

Kandidatnummer 185534



Intertester-reliabilitet ved kliniske
provokasjonstester for skulderimpingement

MANT395 Manuellterapi teori

4. semester, høst 2012

Masterprogram i helsefag

Klinisk masterstudium i manuellterapi for fysioterapeuter
Institutt for samfunnsmedisinske fag, Universitet i Bergen

Antall ord: 10582

INNHOLDSLISTE

Sammendrag	1
Abstract	2
1. Introduksjon	
1.1 Bakgrunn for valg av oppgave	3
1.2 Hensikt med oppgaven	3
1.3 Problemstilling	4
1.4 Oppbygging av oppgaven	4
1.5 Målgruppe for oppgaven	4
1.6 Forfatterens forforståelse	4
2. Teori	
2.1 Patogenese og etiologi	5
2.2 Diagnostikk	6
2.2.1 Artroskopi og bildemodaliteter	6
2.2.2 Klinikk	7
2.2.2.1 Hawkins test	10
2.2.2.2 Neers test	10
2.2.2.3 Jobes test	10
2.2.2.4 Painful arc	10
2.3 Måleteori – reliabilitet	11
2.4 Tidligere forskning på utvalgte testers reliabilitet	11
3. Metode	
3.1 Valg av forskningsdesign	14
3.2 Sammensetning av utvalg	14
3.3 Inklusjons- og eksklusjonskriterier	15
3.4 Utvalgets størrelse	17
3.5 Datainnsamling	17
3.6 Testprosedyre	19
3.7 Standardisering	19
3.8 Analyse av data	20

3.9 Etiske vurderinger	20
4. Resultater	
4.1 Demografiske data	22
4.2 Resultatdata	22
4.2.1 Krysstabeller	22
4.2.2 Kappa-statistikk og prosentvis enighet	25
5. Diskusjon	
5.1 Hovedfunn i studien	27
5.2 Funntilsetninger på tester sett i lys av tidligere studier	27
5.2.1 Hawkins test	27
5.2.2 Neers test	30
5.2.3 Jobes test	32
5.2.4 Painful arc	33
5.3. Metodekritikk og begrensninger ved studien	35
5.4 Områder med behov for videre forskning	37
6. Konklusjon	39
7. Referanseliste	40
8. Vedlegg	
A. Informasjon til pasient og informert samtykke	48
B. Godkjenning fra REK	54

INTERTESTER-RELIABILITET VED KLINISKE PROVOKASJONSTESTER FOR SKULDERIMPINGEMENT

- En tverrsnittsstudie med 33 deltakere i Norge

Klinisk masterstudium i manuellterapi for fysioterapeuter

Institutt for samfunnsmedisinske fag, Universitetet i Bergen

ÅR: 2012

SAMMENDRAG

Bakgrunn for studien: Hvis skulderimpingement ikke diagnostiseres og behandles korrekt kan tilstanden medføre skade på strukturer i det subacromielle rommet, med innskrenket bevegelse og smerte som resultat.

Hensikt: Hensikten med studien var å undersøke intertester-reliabilitet mellom to terapeuter for en rekke kliniske tester som benyttes i diagnostikk av skulderimpingement. Høy reliabilitet er av betydning for best mulig diagnostisering.

Problemstilling: I hvilken grad samsvarer skårene ved bruk av kliniske provokasjonstester i diagnostikk av skulderimpingement, mellom to terapeuter?

Materiale og metode: Et tverrsnittsdesign ble benyttet hvor to blindede testere utførte skuldertestene Hawkins test, Neers test, Jobs test og Painful arc på et utvalg av pasienter med enten skulderimpingement eller annen skulderpatologi samt skulderfriske deltakere. Testerne brukte samme testprotokoll, men det fant ikke sted samtesting i forkant av studien. De vekslet på å teste først og sist, og rekkefølgen på testene vekslet. Skulderpasientene var rekruttert fra en skulderklinikk og gruppetilhørighet ble verifisert ved ultralyd. Ønsket utvalgstørrelse var 50 deltakere. Testene ble skåret som enten positiv eller negativ. Reliabiliteten ble uttrykt gjennom Kappa-verdier og prosentvis enighet.

Resultater: 33 deltakere ble inkludert og testet: Skulderimpingement (n = 13), annen skulderpatologi (n = 4) og skulderfriske (n = 16). Kappa-verdiene lå mellom 0.76 og 0.64, og prosentvis enighet mellom 87,9% og 81,8%. Hawkins test fikk høyest kappa-verdi, etterfulgt av Neers test, Jobs test og Painful arc.

Konklusjon: Intertester-reliabiliteten til samtlige tester var god.

Nøkkelord: Skulderimpingement, SIS, Hawkins test, Jobs test, Neers test, Painful arc, intertester-reliabilitet

AGREEMENT BETWEEN TESTERS CONDUCTING CLINICAL PROVOCATION TESTS FOR SHOULDER IMPINGEMENT

- A cross sectional study with 33 participants in Norway

Master in Health Sciences – Manual therapy

Department of Public Health and Primary Health Care, The University of Bergen

YEAR: 2012

ABSTRACT

Background: If shoulder impingement isn't properly diagnosed and treated this may cause damage to subacromial structures, hence give decreased range of motion and pain. Good reliability is of utmost importance to ensure proper diagnosis.

Aim of study: To examine test-retest reliability of four tests used in the diagnosis of shoulder impingement, as expressed in terms of agreement between two therapists.

Problem: What is the agreement between two therapists conducting clinical provocation tests in the diagnosis of shoulder impingement?

Material and methods: A cross sectional design was used. Two blinded therapists conducted Hawkins test, Neers test, Jobes test and Painful arc on a sample of healthy individuals, and patients with shoulder impingement and other forms of shoulder pathology. The testers used the same test protocol, but no patients were tested together before the study start. The sequences of testers and tests were randomly selected. The shoulder patients were recruited from a shoulder clinic and the diagnosis was verified by ultrasound. The planned sample size was 50. The tests were scored as positive or negative. Reliability was expressed in terms of Cohens Cappa and percentage agreement.

Results: 33 participants were included and tested: Shoulder impingement (n = 13), other pathology in the shoulder (n = 4) and participants without shoulder problems (n = 16). Kappa-values ranged from 0.76 to 0.64, and the percentage agreement from 87,9% to 81,8%. Hawkins test obtained the largest Kappa-value, followed by Neers test, Jobes test and Painful arc.

Conclusions: For all tests, agreement between testers was found to be substantial.

Key words: Shoulder impingement, SIS, Hawkins test, Jobes test, Neers test, Painful arc, intertester-reliability

1. INTRODUKSJON

1.1. Bakgrunn for valg av oppgave

I løpet av et år oppsøker om lag én prosent av den voksne befolkning helsevesenet på grunn av skuldersmerter (Mitchell et al. 2005). Den årlige selvrapporterte prevalensen er imidlertid høyere, og ble i en stor epidemiologisk undersøkelse rapportert å være så høy som 46% (Juel 2007). Av det totale antallet smertefulle skuldre finner en skulderimpingement eller tegn til impingement hos en stor andel av disse. Van der Windt et al. (1996) fant at impingement var den vanligste årsaken til skuldersmerter, og Östör et al. (2005) fant tegn til impingement hos 74% av pasientene som oppsøkte allmennlege på grunn av skuldersmerter. At tilstanden er utbredt er også blitt argumentert for av Szyluk et al. (2008) som hevdet at impingement er den vanligste årsaken til smerter og nedsatt bevegelse i skulderområdet. Med utgangspunkt i tilstandens høye prevalens er det rimelig å anta at tilstanden gir økonomiske og samfunnsmessige konsekvenser.

1.2 Hensikt med oppgaven

Hvis skulderimpingement ikke diagnostiseres korrekt kan tilstanden medføre skade på strukturer i det subacromielle rommet, med bevegelsesinnskrenkninger og smerte som resultat (Szyluk et al. 2008). Hensikten med oppgaven var derfor å undersøke intertester-reliabiliteten for en rekke kliniske tester som benyttes i diagnostikk av skulderimpingement. Høy grad av reliabilitet er av betydning for best mulig diagnostisering.

I oppgaven ble statisk og dynamisk ultralyd benyttet som gullstandard for å vurdere om deltakerne hadde en reell impingementdiagnose eller ei, i tillegg til at også skulderfriske deltakere inngikk i utvalget. Testerne var blindet i forhold til deltakerenes gruppetilhørighet, noe som er viktig ved undersøkelse av reliabilitet hos diagnostiske tester.

Studien er i seg selv frittstående, men inngår som en del av en større studie hvor også test-retest-reliabilitet og diagnostisk validitet ved de samme testene ble undersøkt. Den diagnostiske validiteten ved testene ble kartlagt gjennom å sammenligne funnene fra testene med funn fra ultralyd, siden ultralyd i denne studien ble betraktet som en gullstandard. Test-retest reliabiliteten ble vurdert ved at én terapeut testet det samme utvalget ved to anledninger, med hensikt å kartlegge grad av enighet mellom skårene på to ulike tidspunkter. Disse tre frittstående studiene vil sammen utgjøre grunnlaget for en artikkel som er tenkt publisert, og hvor alle vinklingene sammenfattes og belyses.

1.3 Problemstilling

I hvilken grad samsvarer skårene mellom to terapeuter ved bruk av kliniske provokasjonstester i diagnostikk av skulderimpingement?

1.4 Oppbygging av oppgaven

Innledningsvis vil patoanatomien ved skulderimpingement skisseres, før ulike måter å diagnostisere tilstanden nevnes siden det ikke foreligger konsensus for diagnostiske kriterier (Awerbuch 2008). I fremstillingen av de diagnostiske kriteriene vil det dveles noe ved både ultralyd og bruk av kliniske tester, siden disse elementene blir av spesiell betydning for den videre oppgaven. Deretter vil studiens design og metode presenteres, etterfulgt av resultatene fra studien og en diskusjon av disse funnene sett i lys av tidligere studier og teori. Avslutningsvis vil svakheter ved studien drøftes og behov for videre forskning kort skisseres.

1.5 Målgruppe for oppgaven

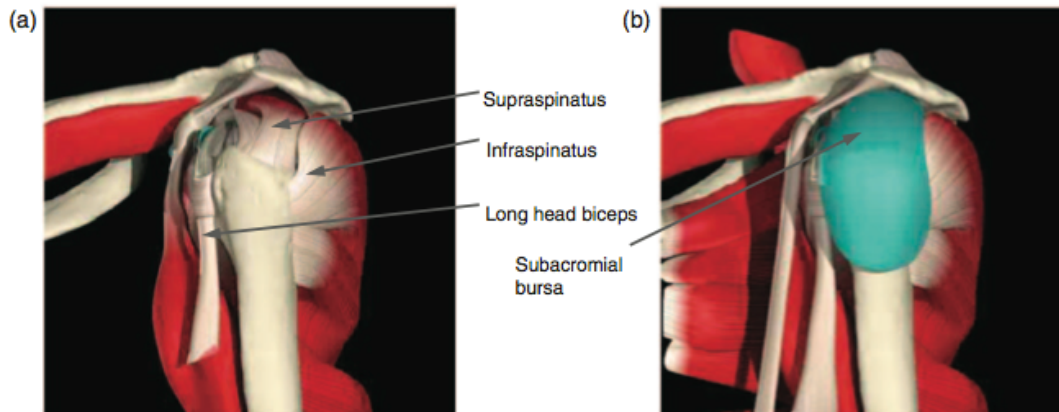
Oppgaven er skrevet for personer med interesse for og kjennskap til skulderanatomi og skulderpatologi. Det forutsettes også at leseren har kjennskap til grunnleggende statistikk.

1.6 Forfatterens forforståelse

Forfatteren er utdannet fysioterapeut og i gang med et masterstudie i manuellterapi. Forfatteren har stor tro på at konservative tilnærminger bør være førstevalget i undersøkelse og behandling av de fleste skulderlidelser.

2. TEORIDEL

2.1 Patogenese og etiologi



Figur 1: Strukturer som kan affiseres ved skulderimpingement (Kelly et al. 2010).

Skulderimpingement er kjennetegnet ved smerter anterolateralt i skulder, som øker når armen eleveres over skulderhøyde (Buss et al. 2009). Tilstanden er ofte assosiert med degenerasjon av skulderstrukturer (Harrison og Flatow 2011) og kan klassifiseres som outlet, non-outlet eller intern (Awerbuch 2008).

Outlet-varianten er den vanligste (Awerbuch 2008) og oppstår i det subacromielle rommet ved abduksjon eller fleksjon av skulderen (Awerbuch 2008) når den subacromielle bursa og rotatorcuffen avklemmes mot enten acromion, ligamentum coracoacromiale (Kelly et al. 2010) eller processus coracoideus (Hughes et al. 2012). Grad av avklemming påvirkes blant annet av fasong og form på acromion (Buss et al. 2009) og hvorvidt det foreligger osteofytter subacromialt (Awerbuch 2008).

Med non-outlet-impingement menes en avklemming som finner sted intraartikulært (Zaslav 2001) på grunnlag av instabilitet i glenohumeralledet, hvorpå caput humeri forskyves kranialt slik at avstand mellom caput humeri og acromion reduseres, og hvor konsekvensen kan bli avklemming av rotatorcuffen (Awerbuch 2008). Denne varianten kan blant annet skyldes ubalanse i skuldermuskulatur (Payne et al. 1997).

Den interne formen for impingement er den siste varianten og skyldes en avklemming av infraspinatus (Awerbuch 2008) mellom tuberculum majus og glenoid når armen er abduisert og utadrottert (Heyworth og Williams 2009), og ses hyppigst blant utøvere av kastidretter (Awerbuch 2008, Behrens et al. 2010).

Impingement er kartlagt biomekanisk i kadaverstudier (Flatow et al. 1994, Wuelker et al. 1995, Yamamoto et al. 2009, Yamamoto et al. 2010, Hughes et al. 2012) men overførbarheten er omdiskutert (Yanai et al. 2006) siden faktorer som ødem (Valadie et al. 2000) og muskulatur (Payne et al. 1997) kan virke inn på belastningsforholdene. En finner imidlertid at kontakt mellom rotatorcuff og acromion eller ligamentum coracoacromiale ikke utelukkende er et fenomen ved impingement, og at større patologiske endringer i rotatorcuff ikke er nødvendig for at kontaktfenomenet subacromialt kan finne sted (Yamamoto et al. 2010).

2.2 Diagnostikk

Det foreligger per i dag ingen konsensus hva angår diagnostikk av skulderimpingement (Awerbuch 2008), herunder hvorvidt tilstanden kan diagnostiseres ved artroskopi eller bildemodaliteter, eller om diagnosen primært er klinisk og heller bør vurderes med utgangspunkt i smerte eller bevegelsesinnskrenkninger.

2.2.1 Artroskopi og bildemodaliteter

Det er blitt argumentert for at både artroskopi (Malhi og Khan 2005, Park et al. 2005, Papadonikolakis et al. 2011), ultralyd (Malvestiti et al. 1997, Papadonikolakis et al. 2011) og magnetresonanstomografi (MR) (Tan 1998, Papadonikolakis et al. 2011) kan benyttes i diagnostikken av skulderimpingement, og den diagnostiske treffsikkerheten ved skulderlidelser er lik eller tilnærmet lik for MR og ultralyd (Teefey et al. 2004, Malhi og Khan 2005). Ultralyd er rask å utføre og kostnadseffektiv (Teefey et al. 2004, Juel 2007) i tillegg til at metoden er uten strålebelastning eller bivirkninger (Juel 2007). Videre gir ultralyd mulighet for ikke-invasiv visualisering av en rekke strukturer i skulderbuens bløtdelsapparat (Moosmayer et al. 2004). Ultralyd blir ofte benyttet i diagnostikk av skulderlidelser i andre land, men utbredelsen er mindre i Norge (Moosmayer et al. 2004, Brox et al. 2010).

Hvis en skal legge bildemodaliteter til grunn for diagnostikken må gitte kriterier for tilstanden foreligge. Avstanden mellom acromion og humerus, benevnet AHD, benyttes ofte som et mål på de subacromiale plassforholdene. Sammenhengen mellom AHD og impingement er imidlertid usikker da det i enkelte studier er funnet at avstanden er redusert ved skulderimpingement (Hébert et al. 2003, Mayerhoefer et al. 2009) mens det i andre studier ikke er funnet en slik sammenheng (Desmeules et al. 2004). Videre kan en vurdere grad av degenerasjon i skulderen (Allen 2008, Kelly et al. 2010) siden degenerasjon kan være en

konsekvens av impingement (Harrison og Flatow 2011). I tillegg til disse statiske vurderingene kan en også vurdere impingement bildemessig ved dynamisk ultralyd, hvor en registrerer bevegelse av bursa og rotatorcuff samtidig som skulderleddet beveges (Bureau et al. 2006, Kim et al. 2007).

Det er sparsomt med intertester-reliabilitetsstudier når det gjelder ultralyd brukt diagnostisk ved skulderimpingement (Awerbuch 2008). AHD har imidlertid vært gjenstand for flere reliabilitetsstudier (Desmeules et al. 2004, Pijls et al. 2010, Kumar et al. 2011) hvor intraclass correlation coefficient, forkortet ICC, ofte har vært benyttet som et uttrykk for grad av enighet mellom testere. ICC er en skala mellom 0 og 1, hvor 1 er uttrykk for perfekt grad av enighet (De Vet et al. 2011) og ICC-verdier over 0.7 blir vurdert som akseptable (De Vet et al. 2011). Graden av enighet i disse tidligere studiene ble funnet å være brukbar, siden ICC-verdiene befant seg i intervallet mellom 0.64 (Pijls et al. 2010) og 0.92 (Desmeules et al. 2010).

Avslutningsvis bør det nevnes at ultralydoperatørens erfaring er blitt trukket frem som en faktor av vesentlig betydning for metodens reliabilitet (Kayser et al. 2005). Brox et al. (2010) hevdet imidlertid at en i Norge har overdreven respekt for lærekurven ved denne bildemodaliteten.

2.2.2 Klinikk

I følge Hegedus et al. (2008) bør kliniske tester være hjørnesteinen i diagnostikk av skulderimpingement. Det er små kostnader forbundet med bruk av slike tester, sammenlignet med bildemodalitetene. Generelt gjelder det at en behøver et kluster av tester som alle peker i den samme retningen, for å gi tilstrekkelig sensitivitet og spesifisitet (Ure et al. 1993, Fodor et al. 2009, Salaffi et al. 2010). Kliniske tester kan også kombineres med injeksjoner subacromialt for å lette differensialdiagnostikken (Calis et al. 2000, Bahr og Mæhlum 2003, Tallia og Cardone 2003).

Det er beskrevet over hundre ulike skuldertester (Moen et al. 2010) og mange av disse er utviklet for å avdekke impingement. I et søk på Pubmed 09.11.11 med søkestrengen "Shoulder Impingement Syndrome/ diagnosis" [MAJR] AND ("test" OR "tests" OR "physical examination") med hensikt å kartlegge studier som omhandlet diagnostisk validitet ved impingementtester gav dette 87 treff. Av disse var det fem studier der den diagnostiske validiteten ved slike tester eksplisitt ble undersøkt.

Sannsynlighetsandeler (Likelihood ratio, LR) er den beste måten å vurdere den kliniske nytteverdien av en test da denne gir et forholdstall mellom sensitivitet og spesifisitet (Hattam og Smeatham 2010), og med sensitivitet mener en her andelen syke som identifiseres som syke mens spesifisitet refererer til andelen friske som identifiseres som friske (Carter et al. 2011). Formler for utregning av sannsynlighetsandeler presenteres i Tabell I. I tolkningen av sannsynlighetsandeler har en at desto lengre bort fra 1 LR+ og desto nærmere 0 LR- denne befinner seg, desto mer informativ er testen (Hattam og Smeatham 2010).

Tabell I: Utregning av sannsynlighetsandeler (LR) (Carter et al. 2011)

$LR + = \frac{\text{Sensitivitet \%}}{(100 - \text{Spesifisitet \%})}$	$LR - = \frac{(100 - \text{Sensitivitet \%})}{\text{Spesifisitet \%}}$
<i>Sannsynlighetsandel, positiv test</i>	<i>Sannsynlighetsandel, negativ test</i>

Testene inkludert i denne studien ble valgt ut på bakgrunn av disse fem studiene og en vurdering av sannsynlighetsandelene testene oppnådde. En oversikt over samtlige tester presentert i studiene, med tilhørende sannsynlighetsandeler, er samlet i Tabell II.

Enkelte av testene, som Hawkins test og Neers test, inngikk i flere studier. Neers test kom relativt dårlig ut i enkelte av studiene, slik som hos Silva et al. (2008) og Kelly et al. (2010) mens den i andre studier kom noe bedre ut, så som hos Michener et al. (2009) og Park et al. (2005). Hawkins test ble skåret relativt likt i alle studiene, og testen kom relativt godt ut. Både Hawkins og Neers ble inkludert i denne studien da disse testene i kadaverstudier har blitt vist å gi en avklemming av strukturer subacromialt (Valadie et al. 2000, Yamamoto et al. 2009), i tillegg til at de er blitt presentert som impingementtester i en rekke lærebøker (Solberg 2002, Bahr og Mæhlum 2003, Magee 2008, Hattam og Steatham 2010) og kom relativt godt ut i den metodisk sterke studien til Park et al. (2005).

Jobs test og Painful arc ble også inkludert i denne studien, til tross for at deres sannsynlighetsandeler var noe sprikende. Dette ble begrunnet i at testene i lærebøker er omtalt som impingementtester (Gross et al. 1996, Bahr og Mæhlum 2003) i tillegg til at også disse kom relativt godt ut i studien til Park et al. (2005).

Tabell II: Oversikt over diagnostiske tester i det innledende litteratursøket og deres sannsynlighetsandeler

Test	Silva et al. 2008		Calis et al. 2000		Kelly et al. 2010		Park et al. 2005		Michener et al. 2009	
	LR+	LR-	LR+	LR-	LR+	LR-	LR+	LR-	LR+	LR-
Neers test	0,98	1,05	1,28	0,37	0,62	∞	2,17	0,46	1,76	0,35
Hawkins test	1,23	0,66	1,23	0,32	1,48	0,52	2,12	0,43	1,63	0,61
Yocum	1,32	0,53								
Jobes test/ Empty can	1,05	0,88			1,56 ₁ 0,78 ₂	0,72 ₁ 1,45 ₂	4,2	0,63	3,9	0,57
Full can					1,79 ₁ 0,46 ₂	0,74 ₁ 2,62 ₂				
Patte	1,45	0,70								
Gerber	1,37	0,63								
Horisontal adduksjon			1,13	0,65			1,25	0,95		
Speeds test			1,54	0,57			2,30	0,74		
Yergasons test			2,66	0,73						
Painful arc			1,67	0,84	0,60	1,41	3,89	0,33	2,25	0,38
Drop arm			2,78	0,96			2,32	0,83		
Abduksjon	0,72 ₃ 0,82 ₄	2,1 ₃ 2,63 ₄			0,76 ₁ 2,21 ₂	1,24 ₁ 0,59 ₂				
Utadrotasjon					0,74 ₁ + ∞ ₂	1,79 ₁ 0,66 ₂			4,39 ₃	0,5 ₃
Infraspinatus muskeltest							4,2	0,65		

₁ = Svakhhet ved test, ₂ = Smerte ved test, ₃ = Mot motstand, ₄ = Passiv bevegelse

I det følgende vil de utvalgte testene Hawkins test, Neers test, Jobes test og Painful arc bli presentert mer utfyllende:

2.2.2.1 Hawkins test

Hawkins test er en provokasjonstest (Yamamoto et al. 2009) og utføres ved at pasienten sitter med overarmen 90 grader flektert i sagittalplanet (Solberg 2002). Scapula på samme side fikseres for å hindre rotasjon av scapula. Overarmen roteres deretter innover til eventuelle smerter oppstår og smerter er kriterier for positiv test (Moen et al. 2010). Hawkins test har vært gjenstand for kadaverstudier og Yamamoto et al. (2009) viste at inneklekking fant sted når tuberculum majus klemte subscapularis-senen mot acromion og ligamentum coracoacromiale. Hawkins test er videre blitt vist å gi inneklekking også hos levende individer, verifisert ved ultralyd (Pappas et al. 2006).

2.2.2.2 Neers test

Neers test er en provokasjonstest (Yamamoto et al. 2009) og gjennomføres ved at pasienten sitter på en stol, mens terapeuten står bak. Terapeuten fikserer scapula på samme side, og flekterer pasientens arm passivt i sagittalplanet (Moen et al. 2010). Det foreligger ikke enighet i forhold til grad av skulderrotasjon ved utførelse (Yamamoto et al. 2009) og testen er positiv ved smerter (Moen et al. 2010). Jia et al. (2011) fant at smerte fra Neers test skyldes kontakt mellom rotatorcuff og glenoid. Neers test er vist å gi inneklekking også hos levende individer, verifisert ved ultralyd (Pappas et al. 2006).

2.2.2.3 Jobs test

Jobs test betegnes også supraspinatus test eller empty can test (Kim et al 2006, Michener et al. 2009), og blir brukt for å diagnostisere patologi i rotatorcuffen (Holtby og Razmjou 2004). Testen utføres ved at pasienten sitter med skulderen 90 grader abduert, 30 grader horisontalt addusert og maksimalt innadrottert mens terapeut plasserer én hånd på oversiden av pasientens overarm. Pasient holder overarmen i ro mens terapeut gir et trykk ned mot pasientens arm (Moen et al. 2010). Testen er nyttig til å diagnostisere rifter i rotatorcuffen (Holtby og Razmjou 2004), som ofte ses ved skulderimpingement (Harris og Flatow 2011). Nedsatt kraft ved test kan skyldes ruptur av supraspinatussene eller nerverotsaffeksjon, mens smerter kan skyldes impingement (Moen et al. 2010).

2.2.2.4 Painful arc

Ved Painful arc føres skulderen aktivt i abduksjon. Testen er beskrevet på ulike måter. Enkelte fører armen i abduksjon i scapulas plan (Park et al. 2005) mens andre fører armen i abduksjon i frontalplanet (Kessel og Watson 1977, Michener et al. 2009). Abduksjon

av armen er i kadaverstudier vist å gi økt trykk subacromialt (Wuelker et al. 1995). Avstanden mellom acromion og rotatorcuff er minst ved 60 til 120 grader skulderabduksjon (Flatow et al. 1994) og testen er positiv hvis pasienten opplever smerter i denne delen av bevegelsesbanen (Kessel og Watson 1977, Bahr og Mæhlum 2003, Park et al. 2005, Michener et al. 2009, Kelly et al. 2010).

2.3 Måleteori - reliabilitet

I denne studien ønsket vi å vurdere impingementtestenes reliabilitet. Med reliabilitet menes om en ville fått de samme resultatene dersom en hadde gjennomført den samme undersøkelsen flere ganger (Jacobsen 2005). Begreper brukt synonymt om reliabilitet er nøyaktighet, stabilitet og konsistens (Carter et al. 2011) og med intertester-reliabilitet spesielt vurderer en konsistensen mellom målinger utført av flere, på samme tid, og på de samme personene (Carter et al. 2011).

Reliabilitet kan kvantifiseres på to måter, som relativ eller absolutt (Carter et al. 2011). Ved relativ reliabilitet undersøkes sammenhengen mellom to eller flere sett med repeterende målinger, mens en ved absolutt reliabilitet undersøker variabiliteten mellom de ulike målingene (Carter et al. 2011). Relativ reliabilitet uttrykkes ved Intraclass Correlation Coefficient (ICC) verdier for intervall og ratioskaler mens absolutt reliabilitet uttrykkes gjennom Standard error of measurement (målefeil), forkortet SEM (Carter et al. 2011). Cohens kappa benyttes ved kategori og ordinale skaler og er et mål som korrigerer for grad av enighet mellom testere som også ville ha funnet sted tilfeldig (De Vet et al. 2011). Både ICC og Cohens kappa uttrykkes ved skalaer mellom 0 og 1, hvor 1 er uttrykk for perfekt grad av enighet (De Vet et al. 2011).

2.4 Tidligere forskning på utvalgte testers reliabilitet

Et søk i PubMed 03.03.12 hadde til hensikt å kartlegge studier som har undersøkt intertester-reliabiliteten for Hawkins test, Neers test, Jobes test og Painful arc. Søke-ordene ("shoulder pain" OR "subacromial impingement" OR "shoulder impingement") AND (inter-tester OR intertester OR interexaminer OR inter-examiner OR inter-observer OR interobserver OR inter-rater OR interrater OR reliability) AND ("jobe" OR "jobes" OR "neer" OR "neers" OR "Painful arc" OR "hawkin" OR "hawkins" OR "test" OR "tests" OR "Physical Examination") gav 104 treff. Det innledende søket ble deretter gruppert i seks kategorier, basert på relevans, som vist i Tabell III.

Tabell III: Oversikt over utvelgelse av relevante studier og artikler

Innledende søkestreng: 104 treff					
A. Relevante skulderstudier eller oversiktsartikler n = 8	B. Ikke relevante skulderstudier n = 41	C. Studier som ikke omhandler skulderregionen n = 2	D. Studier om spørreskjema/målemetoder n = 42	E. Studier med primærfokus bildemodaliteter n = 10	F. Treff som ikke er studie eller oversiktsartikkel n = 1

Av treffene i kategori A i Tabell III fant en åtte relevante skulderstudier eller oversiktsartikler, som alle omhandlet intertester-reliabilitet ved én eller flere av de på forhånd utvalgte skuldertestene Hawkins test, Neers test, Jobes test og Painful arc. En oversikt over testene med tilhørende kappa-verdier for intertester-reliabilitet presenteres i Tabell IV.

De rapporterte kappaverdiene og konfidensintervallene i de ulike studiene ble funnet å sprike mye. For Hawkins test var kappa så lav som 0.1 i studien til Nørregaard et al. (2002) mens den var så høy som 0.91 i studien til Johansson og Ivarson (2009). Det var tilsvarende høyt sprik for Neers test, med verdier fra 0.1 (Nanda et al. 2008) til 1 (Johansson og Ivarson 2009). For Jobes test og Painful arc var spredningen i kappa-verdier noe lavere, i tillegg til at de ble funnet å ligge i den øvre delen av skalaen, fra i underkant av 0.5 til i overkant av 0.9.

En kan tenke seg ulike forklaringer på dette spriket. Én forklaring kan ligge i forskernes yrkesbakgrunn og erfaring. Videre kan metodologiske aspekter trekkes frem, herunder grad av standardisering mellom testerne i forkant av selve studien, hvorvidt rekkefølgen av tester var gitt eller tilfeldig, i tillegg til hvordan selve testen ble utført.

Utvalgene varierte i tillegg mye med tanke på størrelse og sammensetning. Samtlige deltagere i studiene til Michener et al. (2009), Nanda et al. (2008), Johansson og Ivarson (2009) var på forhånd var diagnostisert med skulderimpingement. Deltakerne i studiene til Östör et al. (2005), Nørregaard et al. (2002), Razmjou et al. (2004), Cadogan et al. (2011), Palmer et al. (2000) og Walker-Bone et al. (2002) hadde også skuldersmerter, men for disse studiene var det ikke definert/verifisert at plagene hadde sin årsak i impingement.

I tillegg var studiene også forskjellige med hensyn på om utvalget som ble testet også inkluderte friske individer. Friske individer ble kun inkludert i utvalget til Vind et al. (2011)

mens utvalgene i de andre tidligere studiene utelukkende bestod av deltakere med skuldersmerter. Alle disse nevnte faktorene får konsekvenser med tanke på om utvalgene i de ulike studiene, så vel som rapportert grad av enighet, kunne sammenlignes direkte.

Tabell IV: Oversikt over testene og tilhørende kappa-verdier for intertester-reliabilitet, funnet i tidligere studier

Test	Forfatter	κ og 95% CI	n	Testere
Hawkins test	Nørregaard et al. 2002	0.07 (sag.plan)	86	Ortoped + revmatolog
	Razmjou et al. 2004	0.29	136	Ortoped + fysioterapeut
	Östör et al. 2005	0.31 (0.18-0.43)	136	Sykepleier + revmatolog
	Cadogan et al. 2011	0.38 (0.1-0.63)	40	Fysioterapeuter
	Michener et al. 2009	0.4 (0.13-0.67)	55	Ortoped + fysioterapeut
	Nørregaard et al. 2002	0.4 (scaption)	86	Ortoped + revmatolog
	Nanda et al. 2008	0.55	63	Ortoped + revmatolog
	Vind et al. 2011	0.6 (0.34-0.85)	44	Lite erfarne testere
	Johansson og Ivarson 2009	0.91	33	Fysio + manuellterapeut
Neers test	Nanda et al. 2008	0.1	63	Lege + Lege
	Michener et al. 2009	0.39 (0.12-0.65)	55	Ortoped + fysioterapeut
	Razmjou et al. 2004	0.51	136	Ortoped + fysiot.
	Vind et al. 2011	0.86 (0.71-1) ₁	44	Lite erfarne testere
	Vind et al. 2011	0.95 (0.86-1) ₂	44	Lite erfarne testere
	Johansson og Ivarson 2009	1.0	33	Fysio + manuellterapeut
Jobes test/ Empty can	Nanda et al. 2008	0.44	63	Lege + Lege
	Östör et al. 2005	0.47 (0.44-0.49)	136	Sykepleier + revmatolog
	Michener et al. 2009	0.47 (0.22-0.72)	55	Ortoped + fysioterapeut
	Vind et al. 2011	0.9 (0.76-1)	44	Lite erfarne testere
	Johansson og Ivarson 2009	0.94	33	Fysio + manuellterapeut
Painful arc	Michener et al. 2009	0.45 (0.18-0.72)	55	Ortoped + fysioterapeut
	Nomden et al. 2009	0.47	91	Fysio + manuellterapeut
	Walker-Bone et al. 2002	0.47 (0.33-0.61)	90	Sykepleier + revmatolog
	Nanda et al. 2008	0.48	63	Ortoped kons. + registrar
	Palmer et al. 2000	0.93 (0.71-1)	88	Sykepleier + revmatolog

₁ – Anterior smerte. ₂ – Generell smerte. κ = kappa

3. METODE

3.1. Valg av forskningsdesign

Til denne oppgaven ble det valgt et kvantitativt og ikke-eksperimentelt kohortdesign. Studien kan videre beskrives som en tverrsnittsstudie.

3.2. Sammensetning av utvalg

Utvalget i studien var satt sammen av tre grupper deltakere: En gruppe med diagnostisert skulderimpingement, en gruppe med annen skulderproblematikk, og en gruppe skulderfriske/asymptomatiske individer. Både pasienter med impingement og pasienter med annen skulderpatologi ble inkludert fordi en ønsket å illudere en klinisk situasjon der pasienter oppsøker helsevesenet for diagnostisering av skulderplagene sine, men hvor etiologi er ukjent. En valgte å også inkludere skulderfriske individer i utvalget, siden studien i tillegg til å være frittstående inngår som en del av en større studie hvor også testenes diskriminerende egenskaper i forhold til frisk/syk-dimensjonen ble undersøkt. En ønsket i utgangspunktet at gruppene skulle være like store, og testerne visste ikke om deltakerne hadde skulderplager eller ikke. Pasientene som ble rekruttert til skulderimpingementgruppen og gruppen med annen skulderpatologi hadde alle oppsøkt lege ved en privat skulderklinikk på grunn av skulderplager i tidsrommet fra mai til august 2012, og samtlige pasienter som oppsøkte legen i dette tidsintervallet ble vurdert med tanke på egnethet for å delta i prosjektet.

Dersom legen ved undersøkelse av pasienten fant at pasienten falt innenfor studiens inklusjons- og eksklusjonskriterier fikk pasienten utlevert et informasjonsskriv om studien av legen, hvor pasienten ble gitt anledning til å ta kontakt med forskergruppen for å bli inkludert i studien. Det lot seg ikke gjøre å registrere hvor mange av de inviterte pasientene som faktisk tok kontakt med forskergruppen for å bli inkludert i studien. Forskergruppen bestod av to blindede terapeuter som testet, i tillegg til en tredje terapeut som administrerte og koordinerte testingen. Den sistnevnte terapeuten var den eneste som ikke var blindet for pasientens gruppetilhørighet og var i tillegg til å administrere studien også ansvarlig for blindingen samt å oppbevare koblingsnøkkel og testresultater. Kontakten mellom pasient og forskergruppen fant sted mellom pasient og denne tredje koordinerende terapeuten.

Som en del av inklusjons- og eksklusjonskriteriene for de to skuldergruppene ble ultralyd benyttet for å bestemme hvorvidt skulderplagene og symptomene primært skyldes

impingement eller annen skulderpatologi. Ultralydapparatet ble betjent av en erfaren lege med erfaring innen allmennmedisin og idrett. Hans primære arbeidsoppgaver de siste ti årene har vært diagnostikk av skulderlidelser ved bruk av ultralyd, hvor vedkommende anslagsvis har 2000 slike pasientkonsultasjoner årlig.

For de skulderfriske/asymptomatiske deltakerne ble invitasjonen til å delta i studien fremsatt av det tredje medlemmet av forskergruppen som ikke deltok på selve testingen. Invitasjonen ble gitt dersom individet passet inn i studien i henhold til definerte inklusjons- og eksklusjonskriterier. Da dette er en metodestudie søkte en i utgangspunktet et mest mulig heterogent utvalg, og det skulderfriske utvalget ble matchet opp mot skulderutvalgene med hensyn på kjønn og alder. Rekruttering av individer til studien pågikk parallelt med selve testingen og følgelig fikk en ulikt antall deltakere i gruppene som inngikk i studien. Såfremt det ikke var systematikk i frafall fra invitasjon til inklusjon fra invitasjon til inklusjon i studien kan dette gi en pekepinn på de ulike skulderplagens prevalens.

3.3 Inklusjons- og eksklusjonskriterier

Eksklusjonskriteriene gjaldt uavhengig av gruppetilhørighet, og var definert som:

- Tidligere traume mot skulder
- Tidligere skulderoperert
- Inflammatoriske leddlidelser
- Smerter ved bevegelse av nakken
- Nevropatiske smerter i nakke eller overekstremitet
- Røde flagg som kunne mistenke annen alvorlig patologi

Inklusjonskriteriene var forskjellige for skulderpasientene og de asymptotiske/skulderfriske gruppene. For skulderpasientene gjaldt følgende inklusjonskriterier:

- Skuldersmerter av varighet lengre enn én måned
- Skuldersmerte måtte være hovedproblemet
- Alder 18 – 60 år
- Måtte kunne gi et informert samtykke

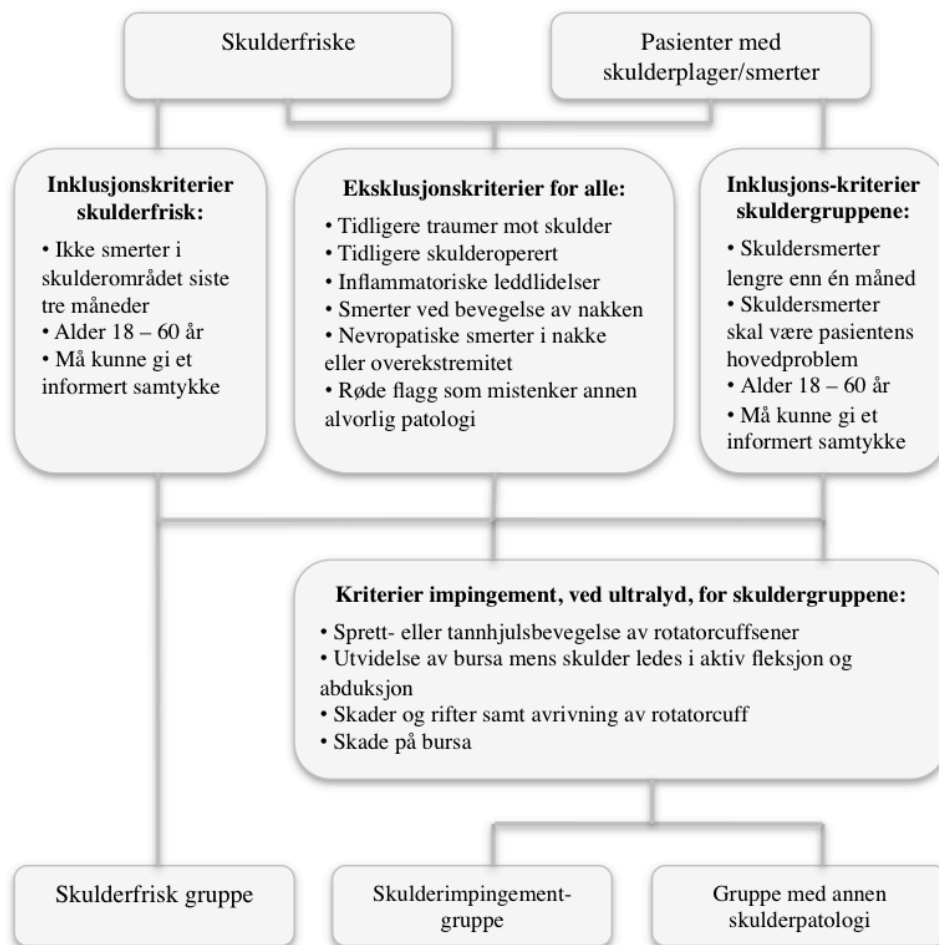
For skulderpasientene falt disse inn under impingementgruppen dersom det forelå tegn til

impingement på ultralyd, mens de falt inn under gruppen med annen skulderpatologi dersom det ikke forelå tegn til impingement på ultralyd. Impingement på ultralyd ble diagnostisert med utgangspunkt i en vurdering av:

- Grad av depresjon av humerus, så vel som sprett eller tannhjulsbevegelse av rotatorcuff-sener eller utvidelse av bursa mens pasientens skulder ble ledet i aktiv fleksjon og abduksjon
- Skader og rifter samt avrivning av rotatorcuff, i tillegg til skade på bursa

Inklusjonskriteriene for de skulderfriske var følgende:

- Ikke smerter i skulderområdet de siste tre månedene
- Alder 18 – 60 år
- Måtte kunne gi et informert samtykke



Figur 3: Karakteristika ved de ulike gruppene i studien

3.4 Utvalgets størrelse

I reliabilitetsstudier blir grad av enighet uttrykt ved hjelp av Cohens kappa dersom data foreligger på nominalnivå, eller ICC dersom en har data på intervallnivå (De Vet et al. 2011). Ved ICC kan en benytte formler for å regne ut det minste antall individer som er nødvendig i utvalget for å oppnå en bestemt bredde på konfidensintervallet (De Vet et al. 2011). Når Cohens kappa benyttes, som i denne studien, er det imidlertid ikke like lett å regne ut et slikt nøyaktig minimumsantall (De Vet et al. 2011). Det er dog alltid ønskelig at utvalget er av en viss størrelse, for at en i større grad skal kunne generalisere til den kliniske populasjonen, samt for å gi økt grad av presisjon for punkttestimatene (Carter et al. 2011). Ved reliabilitetsstudier er det ønskelig med et utvalg bestående av minimum 50 individer (De Vet et al. 2011), og en slik utvalgsstørrelse ble lagt til grunn ved studiens start. En ønsket at det skulle være omtrent like mange deltakere i hver av de tre gruppene deltakere. Innenfor rammene av dette prosjektet lot det seg dessverre ikke gjøre å rekruttere flere enn totalt 33 individer.

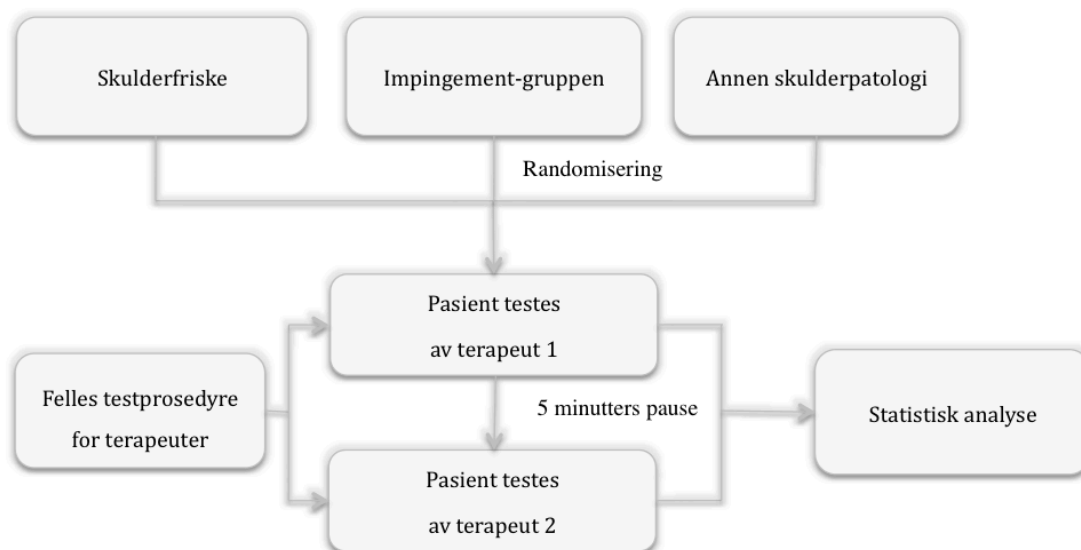
3.5 Datainnsamling

To terapeuter som jobbet på det samme fysikalske instituttet gjennomførte fire kliniske provokasjonstester for skulderimpingement, på det samme utvalget og innenfor samme tidsperiode, våren og høsten 2012. De kliniske testene som ble utført var Hawkins test, Neers test, Jobs test og Painful arc. Disse testene ble utvalgt da en fant disse testene til å ha størst grad av diagnostisk validitet ved et innledende litteratursøk. Begge terapeutene holdt på å avslutte masterstudiet i manuellterapi, hadde relativt lik yrkesbakgrunn og relativt lik klinisk erfaring innenfor muskel/skjelett-området. Selve administreringen av testingen og innsamlede data ble koordinert av en tredje terapeut, og vedkommende var den eneste i studien som ikke var blindet for pasientens sykdomsstatus.

I forkant av testingen ble det utarbeidet en protokoll hvor svarene fra testene ble nedtegnet. Testene ble skåret med utgangspunkt i dikotomien positiv eller negativ test. For alle testene betydde dette at testene ble skåret som positive hvis de gav smerter i skulderområdet. Pasienten måtte oppleve relativt distinkte smerter i skulder for at testen skulle nedtegnes som positiv, andre subjektive opplevelser som tøyning, strekk, diffust ubehag og lignede ble registrert som negativ test. Pasienten kunne for alle testene bare svare tilstrekkelig til at terapeut kunne skåre testen, for alt annet var terapeuten blindet.

Hver test ble først utført én gang av terapeut nr 1. Deretter ble pasienten testet på nytt én gang av terapeut nr 2, etter to minutter. To minutters pause mellom hver test ble valgt for å hindre at den første testen påvirker svarene ved test nummer to. For å sikre ytterligere at testene ikke påvirket hverandre var rekkefølgen på testene også tilfeldig, likeså hvilken terapeut som testet pasienten først og sist. Siden de tre gruppene ble rekruttert samtidig som testingen ble utført var algoritmen for hvilken gruppe deltakeren som til enhver tid ble testet var hentet fra, kun kjent for terapeuten som ikke deltok i selve testingen, og terapeutene som testet var følgelig blindet for deltakernes gruppetilhørighet.

Randomisering av rekkefølge på terapeut og test ble ivaretatt ved at pasienten trakk lodd, først på terapeut og deretter på testrekkefølge. Det var fem minutters pause mellom testing av terapeut én og terapeut to for den samme deltakeren i testen. All testing foregikk i de samme lokalene og med størst mulig grad av standardisering av omgivelsene, og det var kun tester og deltaker som var til stede under testingen.



Figur 4: Flyttdiagram over datainnsamlingen

Skåringsdataene ble nedtegnet av terapeuten som testet, og skårings skjemaet ble umiddelbart gitt til den tredje terapeuten som registrerte dataene i en anonymisert database. Koblingen mellom anonymiserte data og utvalgets identifiserbare personopplysninger ble oppbevart av den tredje uavhengige terapeuten, på et annet sted.

3.6 Testprosedyre

Testprosedyren samlet sett:

- *Hawkins test* = Pasienten satt med overarmen 90 grader flektert i scapulas plan. Terapeut fikserte scapula på samme side for å hindre at scapula ble rotert når testen ble utført. Terapeuten roterte deretter pasientens overarm passivt innover, og testen ble nedtegnet som positiv hvis pasienten fikk smerter.
- *Neers test* = Pasienten satt mens terapeuten stod bak. Terapeuten fikserte pasientens scapula på samme side som ble testet. Fikseringen fant sted for å hindre at skulderen ble protrahert. Terapeuten flekterte deretter pasientens arm passivt i sagittalplanet. Testen ble nedtegnet som positiv hvis pasienten fikk smerter.
- *Jobes test* = Pasienten satt med skulderen 90 grader abduisert, 30 grader horisontal addusert og maksimalt innadrottert. Terapeuten plasserte én hånd på oversiden av pasientens overarm. Pasienten skulle deretter holde overarmen i ro mens terapeuten appliserte et trykk nedover mot pasientens arm. Testen var positiv hvis pasienten opplevde smerter.
- *Painful arc* = Pasienten satt på en stol uten armlener. Pasientens arm ble deretter aktivt ført i abduksjon i scapulas plan. Testen var positiv hvis pasienten opplevde smerter i skulderen når denne var abduisert mellom 60 og 120 grader.



Hawkins test



Neers test



Jobes test



Painful arc

Figur 5: Utgangsstilling ved testutførelse

3.7 Standardisering

Standardisering fant sted ved at terapeutene i fellesskap utarbeidet testprotokollen og la denne til grunn for testingen. Annen standardisering mellom terapeutene ble ikke utført, for å i større

grad bevare den økologiske validiteten. Siden de to terapeutene var studenter på det samme masterstudiet må en dog forutsette at noen grad av standardisering hadde skjedd uvilkarlig i forelesninger og studiesammenheng for øvrig, før testingen fant sted.

3.8 Analyse av data

Testene ble skåret etter dikotomien positiv eller negativ. Denne dikotomien befinner seg på nominalskala-nivå. Skåringene fra de to terapeutene ble plottet inn i en krysstabell. Videre analyser at dataene ble utført ved hjelp av SPSS versjon 19 for mac.

Den relative intertester-reliabiliteten ble vurdert ved Cohens kappa-koeffisient for kategoriske variabler. Cohens kappa-koeffisient er foretrukket i vurdering av samsvar mellom testere utover tilfeldighet (Landis og Koch 1977, Kundel og Polansky 2003) når data er på nominalskala-nivå (Landis og Koch 1977) og flere retningslinjer for tolkning av kappa-koeffisienten er utarbeidet. I henhold til Landis og Koch (1977) har en at styrkeforholdet ved kappa er svakt mellom 0.00 og 0.20, rimelig mellom 0.21 og 0.41, brukbart mellom 0.41 og 0.60, betydelig mellom 0.61 og 0.80 og for verdier mellom 0,81 og 1.00 er styrkeforholdet tilnærmet perfekt. Det finnes også enklere retningslinjer for tolkning av kappa, herunder Fleiss' inndeling hvor kappa under 0.4 anses som dårlig, kappa mellom 0.4 og 0.75 vurderes som brukbar, mens kappa over 0.75 er glimrende (De Vet et al. 2011).

Den absolutte intertester-reliabiliteten ble vurdert med utgangspunkt i kappa-verdiens 95% konfidensintervaller. I tillegg ble prosentvis grad av enighet utregnet. I prosentvis enighet inngår også tilfeldig samsvar mellom testere, i motsetning til ved Cohens kappa hvor denne størrelsen ikke inngår (Kundel og Polansky 2003, De Vet et al. 2011).

3.9 Etiske vurderinger

Prosjektet ble gjennomført i henhold til de etiske prinsippene angitt i Helsinki-deklarasjonen. Hensynet til beskyttelse av forsøkspersonen stod sentralt. Denne studien var hensiktsmessig da det gav økt innsikt i et tema som ikke var klarlagt tidligere og ulempene for individet i utvalget var små. Autonomien til utvalget ble ivaretatt ved at testpersonene valgte å inngå i studien frivillig, de fikk tilstrekkelig informasjon om studiens hensikt og formål, og de kunne trekke seg fra studien når som helst uten å måtte oppgi noen grunn til dette. Forskerne hadde taushetsplikt. Siden studien omfattet medisinsk og helsefaglig forskning på mennesker i henhold til Helseforskningsloven (2008) paragraf 2, ble studien på forhånd søkt og godkjent

av regional etisk komité. REK sør-øst ga sin godkjenning til studien «Reliabilitet og validitet av skuldertester» den 5.mars 2012 (2012/57 D), se vedlegg B.

4. RESULTATER

4.1 Demografiske data

For å få deltakere med både positive og negative funn ble både pasienter diagnostisert med skulderimpingement inkludert, så vel som pasienter med andre skulderlidelser og personer uten skulderplager. Testerne var blindet i forhold til deltakernes status med hensyn til skulderplager før testingen ble gjennomført. Bakgrunnsdata til disse undergruppene så vel som alle deltakerne samlet er presentert i Tabell V.

Tabell V: Demografiske data

Karakteristika	Asymptomatisk/ frisk skulder	Impingement	Annen skulder- patologi	Alle deltakerne samlet
Antall personer, n (%)	16 (48%)	13 (39%)	4 (12%)	33 (100%)
Kjønnsfordeling:				
Kvinner, antall (%)	11 (58%)	6 (31%)	2 (10%)	19 (58%)
Menn, antall (%)	5 (36%)	7 (50%)	2 (14%)	14 (42%)
Alder, år:				
Gjennomsnitt	40.0	46.5	43.8	43.0
Variasjonsbredde	21-60	18-60	33-55	18-60
Standardavvik, SD	15.7	13.4	9.1	14.2
Skulder testet:				
Venstre, antall (%)	5 (33%)	10 (67%)	0 (0%)	15 (100%)
Høyre, antall (%)	11 (61%)	3 (17%)	4 (22%)	18 (100%)
Symptomvarighet, måned:				
Antall registrert		13	4	33
Gjennomsnitt		39.4	16.0	38.4
Median		12	15.5	12
Variasjonsbredde		1-360	9-24	1-360
Standardavvik, SD		97.1	6.8	84.7

4.2 Resultatdata

4.2.1 Krysstabeller

Datamaterialet samlet inn i studien blir i det følgende presentert i krysstabeller. Data fra samtlige grupper sett under ett blir presentert i Tabell VI-I mens de tre undergruppene impingement, annen skulderpatologi samt skulderfriske blir presentert i Tabell VI-II til VI-IV.

Krysstabellene viser grad av av enighet mellom testerne, i tillegg til at de kan si noe om i hvor stor grad de ulike testene totalt sett ble skåret som negativ eller positiv. Dette impliserer at krysstabellen kan gi holdepunkt for å si noe om en test klarer å fange opp en gitt tilstand eller

ei. Videre danner krysstabellene grunnlaget for utregning av både prosentvis enighet mellom testerne, i tillegg til enighet utover tilfeldighet, sistnevnte uttrykt gjennom Cohens kapp.

Tabell VI-I: Samtlige data for de fire testene, $n = 33$

		Tester2					Tester2		
		Negativ	Positiv				Negativ	Positiv	
Tester1	Negativ	15	0	15	Tester1	Negativ	18	2	20
	Positiv	4	14	18		Positiv	2	11	13
Sum		19	14	33	Sum		20	13	33
<i>Hawkins test</i>					<i>Neers test</i>				

		Tester2					Tester2		
		Negativ	Positiv				Negativ	Positiv	
Tester1	Negativ	19	4	23	Tester1	Negativ	16	6	22
	Positiv	1	9	10		Positiv	0	11	11
Sum		20	13	33	Sum		16	17	33
<i>Jobes test</i>					<i>Painful arc</i>				

Tabell VI-I presenterer alle individene sett under ett, uavhengig av gruppetilhørighet. Totalt sett var det relativt stor grad av enighet mellom testerne. Størst prosentvis enighet forelå for Hawkins test og Neers test, hvor testerne var enige i 29 av 33 tilfeller. For Jobes test var det enighet i 28 av 33 tilfeller, og samme tallet for Painful arc var 27 av 33 deltakere. Hawkins test var testen som i størst grad ble vurdert til å være positiv av begge testerne, mens det var minst grad av enighet om positiv test ved Jobes test.

Tabell VI-II: Fordeling av skårer på de enkelte tester i Impingementgruppen, $n=13$

		Tester2					Tester2		
		Negativ	Positiv				Negativ	Positiv	
Tester1	Negativ	1	0	1	Tester1	Negativ	2	1	3
	Positiv	1	11	12		Positiv	0	10	10
Sum		2	11	13	Sum		2	11	13
<i>Hawkins test</i>					<i>Neers test</i>				

		Tester2					Tester2		
		Negativ	Positiv				Negativ	Positiv	
Tester1	Negativ	1	2	3	Tester1	Negativ	0	4	4
	Positiv	1	9	10		Positiv	0	9	9
Sum		2	11	13	Sum		0	13	13
<i>Jobes test</i>					<i>Painful arc</i>				

En ser av Tabell VI-II at det var stor grad av enighet mellom testerne når pasienter tilhørende impingementgruppen ble testet. Spesielt gjaldt dette for Hawkins test og Neers test, hvor en ved Hawkins test var enige ved tolv av tretten tilfeller, og ved Neers var enige i tolv av tretten tilfeller. Noe mindre grad av enighet forelå ved Jobes test, her var det enighet for ti av pasientene. Lavest grad av enighet fant en når Painful arc ble testet, hvor enighet kun forelå for ni av tretten pasienter. For Painful arc skåret tester nummer to alle impingementpasientene som positive, mens tester nummer én definerte testen som negativ kun for ni av tretten.

For samtlige av krysstabellene ser en at brorparten av pasientene ble skåret som positive av begge testerne. For impingementpasientene ble dog Hawkins test oftest positiv hos begge testerne, etterfulgt av Neers test.

Tabell VI-III: Annen skulderpatologi-gruppen, n=4

		Tester2		
		Negativ	Positiv	
Tester1	Negativ	0	0	0
	Positiv	1	3	4
Sum		1	3	4

Hawkins test

		Tester2		
		Negativ	Positiv	
Tester1	Negativ	0	1	1
	Positiv	2	1	3
Sum		2	2	4

Neers test

		Tester2		
		Negativ	Positiv	
Tester1	Negativ	3	1	4
	Positiv	0	0	0
Sum		3	1	4

Jobes test

		Tester2		
		Negativ	Positiv	
Tester1	Negativ	0	2	2
	Positiv	0	2	2
Sum		0	4	4

Painful arc

En ser av Tabell VI-III at for pasientene med annen skulderpatologi var det størst grad av enighet blant testerne ved Hawkins test og Jobes test, og testerne skåret for disse testene deltakerne likt i tre av fire tilfeller. For Hawkins test var det enighet mellom testerne i tre av fire tilfeller, imidlertid ble det gitt en positiv skåre der det var forventet en negativ skåre. For Jobes test var det enighet om en negativ skåre i tre av fire tilfeller. For Painful arc var det enighet blant testerne for bare to av fire deltakere, og for Neers test var enigheten enda mindre, da det kun forelå enighet for én av fire deltakere.

Tabell VI-IV: Den asymptomatiske/friske gruppen, n=16

		Tester2		
		Negativ	Positiv	
Tester1	Negativ	14	0	14
	Positiv	2	0	2
Sum		16	0	16

Hawkins test

		Tester2		
		Negativ	Positiv	
Tester1	Negativ	16	0	16
	Positiv	0	0	0
Sum		16	0	16

Neers test

		Tester2		
		Negativ	Positiv	
Tester1	Negativ	15	1	16
	Positiv	0	0	0
Sum		15	1	16

Jobes test

		Tester2		
		Negativ	Positiv	
Tester1	Negativ	16	0	16
	Positiv	0	0	0
Sum		16	0	16

Painful arc

Tabell VI-IV viser at blant de friske deltakerne hadde en stor grad av enighet blant testerne. For Neers test og Painful arc var det enighet blant testerne for alle seksten deltakerne. Enigheten var også stor for Jobes test og Hawkins test, hvor det var enighet for henholdsvis femten og fjorten deltakere. Videre ser en at andelen falske positive var svært lavt blant de friske deltakerne, og det forelå ingen tilfeller hvor testerne var enige om falsk positiv test.

4.2.2 Kappa-statistikk og prosentvis enighet

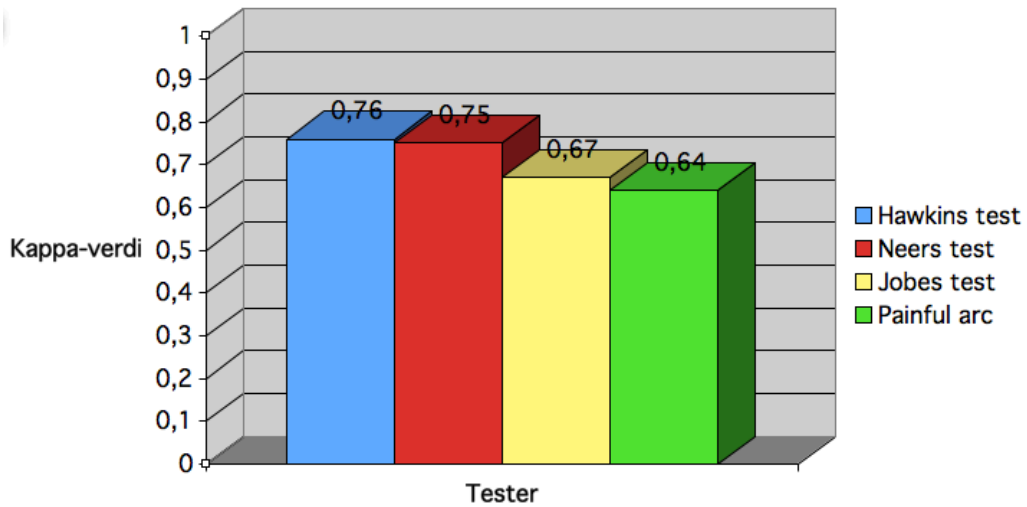
På grunnlag av krysstabellene i det foregående avsnittet presenteres kappa-verdier med tilhørende 95% konfidensintervaller, samt prosentvis enighet, i Tabell VII nedenfor.

Tabell VII: Resultater – kappastatistikk og prosentvis enighet

Test		Kappa (95% CI)	Prosentvis enighet
Alle deltakere	<i>Hawkins test</i>	0.76 (0.55 – 0.98)	87,9%
	<i>Neers test</i>	0.75 (0.51 – 0.98)	87,9%
vurdert samlet	<i>Jobes test</i>	0.67 (0.41 – 0.93)	84,8%
	<i>Painful arc</i>	0.64 (0.40 – 0.88)	81,8%

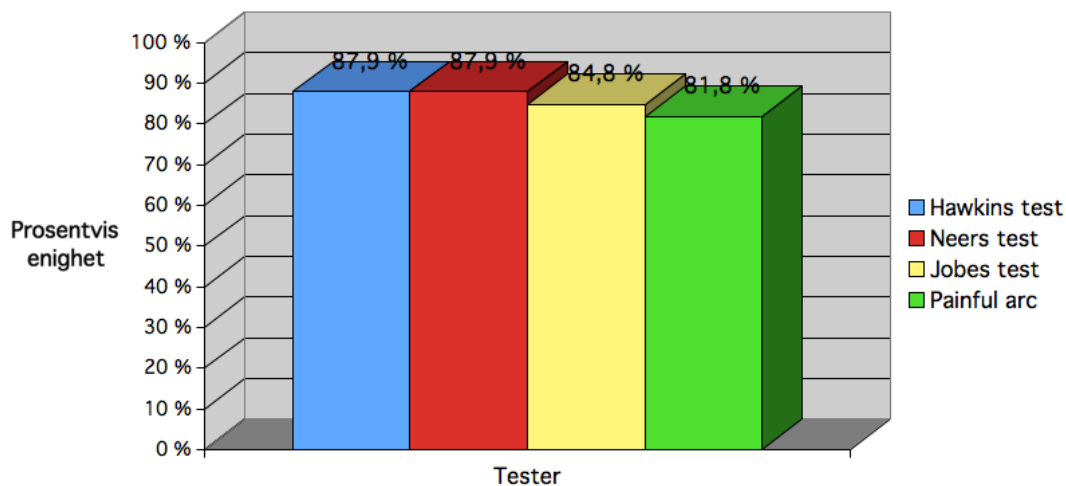
Hver enkelt test fikk relativt høy kappaverdi når skårene til samtlige deltakere ble inkludert i analysene, fra 0.64 til 0.76. Painful arc fikk lavest enighet med kappa lik 0.64 mens Hawkins test test fikk høyest verdi med kappa lik 0.76. Enighet ved Neers test og Jobes test var beliggende mellom Painful arc og Hawkins test, med kappa-verdier på henholdsvis 0.75 og

0.67. Generelt gjaldt det at en fant vide konfidensintervaller, med en bredde på om lag 0.5, for alle testene.



Figur 6: Grad av intertester-reliabilitet undersøkt med Kappa-statistikk, for alle gruppene samlet

Den prosentvise enigheten var generelt høy, i størrelsesorden 80 – 90%. Høyest prosentvis enighet fant en for Hawkins test og Neers test, mens den laveste prosentvise enigheten ble funnet for Painful arc.



Figur 7: Grad av intertester-reliabilitet undersøkt med prosentvis enighet, for alle gruppene samlet

5. DISKUSJON

5.1 Hovedfunn i studien

Hensikten med studien var å undersøke intertester-reliabiliteten for et utvalg tester som benyttes for å diagnostisere skulderimpingement, uttrykt ved grad av samsvar mellom to terapeuter, siden høy grad av reliabilitet er en forutsetning for at en test skal være valid. I studien ble det funnet relativt høye kappa-verdier for samtlige tester, beliggende i intervallet mellom 0.64 og 0.76. Disse verdiene er i henhold til Landis og Koch (1977) uttrykk for at det forelå betydelig enighet mellom testerne for alle testene, mens en ved å legge Fleiss' tolkning til grunn (De Vet et al. 2011) fant god grad av enighet ved Hawkins test og Neers test, men kun brukbar enighet for Jobs test og Painful arc. For alle testene gjaldt det at usikkerheten ble relativt stor, som kan skyldes det lave deltakerantallet i studien. Dette ga seg til uttrykk gjennom vide konfidensintervaller.

I det følgende vil funnene fra studien diskuteres opp mot tidligere studier og annen litteratur. Tidligere studier har vært forskjellige både med tanke på metode og design, noe som gjør det vanskelig å kartlegge hvilke faktorer som er av betydning for enighet. Mulige sammenhenger vil imidlertid trekkes frem. Hvis ikke annet er angitt i diskusjonen refererer enighet så vel som grad av enighet til testens kappa-verdi.

5.2 Funn på tester sett i lys av tidligere studier

5.2.1 Hawkins test

I denne studien ble kappaverdien for intertester-reliabilitet ved Hawkins test funnet å være 0.76 når alle deltakerne ble vurdert samlet. Intertester-reliabiliteten ved Hawkins test ble i det innledende litteratursøket funnet å være undersøkt i ni tidligere studier, og i disse studiene ble det rapportert kappa-verdier mellom 0.07 (Nørregaard et al. 2002) og 0.91 (Johansson og Ivarson 2009). Til tross for at Hawkins test ble funnet å være en av testene med størst spredning i enighet, fant en at hovedtyngden av rapportert enighet i tidligere studier lå i intervallet mellom 0.4 og 0.6.

Studiene som rapporterte høyest grad av enighet (Johansson og Ivarson (2009), Vind et al. (2011)) inkluderte i hovedsak unge individer. Johansson og Ivarson (2009) hadde i sin studie en gjennomsnittlig alder på 32 år mens alderen i studien til Vind et al. (2011) i gjennomsnitt lå på 20 år. Dette stod i kontrast til majoriteten av de tidligere studiene, hvor gjennomsnittlig

alder lå mellom 40 til 50 år, og hvor det samtidig ble rapportert om liten grad av enighet. I seg selv kan dette indikere at alder var av betydning for grad av enighet. Denne studien fant imidlertid en høy grad av enighet, til tross for at gjennomsnittlig alder var på 43 år. En fant følgelig ikke entydige data som tydet på at alder var av betydning for graden av enighet ved Hawkins test. Hvordan resultatene ville vært med deltakere i 60 årene vet vi imidlertid ikke.

Også varighet av skuldersmerter kan tenkes å ha spilt en rolle for grad av enighet. I majoriteten av tidligere studier rapporterte deltakerne om skulderplager av mange års varighet, samtidig som graden av enighet var lav. Deltakerne i studien til Johansson og Ivarson (2009), hvor det ble funnet høyest grad av enighet, rapporterte imidlertid om skulderplager av betydelig kortere varighet. I studien til Vind et al. (2011), som fant den nest høyeste graden av enighet, var sågar ikke skuldersmerter et inklusjonskriterie – eneste kravet til deltakerene var at de måtte være kastutøvere, og det ble i etterkant av inklusjonen bestemt om pasienten hadde skulderimpingement eller ei på bakgrunn av kliniske tester. Utvalgene i studiene med høyest og lavest grad av enighet var følgelig ikke direkte sammenlignbare hva smertevarighet angår. En ser imidlertid at det forelå en tendens til høyere grad av enighet der hvor utvalgene hadde smerter av kortere varighet, og motsatt lavere grad av enighet der hvor en hadde smerter av mer langvarig karakter. Resultatene fra denne studien harmonerer også med en slik trend, da en fant relativt høy grad av enighet samtidig som smertevarigheten var kortere enn for de tidligere studiene hvor det ble rapportert lavere grad av enighet.

Utvalgenes sammensetning kan også ha spilt en rolle. I denne studien var prevalensen av syke 52% mens den i studien til Vind et al. (2011) var 50%. For øvrig rekrutterte de andre studiene utelukkende pasienter med skulderpatologi og følgelig var sykdomsprevalensen 100%. Vi fant i denne studien at det var lettere å klassifisere friske enn syke individer likt mellom testerne, jamfør krysstabellene i resultatdelen av oppgaven. Dette medfører at prosentvis enighet i studiene med utelukkende skuldersyke individer sannsynligvis ville bli lavere enn for tilsvarende studier hvor også friske ble inkludert i utvalget, gitt alt annet likt. Når kappa-verdier fra ulike studier skal vurderes opp mot hverandre er det imidlertid mest interessant å sammenligne kappa-verdiene i studier med omtrent samme sykdomsprevalens. Dersom en stor andel av deltakerne tester negativt (friske) vil konsekvensen kunne bli at prosentvis enighet blir høy, mens kappa samtidig blir lav (Feinstein og Cicchetti 1990, Kundel og Polansky 2003) siden kappa kun viser enighet utover tilfeldighet (Kundel og Polansky 2003). Da sykdomsprevalensen i hovedsak er forskjellig mellom denne studien og majoriteten av de

tidligere studiene kan det være vanskelig å sammenligne kappa-verdiene direkte, siden kappa i stor grad påvirkes av dette (Kundel og Polansky 2003).

Det kan også være nyttig å dvele med hvor utvalgene ble rekruttert fra. Både denne studien, så vel som studiene til Johansson og Ivarson (2009) og Nanda et al. (2008), rekrutterte pasienter fra private klinikker og førstelinjetjenesten. Vind et al. (2011) rekrutterte på sin side tilsynelatende skulderfriske kastutøvere. Felles for disse nevnte studiene var at de fant høy grad av enighet, samtidig som individer med nakkelidelser og revmatiske lidelser ble ekskludert. Majoriteten av studiene som fant vesentlig lavere grad av enighet rekrutterte derimot pasienter fra revmatologiske og ortopediske avdelinger, hvor eksklusjonskriteriene samtidig var mindre snevre. Det kan tenkes at pasientene rekruttert fra spesialisthelsetjenesten presenterte et annerledes og gjerne mer sammensatt symptombylle, og hvor dette igjen gav seg til uttrykk gjennom lavere grad av enighet mellom testere.

Testutførelse kan også ha vært av betydning for grad av enighet. Én variabel som bør trekkes frem er testens utgangsstilling. Hawkins test er en provokasjonstest (Yamamoto et al. 2009) og den antatt viktigste provokative komponenten er innadrotasjon av skulderen, slik at trykket øker mellom rotatorcuff og mediale skulderstrukturer, da spesielt ligamentum coracoacromiale (Valadie et al. 2000). Biomekanikken ved test kan da være forskjellig, avhengig av om pasientens overarm i testposisjonen befinner seg i sagittalplanet eller i scapulas plan. Mens enkelte tidligere studier undersøkte enighet med overarmen i scapulas plan (Razmjou et al. 2004) benyttet majoriteten av tidligere studier, inkludert denne studien, en testposisjon hvor skulderen befant seg i sagittalplanet. Det var kun Nørregaard et al. (2002) som undersøkte enighet ved begge disse testposisjonene, og som gjør det mulig å undersøke konsekvensen av endret testposisjon for enighet direkte. Funnene med tanke på testposisjon var imidlertid sprikende. Nørregaard et al. (2002) fant riktignok at enighet var lavere når testen ble utført i sagittalplanet. Dette kan indikere at høyest grad av enighet blir oppnådd i scapulas plan. På den annen side ble det i studiene hvor det ble testet i sagittalplanet, inkludert denne studien, også funnet gjennomgående høy grad av enighet. I flere av disse studiene fant en videre høyere grad av enighet enn hva Nørregaard et al. (2002) fant i scapulas plan. Det blir vanskelig å si noe om hvorvidt majoriteten av studiene, herunder denne studien, ville fått en annen grad av enighet dersom de hadde valgt å teste i scapulas plan i stedet for sagittalplanet. Siden Nørregaard et al. (2002) imidlertid rapporterte endret enighet ved ulike testposisjoner, indikerer dette imidlertid at testposisjon kan ha betydning for grad av enighet.

Også standardisering kan tenkes å ha vært av betydning. Studiene til Johansson og Ivarson (2009) og Vind et al. (2011) fant høyest grad av enighet, samtidig som de var blant studiene som hadde gjennomført en omfattende standardisering og samtrening før studiestart. Siden alle studiene imidlertid beskrev at standardisering hadde funnet sted, men uten å gjøre nærmere rede for type og omfang, blir det vanskelig å konkludere med konsekvens for enighet. I denne studien valgte en å ikke standardisere utover felles testprotokoll, for å sikre økologiske validitet. Til tross for dette fant en kappa på 0.76. Dette kan indikere at grad av standardisering for Hawkins test er mindre viktig for å få brukbar grad av enighet, men en kan heller ikke se bort fra at enigheten i denne studien ville blitt enda høyere dersom standardiseringen hadde vært mer omfattende.

5.2.2 Neers test

I denne studien fant en at Neers test hadde en kappaverdi på 0.75 når alle deltakerne ble sett under ett. Grad av enighet ved Neers test ble i det innledende litteratursøket funnet å være kartlagt gjennom seks tidligere studier. Disse studiene fant stor grad av spredning. Kappa ble rapportert å ligge mellom 0.1 (Nanda et al. 2008) og 1.0 (Johansson og Ivarson 2009) og med verdier relativt godt spredt utover skalaen. Johansson og Ivarson (2009) og Vind et al. (2011) fant den høyeste graden av enighet, på tilsvarende måte som for Hawkins test, med kappaverdi ved Neers test på henholdsvis 1.0 og 0.86/0.95.

Også for Neers test kom resultatene fra studiene til Johansson og Ivarson (2009) samt Vind et al. (2011) best ut. Det er imidlertid interessant at Nanda et al. (2008) som sammen med Johansson og Ivarson (2009) og Vind et al. (2011) fant høyest grad av enighet ved Hawkins test, kun fant en kappaverdi på 0.1 for Neers test, til tross for at deltakerne ble rekruttert fra primærhelsetjenesten på samme måte som i studien til Johansson og Ivarson (2009). Johansson og Ivarson (2009) rekrutterte imidlertid unge deltakere med en gjennomsnittlig alder på 32 år, mens utvalget i studien til Nanda et al. (2008) var betydelig eldre, med en gjennomsnittlig alder på 52 år. Dette kan indikere at deltakerens alder muligens var av større betydning for Neers test enn for Hawkins test, siden forskjellen i deltakernes alder ikke fikk den samme konsekvensen, i form av redusert enighet, for Hawkins test som for Neers test.

Gitt at forskjell i alder gir endret grad av enighet for Neers test og ikke for Hawkins test, kan dette tentativt begrunnes biomekanisk. Biomekanikken ved Hawkins test og Neers test er ulik (Valadie et al. 2000), til tross for at begge betegnes som provokasjonstester (Yamamoto et al.

2009). Mens Hawkins test hovedsakelig øker belastning mellom subscapularis-sene og ligamentum coracoacromiale (Yamamoto et al. 2009) vil Neers test i større grad gi avklemming av rotatorcuff mellom mediale acromion og glenoid (Valadie et al. 2000). Når en samtidig vet at grad av degenerasjon i rotatorcuffen ses i relasjon med økende alder, enten på grunn av slitasje eller intrinsisk degenerasjon (Mitchell et al. 2005) kan det virke plausibelt at grad av enighet ved Neers test i større grad lar seg påvirke av økende alder enn ved Hawkins test.

Hva angår sammensetning av utvalgene var det kun denne studien, i tillegg til studien til Vind et al. (2011), som for Neers test inkluderte både syke og friske. Begge disse studiene fant blant de høyeste rapporterte gradene av enighet. På samme måte som for Hawkins test kan det være utfordrende å vite hvilken enighet som ville blitt oppnådd i disse studiene dersom utvalg kun bestående av syke deltakere hadde blitt benyttet. Tentativt ville imidlertid prosentvis enighet blitt noe lavere, siden en i denne studien fant høyere grad av prosentvis enighet blant de friske enn de syke isolert sett, jamfør krysstabellene i resultatdelen. Hva gjelder vurdering av kappa-verdier får en samme problematikk som skissert for Hawkins test, ved at det er mest interessant å sammenligne kappa-verdien for utvalg med lik sykdomsprevalens (Kundel og Polansky 2003).

Testutførelse kan også tenkes å ha påvirket enigheten. Denne studien, så vel som Michener et al. (2009), Razmjou et al. (2004), Johansson og Ivarson (2009) og Vind et al. (2011) benyttet en testprosedyre hvor armen ble passivt elevert i sagittalplanet. Nanda et al. (2008) anvendte derimot en prosedyre hvor armen ble elevert i scapulas plan. Siden testutførelsen til Nanda et al. (2008) i stor grad var avvikende fra de andre studiene, og hvor det samtidig ble funnet en betydelig avvikende grad av enighet, er det nærliggende å tro at utførelsen i det minste kan forklare deler av avviket som ble funnet i grad av enighet.

Graden av standardisering var kun i liten grad beskrevet. En hadde imidlertid god beskrivelse av omfattende standardisering for studiene med høyest grad av enighet. Siden denne studien imidlertid fant relativt høy grad av enighet ved Neers test uten at standardiseringen i forkant var særlig omfattende kan dette indikere at for Neers test var standardisering ikke av stor betydning. Dog er det sannsynlig at enigheten ved denne studien hadde blitt enda høyere enn 0.75 dersom en hadde valgt å legge en omfattende standardisering til grunn, siden denne

studien på mange måter er lik studiene til både Vind et al. (2011) og Johansson og Ivarson (2009) som fant en enda høyere grad av enighet.

5.2.3 Jobes test

I denne studien fant en at grad av enighet for testen var relativt høy, med en kappaverdi på 0.67, når alle deltakerne ble vurdert samlet. Grad av enighet ved Jobes test ble i det innledende litteratursøket funnet å være kartlagt gjennom fem tidligere studier. For denne testen fant en noe annerledes kappa-verdier enn for Hawkins test og Neers test, ved at rapportert enighet ikke var spredt fra 0 til 1, men snarere befant seg på den øvre delen av skalaen, fra 0.44 (Nanda et al. 2008) til 0.9 (Vind et al. 2011). I tillegg var rapportert enighet samlet i klustre rundt disse to verdiene.

Enighet kan blant annet være påvirket av hvordan utvalgene var sammensatt, på samme måte som skissert for Hawkins test og Neers test. Også for Jobes test var det kun denne studien, samt studien til Vind et al. (2011), som inkluderte friske deltakere i utvalget. Det kan være vanskelig å skulle sammenligne de forholdsvise høye kappaverdiene til denne studien og studien til Vind et al. (2011) med majoriteten av tidligere studier, siden dette vanskelig kan gjøres på en meningsfull måte da sykdomsprevalensen er forskjellig mellom utvalgene (Kundel og Polansky 2003).

Gjennomsnittlig alder kan også tenkes å ha vært av betydning for grad av enighet. Vind et al. (2011) og Johansson og Ivarson (2009) fant høyest grad av enighet samtidig som utvalgene hadde lavest gjennomsnittsalder. Både denne studien, i tillegg til de andre tidligere studiene, rapporterte lavere grad av enighet samtidig som gjennomsnittlig alder var høyere. Dette understøtter en hypotese om at det for Jobes test var en sammenheng mellom alder og grad av enighet.

Spriket i rapporterte kappa-verdier kan imidlertid også ha oppstått på grunn av andre årsaker enn alder og utvalgets sammensetning, og det er spesielt én interessant forskjell mellom studiene det bør dveles med. Selve utgangstillingen for testen var lik i alle studier. Kriterier for positiv test var imidlertid forskjellig. Nanda et al. (2008) og Michener et al. (2009) definerte positiv test på grunnlag av svakhet, og fant samtidig den laveste graden av enighet. Denne studien, så vel som studiene til Johansson og Ivarson (2009) og Vind et al. (2011), fant imidlertid den høyeste graden av enighet samtidig som testen ble skåret med utgangspunkt i

smerter/ikke smerter. På grunnlag av denne og tidligere studier fant en følgelig en trend hvor grad av enighet ved Jobes test kunne tenkes å være større dersom smerter heller enn svakhet ble lagt til grunn.

Mekanismene bak sistnevnte tentative sammenheng kan diskuteres med utgangspunkt i Neers (1972) inndeling av impingement. Det første stadiet kjennetegnes av hevelse av skulderstrukturer, i det påfølgende stadiet finner en fortykkelse og slitasje av rotatorcuffen, mens en i det siste stadiet har beinete endringer og store rifter eller rupturer av rotatorcuffen (Neer 1972). En konsekvens av denne inndelingen kan bli at dersom en kun vurderer smerte som kriterie for positiv test vil dette enklere kunne fange opp tilstandene i de tidlige fasene, og fortrinnsvis da hos yngre individer hvor degenerasjon av rotatorcuff med påfølgende kraftreduksjon er mindre uttalt. Videre kan en tenke seg at svakhet blir mer gjeldende hos individer hvor det foreligger uttalt degenerasjon og rupturer (Bahr og Mæhlum 2003), og en skåring basert på svakhet kan da tenkes å være bedre egnet hos eldre individer. Datagrunnlaget fra denne og tidligere studier er imidlertid for lite til å kunne trekke konklusjoner om hvorvidt sammenhengen mellom type skåring, utvalgets alder/degenerasjon og grad av enighet er spuriøs eller ei.

Hva standardisering angår ser en at alle tidligere studier har benyttet dette, til tross for at Michener et al. (2009) kun i liten grad har gjort rede for hvilke elementer som har inngått. Denne studien valgte å ikke standardisere i særlig grad, og kappaverdien som ble funnet var da også lavere enn de sammenlignbare studiene til Johansson og Ivarson (2009) og Vind et al. (2011) med tanke på egenskaper ved deltakerne som alder, symptomvarighet og hvor de var rekruttert fra. Det kan følgelig tenkes at denne studien hadde fått en enda høyere grad av enighet dersom standardiseringen hadde vært mer omfattende.

5.2.4 Painful arc

I denne studien fant en kappaverdien for Painful arc til å bli 0.64 når alle deltakerne ble vurdert samlet. Enigheten ved Painful arc ble i det innledende litteratursøket funnet å være undersøkt i fem tidligere studier, og på tilsvarende måte som for Jobes test lå de rapporterte kappaverdiene i intervallet mellom 0.45 (Michener et al. 2009) og 0.93 (Palmer et al. 2000). Imidlertid fant en for Painful arc at majoriteten av kappaverdiene befant seg i området mellom 0.4 og 0.5. Kappaverdien fra denne studien plasserte seg følgelig inne blant de eksisterende

rapporterte verdiene, nær men noe høyere enn graden av enighet rapportert i majoriteten av de tidligere studiene.

Palmer et al. (2000) fant imidlertid en kappaverdi som var vesentlig høyere enn alle andre tidligere studiene, inkludert denne studien. Det vil da være interessant å kartlegge hvilke faktorer som var spesielle for studien til Palmer et al. (2000), for å forklare hvorfor akkurat denne studien fant så stor grad av enighet. En forskjell mellom denne og de andre studiene kan være grad av standardisering. De fleste studiene har beskrevet at en standardisering har funnet sted, men uten å redegjøre for denne. Palmer et al. (2000) benyttet på sin side benyttet en omfattende standardisering, ved at det ble arrangert tolv seanser over seks uker inntil tilstrekkelig konsistens var oppnådd. Når vi i denne studien valgte å standardisere i lite grad, og til tross for dette fant en kappa lik 0.670, kan dette tyde på at høy grad av enighet ikke forutsetter omfattende standardisering. Imidlertid er det uvisst om kappaverdiene i denne studien hadde blitt høyere, dersom standardiseringen hadde vært mer omfattende.

På samme måte som for de andre testene har studiene hatt ulike testprotokoller, og en kan tenke seg at også dette har påvirket graden av enighet. Det er dog et problem at testprotokollene i flertallet av de tidligere studiene var mangelfullt beskrevet, og mens Michener et al. (2009) og Nanda et al. (2008) utførte testen i frontalplanet beskrev verken Nomden et al. (2009), Walker-Bone et al. (2002) eller Palmer et al. (2000) testen i detalj. Walker-Bone et al. (2002) og Palmer et al. (2000) valgte imidlertid begge å legge The Southampton Examination Schedule til grunn for studiene sine. Siden disse to studiene fant svært forskjellig grad av enighet til tross for at den samme protokollen lå til grunn, peker dette på at testprosedyre i seg selv ikke var av avgjørende betydning for rapportert grad av enighet.

Hva gjelder egenskaper ved deltakerene hadde alle studiene som undersøkte Painful arc et utvalg hvor gjennomsnittlig alder var i 40-årene. Smertevarighet var dog noe ulik blant studiene, og for studiene som rapporterte høyest grad av enighet, Palmer et al. (2000) og Nanda et al. (2008), var smertevarigheten ikke oppgitt. Siden studiene med lavest grad av enighet både bestod av deltakere med langvarige smerter (Michener et al. 2009) så vel som deltakere med smerter av kortere varighet (Nomden et al. 2009) blir det vanskelig å konkludere med hvilket bidrag smertevarighet fikk for grad av enighet for Painful arc.

Hvor pasientene ble rekruttert fra er imidlertid interessant, siden studien til Palmer et al. (2000), hvor deltakerne var rekruttert fra revmatologiske og ortopediske avdelinger, fant den største graden av enighet. Dette står i motsetning til Hawkins test, Neers test og Jobes test, hvor graden av enighet generelt ble funnet å være lavere i studiene som rekrutterte fra sekundærhelsetjenesten kontra primærhelsetjenesten. Hvorvidt denne sammenhengen er reell eller spuriøs er det for lite grunnlag i datamaterialet til å si noe om – spesielt siden studiene til Johanson og Ivarson (2009) og Vind et al. (2011), som fant den høyeste graden av enighet ved Hawkins test, Neers test og Jobes test, ikke undersøkte reliabiliteten ved Painful arc.

Avslutningsvis bør det nevnes at for studiene som undersøke enighet ved Painful arc var det kun denne studien som inkluderte skulderfriske/asymptomatiske individer i utvalget. Vi fant for Painful arc en kappaverdi på 0.65, som var noe høyere enn majoriteten av andre studier. Hvorvidt denne differansen stammer fra utvalgenes sammensetning er uvisst, spesielt siden Palmer et al. (2000) fant den desidert høyeste graden av enighet, til tross for at kun deltakere med skulderpatologi ble inkludert i studien. Følgelig kan ikke kappa-verdiene til denne studien og Palmer et al. (2000) sammenlignes direkte siden sykdomsprevalensen er ulik (Kundel og Polansky 2003).

5.3 Metodekritikk og begrensninger ved studien

Denne studien har en rekke begrensninger. En åpenbar begrensning ligger for det første i at skuldersmerter ikke nødvendigvis skyldes enten impingement eller annen skulderpatologi, begge tilstandene kan også foreligge samtidig. På tilsvarende måte er det også mulig at skulderpatologi og/eller degenerasjon foreligger i skulderen, uten at pasienten har plager eller smerter forbundet med dette. Studien til Vind et al. (2011) illustrerer dette, ved at det i denne studien ble rekruttert tilsynelatende skulderfriske individer som senere ble klassifisert som enten frisk eller syke, basert på tester for impingement. Dette viser at de skarpe linjene mellom de tre gruppene i studien kan utgjøre en problematisk forenkling av virkeligheten.

Et annen svakhet ved studien er at det endelige utvalget (n = 33) bestod av 16 skulderfriske deltakere og 17 deltakere med skulderpatologi. Når vi i denne studien ønsket 50 deltakere basert på anbefalingene til De Vet et al. (2011) og ikke klarte å oppnå et slikt antall i løpet av studien, var konsekvensen at graden av presisjon for kappaverdiene ble redusert.

En ønsket videre i utgangspunktet å rekruttere like mange individer til de tre gruppene med skulderfriske, impingement i skulder samt annen skulderpatologi. Et stort antall deltakere med skulderimpingement samt mange deltakere med annen skulderpatologi var begrunnet i et ønske om å etterligne en klinisk situasjon, hvor terapeut ikke vet hvilken patologisk prosess som ligger til grunn for pasientens smerter. Siden rekrutteringen til studien foregikk parallelt med testingen og det ble vanskeligere å styre antallet deltakere i de ulike gruppene, fikk en dessverre kun rekruttert fire pasienter med annen skulderpatologi, mot 13 deltakere med skulderimpingement og 16 skulderfriske deltakere. Dette er en svakhet ved studien, siden en fra krysstabellene ser en tendens til at prosentvis enighet mellom testerne er dårligere blant pasientene med annen skulderpatologi enn for pasientene med impingement.

Det store antallet skulderfriske inkludert i studien ble begrunnet med at denne studien inngikk som del av en større studie hvor også testen diskriminerende egenskaper ble vurdert, sett opp mot ultralyd som en gullstandard. En kan avlese fra krysstabellene i resultatdelen at de skulderfriske i hovedsak ble testet negativt av begge testerne. Konsekvensen ble at de friske deltakerne isolert sett fikk en svært høy prosentvis enighet, som igjen sannsynligvis har ført til at den totale prosentvise enigheten for hele utvalget sett under ett kan ha blitt høyere enn dersom en kun hadde regnet ut prosentvis enighet for deltakerne med skulderimpingement og/eller annen skulderpatologi. Hva kappa-verdier angår vil disse i stor grad påvirkes av sykdomsprevalens (Kundel og Polansky 2003) og det blir følgelig vanskelig å sammenligne kappa-verdier fra denne studien, hvor sykdomsprevalens ble 52%, med andre studier hvor 100% av utvalget var syke. At det blir vanskelig å sammenligne kappa-verdiene i denne studien med majoriteten av tidligere studier direkte basert på ulik sykdomsprevalens illustrerer utfordringen ved å ene og alene bruke kappa-koeffisienter som uttrykk for grad av enighet. Av denne grunnen argumenterte Cicchetti og Feinstein (1990) for at en i vurdering av enighet både bør vurdere kappa-verdi, samt positiv enighet og negativ enighet. Med positiv enighet menes her antallet tester som begge testerne angir som positive, dividert på antall positive tester for begge testerne. Tilsvarende gjelder for negativ enighet, som er antallet tester som begge testerne angir som negative, dividert på antall negative tester for begge testerne. At denne studien kun vurderte enighet med utgangspunkt i kappa-verdier blir følgelig en svakhet.

Det bør videre nevnes at vi fant at det ikke var uproblematisk å skåre testene med utgangspunkt i et dikotomt utfallsmål. Dikotomier er svært grove inndelinger, og det viste seg

i mange tilfeller vanskelig å skåre testene som positiv eller negativ basert på smertekarakteristika og lokalisering.

En annen utfordring ved studien var at det gikk noe tid fra pasienten ble undersøkt med ultralyd, til vedkommende ble inkludert i studien og testet av terapeutene, uten at vi hadde klar kontroll over hva pasienten foretok seg i dette tidsrommet, ei heller hva som kan ha skjedd av spontan tilheling i skulderen i perioden. Tilsvarende ble heller ikke pasientens medikamentbruk eller aktivitetsnivå på tidspunkt for diagnostisering eller testing nedtegnet. Følgelig hadde vi ikke kontroll over hvilke konsekvenser endringer i medikamentbruk eller aktivitetsnivå kan ha fått for smerte eller funksjon på tidspunktene for diagnose og testing.

En ytterligere svakhet knyttet seg til selve rekrutteringen, da det dessverre ikke lot seg gjøre å nedtegne hvor stor andel av pasientene som ble invitert til å delta i studien, som faktisk tok kontakt med forskergruppen for å bli inkludert i denne. Det ville for studiens del vært interessant å kartlegge hvor stor del av pasientene som valgte å ikke delta, og om det eventuelt var store forskjeller mellom gruppene som valgte å takke ja, i forhold til de som ikke ønsket å delta i studien.

Hva generalisering angår så vil resultatene en fant i denne studien strengt tatt kun være gyldige for individene i dette aktuelle kliniske utvalget, men en kan med 95% sikkerhet si at den sanne verdien i den kliniske populasjonen befinner seg innenfor konfidensintervallet som en måtte finne i utvalget. En kan imidlertid ikke generalisere til en større populasjon enn den kliniske populasjonen, da det foreligger en risiko for at ikke alle relevante grupper er inkludert i utvalget og at utvalget følgelig blir skjevt (Jacobsen 2005). I denne forbindelse må det da trekkes frem at pasientene ble rekruttert fra kun én lege på én privat skulderklinikk, antallet individer er relativt begrenset ($n = 33$) rekruttering fant sted i løpet av en kort periode i en varm periode av året, fra mai til september 2012. Mulighet til å generalisere til den større populasjonen blir følgelig begrenset.

5.4 Områder med behov for videre forskning

Tidligere studier er i stor grad av dårlig metodisk kvalitet, og antallet studier som både inkluderer friske og syke deltakere i utvalget er begrenset. Denne studien er den første intertester-reliabilitetsstudien ved kliniske tester for skulderimpingement, som både inkluderer skulderfriske deltakere i utvalget og i tillegg benytter ultralyd for å sikre at det

både er pasienter med impingement og andre skulderlidelser med blant deltakerne. Et slikt design gjør det mulig å sikre seg at pasientens plager skyldes en reell impingement, samtidig som en ved å også inkludere skulderfriske deltakere i utvalget unngår at tester blir systematisk farget av pasientens sykdomsstatus. For fremtiden er det behov for flere forsøk med hensikt å undersøke enkeltvariablers konsekvens for grad av enighet, herunder egenskaper ved både deltaker, tester og testutførelse.

I tillegg til metodiske utfordringer ved tidligere studier er det en utfordring at alle tidligere studier, inkludert denne, har vurdert skuldertestene basert på skåring i form av positiv eller negativ test. Det vil for fremtiden være svært interessant å undersøke om enigheten påvirkes dersom en i større grad velger å bruke data på ordinal- eller skala-nivå i skåringen, herunder gjennom bruk av smertegradering ved VAS-skala, fremfor utelukkende bruk av dikotome skåringer.

Avslutningsvis bør det nevnes at denne studien, så vel som tidligere studier, kun har vurdert intertester-reliabilitet uttrykt ved grad av enighet mellom to testere. Det vil for fremtiden også være interessant å undersøke grad av enighet mellom et større antall testere.

6. KONKLUSJON

I studien ble det funnet god intertester-reliabilitet for Hawkins test, Neers test, Jobs test og Painful arc mellom to testere uttrykt ved Cohens kappa verdier mellom 0.64 (Painful arc) og 0.76 (Hawkins test). Det betyr at to testere som bruker samme standardiserte testprotokoll kan oppnå tilfredsstillende samsvar i skåringer med de fire testene selv om de ikke har gjennomført en omfattende samtesting på forhånd.

Denne studien inngår som en del av en større studie hvor også test-retest og diagnostisk validitet blir søkt kartlagt. For å gi både denne delstudien og den større studien økt styrke, vil datainnsamlingen fortsette slik at en oppnår et høyere antall deltakere. Det vil i den videre datainnsamlingen være spesielt ønskelig å rekruttere flere pasienter med annen skulderpatologi, siden en en behøver et stort antall med så vel impingement som annen skulderpatologi for å kunne undersøke testenes diskriminerende egenskaper i den større studien.

7. REFERANSELISTE

Allen GM. Shoulder ultrasound imaging – integrating anatomy, biomechanics and diseases processes. *Eur J Radiol* 2008; 68(1): 137-46.

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.

Bahr R, Mæhlum S. *Idrettsskader*. 2.utgave. Gazette bok, Oslo 2003.

Behrens SB, Compas J, Deren ME, Drakos M. Internal impingement: a review on a common cause of shoulder pain in throwers. *Phys Sportsmed* 2010; 38(2): 11-8.

Brox JI, Sunde P, Schrøder CP, Engebretsen K, Skare Ø, Ekeberg OM, Juel NG. Non-traumatic shoulder pain. *Tidsskr Nor Laegeforen* 2010; 130(21): 2132-5.

Bureau NJ, Beauchamp M, Cardinal E, Brassard P. Dynamic sonography evaluation of shoulder impingement syndrome. *Am J Roentgenol*. 2006; 187(1): 216-20.

Buss DD, Freehill MQ, Marra G. Typical and atypical shoulder impingement syndrome: diagnosis, treatment, and pitfalls. *Instr Course Lect* 2009; 58: 447-57.

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.

Calis M, Akgun K, Calis H, Tuzun F. Diagnostic values of clinical diagnostic tests in subacromial impingement syndrome. *Am Rheum Dis* 2000; 59: 44-7.

Carter RE, Lubinsky J, Domholdt E. *Rehabilitation Research. Principles and applications*. 4 utgave. Elsevier Saunders, St.Louis 2011.

Cicchetti D, Feinstein A. High agreement but low kappa. II. Resolving the paradoxes. *J Clin Epidemiol* 1990; 43: 551-8.

Desmeules F, Minville L, Riederer B, Côté CH, Frémont P. Acromio-humeral distance variation measured by ultrasonography and its association with the outcome of rehabilitation for shoulder impingement syndrome. *Clin J Sports Med* 2004; 14(4): 197-205.

De Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine*. Cambridge University Press, Cambridge 2011.

Feinstein A, Cicchetti D. High agreement but low kappa. I. The problem of two paradoxes. *J Clin Epidemiol* 1990; 43: 543-9.

Flatow EL, Soslowsky LJ, Ticker JB, Pawluk RJ, Hepler M, Ark J, Mow VC, Bigliani LU. Excursion of the rotator cuff under the acromion. Patterns of subacromial contact. *Am J Sports Med*. 1994; 22(6): 779-88.

Fodor D, Poanta L, Felea I, Rednic S, Bolosiu H. Shoulder impingement syndrome: correlations between clinical tests and ultrasonographic findings. *Ortop Traumatol Rehabil* 2009; 11(2): 120-6.

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

Harrison AK, Flatow EL. Subacromial impingement syndrome. *J Am Acad Orthop Surg* 2011; 19(11): 701-8.

Hattam P, Smeatham A. *Special tests in musculoskeletal examination*. Elsevier Churchill Livingstone, Edinburgh 2010.

Hébert LJ, Moffet H, Dufor M, Moisan C. Acromiohumeral distance in a seated position in persons with impingement syndrome. *J Magn Reson Imaging* 2003; 18(1): 72-9.

Hegedus EJ, Goode A, Campbell S, Morin A, Tamaddoni M, Moorman CT, 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.

Helseforskningsloven. Lov 2008-06-20 nr 44: Lov om medisinsk og helsefaglig forskning (helseforskningsloven). <http://www.lovdata.no/all/hl-20080620-044.html>, besøkt 19.09.2012.

Heyworth BE, Williams RJ 3rd. Internal impingement of the shoulder. *Am J Sports Med* 2009; 37(5): 1024-37.

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(4): 194-200.

Hughes PC, Green RA, Taylor NF. Measurement of subacromial impingement of the rotator cuff. *J Sci Med Sport* 2012; 15(1): 2-7.

Jacobsen DI. Hvordan gjennomføre undersøkelser? 2.utgave. Høyskoleforlaget, Kristiansand 2005.

Jia X, Ji JH, Pannirselvam V, Petersen SA, McFarland EG. Does a positive neer impingement sign reflect rotator cuff contact with the acromion? *Clin Orthop Relat Res.* 2011; 469(3): 813-8.

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.

Juel NG. Norsk fysikalsk medisin. 2.utgave. Fagbokforlaget, Bergen 2007.

Kayser R, Hampf S, Pankow M, Seeber E, Heyde CE. Validity of ultrasound examination of disorders of the shoulder joint. *Ultraschall Med* 2005; 26(4): 291-8.

Kelly SM, Brittle N, Allen GM. The value of physical tests for subacromial impingement syndrome: a study of diagnostic accuracy. *Clin Rehabil.* 2010 Feb; 24(2): 149-58.

Kessel L, Watson M. The Painful arc syndrome. Clinical classification as a guide to management. *J Bone Joint Surg Br* 1977; 59: 166-72.

Kim HA, Kim SH, Seo YI. Ultrasonographic findings of painful shoulders and correlation between physical examination and ultrasonographic rotator cuff tear. *Mod Rheumatol*. 2007; 17(3): 213-9.

Kumar P, Bradley M, Gray S, Swinkels A. Reliability and validity of ultrasonographic measurements of acromion-greater tuberosity distance in poststroke hemiplegia. *Arch Phys Med Rehabil* 2011; 92(5): 731-6.

Kundel HL, Polansky M. Measurement of observer agreement. *Radiology* 2003; 228: 303-8.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-74.

Magee DJ. *Orthopedic physical assessment*. 5. utgave. Saunders Elsevier 2008.

Malhi AM, Khan, R. Correlation between clinical diagnosis and arthroscopic findings of the shoulder. *Postgrad Med J* 2005; 960(81): 657-9.

Malvestiti O, Mariani C, Scorsolini A, Ratti F, Ferraris G, Columbaro G. Subacromial impingement syndrome and rotator cuff tear. Ultrasonography of 140 cases. *Radiol Med* 1997; 94(1-2): 37-42.

Mayerhoefer ME, Breitscheher MJ, Wurnig C, Roposch A. Shoulder impingement: relationship of clinical symptoms and imaging criteria. *Clin J Sports Med* 2009; 19(2): 83-9.

Michener LA, Walsworth MK, Doukas WC, Murphy KP. 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.

Mitchell C, Adebajo A, Hay E, Carr A. Shoulder pain: Diagnosis and management in primary care. *BMJ* 2005; 7525: 1124-8.

Moen MH, de Vos RJ, Ellenbecker TS, Weir A. Clinical tests in shoulder examination: How to perform them. *Br J Sports Med* 2010; 44(5): 370-5.

Moosmayer S, Heir S, Aaser P, Smith HJ. Ultrasound examination of the shoulder – a method description. *Tidsskr Nor Laegeforen* 2004; 124(2): 177-80.

Nanda R, Gupta S, Kanapathipillai P, Liow RYL, Rangan A. An assessment of the inter examiner reliability of clinical tests for subacromial impingement and rotator cuff integrity. *Eur J Orthop Surg Traumatol* 2008; 18: 495-500.

Neer CS 2nd. Anterior acromioplasty for the chronic impingement syndrome in the shoulder: a preliminary report. *J Bone Joint Surg Am.* 1972; 54(1): 41-50.

Nomden JG, Slagers AJ, Bergman GJ, Winters JC, Kropmans TJ, Dijkstra PU. Interobserver reliability of physical examination of the shoulder girdle. *Man Ther* 2009; 14(2): 152-9.

Nørregaard J, Krogsgaard MR, Lorenzen T, Jensen EM. Diagnosing patients with longstanding shoulder joint pain. *Ann Rheum Dis* 2002; 61: 646-9.

Palmer K, Walker-Bone K, Linaker C, Reading I, Kellingray S, Coggon D, Cooper C. The Southampton examination schedule for the diagnosis of musculoskeletal disorders of the upper limb. *Ann Rheum Dis* 2000; 59: 5-11.

Papadonikolakis A, McKenna M, Warme W, Martin BI, Matsen FA 3rd. Published evidence relevant to the diagnosis of impingement of the shoulder. *J Bone Joint Surg Am* 2011; 93(19): 1827-32.

Pappas GP, Blemker SS, Beaulieu CF, McAdams TR, Whalen ST, Gold GE. In vivo anatomy of the Neer and Hawkins sign positions for shoulder impingement. *J Shoulder Elbow Surg* 2006; 15(1): 40-9.

Park HB, Yokota A, Gill HS, El Rassi G, McFarland EG. Diagnostic accuracy of clinical tests for the different degrees of subacromial impingement syndrome. *J Bone Joint Surg Am* 2005; 87(7): 1446-55.

Payne LZ, Deng XH, Craig EV, Torzilli PA, Warren RF. The combined dynamic and static

contributions to subacromial impingement. *American Journal of Sports Medicine* 1997; 25(6): 801-8.

Pijls BG, Kok FP, Penning LI, Guldemond NA, Arens HJ. Reliability study of the sonographic measurement of the acromiohumeral distance in symptomatic patients. *J Clin Ultrasound* 2010; 38(3): 128-34.

Razmjou H, Holtby R, Myhr T. Pain provocative shoulder tests: Reliability and validity of the impingement tests. *Physiotherapy Can* 2004; 56: 229-36.

Salaffi F, Ciapetti A, Carotti M, Gasparini S, Filippucci E, Grassi W. Clinical value of single versus composite provocative tests in the assessment of painful shoulder. *J Clin Rheumatol* 2010; 16(3): 105-8.

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.

Solberg AS. *Klinisk undersøkelse av nakke-skulder*. Høyskoleforlaget, Kristiansand 2002.

Szyluk K, Jasinski A, Koczy B, Widuchowski W, Widuchowski J. Subacromial impingement syndrome – most frequent reason of the painful shoulder syndrome. *Pol Merkur Lekarski* 2008; 25(146): 179-83.

Tallia AF, Cardone DA. Diagnostic and therapeutic injection of the shoulder region. *Am Fam Physician* 2003; 67(6): 1271-8.

Tan RK. A review of the role of magnetic resonance imaging in the evaluation of shoulder impingement syndrome and rotator cuff tendon tears. *Ann Acad Med Singapore* 1998; 27(2): 243-7.

Teefey SA, Rubin DA, Middleton WD et al. Detection and quantification of rotator cuff tears: comparison of ultrasonographic, magnetic resonance imaging, and arthroscopic findings in seventy-one consecutive cases. *J Bone Joint Surg Am* 2004; 86-A: 708-16.

Ure BM, Tiling T, Kirchner R, Rixen D. Reliability of clinical examination of the shoulder in comparison with arthroscopy. A prospective study. *Unfallchirurg* 1993; 96(7): 382-6.

Valadie AL, Jobe CM, Pink MM, Ekman EF, Jobe FW. Anatomy of provocative tests for impingement syndrome of the shoulder. *J Shoulder Elbow Surg* 2000; 9(1): 36-46.

Van Der Windt DAWM, Koes BW, Boeke AJP, Devillé W, De Jong BA, Bouter LM. Shoulder disorders in general practice: prognostic indicators of outcome. *British Journal of General Practice* 1996; 46: 519-23.

Vind M, Bogh SB, Larsen CM, Knudsen HK, Søgaaard K, Juul-Kristensen B. Inter-examiner reproducibility of clinical tests and criteria used to identify subacromial impingement syndrome. *BMJ Open*. 2011 May 18; 1(1): e000042.

Walker-Bone K, Byng P, Linaker C, Reading I, Coggon D, Palmer KT, Cooper C. Reliability of the Southampton examination schedule for the diagnosis of upper limb disorders in the general population. *Ann Rheum Dis* 2002; 61: 1103-6.

Wuelker N, Roetman B, Roessig S. Coracoacromial pressure recordings in a cadaveric model. *J Shoulder Elbow Surg* 1995; 4(6): 462-7.

Yamamoto N, Muraki T, Sperling JW, Steinmann SP, Itoi E, Cofield RH, An KN. Impingement mechanisms of the Neer and Hawkins signs. *J Shoulder Elbow Surg* 2009; 18: 942-7.

Yamamoto N, Muraki T, Sperling JW, Steinmann SP, Itoi E, Cofield RH, An KN. Contact between the coracoacromial arch and the rotator cuff tendons in nonpathologic situations: a cadaveric study. *J Shoulder Elbow Surg* 2010; 19(5): 681-7.

Yanai T, Fuss FK, Fukunaga T. In vivo measurements of subacromial impingement: substantial compression develops in abduction with large internal rotation. *Clin Biomech* 2006; 21(7): 692-700.

Zaslav KR. Internal rotation resistance strength test: A new diagnostic test to differentiate intra-articular pathology from outlet (Neer) impingement syndrome in the shoulder. *J Shoulder Elbow Surg* 2001; 10(1): 23-7.

Östör AJK, Richards CA, Prevost AT, Speed CA, Hazleman BL. Diagnosis and relation to general health of shoulder disorders presenting to primary care. *Rheumatology* 2005; 44: 800-5.

VEDLEGG A: INFORMASJON TIL PASIENT OG INFORMERT SAMTYKKE

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 impingement) 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 [REDACTED] 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 [REDACTED].

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 rekruttert 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 skulderproblemene 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 [REDACTED] ved Dr. [REDACTED]

Datainnsamlingen vil foregå på [REDACTED]. 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.

Dr. [REDACTED] ved [REDACTED] er ansvarlig for rekruttering av pasienter.
[REDACTED] ved [REDACTED] 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 publisering av studien vil alle data være anonymisert.

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

Hvis du sier ja til å delta i studien, har du rett til å få innsyn i hvilke opplysninger som er registrert om deg. 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)

VEDLEGG B: GODKJENNING FRA REK



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

Vår referanse må oppgis ved alle henvendelser



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:
Forskningsansvarlig



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,
Nydaalen, 0484 Oslo

Telefon: 22845511
E-post:
post@helseforskning.etikk.no

Web:

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