

Intertester og test-retest reliabilitet for skåring av smertetegninger ved  
kroppsområde-analyse



Masteroppgave

9253 ord

Masterprogram i helsefag

Studieretning Klinisk masterstudium for fysioterapeuter i manuellterapi

Kandidatnummer: 223183

Universitetet i Bergen

Høst 2014

# Innholdsfortegnelse

---

1	Introduksjon .....	6
1.1	Bakgrunn for studien .....	6
1.2	Teori.....	7
1.2.1	Smerter .....	7
1.2.2	Smertetegning.....	7
1.2.3	Skåringsmetoder .....	8
1.2.4	Reliabilitet .....	9
1.3	Tidligere forskning på området.....	10
2	Hensikt og problemstilling .....	14
2.1	Problemstilling.....	14
3	Metode.....	15
3.1	Valg av forskningsdesign .....	15
3.2	Beskrivelse av materiale.....	15
3.3	Terapeutene .....	17
3.4	Gjennomføring av analyser.....	17
3.5	Reliabilitet.....	19
3.5.1	Kontinuerlige data .....	19
3.5.2	Kategoriske data .....	19
3.6	Etiske betraktninger og egen rolle som forsker .....	20
4	Resultater.....	21
4.1	Bakgrunnsvariabler .....	21
4.2	Reliabilitet ved kroppsområde-skåring av smertetegninger.....	22
4.3	Reliabilitet ved diskriminering mellom korsrygg og bekken .....	23
5	Diskusjon.....	24
5.1	Diskusjon av metode .....	24
5.2	Validitet.....	25
5.2.1	Intern validitet.....	25
5.2.2	Ekstern validitet.....	26
5.3	Sammenligning med tidligere studier.....	26
5.4	Implikasjoner for klinisk praksis .....	28
5.5	Videre forskning .....	29
6	Oppsummering og konklusjon.....	30
7	Litteraturliste .....	31
	Vedlegg 1.....	35
	Vedlegg 2.....	36



## **Sammendrag**

---

Intertester og test-retest reliabilitet for skåring av smertetegninger ved kroppsområde-analyse. Klinisk masterstudium i manuellterapi for fysioterapeuter, Institutt for samfunnsmedisinsk fag, Universitetet i Bergen, 2014

**Bakgrunn og hensikt:** Hensikten med studien var å undersøke intertester og test-retest reliabilitet til skåring smertetegninger, videre undersøkte studien intertester og test-retest reliabilitet til diskriminering av bekken og korsryggsmerter ved bruk av smertetegninger.

**Metode:** Sekundærdata (n=540) fra tidligere studier ble benyttet. Skåring av tegningene ble gjort ved tre anledninger ved bruk av enkel kroppsområde-analyse. Grad av samsvar i skåringer på smertetegningene ble undersøkt ved Intraclass correlation. Kappastatistikk ble benyttet for å undersøke grad av samsvar ved diskriminering av bekken og korsryggsmerter.

**Resultat:** Test retest på skåring av smertetegninger fikk best resultat med ICC verdi på 0,98. Intertester reliabilitet hadde ICC verdi på 0,96. Diskriminering av bekken- og korsryggplager fikk en test-retest kappaverdi på 0,98 og 91% enighet, mens intertester reliabilitet oppnådde en kappaverdi på 0,82 og 87% enighet.

**Konklusjon:** Både intertester reliabiliteten og test-retest reliabiliteten på skåring av smertetegninger i denne studien var meget tilfredsstillende, med svært god samsvar.

**Nøkkelord:** Smertetegning, test-retest reliabilitet, intertester reliabilitet, skåring

## Abstract

---

Intertester and test-retest reliability of pain drawings, using body area analysis.

Clinical program in Manual therapy for physical therapists, Institute for Social Medicine, University of Bergen, 2014

*Background and purpose:* The aim of this study was to investigate the intertester and test-retest reliability of pain drawings using body area analysis, and to investigate the intertester and test-retest reliability of discriminating low back and pelvic-girdle pain using pain drawings.

*Methods:* Secondary data (n=540) from previous studies were used. Analyse of pain drawings were performed on three different occasions using a simple body area method. Degree of agreement, using the body area method for analysing the pain drawings, was assessed by Intraclasscorrelation coefficient. To assess degree of agreement for discriminating between low back pain and pelvic pain, Cohens Kappa was used.

*Results:* Test-retest, using the body area method for analysing, achieved the best result having an ICC value of 0,98. Intertester reliability achieved an ICC value of 0,96. Discrimination of pelvic- and low back pain had a test-retest kappa value of 0,98 and 91% agreement, while intertester reliability for the same purpose had a kappa value of 0,82 and 87% agreement.

*Conclusion:* Reliability of both test-retest and intertester analysing pain drawings in this study were all satisfactory, all achieving a very good agreement.

*Keywords:* Pain drawing, test-retest reliability, intertester reliability, scoring

# 1 Introduksjon

## 1.1 Bakgrunn for studien

Smerter i muskel- og skjelettapparatet er svært utbredt, og 75 - 80% av den norske befolkningen oppgir å ha hatt slike plager i løpet av siste måneden (Ihlebaek et al., 2010). Flere kvinner enn menn rapporterer om plager og man ser generelt en økning i forekomst med økt alder (ibid). Blant dem som har kroniske muskel- og skjelettplager, rapporteres det vanligvis om plager fra tre eller flere kroppsdeler (Sirnes et al., 2003). Økende antall smerteområder er assosiert med nedsatt funksjon og høyere sykefravær og større uføreytelser (ibid). Til tross for en fallende tendens av trygdeytelser blant pasienter med disse lidelsene de siste ti årene, er muskel- og skjelettplager fremdeles den hyppigste medisinske årsaken til sykefravær og uføreytelser (Ihlebaek et al., 2010).

Smerter er fellesnevneren for de fleste pasienter som oppsøker fysioterapeut (Ihlebaek et al., 2010), og det å kartlegge omfanget av plagene er en viktig del av utredningen. Smerte er en subjektiv opplevelse, generert av utberedte, parallelle neurale nettverk i hjernen (Melzack, 1990). Nettverkene baserer sine multiple input fra det sensoriske systemet så vel som fra områder i hjernen som uttrykker tidligere erfaringer, evaluering og smertens betydning. Smerten produseres fra alle disse neurale nettverkene og gjør at dens uttrykk, og dermed smertemålinger, ikke kan sammenlignes mellom pasienter (ibid). Som effektmål på ulike tidspunkt for samme pasient, fungerer målingene derimot bra (Hansen, 2013).

Etter å ha jobbet 5 år på en tverrfaglig rehabiliteringsklinikk for pasienter med muskel- og skjelettplager, ble jeg oppmerksom på hvor god hjelp det var i å benytte ulike skjemaer for å kartlegge smertens utbredelse og intensitet. Skåring av skjemaene ble gjort på en standardisert måte, med unntak av smertetegninger. Her var praksisen ulik fra terapeut til terapeut. Jeg opplevde, tross ulik praksis, god nytteverdi i smertetegningene, men lurte på om det ikke var en felles konsensus på hvordan de skulle skåres. Ettersom pasientene ved dette rehabiliteringssenteret fylte ut smertetegning ved flere anledninger i behandlingsforløpet, syntes jeg det også var en svakhet at de fleste skåringsmetodene ikke fanget opp om smertene forflyttet seg.

Dersom en skåringsmetode skal benyttes i den hensikt å fange opp om smerten forflytter seg, er det viktig at skåringene er reproducerbare.

Tidligere studier av reliabilitet på smertetegninger har benyttet ulike skåringsmetoder på relativt homogene grupper (Jamison et al., 2004). De ulike metodene kombinert med ulike tolkningsmetoder har ført til uenighet i nytteverdien i bruken av smertetegninger (ibid). På bakgrunn av dette, og med tilgang til 540 smertetegninger fra hovedveileder Hilde Stendal Robinson, ønsket jeg i denne studien å se på intertester reliabilitet og test-retest reliabilitet på skåring av smertetegninger. Videre ønsket jeg å bruke en skåringsmetode som kunne diskriminere smertelokalisasjon.

## 1.2 Teori

### 1.2.1 Smerter

Smerte kan defineres som 'en ubehagelig sensorisk eller følelsesmessig opplevelse assosiert med faktisk eller potensiell vevsskade, eller beskrevet i former for slik skade' (IASP, 1996). Smerte er ofte assosiert med vevsødeleggelse men påvirkes også av psykiske og sosiale faktorer, tidligere opplevelser og essensielle faktorer (ibid). Ulike måleredskaper er utviklet for å favne alle aspektene ved smerter. Blant de mest brukte er Visual Analog Scale (VAS), Mc Gill Pain Questionnaire, numerisk- og verbal skala og smertetegning (Hansen, 2013).

### 1.2.2 Smertetegning

Helt siden 1940 har smertetegninger blitt benyttet som en del av undersøkelse av pasienter (Palmer, 1949). Utseendemessig finnes det flere varianter ut i fra hva de ønsker å avdekke diagnostisk (Toomingas, 1999; Geis et al., 2006), men de generiske tegningene er basert på samme prinsipp; en anterior og en posterior tegning av en menneskekropp (Toomey et al., 1983; Uden, 1997; Egloff et al., 2012). Pasienter blir spurt om å skravere de områdene på kroppen hvor de har smerter. Dette kan også gjøres med ulike symboler og ulike farger. Smertetegninger er enkle å administrere og et nyttig verktøy både i klinikk og forskning og kan benyttes for å identifisere og forutse om smertene er av generaliserende og kronifiserende art.

De kan brukes på tvers av diagnoser (Formi, 2013) og er dessuten et viktig instrument for å dokumentere effekt av behandling i tillegg til å være et enkelt verktøy for klinikere til hurtig å innhente en oversikt over pasientenes smertemønster (Jamison et al., 2004).

Smertetegning brukes som et hjelpeverktøy for å diagnostisere og for å utvikle en behandlingsplan (ibid).

Fordelene ved bruk av smertetegninger er mange, noe som også gjenspeiles i litteraturen. Smertetegninger kan vurderes ved kun ett blikk, uten behov for skåring (Uden et al., 1988), og test-retest reliabilitet for skåring av smertetegninger har vist seg å være høy (Margolis et al., 1986; Griffin et al., 1993; Ohnmeiss, 2000) og tegninger av radikulær smerte er vist å korresponderer med computer tomografi (CT) (Ohnmeiss et al., 1997).

Leger har kategorisert pasienter i diagnostiske grupper basert på pasienters smertetegning (Mann et al., 1992). I tillegg har smertetegninger vist seg å være et verdifullt som et psykisk screeningverktøy (Dennis et al., 1981; Hildebrandt et al., 1988; Parker et al., 1995), og for å bestemme uorganiske smertefaktorer (Uden et al., 1988; Mann et al., 1992; Chan et al., 1993).

### 1.2.3 Skåringsmetoder

Til tross for den hyppige bruken av smertetegninger både klinisk og i forskning, har det blitt stilt spørsmål ved nytteverdien når det gjelder undersøkelse av patologi (Brismar et al., 1996). De mange mulighetene som finnes av skårings- og tolkningsmuligheter gjør det vanskelig å vurdere reliabiliteten. Ulike skåringsystemer har blitt benyttet, uten enighet om hvilken metode som er best (Ohlund et al., 1996; Sanders and Mann, 2000).

Metodene som er brukt til å analysere og skåre smertetegninger ved bruk av en transparent mal, har blitt beskrevet som diagramanalyse (Palmer, 1949), topografisk smerterepresentasjon (Cairns et al., 1976) smertefordeling (Toomey et al., 1983) og undersøkelse av smerteutstrekning (Krause et al., 1989) Alle disse analysene benytter en mal som legges over smertetegningene. Malene er et mønster av punkter som gjentar seg periodisk, med samme eller forskjellig avstand til de ulike aksene (Gatchel et al., 1986; Margolis et al., 1986). Antall områder i malene for scoring av tegningene varierer fra 6 til 15000 (Ljunggren, 1983; Toomey et al., 1983; Gatchel et al., 1986; Margolis et al., 1986; Love et al., 1989; Donelson et al., 1991; North et al., 1992; Ohlund et al., 1996). Ratere rådes til å indikere smerte uansett hvor lite område som er skravert i den aktuelle ruten (Margolis et al., 1986).



Analyse av smertetegninger som brukes diagnostisk i forhold til klassifisering ble utviklet av Kvale et al. (2001). En horisontal linje i thoraco-lumbal overgangen avgjør om pasienten plasseres i gruppe 1; smerte over den horisontale linjen, gruppe 2; smerte under den horisontale linjen og gruppe 3; smerte over og under den horisontale linjen (ibid). Analysen gir grunnlag for å bruke tegningen prognostisk, men egner seg dårlig som effektmål (ibid).

En svakhet ved bruk av smertetegning er at ved retest vil en ikke automatisk fange opp eventuell forflytning av smertene.

En annen svakhet ved bruk av smertetegning er at den ikke beskriver pasientens smerteintensitet. Smertetegning blir derfor ofte bruk i kombinasjon med Pain Numeric Rating Scale (PNRS) (Krause et al., 1989).

PNRS er et endimensjonalt mål på smerteintensitet og brukes mye i den voksne populasjonen (McCormack et al., 1988; Jensen et al., 2003).

PNRS er en kontinuerlig skala på en horisontal eller vertikal linje, vanligvis 10cm. For smerteintensitet representerer 0 på skalaen 'ingen smerte', mens 10 representerer 'verst tenkelige' smerte (Hawker et al., 2011).

Nordic Pain Questionnaire (NPQ) er et standardisert spørreskjema utviklet i 1987 (Kuorinka et al., 1987). Spørreskjemaet ble utviklet med tanke på å kartlegge hvorvidt muskel-, skjelettlidelser forekommer i en gitt populasjon, og i hvilke deler av kroppen lidelsene kommer til uttrykk (ibid). Med dette som bakgrunn, ble en figur av en menneskekropp (sett bakfra) delt inn i ni anatomiske regioner. Regionene ble valgt på bakgrunn av hvor symptomer tenderer til å akkumulere. Kun baksiden av kroppen ble valgt som mål på muskel-skjelettlidelser, da andre lidelser ble ansett som årsak dersom plagene var lokalisert foran (ibid).

#### **1.2.4 Reliabilitet**

Reliabilitet handler om dataenes pålitelighet, og forteller i hvor stor grad testen er reproducerbar, eller hvor stor feilmargin metoden har. Dersom flere uavhengige målinger av ett og samme fenomen gir samme eller tilnærmet samme resultat, anses testen som reliabel (Carter et al., 2011). Begreper brukt synonymt om reliabilitet er nøyaktighet, stabilitet og konsistens (ibid).

Årsaken til variabilitet i målinger kan sees ut fra en generaliserbarhetsteori, hvor grunnene til variabilitet ligger hos pasienten (*intersubjektvariabilitet*), hos instrumentet (*instrumentets reliabilitet*), eller mellom ratere (*test- retest reliabilitet og inter tester reliabilitet*) (Carter et al., 2011). Instrumentreliabilitet er beskrevet med begrepene presisjon og nøyaktighet (Altman, 1991) *Presisjon* sier hvor likt samme måleresultat kan oppnås igjen og igjen.

Nøyaktighet er et uttrykk for hvor likt måleresultatet er den sanne verdien (ibid). Intrasubjektvariabilitet er variabiliteten i pasientens resultatoppnåelse ved repeterte målinger (Carter et al., 2011) Hovedkilden til endringer hos et individ er fysiske og psykiske endringer (ibid)

Intertester reliabilitet og test-retest reliabilitet beskriver de to sidene ved raterreliabilitet, hvor intertester reliabilitet representerer grad av enighet som to eller flere tester gjør på et sett av observasjoner. Test-retest reliabilitet gir uttrykk for grad av enighet hos en tester ved repeterte målinger på et sett av observasjoner (ibid).

I denne studien ønsket jeg å vurdere intertester og test-retest reliabilitet ved skåring av smertetegninger.

### 1.3 Tidligere forskning på området

Jeg vil i dette kapittelet presentere noe av den tidligere forskningen som er gjort på reliabilitet ved scoring av smertetegninger.

En rekke studier har sett på intertester og test- retest reliabilitet på skåring av smertetegninger. En utfordring er at det benyttes flere ulike skåringsmetoder i de ulike studiene. Dette gjør direkte sammenligning vanskelig. En annen svakhet er at utvalget som er studert er relativt homogent. Dette begrenser generaliserbarheten til studiene.

Reliabiliteten til skåringsmetoden, som ble presentert av Margolis et al. i 1986 er undersøkt både med tanke på intertester reliabilitet, og test-retest reliabilitet (Margolis et al., 1986; Margolis et al., 1988).

Studien fra 1996, hvor interrater reliabilitet ble kalkulert, besto av 101 smertetegninger som var skravert av pasienter med korsryggsmerter. Pasientene hadde gjennomsnittsalder på 49,1 år og besto av 50,5% kvinner og 49,5% menn. Tegningene ble skåret av sekretærer, som ikke hadde noe forkunnskap med bruk av smertetegninger. Interrater reliabiliteten ble kalkulert ved Pearsons  $r$  til  $r = 0,947$ . Test-retest ble testet ved samme skåringsmetode 2 år senere (Margolis

et al., 1988). 51 smertetegninger skravert av pasienter med kroniske smerter deltok i studien. Pasientene hadde en gjennomsnittsalder på 45,4 år og besto av 53% kvinner og 47% menn. Det kommer ikke frem av studien hvem som skåret tegningene. Test-retest ble kalkulert ved Pearsons  $r$  til  $r = 0,85$ . Metoden som er benyttet her deler kroppen i 45 anatomiske områder. De ulike områdene er vektet etter hvor stor del av kroppen de dekker, og scores deretter. Skåringen er tidkrevende, men kan skilte med både god intertester reliabilitet og test-retest reliabilitet

Uden ;Landin (1987) undersøkte både interrater reliabilitet og test-retest reliabilitet på skåring av smertetegninger. 264 pasienter med korsryggssmerter skraverte tegningene med seks ulike symboler for å karakterisere smerten de opplevde. Hensikten med studien var å finne et kartleggingsverktøy som enkelt kunne identifisere ikke-organisk smerte hos korsryggspasienter. To ratere analyserte smertetegningene med 2 ukers mellomrom. Den ene rateren hadde lang erfaring med analyse av smertetegninger, den andre ingen. Tegningene ble skåret etter antagelse om hva slags type smerte som ble presentert på tegningen. Ut i fra dette ble tegningene kategorisert i fire kategorier: organisk smerte, mulig organisk smerte, mulig ikke-organisk smerte og organisk smerte. Reliabiliteten på studien vurderes til god med henholdsvis 85% (test-retest erfaren rater), 77% (test-retest uerfaren rater) og 71% (intertester) enighet.

Chan et al. (1993) brukte metoden som ble utviklet av Uden ;Landin (1987) på 651 smertetegninger skravert av pasienter med korsryggssmerter. Av de 651 pasientene som hadde skravert smertetegningene var 387 menn og 264 kvinner. Tre ratere skåret tegningene, hvorav to hadde flere års erfaring med skåring av smertetegninger, mens den siste hadde ingen erfaring. Interrater reliabiliteten ble vurdert til god med enighet på mellom 73% og 78%. Studien har også undersøkt forskjeller på smerteutbredelse mellom kjønn, hvor det konkluderes med at det ikke er noen forskjell.

Bryner (1994) sammenlignet i en komparativ studie fire ulike skåringsmetoder av smertetegninger hos pasienter med korsryggssmerter. Av de fire skåringsmetodene var tre av de rutenett-baserte og ett databasert. De tre rutenett-baserte metodene skilte seg fra hverandre ved antall områder kroppen var delt opp i, hhv 46, 200 og 560. Disse ulike metodene er utviklet av Margolis et al. (1986), Gatchel et al. (1986) og Bryner (1994). 17 smertetegninger ble analysert av to undersøkere ved hjelp av disse fire metodene. Alle metodene hadde høy

reliabilitet hvor reliabilitet ble kalkulert ved bruk av Intraclass correlation og Kappa. Korrelasjonskoeffisientene varierte fra 0,46 til 0,94. Den databaserte metoden var mest sensitiv for endringer.

De fleste studier som har sett på reliabilitet for skåring av smertetegninger, har gjort dette på smertetegninger fra ryggpasienter. I 1995 ble tre nye skåringsmetoder introdusert for å avgjøre om det var en psykologisk dysfunksjon hos pasienter med korsryggplager (Parker et al., 1995). 100 pasienter med kroniske ryggsmarter var inkludert i studien. Smertetegninger ble skravert og analysert i etterkant. Studien konkluderer med at smertetegninger, uansett skåringsystem, har lav sensitivitet og spesifisitet, og kan kun sees på som et tilleggs-verktøy i forbindelse med undersøkelse og behandling. Reliabiliteten var høy hos alle skåringsmetodene, med kappaverdier på 0,94, 0,97 og 1 for interrater reliabilitet.

Ohnmeiss (2000) undersøkte fire ulike skåringsmetoder i forhold til inter-tester og test-retest reliabilitet hos pasienter med kroniske korsryggssmerter i en prospektiv test-retest studie fra 2000 (ibid). De ulike metodene bygger på 'penalty point system', visuell inspeksjon, kroppsregioner og rutenett-metoder.

Penalty-point system er et skåringsystem som tar sikte på å differensiere ryggsmarter med og uten psykologiske faktorer (Postacchini, 1999). Visuell inspeksjon er et skåringsystem som er designet for å klassifisere mellomvirvelskiveproblematikk hos pasienter med ryggplager. Skåring ved bruk av kroppsregion og rutenett har evne til å fange opp mer generell problematikk, men i denne studien ble bare regioner forbundet med ryggplager inkludert. 45 pasienter deltok i studien. Reliabilitet ble kun analysert for penalty-point system og visuell inspeksjon hvor det konkluderes med både god intertester- og test-retest reliabilitet.

Interrater reliabilitet ble undersøkt av Lacey et al. (2005).

50 smertetegninger ble skåret ved bruk av 3 transparente maler delt opp i totalt 50 områder. Smertetegningene var skravert av pasienter som svarte på en større spørreundersøkelse om knesmerter. Alle pasientene var over 50 år. Tegningene ble skåret av åtte forskjellige ratere. Syv av raterne hadde ingen klinisk helsefaglig bakgrunn, men alle fikk opplæring i hvordan tegningene skulle skåres. Kappa kalkulert for tilstedeværelse eller fravær av smerte (kategoriske data) ble målt til  $k = 0,98$ . ICC ble målt til 0,99 på bakgrunn av antall positive smerteområder per tegning.

Denne måten å skåre smertetegninger på ble introdusert av Macfarlane i 1996 (MacFarlane et al., 1996). Hensikten var å utvikle et reliabelt verktøy å bruke i epidemiologiske studier og å klare identifisere pasienter med signifikant sykелighet. Studien til Lacey et al. (2005) viste at scoring av smertetegning kan ha høy interreliabilitet blant personer uten klinisk helsefaglig bakgrunn dersom de får grundig opplæring og følger gitte retningslinjer.

Gjennomgang av litteraturen bekreftet god intertester og test-retest reliabilitet på skåring av smertetegninger. Skåringsmetodene som presenteres fanger imidlertid ikke opp om smertene forflytter seg i en retest-situasjon. Det vil være interessant å skåre smertetegninger fra et normalfordelt, heterogent pasientmateriale med en enkel skåringsmetode som baserer seg på kroppsområder.

## 2 Hensikt og problemstilling

Min bakgrunn for å skrive om temaet er min interesse for bruk av kartleggingsskjemaer som hjelpemidler i den kliniske hverdagen. Reliabiliteten til de ulike skjemaene er viktig dersom man skal vurdere om behandlingen har effekt. Er ikke måleinstrumentet reliabelt, så kan en eventuell ending skyldes dette og ikke effekt av behandlingen. Smertetegninger brukes også mye i forskning, hvor det også er viktig med reliable målemetoder (Carter et al., 2011).

Denne studien har som hovedformål å undersøke hvorvidt måleegenskapen til en kroppsområde analyse av smertetegninger er reliabel. Studien vil også undersøke grad av reliabilitet til diskriminering av korsrygg- og bekkensmerter, ved bruk av samme analyse, R på smertetegninger.

### 2.1 Problemstilling

Hvor god er test-retest og intertester reliabilitet til skåring av smertetegninger ved bruk av kroppsområde-analyse?

Sekundær problemstilling: Hvor god er test-retest og intertester reliabilitet til å diskriminere korsrygg- og bekkenplager ved bruk av kroppsområdeanalyse?

## 3 Metode

### 3.1 Valg av forskningsdesign

Denne oppgaven er en reliabilitetsstudie, der jeg har undersøkt hvor god reliabiliteten til en skåringsmetode av smertetegninger er. Studien er kvantitativ, som forutsetter at dataene er målbare og kan gis til uttrykk i tall eller mengdeterminologi (Halvorsen, 2002).

Målet med studien var å undersøke grad av intertesterreliabilitet og test – retest reliabilitet av 540 smertetegninger. For å belyse denne problemstillingen ble det benyttet tidligere innsamlet data fra 4 studier. Samvariasjon ble undersøkt både i en test-retest setting og i en intertester setting. Tegningene ble skåret ved å registrere skravering i forhold til kroppsområder. Kroppsområdene ble valgt ut med bakgrunn i Nordic Pain questionnaire (NPQ) (Kuorinka et al., 1987)

### 3.2 Beskrivelse av materiale

Hilde Stendal Robinson sørget for kopier av de aktuelle smertetegningene i løpet av mai 2014. Smertetegninger som ikke hadde tilhørende demografisk informasjon om alder og kjønn, ble ekskludert.

Smertetegningene som ble analysert var brukt i fire ulike studier. Majoriteten av smertetegningene (326) som ble benytta, var fra doktorgradsprosjektet til Hilde Stendal Robinson, *Pelvic girdle pain and disability during and after pregnancy* (Robinson, 2010). Alle de 326 var gravide kvinner. Disse kvinnene var i utgangspunktet friske, rekruttert til studien ved første kontakt med helsestasjon.

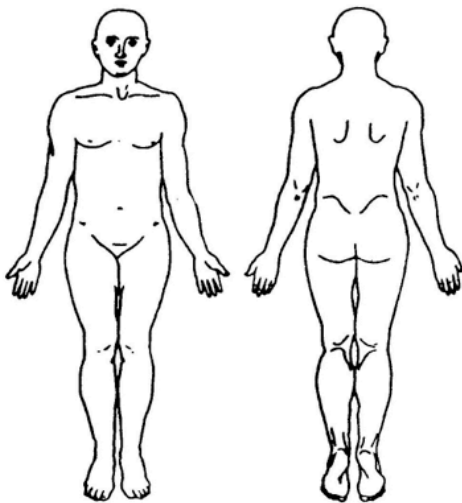
98 smertetegninger var innhentet fra pasienter muskel-skjelettplager. Pasientene var fordelt på tre grupper, pasienter med korsryggsmerter, pasienter med bekkensmerter etter fødsel, samt pasienter med smerter utenfor disse to kategoriene. Alle hentet fra datamaterialet som ble samlet inn til studien: *Assessments of lumbar flexion range of motion: intertester reliability and concurrent validity of 2 commonly used clinical tests* (Robinson and Mengshoel, 2014).

79 smertetegninger kom fra studiene til Britt Stuge, *The efficacy of a treatment program focusing on specific stabilizing exercises for pelvic girdle pain after pregnancy: a randomized*

*controlled trial* (Stuge et al., 2004). Alle pasientene var kvinner med bekkenplager 3 mnd. etter fødsel.

De siste 37 smertetegningene var fra en studie av pasienter med fibromyalgi. Hentet fra deltagere i studien til Anne Marit Mengshoel, *Pain and fatigue induced by exercise in fibromyalgia patients and sedentary healthy subjects* (Mengshoel et al., 1995)

I alt ble 540 smertetegninger analysert. Alle tegningene hadde en forside og en bakside som illustrert på figur 1.



**Figur 1** Smertetegning



Tabell 1 Skjematisk oversikt over de ulike studiene der smertetegningene ble hentet fra

Dataset	Forfatter	År	Antall tegninger	Tittel
1	Hilde Stendal Robinson	2010	326	<i>Pelvic girdle pain and disability during and after pregnancy</i>
2	Hilde Stendal Robinson og Anne Marit Mengshoel	2014	98	<i>Assessments of lumbar flexion range of motion: intertester reliability and concurrent validity of 2 commonly used clinical tests</i>
3	Britt Stuge	2004	79	<i>The efficacy of a treatment program focusing on specific stabilizing exercises for pelvic girdle pain after pregnancy: a randomized controlled trial</i>
4	Anne Marit Mengshoel	1995	37	<i>Pain and fatigue induced by exercise in fibromyalgia patients and sedentary healthy subjects</i>

### 3.3 Terapeutene

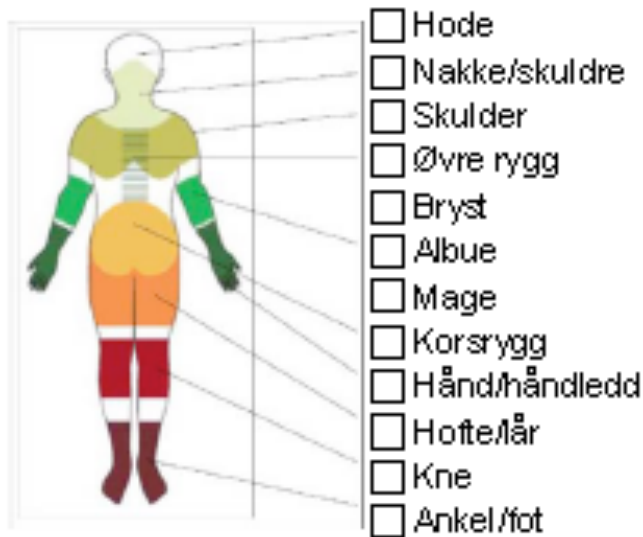
De som analyserte smertetegningene var begge fysioterapeuter med seks års arbeidserfaring. Den ene har jobbet på sykehus og i privat praksis, den andre på et rehabiliteringssenter. Kun fysioterapeuten som har jobbet på rehabiliteringssenter hadde erfaring med skåring av smertetegninger.

### 3.4 Gjennomføring av analyser

Analyse av smertetegningene ble gjort med utgangspunkt i smertetegningen i Nordic Pain Questionnaire (NPQ) Se figur 2 (Kuorinka et al., 1987). Til tross for at bryst og mage var inkludert i beskrivelse av figuren i NPQ, var kun baksiden av figuren inkludert i skjemaet. I tillegg var det ingen diskriminering mellom korsrygg og bekken. Følgende definisjoner og endringer ble derfor gjort i forkant av analysen, både med tanke på tegningenes utseende og med tanke på sekundær problemstilling.

- Bryst – området som representerer 'øvre rygg' på figurens bakside, ble definert som 'bryst' på figurens fremside.
- Mage ble definert som området fra 'bryst' til 'hofte/lår'.

- Korsrygg ble definert som område fra L5 opp til 'øvre rygg'.
- Bekken ble definert fra posterior superior iliac spine (PSIS) til 'hofte/lår'.



Figur 2 Smertetegning fra NPQ

Begge terapeutene skåret alle smertetegningene og dataene ble lagt inn i SPSS. For alle områder indikerte raterne 1 for smerte og 0 ved fravær av smerte. Følgende kriterier ble brukt for å bestemme hvorvidt det skulle indikeres for smerte eller ikke:

- Ethvert område ble skåret med tallet 1, uansett hvor vagt eller lite.
- Områder som var sirklet inn, ble behandlet som om hele området var skravert.
- Markeringer utenfor tegningen ble ignorert.

All analyse ble gjennomført fra mai 2014 til juli 2014. Demografiske data og smertere registreringer ble kodet og lagt inn i statistikkprogrammet SPSS (Statistical package for Social Sciences versjon 22). Tabeller og grafer ble laget i SPSS og analyser ble gjort i SPSS og med kalkulator.

Terapeut 1 gjennomførte test- retest med 2 ukers mellomrom. Terapeut 2 gjennomførte skåringen i etterkant av terapeut 1. Begge var blindet for hverandres resultater. I etterkant av skåringen ble de ni anatomiske områdene representert i NPQ, slått sammen til én sumskår. I tillegg ble skåringen av bekken og korsrygg slått sammen og deretter rekodet som vist i tabell 2.

**Tabell 2 Oversikt over rekoding av bekken og korsryggsmerter**

Sumskår	Etikett
0	Ingen smerte
2	Smerte korsrygg og bekken
3	Smerte korsrygg
4	Smerte bekken

### **3.5 Reliabilitet**

#### **3.5.1 Kontinuerlige data**

ICC ble benyttet på de kontinuerlige dataene. Relativ reliabilitet vurderer sammenhengen mellom to testtilfeller og uttrykkes vanligvis gjennom betraktning av en korrelasjonskoeffisient (ICC). Til tross for at ICC måler reliabilitet, finnes det mange ulike former som kun er passende i et begrensede omstendigheter (Shrout and Fleiss, 1979). I denne studien ble 'two-way mixed', 'absolute agreement' og 'average measures' brukt (ICC 3,1). Som Cohen's Kappa graderes ICC fra 0 til 1, hvor 1 representerer perfekt enighet, og null ingen enighet (ibid).

#### **3.5.2 Kategoriske data**

Kappa ble benyttet på de kategoriske dataene – diskriminering mellom bekken/korsrygg (Altman, 1991). Kappa er en betegnelse for reliabilitetskoeffisienten, med andre ord et tall som sier noe om samsvaret mellom to bedømmere (Nylenna, 2009) Reliabiliteten ble vurdert ved hjelp av kappakoeffisienten med 95% konfidens intervall.

Kappakoeffisient (k), eller reliabilitetskoeffisienten har et maksimum på 1 og et minimum på 0. En kappa på 1 indikerer perfekt enighet og en kappa på 0 indikerer at resultatet ikke er bedre enn ved sjanse (Altman, 1991)

Tabell 3 viser hvordan Kappa- og ICC verdier mellom 0 og 1 kan graderes (Shrout and Fleiss, 1979; Carter et al., 2011)

Tabell 3 Kappa og ICC verdier

<b>K verdi</b>	<b>Enighet</b>
<b>&lt; 0,20</b>	Dårlig
<b>0,21 – 0,40</b>	Svak
<b>0,41 – 0,60</b>	Moderat
<b>0,61 – 0,80</b>	God
<b>0,81 – 1,00</b>	Svært god

### 3.6 Ethiske betraktninger og egen rolle som forsker

Kvantitativ forskning innebærer standardiserte metoder og formaliserte kravspesifikasjoner, som i utgangspunktet bidrar til etisk troverdighet (Befring, 2007).

Tilsiktede feil handler om forskerens etiske grunnholdninger, for å kunne gjennomføre alle ledd i forskningsprosessen på en fullverdig måte. Misvisende resultater oppstår når konklusjoner har større interesse for forskeren enn den vitenskapelige evidens. Denne type tilsiktede feil oppstår først og fremst ved kommersielt-basert forskning og oppdragsforskning (ibid). Denne studien skal ikke brukes kommersielt og ble heller ikke skrevet på oppdrag.

Utilsiktede feil oppstår dersom studien ikke er metodisk troverdig.

Oppdatert og inngående innsikt i aktuelt problemområde er avgjørende for både utvikling av problemstilling og design.

Essensielt for den etiske normen er at forskeren har en akseptabel forskningsetisk-, faglig- og metodisk kompetanse til å legge frem vitenskapelig holdbare og valide konklusjoner (ibid).

Denne studien inneholder utelukkende anonyme opplysninger, og personer kan ikke identifiseres verken direkte, indirekte eller via koblingsnøkkel. Smertetegningene var anonymisert fra de studiene de tidligere var brukt i, og all data vil bli publisert i en ikke-identifiserbar måte. De som har fylt ut smertetegningene har tidligere signert informert samtykke til at den informasjonen. All direkte og indirekte identifiserbare elementer er tatt bort av forfattere av primærstudiene og det var derfor ikke nødvendig å søke Regional Etisk komite (REK) om ny godkjenning.

## 4 Resultater

540 smertetegninger ble skåret av to fysioterapeuter i perioden mai-juli 2014. Alle smertetegningene var innhentet fra 4 forskjellige studier, beskrevet i kapittel 3.2.

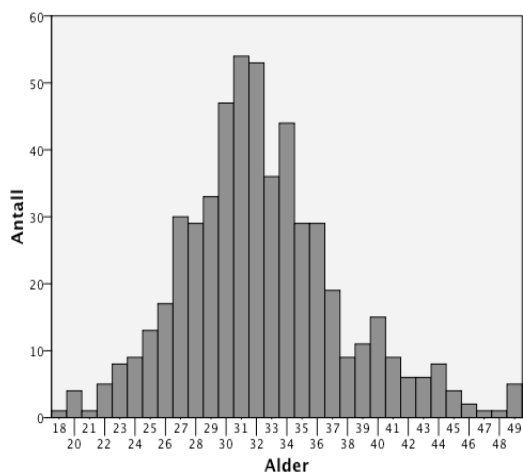
### 4.1 Bakgrunnsvariabler

Tabell 4 Demografiske data

<b>Alder</b>			
	Gjennomsnitt	SD	Min-max
<b>Alle dataset</b>	32	5,3	18-49
<b>Dataset 1</b>	31,4	4,17	18-45
<b>Dataset 2</b>	35,63	7,73	20-49
<b>Dataset 3</b>	32,21	3,86	22-41
<b>Dataset 4</b>	32,33	5,44	20-40
<b>Kjønn</b>			
	Kvinner (%)	Menn (%)	
<b>Alle dataset</b>	95,56	4,44	
<b>Dataset 1</b>	100		
<b>Dataset 2</b>	75,5	24,5	
<b>Dataset 3</b>	100		
<b>Dataset 4</b>	100		
<b>Smertetegninger</b>			
	Positive (%)	Negative (%)	
<b>Alle dataset</b>	85,93	14,07	
<b>Dataset 1</b>	84,4	15,6	
<b>Dataset 2</b>	81,6	18,4	
<b>Dataset 3</b>	91,1	8,9	
<b>Dataset 4</b>	100	0	

Pasientene som hadde skravert de 540 smertetegningene hadde en gjennomsnittsalder på 32 år. Menn var kun representert i dataset 2 og utgjør totalt 4,44% av tegningene.

14,7% av pasientene oppga at de ikke hadde smerter på smertetegningen. I dataset 4 var alle de 37 tegningene skravert.



Figur 3 Oversikt over alder - alle dataset

## 4.2 Reliabilitet ved kroppsområde-skåring av smertetegninger

Test-retest og intertester reliabilitet på kroppsområde-skåring av smertetegning viste seg å være svært god. Best resultat ble oppnådd ved test-retest, med en ICC verdi på  $0,98 \pm 0,003$  (95%CI). Intertest reliabiliteten viste en ICC verdi på  $0,96 \pm 0,021$  (95%CI)

Tabell 4 test-retest kroppsområde-skåring

Intraclass Correlation Coefficient			
	ICC	95% Konfidens intervall	
		Øvre grense	Nedre grense
Gjennomsnittlig mål	0,98	0,98	0,86

Tabell 5 Intertest kroppsområde-skåring

Intraclass Correlation Coefficient			
	ICC	95% Konfidens intervall	
		Øvre grense	Nedre grense
Gjennomsnittlig mål	0,96	0,94	0,98

### 4.3 Reliabilitet ved diskriminering mellom korsrygg og bekken

Test-retest reliabilitet og intertester reliabilitet på den anvendte skåringsmetoden for diskriminering mellom korsrygg og bekken, kom begge godt ut. Test-retest viste en kappaverdi på 0,98 (0,97 – 0,99 ) og prosentmessig enighet på 91%. Intertest viste en kappaverdi på 0,82 (0,79 – 0,86 ) og prosentmessig enighet på 87%.

Tabell 5 Oversikt over kappaverdi ved test-retest. Enighet = 91% (494/540)

N= 540 Kappa= 0,98 CI= 0,974 - 0,988		Retest				Totalt
		Ingen smerte	Smerte korsrygg og bekken	Smerte korsrygg	Smerte bekken	
Test	Ingen smerte	182	0	2	3	187
	Smerte korsrygg og bekken	1	131	2	4	138
	Smerte korsrygg	0	11	37	1	49
	Smerte bekken	4	18	0	144	166
	Totalt	180	168	44	148	540

Tabell 6 Oversikt over kappaverdi ved intertest. Enighet = 87% (471/540)

N= 540 Kappa= 0,82 CI = 0,88 - 0,86		Retest				Totalt
		Ingen smerte	Smerte korsrygg og bekken	Smerte korsrygg	Smerte bekken	
Test	Ingen smerte	173	6	2	6	187
	Smerte korsrygg og bekken	0	127	4	7	138
	Smerte korsrygg	1	12	36	0	49
	Smerte bekken	6	23	2	135	166
	Totalt	180	168	44	148	540

## 5 Diskusjon

Hovedmålet med studien var å vurdere intertester og test-retest reliabiliteten på analyse av smertetegningen ved bruk av kroppsområde-analyse samt undersøke intertester og test-retest reliabilitet på diskriminering av bekken- og korsryggssmerter ved bruk av kroppsområdeanalyse. Analyse av smertetegning ved bruk av kroppsområdeanalyse viste intertester reliabilitet på 0,96. Test-retest reliabilitet ble målt til 0,98. Diskriminering av bekken- og korsryggssmerter viste en intertester reliabilitet på 0,821 med 87% enighet og test-retest reliabilitet ble målt til 0,981 med 91% enighet. Hypotesen kan dermed bekreftes.

### 5.1 Diskusjon av metode

Utformingen på smertetegningen var identiske i dataset 1,2 og 3 (Stuge et al., 2004; Robinson, 2010; Robinson and Mengshoel, 2014) I dataset 4 (Mengshoel et al., 1995) var imidlertid smertetegningene halv størrelse i forhold til de resterende tegningene. De ble allikevel inkludert i studien da størrelsen på datasetet var relativt lite (37), dessuten representerte utvalget en ny type pasientkategori (fibromyalgipasienter). Fibromyalgipasienter diagnostiseres blant annet på bakgrunn av smerteutbredelse, hvor smerter i kroppens fire kvadranter er ett av kriteriene (Wolfe and Hauser, 2011). Dermed presenterte disse tegningene seg med større andel av tegningene skravert enn de resterende tegningene.

Alle dataset ble reliabilitetstestet som ett samlet materiale. Det hadde vært interessant å undersøke statistiske forskjeller mellom de ulike datasetene, for å se om uteseende på tegningen, kjønn eller diagnose/plager hadde signifikant effekt på resultatet. Det er ikke gjort i denne studien.

Ettersom 367 av de 540 smertetegningene ble skåret til å ha smerter i enten korsrygg, bekken, eller begge, sørget dette for at det ble et stort nok utvalg til å reliabilitetsteste sekundær problemstilling.

76 av tegningene var ikke skravert, da en del av utvalget var smertefrie. Det var 100% enighet mellom ratere, og tegningene kunne vært ekskludert.



## 5.2 Validitet

Validitet er kort sagt hvorvidt man klarer å måle det man har til hensikt å måle (Shadish et al., 2001). Det er tolkningen av dataene som validerte, ikke selve målemetoden (ibid). I Carter et al. (2011;s 75) presenteres et system som skisserer hvordan validiteten kan vurderes.

Systemet ble utviklet av Cook (1979), og viser hvordan validitet kan vurderes i forhold til fire typer slutninger; statistiske slutninger (statistisk validitet), kausale slutninger (indre validitet), begrepslutninger (begrepsvaliditet) og generaliseringer (ytre validitet). Ulike feilkilder og trusler inngår i hver av disse slutningene (Carter et al., 2011). Indre validitet og ytre validitet er mest relevant for dette studiet, og vil bli drøftet videre.

### 5.2.1 Intern validitet

Intern validitet ser på om en måler det en på forhånd har antatt, og om studiet gjennomføres på en tilfredsstillende måte (Jacobsen, 2005). I hvilken grad kan jeg stole på resultatene i denne studien og hvilke trusler påvirker disse verdiene.

Tidsaspektet mellom test-retest i denne studien var to uker. Det finnes ingen standard regel for tidsintervallet mellom test-retest, men Altman (1991) anbefaler sunn fornuft, hvor det er god balanse mellom stabiliteten på utvalget og uavhengighet fra test til retest (ibid).

Smertetegningene endret seg ikke fra test til retest, og skåringene ble gjort uavhengige. Dette representerer en styrke til studien.

Datamaterialet er fra 24 menn og 516 kvinner, og til tross for at flere kvinner enn menn rapporterer om muskel- og skjelettplager (Ihlebaek et al., 2010), vil ikke denne fordelingen representere pasientgruppen. Det ville vært ønskelig med et mer heterogent datamateriale kjønnsmessig. Selv om det konkluderes med at det ikke er forskjell på skåring av smertetegninger mellom kjønn i studien til Chan et al. (1993), ville det økt den indre validiteten.

For at den interne validiteten skal bli best mulig er det viktig at utførelse standardiseres med en protokoll så det blir utført likt hver gang (Jacobsen, 2005). Skåringen er standardisert på bakgrunn av teori som foreligger i metodekapittelet. Ingen av raterne trente på skåringsprosedyren før oppstart av studien noe som kan svekke den interne validiteten. Skåringen ble utført i lignende omgivelser. Rater 2 hadde imidlertid noe tidspress på seg. Dette ser imidlertid ikke ut til å ha påvirket resultatet negativt.

I denne studien har jeg hatt flere roller, ettersom jeg både har skåret tegninger og, sammen med veileder, stått for utforming av selve studien. Ideelt sett ville det vært gunstigere om to uavhengige ratere stått for skåringen.

### 5.2.2 Ekstern validitet

Ekstern validitet handler om i hvor stor grad resultatene i en studie kan generaliseres (Lund, 2002; Carter et al., 2011). Dette innebærer hvor stor sannsynlighet det er for at skåringen som ble gjort av de 540 smertetegningene er representativ for befolkningen. Et utvalg på mer enn hundre kan generaliseres ifølge (Laake et al., 2007). Altman (1991) anbefaler 50 for generaliserbarhet, mens Cosmins Checklist (Mokkink et al., 2010) anser et utvalg på mellom 55-99 som 'good'. Etter disse kriteriene vil utvalget på de 540 smertetegningene være svært bra.

Men for å kunne generalisere resultatene, må også terapeutene som skåret tegningene være representative for alle som skal analysere smertetegninger. Til tross for lite kjennskap til smertetegninger, hadde begge raterne i denne studien helsefaglig bakgrunn, samt lik utdanning. Margolis et al. (1986) og Lacey et al. (2005) bruker personer uten helsefaglig bakgrunn til å analysere smertetegninger i sine studier. Altman (1991) anbefaler bruk av flere undersøkere/ratere for å generalisere av resultater. Dette representerer i så måte en svakhet til min studie. Til bruk i klinikk, hvor smertetegninger anbefales benyttet (Chan et al., 1993; Bertilson et al., 2007) og hvor det vanligvis er terapeuter som skårer tegningene, representerer resultatene i min studie en styrke.

### 5.3 Sammenligning med tidligere studier

Skåring ved bruk av de ni kroppsområdene som er representert i NPQ (Kuorinka et al., 1987) er ikke reliabilitetstestet tidligere, til tross for at selve skjemaet er reliabilitetstestet med tanke på test-retest reliabilitet. Dawson et al. rapporterte om gode reliabilitetstall med tanke på test-retest reliabilitet (Dawson et al., 2009).

Tidligere studier har vært ulike både med tanke på metode og design, noe som gjør det vanskelig å kartlegge hvilke faktorer som er av betydning for enighet og gjør det vanskelig å sammenligne resultater..

Mine resultater samsvarer med funn fra i tidligere studier som har undersøkt interrater reliabilitet og test-retest reliabilitet på skåring av smertetegninger. Margolis og medarbeidere undersøkte interreliabilitet og test-retest reliabilitet (Margolis et al., 1986; Margolis et al., 1988). Skåringsmetoden som ble brukt her er beskrevet i kapittel 1 og består av en transparent mal med 45 definerte anatomiske områder. Malen legges over smertetegningen, antall skraverte områder telles, før de vektet etter hvor stor del av kroppen det representerer. Interrater reliabiliteten ble kalkulert ved Pearsons  $r$  til  $r = 0,947$ , test-retest ble kalkulert til 0,85. Pearsons  $R$  kan brukes på data som er på rasionivå og passer i så måte til bruk i studiet (Altman, 1991).

Test-retest ble testet ved samme skåringsmetode 2 år senere (Margolis et al., 1988).

Reliabiliteten ble kalkulert ved Pearsons  $r$  til  $r = 0,85$ .

Ulikhetene mellom disse studiene og min, omhandler først og fremst skåringsmetoden og kalkulering av reliabilitet. Det står ikke noe om tidsbruk på skåringen i studiene, men jeg vil anta at det er mer tidkrevende enn metoden brukt i min studie. Intraclass correlation (ICC) kan også brukes på data på rasionivå (Altman, 1991), og det hadde det vært interessant å se resultatene kalkulert ved bruk av ICC som reliabilitetskoeffisient. Tross ulikheter, samsvarer resultatene godt.

Både Kappa og ICC ble brukt i en annen studie som undersøkte interrater reliabilitet til skåring av smertetegninger (Bryner, 1994).

Kappa kalkulert for tilstedeværelse eller fravær av smerte (kategoriske data) ble målt til  $k = 0,98$ . ICC ble målt til 0,99 på bakgrunn av antall positive smerteområder per tegning. Skåringsmetoden som er brukt i min studie er noe effektivisert i forhold til Margolis et al. (1986), noe som også er studiens hensikt. Fremdeles benyttes transparent mal, og fremdeles er den delt opp i anatomiske områder (50), men skåringen vektet ikke, og vil derfor ta kortere tid.

Mine søk har ikke funnet andre studier som har sett på reliabilitet på diskriminering mellom korsrygg- og bekkensmerter ved bruk av smertetegning.

Diskriminering av bekken og korsryggsmerter ved bruk av smertetegning ble imidlertid introdusert av Fortin et al. (1994b). Gjennom to studier med bruk av smerteprovoserende og lindrende injeksjoner, ble smertemønstre ved smerter i iliosacralleddet kartlagt (Fortin et al., 1994a; Fortin et al., 1994b). Mønsteret hadde en lokalisasjon på 10 cm kaudalt og 3 cm lateralt for posterior superior iliac spine. Den andre studien validerte skjemaet. Pasienter med

smertes lignende det mønsteret som ble presentert i den første studien, fikk provoserende injeksjon for bekreftelse av smertelokalisasjon. Den spesifikke lokalisasjonen av smertene fikk navnet Fortin finger test. Min studie kan vanskelig sammenlignes med Fortins finger test. Til tross for at hensikten var å diskriminere mellom bekken- og korsryggplager, er det ikke smertemønsteret som presenteres av Fortin som ligger til grunn for de områdene som er valgt i denne studien.

Andre studier som har brukt smertetegninger til å se på smertelokalisasjon i forhold til prognostisk og diagnostisk verdi, har hatt til hensikt å skille mellom generaliserte og lokale smerter (Toomey et al., 1983; Parker et al., 1995; Brismar et al., 1996; Kvale et al., 2001) og er således metodisk forskjellige fra min studie.

#### **5.4 Implikasjoner for klinisk praksis**

Resultatene i denne studien tyder på at det er god reliabilitet på skåring av smertetegninger ved kroppsområde-analyse. Dette passer godt med resultater i tidligere studier. Blant de skåringsmetodene som har vist seg å ha god reliabilitet er visuell inspeksjon (Uden and Landin, 1987; Ohnmeiss, 2000). Dette er den metoden som er mest effektiv og som krever minst av terapeut/rater. Smertetegninger har lenge vært ansett som et diagnostisk og prognostisk verktøy (Uden and Landin, 1987; Brismar et al., 1996). Det kan også brukes som effektmål på behandling dersom man benytter smertetegning ved flere anledninger i en rehabiliteringsfase. Med en skåringsmetode som fanger opp om smertene forflytter seg, kan smertetegninger stå enda sterkere som effektmål.

Det økte kravet til dokumentasjon er en utfordring innenfor helsefagene generelt og fysioterapifaget spesielt (Dahm et al., 2002). Det stilles først og fremst krav til dokumentasjon av behandlingseffekt, og til dette er det en forutsetning at det finnes reliable måleinstrumenter (ibid). Smertetegning passer bra til bruk som dokumentasjon, ikke bare ovenfor annet helsepersonell og myndigheter men også ovenfor pasienter.

Sturesson et al. (1997) undersøkte om smertelokalisasjon hos gravide kvinner hadde sammenheng med smerteadferd. Kvinnene ble delt opp i tre grupper basert på smertelokalisasjon og en klinisk undersøkelse av en ortopedisk kirurg. I gruppen med bekkenplager hadde 44 av 121 kvinner et karakteristisk følelse i foten når de gikk, beskrevet

som 'catching of the leg'. Ved å bruke metoden som presenteres i min studie for diskriminering av bekken og korsryggsmerter, kan den være et nyttig diagnostisk verktøy.

## 5.5 Videre forskning

En studie skal ideelt sett gi svar på et forskningsspørsmål, mens man i virkeligheten sitter igjen med flere spørsmål enn svar. Jeg har underveis i arbeidet med studien kommet opp med nye problemstillinger for videre forskning.

Ved oppstart av studien var hensikten å sammenligne den benyttede skåringsmetoden med en annen skåringsmetode. Omfanget gjorde at dette ikke ble gjennomført. Sammenligning med en skåringsmetode som allerede er reliabilitetstestet ville gitt verdifull kunnskap. Bryner (1994) og (Ohnmeiss, 2000) sammenligner i sine studier mellom ulike skåringsmetoder, noe som fremhever styrkene og svakhetene til de ulike studiene godt.

Hovedhensikten med studien var å undersøke reliabiliteten til en skåringsmetode for smertetegninger som registrerer om smerte forflytter seg. Bruk av spørreskjemaet i forkant og etterkant av en pasientintervensjon sammen med for eksempel et spørreskjema ville vært interessant for å undersøke hvor godt metoden fanger opp endring av lokalisasjon.

Selv om utvalget i min studie er stort, er 95,56% kvinner, hvorav 79,8% fremstår med smerter fra bekken og korsrygg. Dersom hensikten er å diskriminere mellom rygg og bekkensmerter, er materialet representativt. Dersom skåringsmetoden skal brukes på andre problemstillinger, ville det vært interessant med et mer heterogent utvalg.

## 6 Oppsummering og konklusjon

*Oppsummering:* Denne studien viste at reliabiliteten på de anvendte skåringsmetodene i denne studien var svært god. Best resultat ble oppnådd ved test-retest reliabilitet på skåring av smertetegninger ved bruk av kroppsområde-analyse. Her ble reliabiliteten målt til en ICC-verdi på 0,98. Intertester reliabiliteten på samme skåringsmetode viste også god reliabilitet med en kappaverdi på 0,96.

Diskriminering mellom korsrygg- og bekkenplager viste test-retest reliabilitet på 0,98, mens intertester reliabiliteten ble målt til en kappaverdi på 0,82.

Hypotesen i denne studien var at det var god test-retest og intertester reliabilitet på skåring av smertetegninger ved bruk av kroppsområde-analyse. Resultatene viser god reliabilitet på alle områder, noe som samsvarer med tidligere studier. En reliabilitetskoeffisient på 0,9 er nærmest perfekt enighet.

*Konklusjon:* Dagens samfunn stiller imidlertid krav til dokumentasjon av effekt, og en analyse av smertetegninger som fanger opp om smerten har forflyttet seg, virker i så måte som et nyttig verktøy.

Sammen med introduksjonen av den biopsykososiale forståelsen av smerter (O'Sullivan, 2005), økte også behovet for kartlegging av den psykososiale delen. Smertetegninger er et slikt verktøy med både prognostisk og diagnostisk verdi.

Hvorvidt skåringsmetoden som presenteres her skal brukes i forskning kommer i stor grad an på forskningsspørsmålet. Dersom en studie har som mål å teste effekt av behandling ved bruk av smertetegning, vil denne metoden være god, da endring av smertelokalisasjon også er en effekt av behandling.

Både i min nåværende og tidligere praksis har jeg sett mange ulike metoder for å skåre smertetegninger. Mitt resonnement har vært at det burde være en gullstandard som kunne brukes generisk. I realiteten handler det om utvalg og formål. Hvem som skraverer smertetegningene og i hvilken hensikt har mye å si for valg av skåringsmetode.

Man skal være varsom med å trekke konklusjoner på bakgrunn av én studie, men dersom resultatene i studien stemmer, vil jeg anbefale å skåre smertetegninger ved bruk av kroppsområde-analyse dersom tegningen skal brukes som effektmål.

## 7 Litteraturliste

- Altman DG. *Practical Statistics for Medical Research*: Chapman and Hall/CRC; 1 edition, 1991.
- Befring E. *Forskningsmetode med etikk og statistikk*. OSLO: Samlaget, 2007.
- Bertilson B, Grunnesjo M, Johansson SE, Strender LE. Pain drawing in the assessment of neurogenic pain and dysfunction in the neck/shoulder region: inter-examiner reliability and concordance with clinical examination. *Pain Med* 2007; 8 (2): 134-146.
- Brismar H, Vucetic N, Svensson O. Pain patterns in lumbar disc hernia Drawings compared to surgical findings in 159 patients. *Acta Orthop Scand* 1996; 67 (5): 470-472.
- Bryner P. Extent measurement in localised low-back pain: a comparison of four methods. *Pain* 1994; 59 (2): 281-285.
- Cairns D, Thomas L, Mooney V, Pace JB. A comprehensive treatment approach to chronic low back pain. *Pain* 1976; 2 (3): 301-308.
- Carter RE, Domholdt E, Lubinsky J. *Rehabilitation Research: Principles and Applications*. St.Luis: Elsevier, 2011.
- Chan CW, Goldman S, Ilstrup DM, Kunselman AR, O'Neill PI. The pain drawing and Waddell's nonorganic physical signs in chronic low-back pain. *Spine (Phila Pa 1976)* 1993; 18 (13): 1717-1722.
- Cook TD. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*: Houghton Mifflin, 1979.
- Dahm K, Holm I, Bakke A, Brox JI. Testing av skulderfunksjon hos pasienter med proksimal humerus fraktur. *Fysioterapeuten* 2002; (11/2002).
- Dawson AP, Steele EJ, Hodges PW, Stewart S. Development and test-retest reliability of an extended version of the Nordic Musculoskeletal Questionnaire (NMQ-E): a screening instrument for musculoskeletal pain. *J Pain* 2009; 10 (5): 517-526.
- Dennis MD, Rocchio PO, Wiltse LL. The topographical pain representation and its correlation with mmpi scores. *Orthopedics* 1981; 5 (4): 432-434.
- Donelson R, Grant W, Kamps C, Medcalf R. Pain response to sagittal end-range spinal motion. A prospective, randomized, multicentered trial. *Spine (Phila Pa 1976)* 1991; 16 (6 Suppl): S206-212.
- Egloff N, Camara RJ, von Kanel R, Klingler N, Marti E, Ferrari ML. Pain drawings in somatoform-functional pain. *BMC Musculoskelet Disord* 2012; 13: 257.
- Formi. Smertemåling. 2013; Available from:  
<http://www.formi.no/helsepersonell/mer/smertemaling/> ;  
[http://www.formi.no/Pasienter/mer/plager\\_flest\\_og\\_koster\\_mest/](http://www.formi.no/Pasienter/mer/plager_flest_og_koster_mest/).
- Fortin JD, Aprill CN, Ponthieux B, Pier J. Sacroiliac joint: pain referral maps upon applying a new injection/arthrography technique. Part II: Clinical evaluation. *Spine (Phila Pa 1976)* 1994a; 19 (13): 1483-1489.
- Fortin JD, Dwyer AP, West S, Pier J. Sacroiliac joint: pain referral maps upon applying a new injection/arthrography technique. Part I: Asymptomatic volunteers. *Spine (Phila Pa 1976)* 1994b; 19 (13): 1475-1482.
- Gatchel RJ, Mayer TG, Capra P, Diamond P, Barnett J. Quantification of lumbar function. Part 6: The use of psychological measures in guiding physical functional restoration. *Spine (Phila Pa 1976)* 1986; 11 (1): 36-42.
- Geis C, Feierabend S, Bohner W, Kares H, Schirmer P, Busche E, Schindler HJ, Siegert J, Hugger S, Turp JC, Hugger A, Sommer C. [Pain drawings of patients with orofacial

- pain. Comparison of acceptance and gain of information]. *Schmerz* 2006; 20 (6): 498-508.
- Griffin JW, Tolley EA, Tooms RE, Reyes RA, Clifft JK. A comparison of photographic and transparency-based methods for measuring wound surface area. *Phys Ther* 1993; 73 (2): 117-122.
- Halvorsen K. *Forskningsmetode for helse- og sosialfag; en innføring i samfunnsvitenskapelig metode*, 2002.
- Hansen H. Smertemåling. 2013; Available from: <http://fysio.dk/fafo/Maleredskaber/Maleredskaber-alfabetisk/Smertemaling/>.
- Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care Res (Hoboken)* 2011; 63 Suppl 11: S240-252.
- Hildebrandt J, Franz CE, Choroba-Mehnen B, Temme M. The use of pain drawings in screening for psychological involvement in complaints of low-back pain. *Spine (Phila Pa 1976)* 1988; 13 (6): 681-685.
- IASP. International association for the study of pain. 1996.
- Ihlebaek C, Brage S, Natvig B, Bruusgaard D. [Occurrence of musculoskeletal disorders in Norway]. *Tidsskr Nor Laegeforen* 2010; 130 (23): 2365-2368.
- Jacobsen DI. *Hvordan gjennomføre undersøkelser?*, 2005.
- Jamison RN, Fanciullo GJ, Baird JC. Usefulness of pain drawings in identifying real or imagined pain: accuracy of pain professionals, nonprofessionals, and a decision model. *J Pain* 2004; 5 (9): 476-482.
- Jensen MP, Chen C, Brugger AM. Interpretation of visual analog scale ratings and change scores: a reanalysis of two clinical trials of postoperative pain. *J Pain* 2003; 4 (7): 407-414.
- Krause SJ, Tait RC, Margolis RB. Pain distribution, intensity, and duration in patients with chronic pain. *J Pain Symptom Manage* 1989; 4 (2): 67-71.
- Kuorinka I, Jonsson B, Kilbom A, Vinterberg H, Biering-Sorensen F, Andersson G, Jorgensen K. Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. *Appl Ergon* 1987; 18 (3): 233-237.
- Kvale A, Ellertsen B, Skouen JS. Relationships between physical findings (GPE-78) and psychological profiles (MMPI-2) in patients with long-lasting musculoskeletal pain. *Nord J Psychiatry* 2001; 55 (3): 177-184.
- Laake P, Hjartaker A, Thelle D, Veierød MB. *Epidemiologiske og kliniske forskningsmetoder*, 2007.
- Lacey RJ, Lewis M, Jordan K, Jinks C, Sim J. Interrater reliability of scoring of pain drawings in a self-report health survey. *Spine (Phila Pa 1976)* 2005; 30 (16): E455-458.
- Ljunggren AE. Descriptions of pain and other sensory modalities in patients with lumbago-sciatica and herniated intervertebral discs. Interview administration of an adapted McGill Pain Questionnaire. *Pain* 1983; 16 (3): 265-276.
- Love A, Leboeuf C, Crisp TC. Chiropractic chronic low back pain sufferers and self-report assessment methods. Part I. A reliability study of the Visual Analogue Scale, the Pain Drawing and the McGill Pain Questionnaire. *J Manipulative Physiol Ther* 1989; 12 (1): 21-25.
- Lund T. *Innføring i forskningsmetodologi*: Unipub forl., 2002.



- MacFarlane GJ, Croft PR, Schollum J, Silman AJ. Widespread pain: is an improved classification possible? *J Rheumatol* 1996; 23 (9): 1628-1632.
- Mann NH, 3rd, Brown MD, Enger I. Expert performance in low-back disorder recognition using patient pain drawings. *J Spinal Disord* 1992; 5 (3): 254-259.
- Margolis RB, Chibnall JT, Tait RC. Test-retest reliability of the pain drawing instrument. *Pain* 1988; 33 (1): 49-51.
- Margolis RB, Tait RC, Krause SJ. A rating system for use with patient pain drawings. *Pain* 1986; 24 (1): 57-65.
- McCormack HM, Horne DJ, Sheather S. Clinical applications of visual analogue scales: a critical review. *Psychol Med* 1988; 18 (4): 1007-1019.
- Melzack R. Phantom limbs and the concept of a neuromatrix. *Trends Neurosci* 1990; 13 (3): 88-92.
- Mengshoel AM, Vollestad NK, Forre O. Pain and fatigue induced by exercise in fibromyalgia patients and sedentary healthy subjects. *Clin Exp Rheumatol* 1995; 13 (4): 477-482.
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010; 19 (4): 539-549.
- North RB, Nigrin DJ, Fowler KR, Szymanski RE, Piantadosi S. Automated 'pain drawing' analysis by computer-controlled, patient-interactive neurological stimulation system. *Pain* 1992; 50 (1): 51-57.
- Nylenna M. Kappa. 2009; Available from: <http://sml.snl.no/kappa>.
- O'Sullivan P. Diagnosis and classification of chronic low back pain disorders: maladaptive movement and motor control impairments as underlying mechanism. *Man Ther* 2005; 10 (4): 242-255.
- Ohlund C, Eek C, Palmblad S, Areskoug B, Nachemson A. Quantified pain drawing in subacute low back pain. Validation in a nonselected outpatient industrial sample. *Spine (Phila Pa 1976)* 1996; 21 (9): 1021-1030; discussion 1031.
- Ohnmeiss DD. Repeatability of pain drawings in a low back pain population. *Spine (Phila Pa 1976)* 2000; 25 (8): 980-988.
- Ohnmeiss DD, Vanharanta H, Ekholm J. Degree of disc disruption and lower extremity pain. *Spine (Phila Pa 1976)* 1997; 22 (14): 1600-1605.
- Palmer H. Pain charts; a description of a technique whereby functional pain may be diagnosed from organic pain. *N Z Med J* 1949; 48 (264): 187-213.
- Parker H, Wood PL, Main CJ. The use of the pain drawing as a screening measure to predict psychological distress in chronic low back pain. *Spine (Phila Pa 1976)* 1995; 20 (2): 236-243.
- Postacchini F. Management of herniation of the lumbar disc. *J Bone Joint Surg Br* 1999; 81 (4): 567-576.
- Robinson HS. Pelvic girdle pain and disability during and after pregnancy. 2010.
- Robinson HS, Mengshoel AM. Assessments of lumbar flexion range of motion: intertester reliability and concurrent validity of 2 commonly used clinical tests. *Spine (Phila Pa 1976)* 2014; 39 (4): E270-275.
- Sanders NW, Mann NH, 3rd. Automated scoring of patient pain drawings using artificial neural networks: efforts toward a low back pain triage application. *Comput Biol Med* 2000; 30 (5): 287-298.
- Shadish WR, T.D. C, Campbell DT. *Experimental and Quasi-experimental Designs for Generalized Causal Inference*, 2001.

- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86 (2): 420-428.
- Sirnes E, Sodal E, Nurk E, Tell GS. [Occurrence of musculoskeletal complaints in Hordaland]. *Tidsskr Nor Laegeforen* 2003; 123 (20): 2855-2859.
- Stuge B, Laerum E, Kirkesola G, Vollestad N. The efficacy of a treatment program focusing on specific stabilizing exercises for pelvic girdle pain after pregnancy: a randomized controlled trial. *Spine (Phila Pa 1976)* 2004; 29 (4): 351-359.
- Sturesson B, Uden G, Uden A. Pain pattern in pregnancy and "catching" of the leg in pregnant women with posterior pelvic pain. *Spine (Phila Pa 1976)* 1997; 22 (16): 1880-1883; discussion 1884.
- Toomey TC, Gover VF, Jones BN. Spatial distribution of pain: a descriptive characteristic of chronic pain. *Pain* 1983; 17 (3): 289-300.
- Toomingas A. Characteristics of pain drawings in the neck-shoulder region among the working population. *Int Arch Occup Environ Health* 1999; 72 (2): 98-106.
- Uden A. Pain drawing in lumbar disc hernia. *Acta Orthop Scand* 1997; 68 (2): 182.
- Uden A, Astrom M, Bergenudd H. Pain drawings in chronic back pain. *Spine (Phila Pa 1976)* 1988; 13 (4): 389-392.
- Uden A, Landin LA. Pain drawing and myelography in sciatic pain. *Clin Orthop Relat Res* 1987; (216): 124-130.
- Wolfe F, Hauser W. Fibromyalgia diagnosis and diagnostic criteria. *Ann Med* 2011; 43 (7): 495-502.

## Vedlegg 1

Smertetegning med beskrivelse av prosedyre for skravering.

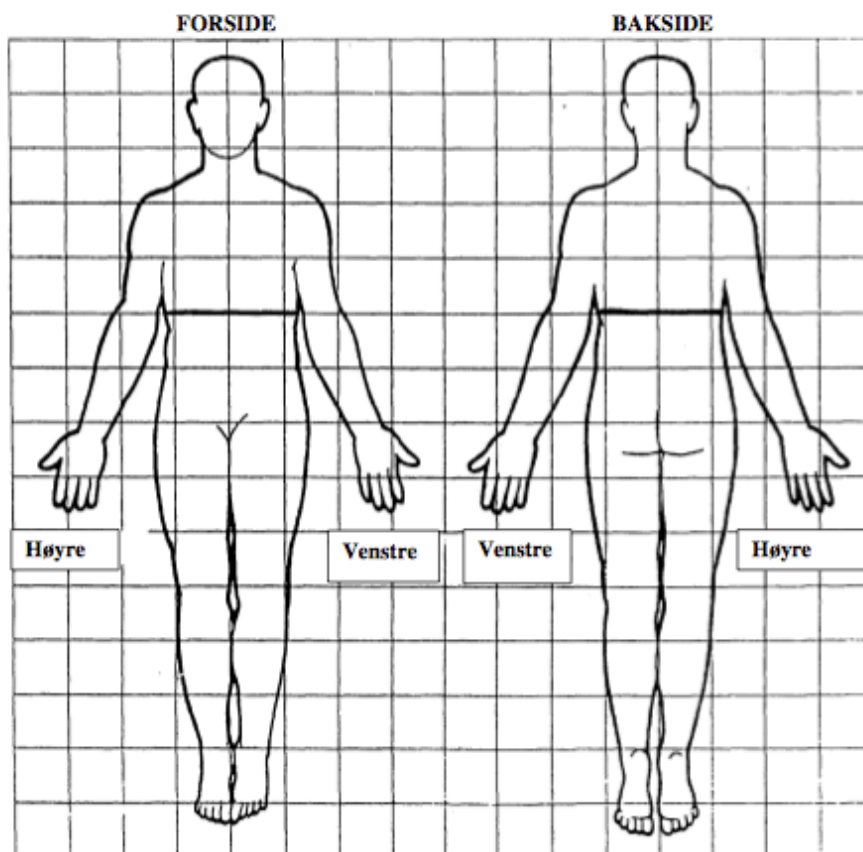
### Smerteskjema

Dato:

Navn:

Født:

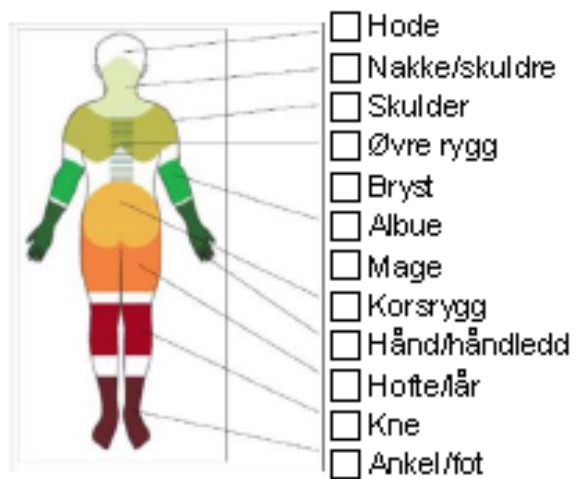
Skraver med kulepenn områdene på kroppen hvor du har hatt smerter i løpet av de siste 14 dager.



## Vedlegg 2

Kopi av NPQ.

12. Smerteutbredelse: Kan du sette kryss i boksen eller boksene for de områdene der du har smerte, eller har hatt smerte de siste to ukene



The diagram shows a human figure with various body parts highlighted in different colors: head (white), neck/shoulders (yellow), shoulders (green), upper back (yellow), chest (orange), elbows (green), stomach (orange), lower back (yellow), hands/wrists (green), hips/thighs (red), knees (red), and ankles/feet (red). Lines connect these regions to a checklist on the right.

- Hode
- Nakke/skuldre
- Skulder
- Øvre rygg
- Bryst
- Albue
- Mage
- Korsrygg
- Hånd/håndledd
- Hofte/lår
- Kne
- Ankel/fot