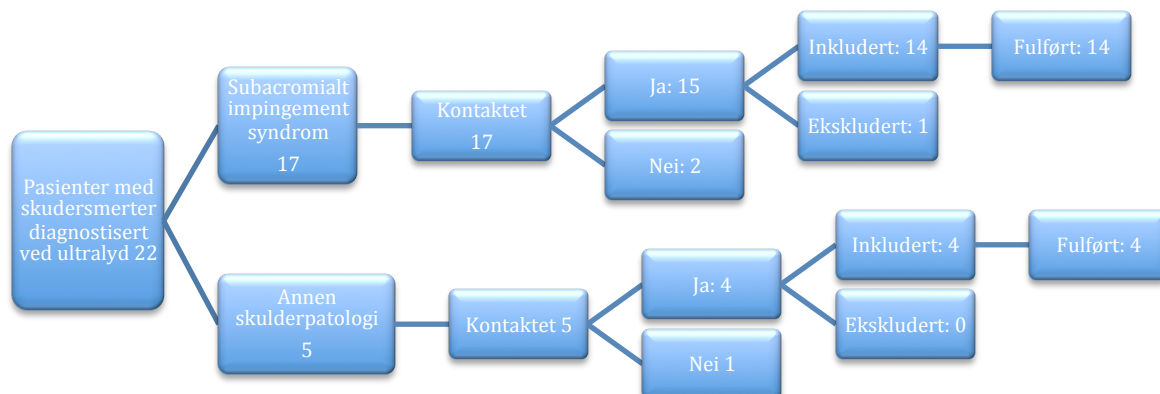
Rekrutteringsprosessen blir belyst i to flytskjemaer under, figur 1 og 2.

Figur 1: Rekrutteringsprosessen i underutvalgsgruppen friske.



I gruppen med friske ble totalt 20 personer kontaktet. 17 av disse sa seg villige til å bli med på studien. Av disse 17 viste det seg at en av dem hadde utviklet skulderpatologi før oppstart av testing. Hun ble derfor ekskludert. Av 16 inkluderte i friskgruppen fullførte samtlige både test og retest.

Figur 2: Rekrutteringsprosessen i underutvalgsgruppene impingement og annen skulderpatologi.



I gruppen med diagnostisert impingement på ultralyd ble 17 pasienter rekruttert fra Per Sunde ved Skulderklinikken hvorav to sa nei før oppstart av testing, mens en pasient ble ekskludert grunnet oppstart av behandling og injeksjon i skulder. Av 14 inkludert fullførte samtlige både test og retest. I gruppen annen skulderpatologi ble kun 5 pasienter rekruttert fra Per Sunde. En pasient sa nei til å delta mens de fire andre fullførte test og retest.

Resultatene fra test og retest er fremvist i tabeller under. Tabell 1 tar for seg underutvalgenes størrelse og sammensetning med tanke på kjønn, alder, sykdomsvarighet samt hvilken side som ble testet.

Tabell 1. Oversikt over bakgrunnsdata på deltakerne i de tre gruppene som ble testet

	<b>Gruppe 1: Impingement</b>	<b>Gruppe 2: Annen skulderpatologi</b>	<b>Gruppe 3: Friske</b>
<b>Antall (%)</b>	14/34 (41%)	4/34 (12%)	16/34 (47%)
<b>Kjønn (%)</b>	K: 6/19 (32%)	K: 2/19 (11%)	K: 11/19 (57%)
<b>K=kvinne, M=mann</b>	M: 8/15 (54%)	M: 2/15 (13%)	M: 5/15 (33%)
<b>Gj.snitt alder, (SD)</b>	46 år (13)	44 år (9)	40 år (16)
<b>Variasjonsbredde</b>	(18-60 år)	(33-55 år)	(21-60 år)
<b>Sykdomsvarighet, mndr: Gj.snitt (SD), median, Variasjonsbredde</b>	37 mndr, (94), Median 12* (1-360 mndr)	16 mndr, (6.8), Median 15,5 mndr, (9-24 mndr)	-
<b>Aktuell side testet (%)</b>	H: 4/19 (21%) V: 10/15 (67%)	H: 4/19 (21%) V: 0/15 (0%)	H: 11/19 (58%) V: 5/15 (33%)

\*Stor forskjell på median og gjennomsnitt grunnet en deltager med smerter i 360 mnd.

Ut fra tabellen kan vi lese at vi fikk en relativt homogene grupper med hensyn til alder som er den viktigste variabelen. Kjønnsspredning er relativt lik totalt med 19 kvinner og 15 menn. Det er dog litt sprik gruppene mellom. Aktuell side testet er også relativt likt totalt med 19 høyre og 15 venstre, selv om vi ser at det er litt sprik innad i gruppene. Sykdomsvarigheten blir veldig upresis grunnet en deltager med smerter i 30 år, median er derfor riktigst å henvende til i gruppen med impingement.

Dessverre lot det seg ikke gjøre innenfor tidsrammene å rekruttere flere enn fire pasienter til gruppe 2, annen skulderpatologi.

I tabell 2-4 er testresultatene fra test og retest i de tre gruppene gjengitt i krysstabeller. Krysstabellene gir en oversikt over antall positive og negative målinger i de tre ulike

gruppene, og viser hvor mange målinger som er like fra test til retest. Tabell 5 gir en samlet oversikt over alle gruppene.

Tabell 2: Krysstabell over de fire testene i underutvalget med impingement. Tabellen gir en oversikt over hvor mange tester som skåres likt/ulikt fra test til retest.

		Neers			Hawkin				
		Negativ	Positiv	Total			Retest		
					Negativ	Positiv	Total		
Neers	Neg.	3	0	3	Hawkin	Neg.	2	1	3
Test	Pos.	0	11	11	Test	Pos.	2	9	11
Total		3	11	14	Total		4	10	14

		Jobe			P.A.				
		Negativ	Positiv	Total			Retest		
					Negativ	Positiv	Total		
Jobe	Neg.	3	0	3	P.A	Neg.	1	0	1
Test	Pos.	1	10	11	Test	Pos.	3	10	13
Total		4	10	14	Total		4	10	14

I underutvalget av deltakere med impingement viste Neers test perfekt enighet fra test til retest. Jobe viste endring i skåre hos en deltaker. Hawkins-Kennedy test og Painful arc har flest ulike målinger fra test til retest med endring i testresultat hos 3 deltakere.

Tabell 3: Krysstabell over de fire testene i underutvalget med annen skulderpatologi. Tabellen gir en oversikt over hvor mange tester som skåres likt/ulikt fra test til retest.

		Neers			Hawkin				
		Negativ	Positiv	Total			Retest		
					Negativ	Positiv	Total		
Neers	Neg.	1	1	2	Hawkins	Neg.	1	0	1
Test	Pos.	0	2	2	Test	Pos.	0	3	3
Total		1	3	4	Total		1	3	4

		Jobe			P.A.				
		Negativ	Positiv	Total			Retest		
					Negativ	Positiv	Total		
Jobe	Neg.	3	0	3	P.A	Neg.	1	0	1
Test	Pos.	0	1	1	Test	Pos.	2	1	3
Total		3	1	4	Total		3	1	4

I underutvalget annen skulderpatologi viste testene Jobe og Hawkins perfekt enighet fra test til retest. Ved Neers test var det endring i skåre hos en deltaker, mens Painful arc har flest ulike målinger fra test til retest med endring i testresultat hos 2 deltakere.

Tabell 4: Krysstabell over de fire testene i underutvalget med friske deltakere. Tabellen gir en oversikt over hvor mange tester som skåres likt/ulikt fra test til retest.

		Neers			Retest			Hawkin		
		Negativ	Positiv	Total				Negativ	Positiv	Total
Neers	Neg.	15	0	15	Hawkins	Neg.	16	-	16	
Test	Pos.	0	1	1	Test	Pos.	-	-	-	
Total		15	1	16	Total	16	-	-	16	

		Jobe			Retest			P.A.		
		Negativ	Positiv	Total				Negativ	Positiv	Total
Jobe	Neg.	16	-	16	P.A	Neg.	15	1	16	
Test	Pos.	-	-	-	Test	Pos.	-	-	-	
Total		16	-	16	Total	15	1	1	16	

Ved testene Neers, Hawkins og Jobe fikk samtlige deltaker samme skåre fra test til retest i underutvalget friske. Ved testen Painful arc var det endring i skåre hos en deltaker fra test til retest. Totalt 3 av 128 tester kommer ut falsk positiv i friskgruppen.

Tabell 5: Krysstabell over de fire testene med totalt antall deltakere. Tabellen gir en oversikt over hvor mange tester som skåres likt/ulikt fra test til retest.

		Neers			Retest			Hawkin		
		Negativ	Positiv	Total				Negativ	Positiv	Total
Neers	Neg.	20	1	21	Hawkins	Neg.	19	1	20	
Test	Pos.	0	13	13	Test	Pos.	2	12	14	
Total		20	14	34	Total	21	13	13	34	

		Jobe			Retest			P.A.		
		Negativ	Positiv	Total				Negativ	Positiv	Total
Jobe	Neg.	21	0	21	P.A	Neg.	17	1	18	
Test	Pos.	1	12	13	Test	Pos.	5	11	16	
Total		22	12	34	Total	22	12	12	34	

Det var stor enighet i skårer mellom test og retest. I det totale utvalget var det ved testene Neers og Jobe en deltaker som fikk ulik skåre fra test til retest. Hawkins-Kennedy test viste totalt tre ulike testresultat mens Painful arc hadde flest ulike skåringer fra test til retest med endring i testresultat hos 6 deltakere.

Tabell 6 gir en samlet oversikt over antall positive og negative tester ved test og retest. Tabellen er ikke delt opp i de ulike gruppene.

	Test		Retest	
	Positiv	Negativ	Positiv	Negativ
<b>Neer impingement</b>	13 (38%)	21 (62%)	14 (41%)	20 (59%)
<b>Jobe supraspinatus test</b>	13 (38%)	21 (62%)	12 (35,3%)	22 (64,7%)
<b>Hawkins-Kennedy test</b>	14 (41%)	20 (59%)	13 (38%)	21 (62%)
<b>Painful arc</b>	17 (50%)	17 (50%)	12 (35,3%)	22 (64,7%)
<b>Totalt</b>	57 (42%)	79 (58%)	51 (37,5%)	85 (62,5%)

Fra tabell 6 kan en lese at det i snitt ble skåret flere positive tester ved første test enn ved retest. Totalt ble 57 (42%) av 136 tester skåret positivt første test mot 51 (37,5%) av 136 tester ble skåret positivt ved retest en uke senere. Mulige årsaker til dette blir videre tatt opp i diskusjonskapittelet. Spesielt er det testen Painful arc som er med på å lage dette skillet. Den alene ble skåret positiv 17 ganger ved test en mot 12 ganger ved retest.

Tabell 7. Oversikt over Kappa verdier og prosentmessig enighet på de fire impingementtestene ved test og retest. N=34

Studie	Tester	Kappa koeffisient (95 % KI)	Prosentmessig enighet test-retest
Antall: n=34	Neers tegn	.939 (.060)	97,058 %
	Jobe supraspinatus test	.937 (.062)	97,058 %
	Hawkins- Kennedy test	.816 (.101)	91,176 %
	Painful arc	.641 (.129)	82,352 %

Test retest reliabiliteten på de fire anvendte testene i denne studien kommer alle godt ut. Best resultat ble oppnådd ved testene Jobe supraspinatustest og Neers tegn. De kom best ut med en kappaverdi på 0,93-0,94 og prosentmessig enighet på 97,058%. Hawkins test ble skåret med en k-verdi på 0,816, 91,8 % enighet, mens painful arc kom dårligst ut med en k-verdi på 0,641 og prosentmessig enighet på 82,35 %. For å justere resultatene med 95% konfidensintervall må tallene i parentes i tabell 7 ganges med 1,96, som er 2 standardavvik (Landis og Koch 1977, Altman 1997). K- verdiene med 95 % KI blir da for Neers tegn 0,82-1, for Jobe supraspinatus test 0,82-1, for Hawkins- Kennedy test 0,62-1 og for Painful arc 0,39-0,89.

## 5.0 DISKUSJON

Nedsatt plassforhold og eller irritasjon av strukturene mellom acromion og humerushodet kan føre til subacromialt impingement (SIS). Målet med studien var å måle test-retest reliabiliteten på fire av de mest anvendte testene som brukes for å avdekke denne tilstanden. Tre av fire tester viste i denne studien en Kappaverdi bedre enn 0,8 som betyr at det var meget godt samsvar mellom repeterte skåringer hos den samme testeren. Best resultat ble oppnådd ved Neers tegn og Jobe supraspinatus test, med  $\kappa = 0,94$ . Prosentmessig var enigheten på disse testene 97 %. Kun en pasient ble testet ulikt fra test til retest.

Hawkins-Kennedy test kom også godt ut med tanke på test-retest reliabilitet. Kappaverdien her ble målt til 0,82 og prosentmessig enighet var 91 %. Totalt testet tre pasienter her ulikt fra test til retest. Painful arc er den testen som kom dårligst ut. Den fikk en Kappaverdi på 0,64, hvilket går under kategorien godt samsvar. Prosentmessig enighet var her 82,4 %. Totalt 6 deltakere skåret ulikt fra test til retest.

Ved første test ble totalt 57 tester (42 %) skåret positivt, mot 51 tester (37,5 %) ved retest. Det er verdt å merke seg at det ved test 1 ble utført to testrunder (2 x 4 tester), en av meg og en av tester to for intertesterreliabilitet. Retest en uke senere ble bare utført av undertegnede for test retest reliabilitet (1 x 4 tester). Videre vil det bli diskutert hva endringene fra test til retest skyldes. Kan pasientene ha endret symptombildet og blitt bedre? Skyldes endringen for stor testbelastning ved første test. Kan det ha vært en endring i utførelse eller standardisering, eller har andre faktorer spilt inn?

Jeg vil først diskutere vår studie opp mot lignende studier, deretter diskutere utførelse av testene, før jeg til slutt diskuterer hva som påvirker studiens validitet og oppsummerer studiens styrker og svakheter.

### 5.1 Studien opp mot lignende studier

Etter å ha utført grundige søk i Pubmed fant jeg kun to tidligere studier som direkte har tatt for seg test-retest reliabilitet av impingementtester i skulder, mens flere studier er blitt utført med tanke på intertester-reliabilitet. De Wilde et al. (2003) rapporterte i sin studie om gode reliabilitetstall både med tanke på test- retest og intertester reliabilitet. I studien ble en modifisert Hawkins-Kennedy impingement test benyttet. Fem testere utførte tre målinger hver på skuldre hos fem friske individer, og det ble brukt en liggende utgangsstilling. De oppnådde svært godt samsvar mellom test og retest, og ved intertester reliabilitet på henholdsvis (ICC

0,97 og 0,93). De fem testerne hadde i forkant trent på og samkjørt prosedyren ved testingen. De Wilde et al. undersøkte i tillegg til smerteprovokasjon ved testene også grad av subacromiale plassforhold. De konkluderer med at Hawkins-Kennedy test er en god test for å tallfeste subacromiale plassforhold. Studien viser også at det i testens mest provoserende stilling skapes en benet kontakt mellom tuberositas major på humerus og anterolaterale del av acromion, hvilket også underbygges i studien til Beaulieu et al. (2006).

Johannson et al. (2009) oppnådde svært gode resultater i sin studie der fire skuldertester ble reliabilitetstestet både med tanke på test- retest og intertester-reliabilitet. Resultatet viste en perfekt kappa enighet ved test og retest. Studien lar seg godt sammenligne med vår studie. Dette i motsetning til studien til De Wilde et al. (2003) der utvalget kun var friske pasienter og utgangsstillingen samt utførelse var annerledes. Johannson et al. (2009) tok for seg tre av de fire testene vi undersøkte i vår studie (Neers, Hawkins og Jobe). De utførte testene likt, hadde like mange deltakere og de hadde en uke mellom test og retest. Det som skiller studien mest fra vår er at de kun testet pasienter med skulderplager. De hadde ingen kontrollgruppe med friske (og ingen gullstandard for diagnostisering). I denne studien ble det brukt mye tid på standardisering og trening i forkant, i motsetning til i vår studie. Det kan derfor argumenteres for at vår oppgave er mer lik normal klinisk hverdag, da den i større grad ivaretar den økologiske validiteten beskrevet av Kruuse (2007) i metodedelene. En omfattende standardisering i forkant av studien ville kunnet bedret resultatene. På den annen side anvendes testene vanligvis i daglig praksis av terapeuter som ikke har øvd sammen i forkant, og som kun har testbeskrivelsen og erfaring å gå ut i fra.

En annen faktor ved studiene som er interessant å diskutere er valg av tid/pause mellom testene og mellom testere. Vi brukte i vår studie to minutter mellom hver test og tre minutter mellom tester 1 og tester 2. Johannson et al. (2009) brukte en time mellom tester 1 og tester 2, og de brukte Visuell Analog Smerteskala (VAS) mellom hver test. De vil si at pasienten alltid skulle være tilbake til samme skåre på VAS før neste test ble utført. Dette sikrer i større grad deres studie mot falsk positive skåringer i forhold til vår studie. En av årsakene til at det ble skåret 6 flere positive tester ved første test i vår studie kan ha vært at pausene ble for korte og testbelastningen med 8 tester dermed ble for stor. Ved re-testen en uke senere var det bare jeg som testet og kun fire tester som ble utført.

I studien til Johannson et al. (2009) kan en lese at inklusjonskriteriet med tanke på alder var 18-50 år. Et så ungt utvalg reduserer faren for komorbiditet. Utvalget er yngre enn samtlige

andre studier oppsummert i denne oppgaven og kan delvis være med på å forklare denne studiens gode resultater. Vi hadde til sammenligning inkludert deltakere mellom 18-60 år og hadde snittalder på henholdsvis 46 år (impingement), 44 år (annen skulderpatologi) og 40 år (friske).

Vår studie skiller oss fra de to andre test-retest studiene med tanke på testerne. I studien til De Wilde (2003) var det fem erfarne terapeuter som gjennomførte testene. I studien til Johannson et al. (2009), var det en erfaren og en relativt nyutdannet terapeut som utførte testene. I vår studie har vi vært to relativt ferske terapeuter med omtrent 5 års klinisk bakgrunn. Det kan derfor argumenteres for at studien til De Wilde (2003) med bare erfarne terapeuter har bedre forutsetninger for godt samsvar grunnet lenger klinisk praksis. God og lang klinisk erfaring øker også sjansene for at testene er bedre innøvd og at man klarer å utføre dem likt ved test og retest. Jeg som relativt uerfaren terapeut har ikke samme kliniske erfaring og trygghet.

## 5.2 Diskusjon av testene og utførelse

I metoddelen ble det gitt en nøye beskrivelse av testene med tekst og bilder. Personlig opplevde jeg at testene Neers og Jobe var de enkleste testene både å utføre og å skåre. Det var også disse testene som endte opp med best  $\kappa$ -verdier på 0,94. Impingement-testen Hawkins-Kennedy var også en grei test å gjennomføre. Skåringen var dog noe vanskeligere her da flere av deltakerne var usikre på om de opplevde økt smerte eller bare ubehag grunnet strekk i ytterposisjon. Da vi i vår standardisering kun valgte å teste en gang ble noen av skåringene beheftet med usikkerhet. Johannson et al. (2009) gjorde hver test to ganger for å få et sikrere svar på hver test, det er mulig vi også burde gjort det.

Painful arc som er en aktiv test var den vanskeligste testen å skåre. Ved utførelse opplevde mange deltakere på denne testen litt mer diffust ubehag og liten økning i smerte. Utfordringen ble da om jeg skulle gi testen en positiv skåre, eller vurdere det dit hen at ubehaget eksempelvis var knyttet til strekksmerter som skulle skåres negativt. Et annet vesentlig moment med denne testen er at den skal være positiv innenfor intervallet 60-120 graders abduksjon. Smerter over eller under dette intervallet skulle registreres negativt. I en testsituasjon der lite øving og standardisering har foregått i forkant kan man glemme å spørre eksakt hvor eventuell smerte oppsto. Dermed vil feil skåring av testen kunne oppstå.



Skåringer, spesielt på denne testen, ble i noen tilfeller basert på «magefølelsen». Det er da også denne testen som kommer dårligst ut i vår studie. Grunnet liten standardisering i forkant kan jeg ha vurdert situasjonen for pasienten ulikt fra gang til gang, og samtidig blitt bedre til å utføre og skåre testen i løpet av studien.

Samtlige fire tester er gode til å skille syk fra frisk. Kun 3 av totalt 128 tester i friskgruppen endte opp med falsk positiv skår. Dette blir grundig belyst i oppgavedelen til min kollega som undersøkte testenes diagnostiske nøyaktighet. Verdt å merke seg er da også at alle de positive skåringen kommer fra samme pasient, nr 27. Der kom det frem under retesten at hun i lengre tid hadde vært plaget med motsatt sides skulder og at hun nå begynte å få symptomer i aktuell skulder.

Hun skåret ved retesten positivt på to av fire tester mot en av fire tester første testrunde. Den første uken skåret hun også positivt på den siste testen, altså den 8, hvilket kan være et resultat av for liten pause mellom testene og en lett begynnende irritasjon i skulder som gav smerter ved den siste testen.

### **5.3 Validitet**

Med validitet menes kort sagt om en test måler det som den er ment til å måle (Shadish, Cook & Campbell, 2002) I den videre diskusjonen er det viktig å se på forhold som kan øke eller minske studiens interne og eksterne validitet. Validitet er et uttrykk for gyldigheten i resultatene og slutningene som trekkes på grunnlag av dette. Campbell et al. (2002) har utviklet et validitetssystem som skisserer hvordan validiteten kan vurderes i forhold til fire typer slutninger; statistiske slutninger (statistisk validitet), kausale slutninger (indre validitet), begrepslutninger (begrepsvaliditet) og generaliseringer (ytre validitet). Til hver av disse slutningene inngår flere mulige feilkilder eller trusler mot validiteten (Shadish, Cook & Campbell, 2002; Lund, 2002). Med utgangspunkt i dette systemet vil jeg drøfte aktuelle trusler mot de to underpunktene intern og ekstern validiteten som er mest relevant for dette studiedesignet.

#### **5.3.1 Intern validitet:**

Intern validitet har å gjøre med i hvilken grad vi kan si at det eksisterer et kausalitetsforhold. Dvs. i hvilken grad vi kan konkludere med at en effekt/resultat kan tilskrives den årsaken vi tror, eller om det kan være andre utenforliggende faktorer som er like sannsynlig (Jacobsen

2005). Viktige spørsmål ved intern validitet i dette studiet er om jeg kan stole på resultatet, les reliabilitets og kappaverdiene og hvilke trusler som påvirker disse verdiene.

I min del av studien var det en tester som gjorde alle målingene. I studier med en tester vil det alltid kunne stilles spørsmål ved i hvilken grad det er forhold ved tester som innvirker på studiens måleresultater. Et av hovedpunktet som dukker opp med tanke på test-retest reliabilitet er om endringene i resultat skyldes ulik utførelse av testene eller symptomendring hos pasienten fra test til retest.

Tidsaspektet med en uke mellom test og retest og mulig endring av symptombildet er den største feilkilden i min del av studien. For min kollega som ser på intertesterreliabilitet er det derimot ulik utførelse av testene mellom testerne og eventuell skåring som er den største feilkilden, da pasientene testes samme dag. Det er ikke utenkelig at symptomene i skulderen endrer seg i løpet av en uke, og at dette kan innvirke på resultatene. I den forbindelse er det en svakhet med studien er at vi ikke har kartlagt medisinbruk. Pasientene i diagnosegruppene skulle ikke gå i behandling under testperioden, men de kan ha gått på smertestillende, og verre, de kan ha tatt smertestillende ved test og ikke ved retest, eller motsatt. Denne faktoren kan dermed ha bidratt til å svekke resultatene.

Med intern validitet menes videre om en har dekning for de slutningene en trekker (Jacobsen 2005). Lund (2002) beskriver at det er færre feilkilder ved standardiserte fysiske tester som i vår studie enn ved intervju og spørreskjemaer som åpner for flere mulige svar/fortolkninger og medfølgende feilkilder. Likevel er det flere andre elementer som kan påvirke den interne validiteten i denne studien, blant annet grad av standardisering av tester og utførelse mellom terapeuter. For at den interne validiteten skal bli så stor som overhodet mulig er det viktig at tester og utførelse standardiseres med en protokoll så det blir utført likt hver gang og fra pasient til pasient (Jacobsen 2005). Vi har i studien standardisert testene på bakgrunn av teori og beskrivelser som foreligger. Vi valgte dog ikke å trene spesifikt sammen på testene før oppstart av studien, da dette ikke ville gi en klinikk-lik standardisering. Det kan dermed argumenteres for at den interne validiteten svekkes noe. På den annen side styrker det studien, fordi det øker sjansene for at andre terapeuter i en klinisk hverdag ville fått de samme resultatene.

Kjemi eller forholdet mellom terapeut og pasient, eventuell læringseffekt, eventuelle frafall, samt tidspunkt og sted hvor test foretas er også trusler mot den indre validiteten.

Vi ønsket å gjøre testsituasjonene mest mulig like for å sikre en så høy intern validitet som mulig, blant annet ved at testene foregikk like etter hverandre i tid, i de samme omgivelser, samt at et minimum av standardisering har funnet sted mellom terapeuter. Her kan en dog

argumentere for at dette kan ha svekket den interne validiteten da to testserier tett opp mot hverandre kan føre til at læringseffekten blir større. For å møte denne utfordringen endret vi alltid testrekkefølgen ved hjelp av loddrekning, samt at vi vekslet på hvem som testet først og sist.

Ulik kjemi mellom terapeut og ulikhet mellom testere diskuteres ikke her da det er viktigere for intertester-reliabiliteten. Et faktum som kan ha påvirket mine skåringer under de to testene er deltakernes og egen læringseffekt. Selv om testene i min del av studien ble utført med en ukes mellomrom kan deltakerne huske noen av testene og hvilke svar de gav. Ved retesten var det tredje gang deltakerne ble testet og en viss læringseffekt kan bidra til kunstig gode resultater. Spesielt er det lett for de som skåret enten negativt eller positivt på alle testene hver gang å gi samme svar. For egen del kan jeg også bevisst eller ubevisst skjønne hvilket utvalg deltakeren kom fra og dermed vært ”snillere” eller ”tøffere” i testutførelsen. Grunnet noe kort klinisk erfaring kan jeg også underveis i testingen ha blitt bedre til å utføre testene, og dermed flinkere til å score dem riktig, hvilket kan ha ført til endring fra test til retest.

### **5.3.2 Ekstern validitet:**

Ekstern validitet betegner muligheten for å generalisere resultatene av en test (Lund, 2002). En undersøkelse har god ekstern validitet i den grad den gjør det mulig å foreta ikke-statistiske generaliseringer til eller over relevante individer, situasjoner og tider med rimelig sikkerhet (Ringdal 2009). Trusler mot ekstern validitet er systematiske forhold som vanskeliggjør slike slutninger (Lund 2002).

Vår største trussel i denne studien var om utvalgene våre var store nok til å trekke generaliserte slutninger. Hadde vi et stort nok utvalg og hadde vi et representativ utvalg? I metoden blir det belyst at slike studier er avhengige av minimum 30 deltakere, men helst 50 for å kunne generalisere resultatene (Straus et al 2005). Vi befinner oss over minimumsgrensen med 34 deltakere men ser jo helt klart at et større utvalg ville kunne gitt oss mer pålitelige resultater. Spesielt er det en stor svakhet at vi bare fikk 4 pasienter i gruppen med annen skulderpatologi. Generaliseringer i denne gruppen er umulig. Likevel har den vært viktig å ha med gruppen med annen skulderpatologi da den gjorde blindingen av deltakerne bedre. Testpersonene visste ikke i forkant hvor mange deltakere som befant seg i hvert utvalg. Det var ønskelig å inkludere både pasienter med SIS og andre skulderlidelser for at datamaterialet skulle ligne det en terapeut møter i sin kliniske praksis. Målet var å oppnå en stor andel som skåret både positivt og negativt på testene, som også var en forutsetning for å

undersøke reliabilitet med kappa statistikk. I tillegg ble det inkludert skulderfriske personer, først og fremst for å undersøke diagnostiske egenskaper. Det kan argumenteres for at pasientene som var diagnostisert med SIS og de skulderfriske kunne være enklere å teste og få klarere svar fra enn om man inkluderte en uselektert gruppe av pasienter med skulderplager. Dette kan ha bidratt til kunstig gode resultater.

Videre er det interessant å diskutere om resultatene kan være gjeldende for resten av befolkningen, og om det var stor nok spredning i alder og kjønn?

Om resultatene er gjeldende for resten av befolkningen er selvsagt svært interessant å diskutere. I og med at det kun er snakk om fire fysiske tester, og testene er standardisert etter originalutførelse, er overføringsverdien til resten av befolkningen god. Dette støttes av Lund (2002) som beskriver at standardiserte fysiske tester har god overføringsverdi. Det er her ikke snakk om å fange opp meningsforskjeller eller holdninger som ved eksempelvis spørreskjema, der ulike landsdeler kan ha forskjellige meninger (ibid). De anatomiske forhold i skulderen viser i følge litteraturen ikke geografiske forskjeller og en har derfor god dekning for å si at utfallet av testene vil være likt i ulike deler av landet. Et annet viktig spørsmål i forhold til studien og ytre validitet er om andre testere ville få de samme resultatene som jeg. På bakgrunn av relativt liten standardisering og øving i forkant mener jeg at sannsynligheten for det økes. Vi har prøvd å bevare den økologiske validiteten ved ikke å trene i forkant. Dette nettopp for å gjøre testsituasjonen så klinikk-lik som mulig.

Når det gjelder snittalder i utvalgene ønsket vi at den skulle være så lik som mulig. Vi anså stor spredning i alder som en større feilkilde enn kjønn og hvilken sides skulder som ble testet. Snittalderen i de tre gruppene var 46 år (impingement), 44 år (annen skulderpatologi) og 40 år (friske). At gruppene ble såpass homogene med tanke på alder er vi godt fornøyd med og det styrker den ytre validiteten. Kjønnsspredning var relativt lik med 19 kvinner og 15 menn. Det var dog noe sprik mellom gruppene. Aktuell side testet var også relativt likt med 19 høyre og 15 venstre, selv om det også her var litt sprik innad i gruppene.

God ekstern validitet er sentralt, da resultatene skal kunne generaliseres. En tilfredsstillende intern validitet er irrelevant hvis resultatene ikke kan generaliseres til en populasjon. Derved vil forskningsproblemet bli tilsvarende dårlig belyst og oppgaven som helhet bli svekket (Lund, 2002). Oppgavens omfang og tidsramme gjør at studiens største svakhet er utvalgsstørrelsen som igjen direkte påvirker den eksterne validiteten.

#### 5.4 Styrker og svakheter ved studien

Jeg har gjennom diskusjonen prøvd å belyse styrker og svakheter med studien. Under dette avsnittet vil disse punktene raskt bli oppsummert uten noen større diskusjon.

##### **Først studiens styrker:**

En av de største styrkene med denne studien sammenlignet med andre studier er at vi sikret at en andel av deltagerne hadde diagnosen SIS og en andel ikke hadde diagnosen gjennom de tre utvalgene, og at de to testpersonene var blindet for hvilken gruppe deltakerne tilhørte.

Selve rekrutteringen av både dem med skulderpatologi og friske skjedde av andre enn testpersonene, hvilket styrker studiens validitet. Det er også en styrke for studien at alderssammensetningen i de tre utvalgene er omtrent lik.

Samarbeidet med en erfaren lege som har spesialisert seg på ultralydundersøkelse, og jobber på en stor ved skulderklinikk er en annen styrke ved studien. Han rekrutterte og diagnostiserte pasientene med skulderpatologi opp mot gullstandarden ultralyd. Dette løfter studien som helhet og gir også mulighet til å undersøke testenes diagnostiske nøyaktighet.

Det kan også nevnes som en styrke at det ikke ble øvd mye på testene og testprosedyrene i forkant av studien. Studien har derfor en større overførbarhet til klinisk hverdag enn mange lignende studier.

##### **Studiens svakheter:**

Studiens største svakhet er antall deltakere. Vi hadde et mål om å komme opp på 50 deltakere men klarte innenfor rammene og tidsfristene bare å rekruttere 34 deltakere. Dette er over minimumskravene for å kunne bruke kappastatistikk og trekke generaliserte slutninger, men i minste laget til å kunne stole helt på resultatene. Flere deltakere ville styrket studiens validitet. Det er også en klar svakhet at vi bare fikk fire deltakere i gruppen med annen skulderpatologi. Spesielt er dette negativt for min kollega som skulle se på diagnostisk nøyaktighet da det er umulig å trekke slutninger fra et så lite utvalg.

Vi burde helt klart også ha kartlagt deltakernes eventuelle bruk av medisiner eller smertestillende, og gitt beskjed om at de skulle bruke samme dosering ved begge testtidspunkt. Endring i medisinbruk eller smertestillende fra test til retest kan ha svekket resultatene.

Grad av standardisering kan også nevnes med negativt fortegn. Vi har fra lignende studier sett svært gode reliabilitetstall, etter at testpersonene hadde øvd mye i forkant av sine studier. Det

er derfor mulig at jeg kunne fått bedre resultater med mer øving og strengere standardisering i forkant. Lenger erfaring som fagpersoner kunne kanskje også ha bedret resultater.

Valg av pausetid mellom testene og testere kan ha vært for kort, og muligens ført til falske positive testskåringer.

### **5.5 Videre forskning**

På grunn av tidsfrister for å bli ferdige med masterenoppgaven måtte vi sette grense ved 34 deltakere i studien. Utover høsten-12 vil vi fortsette samarbeidet med rekruttering av deltakere slik at forskningsprosjektet som helhet blir styrket. Vi har et mål å rekruttere minimum 50 deltakere fordelt på de tre utvalgene, dette i tråd med teori belyst i oppgaven (Straus et al., 2005, Carter et al., 2011). Målet med det videre forskningsarbeidet er å skrive en artikkel og få publisert studien. Det er tidligere lite forskning på test-retest reliabilitet av skuldertester og derfor stort behov for metodisk sterke studier på dette området.

Manuellterapeuter har i disse dager fått godkjent kurs innen injeksjonsteknikk som del av utdannelsen. Det vil medføre at manuellterapeuter de nærmeste årene kan sette smerteblokade og kortisonsprøyter. Det ville i den forbindelse være veldig interessant å gjennomføre en effektstudie der en blant annet bruker de testene som kommer best ut i denne studien, og utfører tester før og etter injeksjonsbehandling. En vil da kunne få gode svar på om injeksjonsbehandlingen har virket.

## 6.0 KONKLUSJON

Test retest reliabiliteten på de fire anvendte testene i denne studien var god til meget god. Best resultat ble oppnådd ved testene Jobe supraspinatustest og Neers tegn. De endte opp med en kappaverdi på 0,93-0,94 og prosentmessig enighet på 97 %. Hawkins test ble skåret med en  $\kappa$ -verdi på 0,82 og 91,8 % enighet, mens Painful arc kom dårligst ut med en  $\kappa$ -verdi på 0,64 og prosentmessig enighet på 82,35 %. Tre av fire tester i denne studien viste dermed nesten perfekt samsvar mellom test og retest med  $\kappa$ -verdier over 0,8, mens Painful arc viste godt samsvar med  $\kappa$ -verdi  $> 0,6$ .

## REFERANSELISTE

1. Altman DG. Practical statistics for medical research. London: Chapman & Hall, 1997.
2. Awerbuch MS. The clinical utility of ultrasonography for rotator cuff disease: shoulder impingement syndrome and subacromial bursitis. *Med J Aust* 2008;188(1):50-3.
3. Bahr R, Mæhlum S. Idrettsskader. Oslo: Gazette bok, 2003.
4. Beaudreuil, J., Nizard, R., Thomas, T., Peyre, M., Liotard, J.P., Boileau, P., Marc, T., Dromard, C., Steyer, E., Bardin, T., Orcel, P., Walch, G.: Contribution of clinical tests to the diagnosis of rotator cuff disease: a systematic literature review. *Joint Bone Spine* 2009;76(1):15-9.
5. Beaulieu CF, Pappas GP, Blemker SS, McAdams TR, Whalen ST, In vivo anatomy of the Neer and Hawkins sign positions for shoulder impingement. Medical Center, Stanford University School of Medicine, 2006.
6. Benestad HB, & Laake P. *Forskningsmetode i medisin og biofag*. Gyldendal Akademisk. 2004.
7. Birkler J, *Videnskabsteori – en grundbog*. København: Munksgaard Danmark. 2005.
8. Buchbinder, R., Green, S., Hetrick S. E., Systematic Review. Physiotherapy interventions for shoulder pain. Editorial Group: Cochrane Musculoskeletal Group, 2008.
9. Brodal P, Sentralnervesystemet. Universitetsforlaget, 2007.
10. Brukner P., Khan, K.: Clinical sports medicine. McGrawHill, Sydney 2009.
11. Cadogan A., Laslett, M., Hing, W.A., McNair, P.J., Coates, M.H.: A prospective study of shoulder pain in primary care: prevalence of imaged pathology and response to guided diagnostic blocks. *BMC Musculoskeletal Disorders* 2011, 12:119.
12. Cadogan A., Laslett M., Hing W., McNair P., Williams M.: Interexaminer reliability of



orthopaedic special tests used in assessment of shoulder pain. *Man Ther* 2011 Apr; 16(2), 131-5.

13. Calis M., Akgun K., Calis H., Tuzun F.: Diagnostic values of clinical diagnostic tests in subacromial impingement syndrome. *Am Rheum Dis* 2000; 59; 44-47.

14. Carter R.E., Lubinsky J., Domholdt E.: *Rehabilitation Research. Principles and applications.* Elsevier Saunders, St.Louis 2011.

15. Dahl H.A, Rinvik E. *Menneskets funksjonelle anatomi: med hovedvekt på bevegelsesapparatet.* Cappelen akademiske, 2010.

16. Dalland O, *Metode og oppgaveskriving for studenter,* Gyldendal akademisk, 4. utg. 2007.

17. De Wilde L, Plasschaert F, Berghs B, Van Hoecke M, Verstaete K, Verdonk R, Quantified measurement of subacromial impingement. *Journal of Shoulder and Elbow Surgery*, 12 (2003), 346–349.

18. Domholdt E, *Physical Therapy Research: Principles and Applications* Publisher Saunders (W.B.) Co Ltd 2000.

19. Friis R.H., Sellers T.A.: *Epidemiology for public health practice.* Jones and Bartlett, Boston 2009.

20. Green S., Buchbinder R., Forbes A., *Systematic review Interventions for shoulder pain* Editorial Group: Cochrane Musculoskeletal Group, 2006.

21. Gross J., Fetto J., Rosen E.: *Musculoskeletal examination.* Blackwell Science, Massachusetts 1996.

22. Hattam P., Smeatham A.: *Special tests in musculoskeletal examination.* Churchill Livingstone 2010.

23. Halvorsen K.: *Å forske på samfunnet: En innføring i samfunnsvitenskapelig metode.*

Cappelen akademisk forlag Oslo 2008.

24. Halvorsen K.: Forskningsmetode for helse- og sosialfag. Cappelen akademisk forlag, Oslo 1996.

25. Hegedus E.J., Goode A., Campbell S., Morin A., Tamaddoni M., Moorman C.T., Cook C.: Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests. Br J Sports Med 2008;42(2):80-92.

26. Holtby R, Razmjou H: Validity of the supraspinatus test as a single clinical test in diagnosing patients with rotator cuff pathology. J Orthop Sports Phys Ther 2004; 34: 194–200.

27. Hopkins W. G. A new view of statistics. Internet Society for Sport Science: <http://www.sportsci.org/resource/stats/> 2000.

28. Hughes P.C., Taylor N.F., Green R.A.: most clinical tests cannot accurately diagnose rotator cuff pathology: a systematic review. Aust J Physiother 2008;54(3):159-70.

29. Itoi E, Minagawa H, Yamamoto N, Seki N, Abe H. Are pain location and physical examinations useful in locating a tear site of the rotator cuff? Am J Sports Med 2006; 34: 256–64.

30. Jacobsen D.I.: Hvordan gjennomføre undersøkelser? Høyskoleforlaget, Kristiansand, 2.utgave 2005.

31. Jamtvedt G., Hagen K.B., Bjørndal A.: Kunnskapsbasert fysioterapi. Metoder og arbeidsmåter. Gyldendal akademisk, Oslo 2003.

32. Johansson K., Ivarson S.: Intra- and interexaminer reliability of four manual therapy shoulder maneuvers used to identify subacromial pain. Man Ther. 2009 Apr; 14(2):231-9.

33. Juel N.G.: Norsk fysikalsk medisin, 2.utgave. Fagbokforlaget, Bergen 2007.

34. Kelly S.M., Brittle N., Allen G.M.: The value of physical tests for subacromial impingement syndrome: a study of diagnostic accuracy. *Clin Rehabil.* 2010 Feb; 24(2), 149-58.
35. Kessel L, Watson M. The painful arc syndrome. Clinical classification as a guide to management. *J Bone Joint Surg Br.* 1977 May;59(2):166-72.
36. Kunnskapssenteret. Ordliste med forklaringer, vedlegg 3, 2010. [www.kunnskapssenteret.no/verktøy](http://www.kunnskapssenteret.no/verktøy).
37. Kruuse E.: Kvantitative forskningsmetoder – i psykologi og tilgrænsende fag. Danske psykologisk forlag, 6 utgave 2006.
38. Landis J.R., Koch G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics.* 33, 159–174.
39. Lewis JS. Rotator cuff tendinopathy/subacromial impingement syndrome: is it time for a new method of assessment? *Br J Sports Med* 2009;43:259–64.
40. Lund T., Kleven T.A., Kvernbekk T., Christophersen K.A.: Innføring i forskningsmetodlogi. Akademika, Oslo 2002.
41. MacDonald P.B., Clark P., Sutherland K.: An analysis of the diagnostic accuracy of the Hawkins and Neer subacromial impingement signs. *J Shoulder Elbow Surg* 2000;9(4):299-301.
42. Magee D.J.: *Orthopedic physical assessment*, ed 5. Saunders Elsevier 2008.
43. Malhi A.M, Khan R.: Correlation between clinical diagnosis and arthroscopic findings of the shoulder. *Postgrad Med J* 2005;960(81):657-9.
44. Malterud K.: Kvalitative metoder i medisinsk forskning – forutsetninger, muligheter og begrensinger. *Tidsskr Nor Lægeforen* 2002;122:2468-72.

45. Michener L.A., Walsworth M.K., Doukas W.C., Murphy K.P.: Reliability and diagnostic accuracy of 5 physical examination tests and combination of tests for subacromial impingement. *Arch Phys Med Rehabil.* 2009 Nov; 90(11), 1898-903.
46. Meislin RJ, Sperling JW, Stitik TP. Persistent shoulder pain: epidemiology, pathophysiology, and diagnosis. *American Journal Orthopedics* (2005) 34:5-9.
47. Mitchell C., Adebajo A., Hay E., Carr A.: Shoulder pain: diagnosis and management in primary care. *BMJ* 2005;7525:1124-8.
48. Moen M.H., De Vos R.J., Ellenbecker T.S., Weir A.: Clinical tests in shoulder examination: how to perform them. *Br J Sports Med* 2010;44(5):370-5.
49. Nørregaard J., Krogsgaard M.R., Lorenzen T., Jensen E.M.: Diagnosing patients with longstanding shoulder joint pain. *Ann Rheum Dis* 2002;61:646-9.
50. Park H.B., Yokota A., Gill H.S., El Rassi G., McFarland E.G.: Diagnostic accuracy of clinical tests for the different degrees of subacromial impingement syndrome. *J Bone Joint Surg Am* 2005;87(7):1446-55.
51. Ringdal K.: Enhet og mangfold, Samfunnsvitenskapelig forskning og kvantitativ metode, 2. utgave. Kunnskapsforlaget, 2009.
52. Seeger L.L., Gold R.H., Bassett L.W., Ellman H.: Shoulder impingement syndrome: MR findings in 53 shoulders. *Am J Roentgenol* 1988; 150(2):343-7.
53. Silva L, Andreu JL, Munoz P, Pastrana M, Millan I, Sanz J, Barbadillo C, Fernandez-Castro M. Accuracy of physical examination in subacromial impingement syndrome. *Rheumatology* 2008; 47(5): 679-83.
54. Solberg A.S.: Klinisk undersøkelse av nakke-skulder. Høyskoleforlaget, Kristiansand

2002.

55. Straus S.E., Richardson W.S., Glasziou P., Haynes R.B.: Evidence-based medicine. How to practise and teach EBM. Elsevier Churchill Livingstone 2005.

56. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC., Quality criteria were proposed for measurement properties of health status questionnaires. VU University Medical Center, 2007 Jan;60(1):34-42.

57. Undheim J.O.: Innføring i statistikk og metode for samfunnsvitenskapelige fag. Universitetsforlaget, Oslo 1996.

58. Valadie A.L., Jobe C.M., Pink M.M., Ekman E.F., Jobe F.W.: Anatomy of provocative tests for impingement syndrome of the shoulder. J Shoulder Elbow Surg 1000;9(1):36-46.

59. Yamamoto N., Muraki T., Sperling J.W., Steinmann S.P., Itoi E., Cofield R.H., An K.N.: Impingement mechanisms of the Neer and Hawkins signs. J Shoulder Elbow Surg 2009;18:942-7.

60. Östör A.J.K., Richards C.A., Prevost A.T., Speed C.A., Hazleman B.L.: Diagnosis and relation to general health of shoulder disorders presenting to primary care. Rheumatology 2005; 44:800-5.

## Vedlegg 1:



<b>Region:</b> REK sør-øst	<b>Saksbehandler:</b> Øyvind Grønlie Olsen	<b>Telefon:</b> 22845522	<b>Vår dato:</b> 05.03.2012	<b>Vår referanse:</b> 2012/57/REK sør-øst D
			<b>Deres dato:</b> 17.01.2012	<b>Deres referanse:</b>

Vår referanse må oppgis ved alle henvendelser

Att: Liv Inger Strand

### 2012/57 D Reliabilitet og validitet av skuldertester

Vi viser til søknad om forhåndsgodkjenning av ovennevnte forskningsprosjekt. Søknaden ble behandlet av Regional komité for medisinsk og helsefaglig forskningsetikk i møtet 09.02.2012.

Prosjektleder: Liv Inger Strand  
Forskningsansvarlig: Universitetet i Bergen

#### Prosjektomtale

*Hensikten med studien er tredelt:*

1. Undersøke den diagnostiske validiteten ved fire kliniske tester som brukes for å stille en impingement-diagnose
2. Undersøke intertester-reliabiliteten ved de samme testene
3. Undersøke test-retest-reliabiliteten ved de samme testene.

*Skuldersmerter generelt er et omfattende problem med konsekvenser for individet og samfunnet. En stor andel av smertene skyldes avklemming av strukturer i øvre deler av skulderleddet, kalt impingement syndrom. For å stille diagnosen brukes i dag blant annet MR, ultralyd og røntgen i tillegg til en rekke kliniske tester. De første tre testene regnes som ganske sikre tester for å påvise skulder impingement. Det er ønskelig å kunne påvise tilstanden ved hjelp av enkle diagnostiske tester i vanlig klinisk praksis. Dette fordrer at testene er diagnostisk valide og at de har høy reliabilitet.*

*Testene som skal utføres er Hawkins test, Neers test, painful arc og Jobs test. Testene blir scoret med utgangspunkt i dikotomien positiv eller negativ test. For Hawkins, Neers og painful arch, betyr dette at testen scores som positiv hvis den primært gir smerter i skulderområdet.*

*Totalt deltar 50 pasienter med skulderplager. Halve gruppen har fått indikasjon på skulderimpingement ved hjelp av ultralydundersøkelse mens den andre halvparten har fått indikasjon på at skulderplagene ikke skyldes impingement.*

#### Vurdering

Komiteen anser prosjektet som en prospektiv sammenligning av etablerte metoder. Prosjektets formål er etter komiteens mening mer enn en ren evaluering av praksis. En effektivisering av behandlingsskjeden kan føre til at pasienter blir gitt riktig diagnose på raskere tid – noe som kan føre til at riktig behandling til vedkommende pasient gis på et tidligere tidspunkt. Forskningsprosjektet er etter dette fremleggelsespliktig, jf. helseforskningsloven § 10.

**Besøksadresse:**  
Gullhaug torg 4A,  
Nydalene, 0484 Oslo

**Telefon:** 22845511  
**E-post:**  
post@helseforskning.etikkom.no

**Web:**  
www.etikkom.no

All post og e-post som inngår i saksbehandlingen, bes adressert til REK sør-øst og ikke til enkelte personer

Kindly address all mail and e-mails to the Regional Ethics Committee, REK sør-øst, not to individual staff

Informasjonssikkerhet

Komiteen kan ikke se at det er oppgitt databehandlingsansvarlig for studien. Det forutsettes at Universitetet i Bergen er databehandlingsansvarlig for prosjektet.

Informasjonsskriv

I informasjonsskrivet benyttes det til dels vanskelige medisinske ord og uttrykk. I den grad det er mulig skal innholdet i informasjonsskrivet presenteres på en lettfattelig måte. Medisinske ord og uttrykk som må erstattes med alminnelige ord og uttrykk.

Det er et uttalt krav at presentasjonen av forskningsprosjektets innhold skal være nøytralt. Det innsendte informasjonsskrivet inneholder formuleringer som ikke tilfredsstillende dette kravet. Før rekruttering til forskningsprosjektet gjennomføres forutsetter komiteen at informasjonsskrivet bearbejdes slik at opplysninger om prosjektets innhold presenteres i nøytral form. Revidert informasjonsskriv ettersendes komiteen til orientering.

**Vedtak**

Komiteen har vurdert søknaden og godkjenner prosjektet med hjemmel i helseforskningsloven § 9 jf. § 33. Godkjenningen er gitt under forutsetning at de vilkår som er stilt innarbejdes i protokoll og informasjonsskriv.

Godkjenningen av prosjektet gjelder til 31.12.2014. Av dokumentasjonshensyn skal opplysningene likevel bevares inntil 31.12.2019. Opplysningene skal deretter slettes eller anonymiseres, senest innen 1.6.2020.

Opplysningene skal lagres aidentifisert, det vil si adskilt i en nøkkel- og en opplysningsfil.

Forskningsprosjektets data skal oppbevares forsvarlig, se personopplysningsforskriften kapittel 2, og Helsedirektoratets veileder for «Personvern og informasjonssikkerhet i forskningsprosjekter innenfor helse- og omsorgssektoren».

Prosjektet skal sende sluttmelding til REK sør-øst D på fastsatt skjema senest 1.6.2015.

Tillatelsen er gitt under forutsetning av at prosjektet gjennomføres slik det er beskrevet i søknaden og protokollen, og de bestemmelser som følger av helseforskningsloven med forskrifter.

Dersom det skal gjøres vesentlige endringer i prosjektet i forhold til de opplysninger som er gitt i søknaden, må prosjektleder sende endringsmelding til REK.

Komiteens vedtak kan påklages til Den nasjonale forskningsetiske komité for medisin og helsefag, jfr. helseforskningsloven § 10, 3 ledd og forvaltningsloven § 28. En eventuell klage sendes til REK Sørøst D. Klagefristen er tre uker fra mottak av dette brevet, jfr. forvaltningsloven § 29.

Vennligst oppgi vårt referansenummer i korrespondansen.

Med vennlig hilsen,

Stein A. Evensen (sign.)  
dr. med.  
leder

Øyvind Grønlie Olsen  
saksbehandler

**Vedlegg 2:**

Gruppe/Diagnose	Deltager	Kjønn	Alder	Sykdomsvarighet i måneder	Testside
Gruppe 1/Impingement	1	K	49	12	H
Gruppe 1/Impingement	2	M	55	1	V
Gruppe 1/Impingement	3	M	60	9	V
Gruppe 1/Impingement	13	K	28	9	V
Gruppe 1/Impingement	15	K	46	360	V
Gruppe 1/Impingement	17	K	53	12	H
Gruppe 1/Impingement	19	M	44	6	V
Gruppe 1/Impingement	20	M	60	18	V
Gruppe 1/Impingement	21	M	57	12	V
Gruppe 1/Impingement	22	M	60	12	V
Gruppe 1/Impingement	24	K	33	48	V
Gruppe 1/Impingement	25	M	18	1	H
Gruppe 1/impingement	33	K	42	12	V
Gruppe 2/Annen skulderpatologi	12	K	33	24	H
Gruppe 2/Annen skulderpatologi	23	M	42	9	H
Gruppe 2/Annen skulderpatologi	26	M	45	12	H
Gruppe 2/Annen skulderpatologi	28	K	55	19	H
Gruppe 3/friske	4	K	31	-	H



Gruppe 3/friske	5	M	35	-	V
Gruppe 3/friske	6	K	30	-	H
Gruppe 3/friske	7	M	30	-	V
Gruppe 3/friske	8	K	21	-	H
Gruppe 3/friske	9	M	60	-	H
Gruppe 3/friske	10	K	22	-	H
Gruppe 3/friske	11	K	23	-	V
Gruppe 3/friske	14	K	59	-	H
Gruppe 3/friske	16	K	30	-	H
Gruppe 3/friske	18	K	23	-	V
Gruppe 3/friske	27	K	46	-	H
Gruppe 3/friske	29	M	60	-	H
Gruppe 3/friske	30	M	55	-	H
Gruppe 3/friske	31	K	55	-	V
Gruppe 3/friske	32	K	60	-	H

## Vedlegg 3:

### Forespørsel om deltakelse i forskningsprosjektet

### *”Gyldighet av skuldertester”*

#### **Bakgrunn og hensikt**

Mellom 30-40 prosent av befolkningen opplever skuldersmerter hvert år. Vanligst er plager knyttet opp mot musklene rundt skulderen og avklemminger lokalt. Nedsatt plassforhold av slimpose og/eller muskulatur i skulderen (subacromialt impingment) er den vanligste skulderlidelsen i befolkningen, og representerer 44-65% av alle registrerte skulderplager i kliniske praksis.

Dette er et spørsmål til deg om å delta i en studie for å undersøke gyldigheten av tester som vi bruker for å diagnostisere smertetilstander i skulder. Vi ønsker å undersøke om testene våre er nøyaktige nok.

#### **Hva innebærer studien?**

Dersom du takker ja til å delta i studien, vil du bli kalt inn til selve testprosedyren 2 ganger med en ukes mellomrom. Det vil da bli gjennomført 4 enkle kliniske tester av to ulike terapeuter første gang, og av kun en terapeut den ene andre gangen. Testene som utføres brukes for å se om du har det gitte skulderproblemet eller ikke, og er helt ufarlige. Det vil ta ca. 20 minutter hver gang.

#### **Mulige fordeler og ulemper**

Fordeler:

Det vil bli gitt informasjon, råd og behandlingsforslag til deg som deltar.

Ulemper:

Du må komme til testing to ganger.

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

Prøvene tatt av deg og informasjonen som registreres om deg skal kun brukes slik som beskrevet i hensikten med studien. Alle opplysningene og prøvene vil bli behandlet uten navn og fødselsnummer eller andre direkte gjenkjennende opplysninger. En kode knytter deg til dine opplysninger og prøver gjennom en navneliste.

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 .....

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

**Ytterligere informasjon om biobank, 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**

### **Kriterer for deltagelse:**

Det vil bli rekrutert pasienter i 2 forskjellige utvalg. Et utvalg der ultralydundersøkelse har vist nedsatt plassforhold i skulderen (positiv impingement) og et utvalg der skulderplagene skyldes andre forhold.

For å være med i gruppen med nedsatt plassforhold i skulderen gjelder følgende kriterier:

Diagnostisert med nedsatt plassforhold i skulderen ved ultralyd

Skuldersmerter med varighet mer enn én måned

Skuldersmertene må være pasientens aktuelle hovedproblem

Må kunne gi et informert samtykke

Mellom 18 og 60 år

Du kan ikke inngå i gruppen med redusert plassforhold i skulderen ved:

Andre kjente plager i skulder

Tidligere skade mot skulder

Tidligere skulderoperert

Betennelsesdannende leddlidelser

Smerter ved bevegelse av nakken

Smerter i nakke med utstrålende plager ut til fingrene

Mistenke om alvorlig sykdom

For å være med i gruppen for andre skulderplager, gjelder følgende kriterier:

Diagnostisert med andre skulderplager ved hjelp av ultralyd

Skuldersmerter med varighet mer enn én måned

Skuldersmertene må være pasientens aktuelle hovedproblem

Må kunne gi et informert samtykke

Mellom 18 og 60 år

Du kan ikke inngå i gruppen med andre skulderplager ved:

Andre kjente plager i skulder

Tidligere skade mot skulder

Tidligere skulderoperert

Betennelsesdannende leddlidelser

Smerter ved bevegelse av nakken

Smerter i nakke med utstrålende plager ut til fingrene

Mistenke om alvorlig sykdom

### **Bakgrunnsinformasjon om studien:**

Vi møter i vår hverdag mange pasienter med skulderproblemer. For å skille de ulike skulderproblemer fra hverandre bruker vi ulike tester for å kunne stille riktig diagnose. Det er derfor svært sentralt at testene vi bruker er pålitelige og diagnostisk nøyaktige. Hensikten med denne studien er nettopp å undersøke kvaliteten på fire tester brukt for å diagnostisere nedsatt plassforhold av slimpose og/eller muskulatur i skulderen.

### **Tidsskjema:**

Rekrutteringen av pasienter til mastergradstudiet vil skje på .....

Datainnsamlingen vil foregå på ..... Det vil bli gjort målinger på testene i skulderen, og det vil registreres om den gitte testen fremprovoserer smerte i skulderen eller ikke.

Datainnsamlingene vil foregå vår/høst 2012 og bearbeidelse høst 2012. Tidsfristen for ferdig prosjekt er 1. november 2012. Studiet vil bli gjennomført som en test- retest analyse. De kliniske testene som vil gjennomføres er Neers test, Hawkins-Kennedy test, Jobe/empty can-test og painful arc.

**Mulige fordeler:**

Du vil delta i et prosjekt som kan bedre kvaliteten på tester som brukes ved diagnostisering av skulderproblemer. Du vil også få informasjon og råd vedrørende dine skulderplager.

**Mulige bivirkninger:**

Det er ingen bivirkninger da det ikke gis medikamenter eller tas blodprøver etc.

**Mulige ubehag:**

Testene kan gi lett forbigående ubehag i skulder. Det legges derfor opp til pauser mellom hver test slik at eventuelle ubehag er borte før utførelse av hver test.

**Pasientens/studiedeltakerens ansvar:**

Eneste krav til deg som deltager i studiet er at du møter opp til avtalt tid 2 ganger med en ukes mellomrom. Ved ønske om eventuell videre behandling avtales dette på klinikken.

**Kompensasjon/utgifter:**

Eventuell kompensasjon til og dekning av utgifter for deltakere gis ikke. Informasjon, råd og treningsprogram med øvelser er gratis. Eventuell videre behandling vil følge vanlig takstsystem.

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

**Personvern**

Opplysninger som registreres om deg er navn, kjønn, alder, medikamentbruk, diagnose og testens resultat. Alle som får innsyn i studien har taushetsplikt og ingen data vil bli misbrukt. Hver deltager vil bli nummerert og testerne får kun oppgitt nummeret og ingen annen informasjon.

..... er ansvarlig for rekruttering av pasienter.

..... er ansvarlig for innsamling av data og rekruttering av pasienter.

**Utlevering av materiale og opplysninger til andre**

Ingen persondata vil i ettertid kunne bli misbrukt. Ved publikasjon av studien vil alle data være anonymisert.

 **Rett 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. Du har videre rett til å få korrigert eventuelle feil i de opplysningene vi har registrert. Dersom du trekker deg fra studien, kan du kreve å få slettet innsamlede

opplysninger, med mindre opplysningene allerede er inngått i analyser eller brukt i vitenskapelige publikasjoner.

**Økonomi**

Det er ingen sponsor som finansierer studiet. Det er derfor ikke mulig å gi kompensasjon for eventuelle utgifter deltageren måtte ha.

**Forsikring**

Forsikring på vanlig måte gjennom pasientskadeloven.

**Informasjon om utfallet av studien**

Alle deltakerne i masterstudiet har rett til å få informasjon om utfallet/resultatet av studiet.

## Samtykke til deltakelse i studien

Jeg er villig til å delta i studien

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

Jeg bekrefter å ha gitt informasjon om studien

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