

Marianne Bakke Johnsen

Reliabilitet av isokinetiske muskelstyrkemålinger og hinketester for barn

- En metodestudie

Masteroppgave i idrettsfysioterapi

Seksjon for idrettsmedisinske fag
Norges idrettshøgskole, 2011

Forord

Masterstudiet i idrettsfysioterapi har vært en utrolig lærerik og inspirerende tid hvor jeg har truffet mange hyggelige, og ikke minst dyktige mennesker.

Først av alt vil jeg takke min hovedveileder, May Arna Risberg, for å ha guidet meg gjennom hele prosessen med å gjennomføre et studie og til slutt føre alt ned på papiret! Jeg har lært utrolig mye underveis. Takk for at du alltid har vært tilgjengelig og hatt god tid til veiledning underveis i prosessen.

En stor takk også til min biveileder, Håvard Moksnes, som har bidratt hele veien med gode tilbakemeldinger, praktiske råd, og ikke minst – testet alle barna mine!

Takk til Ingrid Eitzen som også testet for prosjektet mitt. Jeg setter stor pris på alle timene du har bidratt med! Takk for faglige råd og hjelp med gjennomgang av data.

Takk til Kristin Bølstad som har vært til stor hjelp med gjennomføringen av studien på Hjelp24 NIMI Ullevål. Du har gjort alt så mye lettere!

En stor takk til alle barna fra Ullern og Heming som ble med på testingen! Takk også til trenere og foreldre. Takk til Arnhild, Carina, Maren, og Nina for hjelp på testdagene!

Takk til Ingar Holme for gode statistiske råd underveis.

Jeg vil også gjerne få takke arbeidsgiveren min ved Oslo universitetssykehus Ullevål, og da spesielt Mona Mortensen, for å ha gitt meg stor fleksibilitet gjennom hele masterstudiet. Takk til alle mine medstudenter som har vært med på å gjøre denne studietiden mer interessant, og ikke minst morsom! Til slutt en stor takk til venner og familie som alltid støtter meg og har tro på meg.

Oslo, mai 2011.

Marianne Bakke Johnsen

Sammendrag

Introduksjon: Hinketester og isokinetiske muskelstyrkemålinger blir mye brukt for å evaluere knefunksjon etter ACL skade. Målemetodene har demonstrert høy reliabilitet ($ICC > 0,80$), men studiene er i hovedsak gjennomført på voksne. Eldre studier har vist at isokinetiske muskelstyrkemålinger er en reliabel målemetode også på barn og unge (6-15 år). Ingen studier har undersøkt reliabiliteten av hinketestene på barn. I en pågående kohortstudie ved NAR brukes hinketester og isokinetiske muskelstyrkemålinger i evalueringen av knefunksjon på barn med ACL skade. Formålet med metodestudien var å undersøke intra- og interrater reliabiliteten av disse testene på et utvalg av barn uten skade. Vi ønsket i tillegg å gi en indikasjon på minste reelle endring (SRD) for målemetodene basert på reliabilitetsresultatene.

Metode: Vi inkluderte 29 fysisk aktive barn ($12,4 \pm 0,3$ år). Barna ble testet tre ganger fordelt på to testdager, med en ukes intervall. To fysioterapeuter testet barna. Hinketestene ble målt i cm og sekunder og styrkevariablene i Nm og joule. Begge ben ble testet i randomisert rekkefølge. Reliabiliteten ble beregnet ved bruk av ICC, SEM og LOA. Vi beregnet SRD som en indikasjon på minste reelle endring til målemetodene.

Resultater: 26 barn ble inkludert for statistiske analyser. Hinketestene viste ICC verdier fra 0,62-0,91. Tilsvarende verdier ble funnet for de isokinetiske muskelstyrkemålingene ($ICC 0,76-0,94$), med unntak av TQ30 som viste lavere reliabilitet ($ICC 0,48-0,71$). SEM % varierte for styrkemålingene fra 5,1-14,5 % og fra 3,9-7,4 % for hinketestene. Minste reelle endring (SRD %) varierte fra 10,8-20,5 % for hinketestene og fra 14,0-25 % (TQ30 40,3 %) for styrkevariablene.

Konklusjon: Hinketestene og de isokinetiske muskelstyrkemålingene viste fra moderat til høy reliabilitet, unntaket var TQ30 som viste fra lav til moderat reliabilitet. Størrelsen på SEM indikerte at målemetodene var reliable testet på et utvalg av aktive friske barn. Basert på reliabilitetsresultatene indikerte SRD at måleforskjeller på over 20 % for hinketestene og over 25 % for de isokinetiske muskelstyrkemålingene potensielt kan vurderes som en reell endring i knefunksjon.

Forkortelser

ACL	Anterior cruciate ligament; fremre korsbånd
ANOVA	Variasjonsanalyse
AVG PT	Average peak torque
CV	Variasjonskoeffisient
ICC	Intraclass correlation coefficient
ICF	International classification of functioning, disability and health
J	Joule
KI	Konfidensintervall
LOA	Limits of agreement
LSI	Leg symmetry index
NIH	Norges idrettshøgskole
NAR	Norsk forskningssentre for aktiv rehabilitering
Nm	Newton meter
OUS	Oslo universitetssykehus
PT	Peak torque
RMSE	Residual mean square error
SEM	Standard error of measurement
SRD	Smalles real difference
TQ30	Torque ved 30 grader
Tw	Total work

Tabelloversikt

Tabell 1: Eksempler på ICF målenivåer og forslag til målemetoder etter ACL skade..	19
Tabell 2: Oversikt over reliabilitetsstudier for isokinetiske muskelstyrkemålinger på barn uten skade.	34
Tabell 3: Oversikt over reliabilitetsstudier for isokinetiske muskelstyrkemålinger på voksne friske og med ACL skade.....	35
Tabell 4: Oversikt over reliabilitetsstudier for hinketestene på voksne med og uten ACL skade.	36
Tabell 5: Antropometriske data for barna i metodestudien (n= 28).....	52
Tabell 6: Gjennomsnitt \pm SD og p-verdier for konsentriske isokinetiske muskelstyrkemålinger med hensyn til intrarater reliabiliteten (n= 26).	54
Tabell 7: Gjennomsnitt \pm SD og p-verdier for konsentriske isokinetiske muskelstyrkemålinger med hensyn til interrater reliabiliteten (n= 26).	55
Tabell 8: Gjennomsnitt \pm SD og p-verdier for hinketestene med hensyn til intrarater reliabiliteten (n= 26).	56
Tabell 9: Gjennomsnitt \pm SD og p-verdier for hinketestene med hensyn til interrater reliabiliteten (n= 26).	57
Tabell 10: Relativ og absolutt intrarater reliabilitet presentert for de ulike isokinetiske muskelstyrkemålingene (n= 26).	58
Tabell 11: Relativ og absolutt interrater reliabilitet for de ulike isokinetiske muskelstyrkemålingene (n= 26).	59
Tabell 12: Relativ og absolutt intrarater reliabilitet for hinketestene (n=26).	60
Tabell 13: Relativ og absolutt interrater reliabilitet for hinketestene (n= 26).	60

Figuroversikt

Figur 1: Flytskjema over rekrutteringsprosedyrer.....	39
Figur 2: Flytskjema som viser inklusjon/ekslusjon og frafall ved testing.....	41
Figur 3: Biodex protokoll.....	43
Figur 4: Presentasjon av hinketestene.....	44
Figur 5: Tidstabell for den totale undersøkelsesprotokollen.....	44
Figur 6: Isokinetisk muskelstyrke målt med et Biodex 6000 dynamometer.....	47
Figur 7: Bland-Altman plott med hensyn til interrater reliabilitet (n= 26).....	63

Innhold

Forord	3
Sammendrag	4
Forkortelser	5
Tabelloversikt	6
Figuroversikt	7
Innhold	8
1. Innledning	11
1.1 Bakgrunn for oppgaven.....	11
1.2 Formål med studien og problemstilling	14
1.3 Begrepsavklaringer & definisjoner	14
2. Teori	16
2.1 Hvorfor måler vi?	16
2.2 Hva måler vi?	16
2.2.1 Funksjon	16
2.3 Hvem måler vi?	17
2.3.1 Målemetoder etter ACL skade.....	18
2.4 Målemetoder for å vurdere knedefunksjon.....	19
2.4.1 Isokinetiske muskelstyrkemålinger – ICF kroppsfunksjoner	19
2.4.2 Hinketestene – ICF aktivitet	21
2.5 Målemetodenes egenskaper.....	22
2.5.1 Reliabilitet	23
2.5.2 Kilder til variasjon	23
2.5.3 Kvantifisering av reliabilitet	26
2.5.4 Standard for størrelsen av reliabilitetskoeffisienten	28
2.5.5 Validitet & responsivness	29
2.6 Studier på reliabilitet.....	30
2.6.1 Reliabilitet av isokinetiske muskelstyrkemålinger.....	31
Reliabilitet av hinketestene	33
2.7 Testing av barn.....	37
3. Materiale og metode	38
3.1 Bakgrunn for valg av deltagere	38

3.2	Utvalg	38
3.2.1	Rekrutteringsstrategier.....	38
3.3	Testere	40
3.4	Prosedyrer for datainnsamlingen	41
3.5	Studiedesign	41
3.5.1	Målemetodene	42
3.6	Datainnsamling	44
3.6.1	Testrekkefølge	45
3.6.2	Registreringsskjema.....	45
3.6.3	Organisert aktivitet & idrett.....	46
3.6.4	Randomisering av startben	46
3.6.5	Antropometriske data.....	46
3.7	Testprosedyrer	46
3.7.1	Isokinetisk muskelstyrketesting	46
3.7.2	Hinketester.....	47
3.8	Databehandling	48
3.8.1	Målevariabler.....	48
3.8.2	Utvalgsstørrelse	48
3.8.3	Statistiske analyser.....	49
3.9	Etikk	50
4.	Resultater	52
4.1	Karakteristikk av utvalget	52
4.2	Deskriptive data til målemetodene	53
4.2.1	Isokinetiske muskelstyrkemålinger	54
4.2.2	Hinketester.....	56
4.2.3	Oppsummering av deskriptive data	57
4.3	Reliabiliteten til målemetodene	57
4.3.1	Isokinetisk muskelstyrke	58
4.3.2	Hinketestene	59
4.3.3	Bland – Altman plott	61
4.3.4	Oppsummering av reliabilitet	63
4.4	Responsivness	63
5.	Diskusjon	64
5.1	Resultater	64
5.1.1	Reliabilitet av isokinetisk muskelstyrkemålinger	65
5.1.2	Reliabilitet av hinketestene.....	69
5.1.3	Responsivness til målemetodene	72
5.2	Metode	73
5.2.1	Utvalget	73
5.2.2	Testere	74
5.2.3	Studiedesign.....	75

5.2.4	Datainnsamling og testprosedyrer	76
5.2.5	Målemetodene	78
5.3	Valg av statistiske metoder	80
5.3.1	Relativ reliabilitet	80
5.3.2	Absolutt reliabilitet	81
5.3.3	Måling av endring i funksjon.....	82
5.4	Kliniske implikasjoner	82
6.	Konklusjon.....	84
	Referanser.....	85
	Vedlegg	94

1. Innledning

1.1 *Bakgrunn for oppgaven*

Dette masterprosjektet er del av en pågående studie ved Norsk forskningssenter for aktiv rehabilitering (NAR), som er et forskningssamarbeid mellom Norges idrettshøgskole, Ortopedisk avdeling–Oslo universitetssykehus Ullevål og Hjelp24 NIMI. Studien har tittelen: ”*Barn med fremre korsbåndskade. En prospektiv kohortstudie av barn under 13 år*”.

Det har blitt rapportert en økende forekomst av korsbåndsskader (ACL skader) blant barn og unge de to siste tiårene (Bales, Guettler, & Moorman, III, 2004; Shea, Pfeiffer, Wang, Curtin, & Apel, 2004; Mohtadi & Grant, 2006). Flere deltar i dag tidligere i idrettsspesifikk trening og konkurranser og som konsekvens har dette ført til en økt risiko for idrettsrelaterte skader blant denne gruppen (Caine, Maffulli, & Caine, 2008; Shanmugam & Maffulli, 2008). I Norge har det blitt registrert 979 operasjoner på barn og unge i aldersgruppen 11-16 år siden oppstarten av Korsbåndregisteret i 2004 og frem til og med 2009 (Granan, muntlig meddelelse, 18.mars, 2011). Dette medfører et særskilt behov for å utvikle gode rehabiliteringsprogram, og også et behov for å utvikle reliable og valide utfallsmål for barn med ACL skade.

Det er viktig å inkludere flere forskjellige utfallsmål for å fange opp de ulike aspektene og konsekvensene av en ACL skade (Bent, 2009; Wright, 2009). Vi kan klassifisere funksjon etter en ACL skade ut i fra retningslinjene til International Classification of Functioning, Disability and Health (ICF), som relateres til kroppsstrukturer, -funksjoner, aktivitet og deltagelse (WHO, 2001; Lysholm & Tegner, 2007). Innen fysioterapien blir det lagt vekt på å bruke standardiserte målemetoder for å undersøke, og evaluere klinisk praksis (Clark, 2001). Målemetodene vi bruker må i tillegg til å være klinisk relevante, inneha egenskaper som å være reliable, valide og kunne fange opp endringer over tid (Shaw, Chipchase, & Williams, 2004; Wright, 2009).

Hinketester og isokinetiske muskelstyrkemålinger er eksempler på målemetoder som brukes for å evaluere knefunksjon hos voksne etter ACL skade (Barber, Noyes, Mangine, McCloskey, & Hartman, 1990; Noyes, Barber, & Mangine, 1991; Eastlack, Axe, & Snyder-Mackler, 1999; Fitzgerald, Axe, & Snyder-Mackler, 2000).

Hinketestene (ICF aktivitet) er vanlig å bruke som et praktisk, prestasjonsbasert utfallsmål for å kvantifisere funksjon i underekstremitetene (Reid, Birmingham, Stratford, Alcock, & Giffin, 2007). Isokinetiske muskelstyrkemålinger (ICF kroppsfunksjoner) er en foretrukket målemetode for å evaluere muskelstyrke både hos friske (uten skade) og etter ACL skade (Eitzen, Moksnes, Snyder-Mackler, Engebretsen, & Risberg, 2010). Andre tester som ofte brukes i tillegg er KT-1000 målinger (fremre tibiofemoral translasjon) (ICF kroppstrukturer) og selvrapporing i form av ulike spørreskjemaer (f.eks. ICF aktivitet) (Bent, 2009; Shaw et al., 2004). Disse testene er ikke validert eller utviklet for barn, og brukes som oftest direkte adoptert fra voksne.

Reliabiliteten av hinketester og isokinetisk muskelstyrke er godt beskrevet i litteraturen, men disse studiene er i hovedsak gjennomført på voksne. Hinketestene har demonstrert høy intrarater reliabilitet blant friske voksne (ICC 0,80-0,97) (Bolglia & Keskula, 1997; Munro & Herrington, 2011; Ross, Langford, & Whelan, 2002) og etter ACL skade (ICC 0,82-0,92) (Hopper et al., 2002; Reid et al., 2007). Ingen studier, så langt vi har funnet, har undersøkt reliabiliteten av hinketestene for barn.

Reliabilitetsstudier på voksne indikerer også at isokinetisk muskelstyrkemålinger (for kneekstensjon og fleksjon) er en reliabel metode på friske (Sole, Hamren, Milosavljevic, Nicholson, & Sullivan, 2007) og etter ACL skade (Ross, Irrgang, Denegar, McCloy, & Unangst, 2002). For begge grupper er det rapportert høy reliabilitet med ICC verdier fra 0,81-0,99. To studier har undersøkt interrater reliabiliteten (Hartmann, Knols, Murer, & De Bruin, 2009; Keskula, Dowling, Davis, Finley, & Dell'omo, 1995). De fant tilsvarende høye ICC verdier som ved intrarater reliabiliteten (ICC > 0,90).

Eldre studier har vist at isokinetiske muskelstyrkemålinger er en reliabel målemetode også på barn og unge fra 6-15 år (Merlini, Dell'Accio, & Granata, 1995; Molnar, Alexander, & Gutfeld, 1979; Mohtadi, Kiefer, Tedford, & Watters, 1990). To nyere studier har angitt moderat til høy reliabilitet (ICC 0,55-0,98) for kneekstensjon og fleksjon testet med hastigheter fra 30-180°/sek. (Kellis, Kellis, Gerodimos, & Manou, 1999; Deighan, De Ste Croix, & Armstrong, 2003). I studiene på barn har peak torque blitt testet, mens noen av studiene på voksne i tillegg har testet total work (Keskula et al., 1995; Sole et al., 2007) og average peak torque/-power (Hartmann et al., 2009;

Maffiuletti, Bizzini, Desbrosses, Babault, & Munzinger, 2007). En nyere studie fra NAR refererer til størst quadriceps deficitt under 40° knefleksjon etter ACL skade (Eitzen, Eitzen, Holm, Snyder-Mackler, & Risberg, 2010). Peak torque ved 30° kan derfor være en interessant variabel å undersøke. På bakgrunn av dette inkluderte vi denne variabelen som en del av styrkevariablene vi reliabilitetstestet.

Hinketester og isokinetiske muskelstyrkemålinger inngår i vurdering av knefunksjon etter ACL skade i den pågående kohortstudien på barn (Moksnes, Engebretsen, & Risberg, 2008). Testene er også presentert og publisert i tidligere ACL studier av forskere ved NAR (Eitzen et al., 2010; Moksnes & Risberg, 2009; Moksnes et al., 2008; Oiestad, Holm, Gunderson, Myklebust, & Risberg, 2010).

Selv om målemetodene har vist seg reliable på den voksne populasjonen er det ingen selvfølge at disse resultatene er direkte overførbare til barn. En test er ikke reliabel i seg selv, men er nært knyttet opp til utvalget den har blitt testet på og forholdene for testingen (Streiner & Norman, 2008). Vi ønsket på bakgrunn av at dette å undersøke reliabiliteten av hinketestene og isokinetiske muskelstyrkemålinger på et utvalg av barn. For å få et estimat på forventet ”normalvariasjon” til målemetodene ønsket vi å innhente reliabilitetsdata fra et utvalg av barn uten skade.

I den pågående kohortstudien er ofte flere testere involvert og vi ønsket derfor å undersøke både intra- og interrater reliabiliteten til målemetodene som blir benyttet. Undersøkelse av reliabilitet er viktig for videre å kunne evaluere validiteten til en målemetode (Portney & Watkins, 2000). Barn kan tenkes å variere mer i prestasjon på bakgrunn av andre fysiske og mental forutsetninger enn voksne. Det er derfor viktig å undersøke testprotokollene på det aktuelle utvalget, for å se om det er endringer vi kan gjøre for å forbedre reliabiliteten til målemetodene i vår praksis. Reliabiliteten kan være med på å gi oss et estimat på hvor stor forskjell i score som er nødvendig for å kunne fastslå at en reell klinisk endring har skjedd.

1.2 Formål med studien og problemstilling

Hovedformålet med metodestudien var å undersøke reliabiliteten av hinketester og isokinetiske muskelstyrkemålinger brukt i vurderingen av knefunksjon hos barn med ACL skade, på et utvalg av barn uten skade.

Følgende problemstilling ønskes besvart:

- *Hva er intra- og interrater reliabiliteten for to testere for isokinetiske muskelstyrkemålinger og hinketester for barn uten skade?*

Vi undersøkte reliabiliteten av fire hinketester (Noyes et al., 1991), og så i tillegg på ulike isokinetiske styrkevariabler for å få en indikasjon på hvilke variabler som var mest reliable for dette utvalget. Reliabilitetsresultatene ble brukt for å beregne minste reelle endring, for fremtidig å kunne bruke dataene til å evaluere endring i funksjon hos barna med ACL skade i den pågående kohortstudien, og for andre tilsvarende studier.

Del-problemstilling:

- *Hva er minste reelle endring for isokinetiske muskelstyrkemålinger og hinketestene basert på reliabilitetsdata fra et utvalg av barn uten skade?*

1.3 Begrepsavklaringer & definisjoner

Aktivitet – defineres i denne oppgaven som fysisk aktivitet, som omfatter all slags kroppsbevegelse utført av skjelettmuskulatur som øker energiforbruket (Bahr, 2009).

Barn – i denne oppgaven forstås termen *barn* som jenter og gutter med kronologisk alder til og med 13 år (i henhold til inklusjonskriteriet av barn i den pågående kohortstudien).

Frisk - betyr her 'uten skade' i underekstremitetene, eller annen skade (se def. skade) som kan hindre deltagelse i hinketester og isokinetiske muskelstyrketester.

Funksjon – en term som omfatter alle kroppsfunksjoner, aktiviteter og deltagelse (WHO, 2001).

Insidens – antall nye tilfeller (feks skadetilfeller) i en spesifisert populasjon over en definert tidsperiode (Kunnskapssenteret, 2010).

Isokinetikk - dynamisk muskelkontraksjon hvor hastigheten av bevegelsen er kontrollert og holdt konstant ved hjelp av et dynamometer (Holm, 1996).

Knefunksjon - kan ut i fra ICF klassifiseringen inkludere aspekter som omfatter både kroppsstrukturer, kroppsfunksjoner og aktivitet og deltagelse (WHO, 2001), som f.eks. muskelstyrke (ICF kroppsfunksjon), leddstabilitet (ICF kroppsstruktur/-funksjon), smerte (ICF kroppsfunksjon), gangfunksjon (ICF aktivitet) m.fl.

Metodebarna – barna som ble testet i denne metodestudien.

Muskelstyrke – den maksimale spenning en muskel eller muskelgruppe kan utøve ved en frivillig handling under en spesifikk testsituasjon (Jones & Stratton, 2000).

Responsivness - evnen til å fange opp små endringer over tid (Portney & Watkins, 2000). Målt i denne studien som 'smallest real difference' (SRD), betegnet som minste reelle endring.

Skade – forstås her som en tilstand som krever tid borte fra trening/aktiviteter, og/eller behov for å oppsøke helsepersonell pga det aktuelle (van Mechelen, Hlobil, & Kemper, 1992). Spørsmål om skade ble brukt i registreringsskjema ved inklusjon av deltagere.

Utfallsmål – i studier må det foreligge en klar definisjon av utfallene og en god beskrivelse av hvordan de måles. Endepunktene kan være dikotome eller kontinuerlige (Kunnskapssenteret, 2010).

2. Teori

2.1 *Hvorfor måler vi?*

En stor del av fysioterapien består i å kunne vurdere pasienter eller friske individer i rehabilitering eller idrettslig sammenheng. Hos friske kan det være et redskap for å avdekke asymmetrier som potensielt kan predisponere for skade (Bandy, Rusche, & Tekulve, 1994; Hopper et al., 2002). Hos de med skade eller sykdom kan det være for å vurdere utfall av eventuelle dysfunksjoner, fremgang i rehabilitering eller predikere videre forløp av behandling. Der det ikke eksisterer eksplisitte definisjoner av hva som er normal funksjon og hva som regnes som avvik, vil utfallet av en slik vurdering ofte være påvirket av den som vurderer. Det er derfor viktig å utvikle, og å bruke, gode målemetoder basert på operasjonelle (målbare) definisjoner av det som er fokusområdet til en målemetode. Måleresultatene kan brukes som et redskap i kommunikasjon med kollegaer, på tvers av faggrupper og i kvalitetssikring og forskning (Beyer & Magnusson, 2003).

2.2 *Hva måler vi?*

For å bestemme hva vi skal måle, kan det være nyttig å ta utgangspunkt i International Classification of Functioning, Disability and Health (ICF) sin begrepsmessige klassifisering av helse og helserelevante tilstander, som omfatter to hovedområder; 1) funksjon og funksjonshemming og 2) kontekst. Første hovedområde (1) omhandler kroppen, hvor funksjon og struktur beskrives som kroppsfunksjoner og kroppsstrukturer, i tillegg til aktivitet og deltagelse. Kontekstuelle faktorer består av miljøfaktorer og personlige faktorer (WHO, 2001). De kontekstuelle faktorene blir ikke belyst i denne oppgaven. ICF er først og fremst et deskriptivt system, og ikke et måleredskap i seg selv. Hvilket nivå (ICF kroppsstrukturer, -funksjoner, aktivitet og deltagelse) man måler bør baseres på hva som er relevant for pasienten og stadiet vedkommende befinner seg på i rehabiliteringen og klinisk skjønn (Bent, 2009).

2.2.1 *Funksjon*

Nøytrale (*”problemfrie”*) sider av kroppsfunksjoner, aktivitet og deltagelse kan beskrives under samlebetegnelsen *funksjon*. Kroppsfunksjoner refererer til kroppens fysiologiske (og mentale) funksjoner, og kroppsstrukturer refererer til de ulike anatomiske delene av kroppen (WHO, 2001). Signifikante avvik fra ”normalen” kan

forklares ut i fra de fire kategoriene av funksjon og funksjonshemming hvor kroppstrukturer; referer til selve skaden, kroppsfunksjoner; til symptomer og tegn fra skaden, begrensninger i aktiviteter; til vanskeligheter med å utføre funksjonelle aktiviteter og deltagesinnskrenkninger; til problemer en person kan oppleve ved deltagelse i situasjoner i det dalige liv (Lysholm & Tegner, 2007).

Knefunksjon kan da ut i fra ICF klassifiseringen inkludere aspekter som omfatter både kroppstrukturer, -funksjoner, aktivitet og deltagelse (WHO, 2001). Strukturer som er med på å opprettholde funksjon i underekstremitetene er blant annet ligamenter, menisker, muskulatur/sene og benstrukturer. Vi kan tenke oss at muskelstyrke, -kraft, -fleksibilitet, leddmobilitet, proprioepsjon, nevromuskulær kontroll, dynamisk/statisk balanse, smidighet og evne til å utføre en handling/oppgave alle er komponenter som kan inngå i termen knefunksjon.

2.3 Hvem måler vi?

Barn og ACL skader vil bli nevnt i dette avsnittet da bakgrunnen for denne oppgaven var å belyse noen av målemetodene (hinketester og isokinetiske muskelstyrkemålinger) som brukes i vurderingen av knefunksjon hos barn med ACL skade i en pågående prospektive kohortstudien ved NAR.

Blant alvorlige kneskader har det blitt registrert en økende forekomst av ACL skader gjennom de to siste tiårene for barn og unge (Bales et al., 2004; Caine et al., 2008; Mohtadi & Grant, 2006; Shea et al., 2004). Barn deltar i dag i større grad i idrettsrelatert trening og konkurranser. Dette har ført til en økt risiko for idrettsrelaterte skader også blant denne aldersgruppen (Caine et al., 2008; Shanmugam & Maffulli, 2008). De fleste av skadene er lokalisert til underekstremitetene, hvor ankel og kne er de vanligste skadeområdene. Skadetyper som oftest blir rapportert er forstuvning, strekk- og støtskader (Caine, Caine, & Maffulli, 2006). Økt insidens av ACL skader kan være et resultat av større fysiske krav og belastning ved økt idrettsdeltagelse. Store krav til fart, teknikk og utstyr er med på å gi økt stressbelastning av ledd- og båndstrukturer (Bales et al., 2004). Samtidig kan forbedret billeddiagnostikk og økt interesse og oppmerksomhet rundt det faktum at ACL skader også forekommer hos barn ha bidratt til økt rapportering.

Selv om ACL skader hos barn ser ut til å være knyttet opp mot idrettsaktivitet. Er det bare en studie hittil som har rapportert bruk av prestasjonsbaserte tester (f.eks. hinketester og isokinetisk muskelstyrkemålinger) som utfallsmål for å klassifisere knefunksjon hos denne gruppen (Moksnes et al., 2008). Klassifiseringen av såkalte 'copers' og 'non-copers' er basert på behandlingsalgoritmer som er blitt introdusert for voksne etter ACL skade (Fitzgerald et al., 2000). En lignende behandlingsalgoritme for å guide videre behandling finnes per i dag ikke for barn (Moksnes et al., 2008). For å vite om målemetodene kan brukes på barn til samme formål, må man undersøke om testene er reliable og valide i samme grad for denne populasjonen. Målemetodene vi velger å bruke må være tilpasset barns forutsetninger på ulike nivå, det være seg fysiske og/eller mentale forutsetninger for å kunne gjennomføre en test/målemetode.

2.3.1 Målemetoder etter ACL skade

Det er viktig etter en ACL skade/rekonstruksjon å bruke forskjellige målemetoder for å identifisere elementer av redusert knefunksjon som kan være knyttet opp til ulike ICF nivåer. Eksempler på dysfunksjon kan være: ACL skaden i seg selv (kroppstrukturer), redusert muskelstyrke og smerte (kroppsfunksjoner), redusert gangkapasitet og/eller hinkedistans (begrensninger i aktivitet) og som følge av dette redusert deltagelse i idrett eller sosiale settinger (innskrenkninger i deltagelse). Som en del av ett testbatteri for å vurdere knefunksjon etter ACL skade, inngår ofte både selvrapporterte målemetoder (i form av spørreskjemaer), prestasjonsbaserte tester (hinketester og isokinetiske muskelstyrkemålinger) og kliniske tester (KT1000 målinger).

Målemetodene vi har testet i denne studien er knyttet opp til ICF kroppsfunksjoner (isokinetisk muskelstyrke) og ICF aktivitet (hinketester). Disse er eksempler på tester ofte omtalt i litteraturen som *prestasjonsbaserte tester* (Moksnes et al., 2008; Reid et al., 2007). De er beskrevet som fysiske tester som gir et kvantitativt mål på knefunksjon og samlebetegnelsen prestasjonsbaserte tester kan i så måte inkludere tester som gjenspeiler ulike ICF nivåer.

En av hoveddysfunksjonene etter ACL skade/rekonstruksjon er redusert quadriceps styrke og dynamisk instabilitet (Eitzen, Holm, & Risberg, 2009). Quadriceps muskelarbeid er en viktig bidragsyter for dynamisk knestabilitet gjennom hele bevegelsesbanen ved gange. Isokinetiske muskelstyrkemålinger er en foretrukket

målemetode for å evaluere muskelstyrke både hos pasienter etter en ACL skade og hos friske individer (Eitzen et al., 2010).

Tabell 1: Eksempler på ICF målenivåer og forslag til målemetoder etter ACL skade.

ICF nivå	Eksempel	Målemetode
Kroppsstrukturer	ACL skade	MRI
Kroppsfunksjoner	Nedsatt muskelkraft	Isokinetisk muskelstyrke
Aktivitet	Nedsatt hoppdistanse	Hinketester
Deltagelse	Ute av idrett	Idrettsspesifikke tester

MRI = magnetic resonance image. Tabellen er hentet fra artikkelen til Bent (2009).

Målemetoder kan brukes i ulike sammenhenger og resultatene kan dermed benyttes til ulike formål. Vi kan inndele måletoder i tre overordnede kategorier: 1) diagnostiske målemetoder; som har til hensikt å klassifisere forskjellige problemer eller kontrollere en bestemt befolkningsgruppe, 2) målemetoder til å evaluere endringer over tid; som har til hensikt å dokumentere resultater ved hjelp av utfallsmål (dvs. redskaper for måling av resultater/effekt) og 3) prognostiske målemetoder; som har det formål å forutsi sannsynlighet for en hendelse/utvikling vil kunne skje i fremtiden (Beyer & Magnusson, 2003).

2.4 Målemetoder for å vurdere knefunksjon

Jeg vil her presentere de målemetodene vi har reliabilitetstestet i denne metodestudien, og som brukes i pågående og publiserte studier ved NAR. Andre tester som eventuelt inngår i testbatteriet for vurdering av knefunksjon blir ikke belyst i denne oppgaven.

2.4.1 Isokinetiske muskelstyrkemålinger – ICF kroppsfunksjoner

Muskelfunksjon blir typisk undersøkt som styrke, enten under isokinetisk, isotonisk, eller isometriske betingelser. Isokinetisk dynamometri har lenge blitt ansett som et av de mest valide redskapene vi har for å undersøke muskelfunksjon, og er mye brukt både i klinikken og i forskning (Jones & Stratton, 2000). Målingene fra isokinetisk muskelstyrke tillater kvantifisering av variabler relatert til *torque* (peak og average) som er en funksjon av muskelstyrke, *power* (kraft), *work* (arbeid), leddvinkel og utholdenhet (De Ste Croix, Deighan, & Armstrong, 2003). Isokinetiske dynamometre gir en konstant angulær hastighet med tilpassende motstand gjennom en satt bevegelsesbane (ROM) (Drouin, Valovich-mcLeod, Shultz, Gansneder, & Perrin,

2004). De fleste isokinetiske dynamometre tillater testing av både konsentrisk og eksentrisk muskelfunksjon (Jones & Stratton, 2000).

Isokinetisk muskelstyrke kan testes under ulike hastigheter. Ved lavere hastigheter (30-60°/sek.) utføres det vanligvis et lavt antall maksimale repetisjoner (4-5 rep.), i motsetning til ved høyere hastigheter (180-300°/sek.) hvor det er vanlig med mange repetisjoner (20-30 rep.) (Dvir, 2003). For å måle quadriceps muskelstyrke etter en ACL skade er isokinetisk testing ved 60°/sek. ofte referert til som en relevant og valid måling (Dvir, 2003; Eitzen et al., 2010). En angulær hastighet på 60°/sek. blir ansett som en ”langsom” hastighet hvor muligheten for å generere maksimalt kraft er størst (Ross et al., 2002). Isokinetisk testing av barn har typisk blitt utført ved hastigheter mellom 0-240°/sek. (Jones & Stratton, 2000).

Den mest brukte variabelen er *peak torque* (PT) som har blitt anbefalt til bruk både i forskning og i klinikken. PT er den høyeste verdien av *torque* som er utviklet igjennom en gitt bevegelsesbane, målt i Newton meter (Nm). *Total work* er summen av arbeid utført i alle repetisjonene i en serie, oppgitt i måleenheten joule (J) (Kannus, 1994; Dvir, 2003). Torque ved 30° (TQ30) er foreslått som en interessant variabel på bakgrunn av funn etter ACL skade som viser størst quadriceps defisitt under 40° knefleksjon (Eitzen et al., 2010). Noen studier oppgir også en gjennomsnittsverdi av PT (Hartmann et al., 2009).

Fordelen ved bruk av isokinetisk testing er at metoden tilbyr trygge og kontrollerte testforhold og at ingen ekstern motstand blir gitt så snart bevegelsen stopper. Risikoen for muskel- og leddskader er minimal, noe som kan forekomme hyppigere ved å prøve å kontrollere en ekstern motstand ved maksimal testing (1 RM) med frivekter (De Ste Croix et al., 2003). Man får også undersøkt muskelfunksjonen isolert. Bakdelene kan være kostnader, kompleksiteten av selve testingen og at testapparatet/dynamoteteret primært er designet til å passe voksne, og ikke barn (Jones & Stratton, 2000).

Kritikken mot isokinetisk testing er at selve testbevegelsen er ”unormal” i forhold til bevegelse i daglige aktiviteter. I en isokinetisk maskin testes som regel en bevegelse isolert til ett ledd, i en åpen kjede og med konstant angulær hastighet. I daglige aktiviteter foregår bevegelse ofte over flere ledd, i en lukket kjede og innebærer

konsentrisk og eksentrisk muskelarbeid hvor den angulære hastigheten forandres i forhold til leddvinkelen. Det stilles derfor spørsmål ved om isokinetiske prestasjoner nøyaktig kan predikere/garantere funksjonell kapasitet ved praktiske oppgaver som for eksempel hopping, kasting og løping (Kannus, 1994).

2.4.2 Hinketestene – ICF aktivitet

Hinketestene er designet for å imitere funksjonelle krav man vil kunne møte i idrett og sportsaktiviteter. Hinketestene er enkle å gjennomføre uten krav til mye ekstra utstyr, kostnader eller plass utenom en vanlig klinisk setting. Det er foreslått at hinketestene reflekterer og kan gi en indikasjon på muskelstyrke og kraft, nevromuskulær kontroll og evne/tro på at man kan gjennomføre oppgaven (Shaw et al., 2004; Reid et al., 2007). Testene gir et generelt kvantitativt mål på funksjon i underekstremitetene målt i distanse og tid (Clark, 2001).

En kombinasjon av de fire ulike hinketester vi bruker i denne studien, og i de foregående studiene ved NAR, er beskrevet av Noyes et al. (1991). De aktuelle hinketestene er *single hop* (SH), *crossover hop* (CH) og *triple hop* (TH) for distanse, og *6 meter hop* (6m) på tid. Testene omfatter en variasjon av bevegelsesprinsipper som retningsendring, fart og akselerasjon/reduksjon (Reid et al., 2007). Hinketester er mye brukt som et mål på knefunksjon etter ACL skade (Hopper et al., 2002). Testene kan også brukes på friske individer for å evaluere ben symmetri og predikere styrke og kraftutvikling. Limb symmetry index (LSI) blir ofte kalkulert for å rapportere resultater fra hinketestene. LSI gir en prosentverdi av forskjeller mellom høyre og venstre ben, eventuel mellom skadet- og kontrollben ved ACL skade (Munro & Herrington, 2011). Studier har foreslått at $LSI \geq 85\%$ (Barber et al., 1990), eller $\geq 90\%$ (Daniel, Malcolm, & Stone, 1982; Munro & Herrington, 2011) antyder at det foreligger ”normal” knefunksjon.

Kritikk mot hinketestene har vært deres begrensning til å avdekke en underliggende årsak til nedsatt prestasjon, og dermed avdekke spesifikke abnormaliteter. Hinketestene er en forutbestemt og planlagt handling, noe som kan svekke sammenligningen med idrettsaktiviteter hvor kravene til bevegelse er uforutsigbare, noe som krever en automatisk respons til stadige forandringer (Shaw et al., 2004). Ved testing av barn er hinketestene trygge å bruke da de ikke er ekstreme av noe slag og inneholder elementer

av bevegelse som barn kjenner igjen og normalt utsetter seg for i lek, idrett og aktiviteter til daglig.

Kritikken mot LSI (gjelder både for hinketester og isokinetiske muskelstyrkemålinger) har vært at man tar utgangspunkt i at kontrollbenet er friskt, og at et eventuelt nedsatt aktivitetsnivå ikke har påvirket funksjonen av dette benet. Man ser også bort i fra en effekt av dominant ben (Clark, 2001). LSI kan maskere eventuelle endringer i prestasjon hvis forholdet mellom bena er stabilt, selv som den absolutt score for hvert av bena kan ha endret seg (Reid et al., 2007).

2.5 Målemetodenes egenskaper

For at vi skal kunne stole på resultatene fra en målemetode enten i forskningsmessig eller praktisk sammenheng, er vi avhengig av at målemetoden vi bruker innehar egenskaper som reliabilitet, validitet og evne til å fange opp endringer over tid (Portney & Watkins, 2000; Lysholm & Tegner, 2007).

Hvor strengt man setter grensene for hver av disse egenskapene må ses ut i fra bruksområdet (råd/beslutninger på gruppenivå, individnivå, diagnostisering m.m.), variabelen som testes (stabil/ustabil) og egenskapene til utvalget som testes (friske, syke, trente/sedate, unge, gamle m.fl.). Målemetoder tiltenkt klinisk bruk og beslutninger på individnivå krever en høyere reliabilitet enn ved forskningsmessig bruk på et større utvalg (Streiner & Norman, 2008). Signifikant, eller ikke-signifikante funn man finner ved statistiske analyser i forskningen må tolkes for å bestemme hvilken betydning de har i praksis. Kjennskap til målemetoden og måleresultatene/-enheten er derfor en forutsetning for å kunne vurdere om funnene har klinisk verdi.

Målemetodene vi ønsker å bruke må være akseptable for personene vi tester med tanke på eventuell risiko for skade ved testing, tidsbruk og annen fysisk og eventuell mental påkjenning testingen kan medføre. Metoden må i tillegg være gjennomførbar i forhold til ressurser, tid og kostnader, sett i lys av hva hensikten med testen er (Bent, 2009).

2.5.1 Reliabilitet

Reliabilitet refererer til i hvilken grad en målemetode er fri for målefeil (Domholdt, 2005; Portney & Watkins, 2000) og produserer nøyaktige og stabile målinger ved testing av den samme variabelen under identiske forhold (Clark, 2001). Realistisk sett, så vil det alltid forekomme en viss grad av målefeil. Reliabilitet kan på bakgrunn av det anses som andelen av målefeil som er akseptert for effektiv og praktisk bruk av en bestemt målemetode (Atkinson & Nevill, 1998). Ulike termer blir tatt i bruk for å beskrive begrepet 'reliabilitet', som feks reproduserbarhet, enighet, repeterbarhet og samsvar mellom målinger. Disse termene er blitt kritisert for å kun ta hensyn til et av aspektene ved reliabilitet, som er 'målefeilene', og ignorere den like viktige delen (av nevneren i brøken) som er variabiliteten mellom forsøkspersonene (Streiner & Norman, 2008).

Det er tre aspekter av reliabilitet som vanligvis er av interesse: test-retest, intra- og interrater reliabilitet (Rousson, Gasser, & Seifert, 2002). Test-retest reliabilitet brukes til å evaluere om et instrument har evnen til å måle en variabel konstant over tid. Mest typisk blir en gruppe individer utsatt for samme test/oppgave ved to separate anledninger (Portney & Watkins, 2000). Noen omtaler test-retest der hvor testere (raters) ikke er involvert i målingen, som feks ved selv-rapportering i form av spørreskjemaer (Portney & Watkins, 2000; Streiner & Norman, 2008), mens andre henviser til at test-retest kan uttrykkes også som intra- og interrater reliabilitet med varierende grad av involvering fra tester (-e) (Domholdt, 2005; Lysholm & Tegner, 2007). Test-retest intervallet må velges med omhu for å unngå tretthet, lærings- eller hukommelseeffekt. Samtidig må ikke intervallet bli for langt slik at reelle forandringer forekommer i målevariabelen. Intrarater reliabilitet refererer til stabilitet av data registrert av en tester (rater) på bakgrunn av to eller flere målinger. Interrater reliabilitet refererer til variasjon mellom to eller flere testere som måler den samme gruppen av individer (Portney & Watkins, 2000).

2.5.2 Kilder til variasjon

Begrepet 'reliabilitet' er en grunnleggende måte å reflektere mengden av feil (error), både systematisk og tilfeldig, som er i enhver måling (Streiner & Norman, 2008). Den klassiske reliabilitetsteorien tar hensyn til at hver måling, eller observert score, består av en feilkomponent (error) og en *sann score*. Den sanne scoren er en satt verdi som

eksisterer uavhengig av forholdene rundt målingen. All forskjell mellom observert score og den sanne scoren tilegnes derfor en komponent av tilfeldige målefeil (random error). Teorien antar at målefeilene ikke er differensierbare, dvs. at de kommer fra mange forskjellige kilder i en ikke-systematiske (biased) form (Portney & Watkins, 2000).

I den nyere generaliserbarhetsteorien regnes ikke testforholden som en uavhengig faktor. Teorien tar hensyn til at en sann score er avhengig av den underliggende teoretiske komponenten som bare eksisterer under visse forhold, i en viss kontekst. Det betyr at all variasjon fra en måling til en annen ikke kan tilskrives tilfeldige målefeil alene. Dersom vi kan identifisere relevante testforhold (kilder/fasetter) som påvirker måleresultatene, så har vi forutsetning for bedre å kunne forutsi en større andel av forandringer i ett sett av målinger (Portney & Watkins, 2000). Det betyr videre at det ikke er noe som kan kalles reliabiliteten av en målemetode, da reliabiliteten er sterkt knyttet opp til de ulike forholdene rundt hver testing (Streiner & Norman, 2008). Intraklasse korrelasjonskoeffisienten (ICC) tar til en viss grad hensyn til noe av dette i beregningen, da den kan ta hensyn til både systematiske og tilfeldige målefeil som finnes i ett sett målinger (Atkinson & Nevill, 1998). Videre analyser brukt i generaliserbarhetsteorien er utenfor omfanget til denne oppgaven.

Systematiske målefeil (bias)

Systematiske målefeil har tendens til å påvirker målingene i en bestemt retning, enten positivt eller negativt, som for eksempel ved læring i en test-retest situasjon, ved for kort restitusjonstid i mellom testing eller ved systematisk feil-kalibrering av et instrument (Portney & Watkins, 2000; Atkinson & Nevill, 1998). Systematiske målefeil er blitt foreslått å være et 'naturlig fenomen' og bidrar derfor ikke som en trussel mot reliabiliteten per se i en test-retest situasjon. Noen mener på bakgrunn av det at bare tilfeldige målefeil burde tas hensyn til i reliabilitetsutregningen (Weir, 2005). Vanskeligheten er å bestemme hvilke av de to (eller flere) "sanne målingene" som genuint representerer kvantiteten av det vi ønsker å måle, m.a.o representerer validiteten av testresultatene (Baumgartner, 2003; Dvir, 2003).

Tilfeldige målefeil (random errors)

Den andre komponenten av variasjon er tilfeldige målefeil (random errors) som refererer til ”støy” i målingene (Batterham & George, 2003; Streiner & Norman, 2008). Tilfeldige målefeil er utforutsigbare og kan påvirke resultatet forskjellig fra gang til gang. Det kan være et resultat av biologiske (forsøksperson eller tester) eller mekaniske variasjoner (instrumentet), eller et resultat av en ikke-standardisert testprotokoll. Komponenten av tilfeldige målefeil er vanligvis større enn systematiske feil (Atkinson & Nevill, 1998). Mange anser derfor ’within-subject’ variasjoner som den viktigste typen reliabilitetsmåling siden det sier oss noe om en målemetodes nøyaktighet. Jo mindre ’within-subject’ variasjon, jo bedre er målemetoden (Batterham & George, 2003).

Kildene til målefeil kan generelt tilskrives de tre ulike komponentene: tester (rater), instrumentet og variasjonen av fenomenet/subjektet som blir testet. Det er også viktig å sikre at settingen rundt testingen er lik for hver gang. (Batterham & George, 2003). I mange tilfeller kan det være vanskelig å fullt ut separerer de ulike komponentenes innvirkning på målefeilen (Domholdt, 2005; Portney & Watkins, 2000).

Variasjonskomponenter ved hinketester og isokinetiske muskelstyrkemålinger

For isokinetisk styrke kan mange faktorer påvirke måleresultatene. Dynamometeret kan bli kalibrert ulikt fra gang til gang, noe som kan gjenspeile testprotokollen og/eller testansvarlig. Stabilisering og posisjon av kroppen til testsubjekteter kritiske faktorer som kan påvirke både reliabilitet og validitet av målingene. Inadekvat stabilisering av trunkus og/eller proksimale segmenter kan påvirke kreftene nærliggende muskulatur gir til mer distale segmenter, og dermed resultere i en ikke-optimal prestasjon. Andre faktorer som kan påvirke test-retest resultatet er for eksempel testhastighet (°/sek.), påvirkning/korrigerende av tyngdekraften, testtid på dagen, testrekkefølge, verbal oppmuntring og personlig motivasjon (Kannus, 1994). Isokinetiske dynamometre kan også inneha naturlige test-retest variasjoner som kan variere fra 4-11% (Pincivero, Heller, & Hou, 2002). Korreksjon av tyngdekraften er spesielt viktig når man tester barn fordi mindre kraft blir registrert, noe som fører til en større feilprosent. Det er derimot ikke utarbeidet en validert og spesifikk prosedyre for barn (Jones & Stratton, 2000).

For hinketestene er det også viktig å følge en standardisert protokoll med klare scoringsprosedyrer. Variasjonen kan ligge hos testeren i forhold til om hvor strengt protokollen blir fulgt med hensyn til verbal informasjon, gyldig hopp, nøyaktighet i distanse og tid målt osv. For forsøkspersonene kan fatigue/tretthet under testingen forekomme dersom pausene mellom hinkene er for kort, eller ved mange hopp totalt på hvert ben i form av prøvehopp og selve testingen. Også her kan en læringseffekt forekomme å påvirke resultatene.

Barn vil potensielt kunne tenkes å variere mer i en testsituasjon enn voksne pga konsentrasjon, forståelse av oppgaven/utførelse av testen og motivasjon. Dersom det ikke blir gitt tilstrekkelig tid til tilvenning kan læringseffekten være med på å gi forbedret resultat fra gang til gang (Jones & Stratton, 2000). I likhet vil man forvente at de som er skadet vil variere mer i prestasjon enn de som ikke er det pga smerte eller andre begrensende faktorer i forbindelse med skaden.

Nøyaktigheten og stabiliteten til testerne avhenger blant annet av kjennskap og erfaring til målemetodene, dagsform og motivasjon. Som for forsøkspersonene som blir målt, kan også det også forekomme en læringseffekt blant testerne. Det vil derfor være viktig å inkludere en form for opplæring eller pilottesting i design av studien med tanke på dette.

2.5.3 Kvantifisering av reliabilitet

Reliabilitet omfatter både relativ og absolutt reliabilitet. Relativ reliabilitet undersøker forholdet mellom to eller flere sett av målinger, mens absolutt reliabilitet undersøker variasjonen av score fra måling til måling (Domholdt, 2005).

Relativ reliabilitet

Denne typen reliabilitet måles gjerne med en korrelasjonskoeffisient. ICC er en univariat reliabilitetskoeffisient som kan brukes ved to eller flere sett av målinger av samme variabel (Atkinson & Nevill, 1998). ICC reflekterer graden av relasjon og enighet blant målingene, og gir således informasjon om en målemetodes evne til å skille mellom individer (Stratford & Goldsmith, 1997; Portney & Watkins, 2000).

Den relative naturen til ICC reflekteres i det faktum at størrelsen av ICC er avhengig av variasjonen mellom individene (between-subject). Hvis forsøkspersonene varierer lite i forhold til hverandre vil størrelsen på ICC bli liten selv om test-retest variasjonene er små. Likeså, hvis forsøkspersonene varierer mye i forhold til hverandre kan ICC bli stor selv om test-retest variasjonene er store (Weir, 2005). Det finnes flere ulike modeller av ICC (Shrout & Fleiss, 1979; McGraw & Wong, 1996). Det er viktig å se på hensikt, design av studie og type målinger man skal gjøre før man bestemmer hvilken ICC man skal bruke.

Analyser for å beregne ICC baserer seg på en variansanalyse av målingene. En *two-way ANOVA* gir oss mulighet til å undersøke to eller flere uavhengige faktorer for en avhengig faktor på samme tid. I en reliabilitetsstudie vil forsøkspersonene være en av de uavhengige faktorene. Ved et interraterstudiet er testerne (raters) den andre faktoren, og ved intrarater er testrundene (trials) den andre faktoren. Graden av sammenheng mellom målinger kan uttrykkes som *consistency* eller *agreement*. Ved *consistency* er systematiske målefeil mellom testere/målinger ekskludert, mens for *agreement* er disse variasjonene tatt med i bergningen. ICC estimatet kan oppgis som *single measure* reliabilitet, basert på enkeltmålinger, eller som *average measures* reliabilitet som er gjennomsnittet av k målinger (McGraw & Wong, 1996). ICC kan variere fra 0,00 til 1,00, hvor 0 indikerer ingen reliabilitet og 1,00 foreslår perfekt reliabilitet (Denegar & Ball, 1993).

ICC sier ingenting om presisjonen av målingene uttrykt i en gitt målenhet (Denegar & Ball, 1993). På bakgrunn av at ICC også i stor grad påvirkes av heterogeniteten til utvalget, er det ikke anbefalt å bruke ICC alene som eneste mål på reliabilitet (Domholdt, 2005). Man bør i tillegg bruke et mål på absolutt reliabilitet som kan ha en større praktisk nytteverdi.

Absolutt reliabilitet

Metoder brukt for å måle absolutt reliabilitet er *standard error of measurement* (SEM/SEM %), variasjonskoeffisienten (CV) og *limits of agreement* (LOA). SEM kvantifiserer målefeil i samme enhet som den original måleenheten til målemetoden som testes (Stratford & Goldsmith, 1997). Jo mindre SEM er, jo bedre er reliabiliteten til målingene (Atkinson & Nevill, 1998). Små variasjoner innen hvert individ (within-

subject) legger til rette for å oppdage små, men likevel potensielt klinisk meningsfulle endringer i variabelen vi måler (Batterham & George, 2003).

CV og SEM % er målevvariabilitet uttrykt som en variasjonskoeffisient oppgitt i prosent. Fordelen ved bruk av en CV og SEM % er at reliabiliteten for ulike målemetoder kan sammenlignes (Lexell & Downham, 2005). I motsetning til ICC er størrelsen på SEM/SEM %, CV og LOA i mindre grad påvirket av heterogeniteten til utvalget (Batterham & George, 2003). Begrensninger for SEM/SEM % og CV er at de bare representerer 68 % sannsynlighet for at den "sanne score" finnes innen ± 1 SEM av en observert score. Bland og Altman introduserte metoden LOA som gir 95 % sikkerhet for at den sanne scoren finnes innen $\pm 1,96$ SEM. Metoden baserer seg på å analysere forskjellene mellom parvise score fra test og retest for hvert individ (test 1-test 2). Disse forskjellene blir plottet mot de respektive individuelle gjennomsnittene i et Bland-Altman plot. En grov indikasjon på systematisk og tilfeldige målefeil blir gitt ved å undersøke retningen og størrelsen av spredningen rundt null-linjen. Standardavviket (SD) til differansen mellom målingene brukes for å danne øvre og nedre grense (LOA) for hvor de to målingene/testene er enige (Atkinson & Nevill, 1998; Dvir, 2003).

2.5.4 Standard for størrelsen av reliabilitetskoeffisienten

Hva som defineres som et "akseptabelt" nivå av reliabilitet må sees i sammenheng med faktorer som valg av ICC modell, nøyaktigheten til måleinstrumentet, fenomenet/subjektene vi måler og hva vi skal bruke resultatene til. Ulike ICC 'cut-off points' for hva som regnes som høy og lav reliabilitet kan dermed variere. Reliabiliteten til målemetoder som skal brukes til diagnostikk og råd for videre behandling på individnivå krever som regel høyere reliabilitet enn for tester som skal brukes til et beskrivende formål alene (Streiner & Norman, 2008). En ICC større eller lik 0,90 er foreslått som indikasjon på "høy" målereliabilitet og er foretrukket for de fleste kliniske forsøk. En generell retningslinje, selv om slike grenser må sees på som skjønnsmessige, er å betrakte ICC verdier under 0,50 som dårlig, verdier fra 0,50-0,75 som moderate og ICC verdier over 0,75 som god/høy reliabilitet (Portney & Watkins, 2000). Fysiske prestasjonsbaserte tester med en veldefinert scoringsprosedyre tenderer til å gi målinger med reliabilitet som ligger i øvre sjiktet av 0,70 til den lavere enden av 0,90, for friske, unge og voksne i alderen 10-30 år (Baumgartner, 2003). Om en målemetode anses som reliabel kommer derfor an på om målefeilen til det aktuelle instrumentet blir vurdert

som akseptabelt ut i fra praktisk eller forskningsmessig bruk på en aktuell populasjon under gitte testforhold.

2.5.5 Validitet & responsivness

Validitet

Forholdet mellom reliabilitet og validitet er unidireksjonelt. Det betyr at reliabilitet er en forutsetning, men ingen garanti for validitet (Portney & Watkins, 2000). Dersom et måleinstrument ikke produserer stabile målinger under like forhold så kan det heller ikke anses å være valid (Batterham & George, 2003). Validitet beskriver en målemetodes evne til å måle det den har til hensikt å måle (Portney & Watkins, 2000; Streiner & Norman, 2008). Validitet begrenser seg til *responsivness* i denne oppgaven.

Responsivness

Dersom vi ønsker å bruke en målemetode til å evaluere, er vi avhengig av at den har evnen til å fange opp reelle endringer over tid (*responsivness*). Dette er en viktig egenskap når man vil måle effekten av en intervensjon, som for eksempel endring i status fra pre- til post-test. Samtidig må metoden være stabil når en endring ikke har skjedd (Portney & Watkins, 2000). Det er flere statistiske termer som blir brukt for å definere en reelle/faktisk endring: *minimal detectable change* (MDC), *smallest detectable change* (SDC) *smallest real change* (SRC) (de Vet et al., 2006) og *smallest real difference* (SRD) (Lexell & Downham, 2005). Disse representerer den minste verdi vi kan kalle en endring utover den forventede målefeilen til et instrument (de Vet et al., 2006). Jeg vil videre bruke termen minste reelle endring om SRD.

Responsivness er relatert til både reliabilitet og validitet til en målemetode. På et begrepsmessig nivå er responsivness et aspekt av validitet, nærmest beslektet til *kriterievaliditet*, som går ut på om endringene vi registrerer korrelerer med endringer målt med en annen målemetode (Streiner & Norman, 2008). I beregning av SRD er SEM den mest vanlige statistikken å bruke, som relaterer responsivness til reliabiliteten til en målemetode. En '*distribution-based approach*' er basert på statistiske karakteristikk til utvalget som testes, hvor SEM kommer inn som et mål på spredningen (de Vet et al., 2006). En liten SEM gjør at størrelsen på den kliniske

relevante endringer blir mindre. Det betyr at instrumentet i større grad kan fange opp endring i prestasjon/funksjon (Portney & Watkins, 2000).

En alternativ tilnærming til SRD er LOA. Ved å bruke 95 % sikkerhet i beregning av SRD vil det tilsvare øvre og nedre LOA (Bent, 2009; Lexell & Downham, 2005). Vi kan tenke oss endring over tid som et kontinuum, hvor SEM representerer målefeilen til instrumentet og SRD den *minste* verdi vi kan kalle en endring utover den forventede målefeilen til instrumentet. For å bestemme hvor stor endringen må være for at den kan kalles ”viktig” ligger i begrepet ’minimal clinically important change’ (MCID) (Portney & Watkins, 2000). På kontinuumet burde en valid MCID være minst like stor, eller større enn SRD (Copay, Subach, Glassman, Polly, Jr., & Schuler, 2007). Jeg har i denne oppgaven brukt SRD, som et første mål på responsivness, og kommer derfor ikke til å nevne MCID videre.

Reliabiliteten vi finner i denne metodestudien kan således være med på å gi et estimat på hvor stor endring vi trenger å oppdage hos kohortbarna for at vi skal kunne kalle det en reell endring i funksjon og ikke bare et resultat av målefeil. Endringer mindre enn SRD vil mest sannsynlig ikke kunne kalles en klinisk reell endring.

2.6 Studier på reliabilitet

Jeg har sett på reliabilitetsstudier for hinketestene og isokinetisk muskelstyrkemålinger som omhandler friske og personer med ACL rekonstruksjon. Studiene på ACL populasjonen er bare utført på voksne. Det finnes heller ingen reliabilitetsstudier for hinketestene på barn. Jeg har primært valgt studier som har testet isokinetiske muskelstyrkemålinger ved samme hastighet (60°/sek.) som oss. På barn finnes det færre reliabilitetsstudier, så de utvalgte studiene er tatt med for å belyse utviklingen. For hinketestene har jeg sett på studier som har brukt hinketestene (Noyes et al., 1991) som vi testet. Studiene er presentert i tabell 2-4.

2.6.1 Reliabilitet av isokinestiske muskelstyrkemålinger

Instrumentet

Ulike studier har demonstrert akseptabel intra- og intersession reliabilitet for ulike isokinestiske maskiner som Biodex, Cybex, Kin Com og Merac (Maffiuletti et al., 2007). Instrumentell reliabilitet av Biodex System 3 isokinestisk dynamometer har blitt undersøkt for ulike vinkelposisjoner, isometrisk torque og konsentrisk hastighet. Det ble funnet høy reliabilitet, både innen samme dag og mellom to testdager (ICC 0,99-1,00) for alle variablene som ble testet (Drouin et al., 2004). Ulike begrensninger og kilder til variasjon kan derimot gjøre det vanskelig å sammenligne på kryss av ulike maskinmerker og modeller (Jones & Stratton, 2000).

Reliabilitetsstudier på barn

Forskning på reliabilitet av isokinestiske styrkemålinger på barn og unge (6-16 år) er begrenset. Studiene er vanskelige å sammenligne på bakgrunn av ulike design og metode. Ledd- og bevegelsesutslag (ROM), hastighet (°/sek.) og metode for statistiske analyser varierer mellom studiene (Jones & Stratton, 2000). Dette gjelder også for studiene jeg så på som er presentert i tabell 2.

To studier har oppgitt ICC som et mål på relativ reliabilitet som viser moderat til høy intrarater reliabilitet (ICC 0,63-0,98) (Deighan et al., 2003; Kellis et al., 1999). Disse oppga i tillegg LOA. Ingen har oppgitt SEM verdier som et mål på absolutt reliabilitet, men i tre av studier blir en CV rapportert (Deighan et al., 2003; Merlini et al., 1995; Mohtadi et al., 1990). Utvalgssørrelsen var lav i alle studiene (n = 5-13) og bare gutter er blitt testet, med unntak av i studien til Molnar et al. (1979).

To studier testet bare høyre ben (Deighan et al., 2003; Merlini et al., 1995). Alle studiene har testet styrkevariabelen PT. Tre av studiene testet både konsentriske og eksentriske målinger (Deighan et al., 2003; Kellis et al., 1999; Mohtadi et al., 1990). Ett prøverunde (tilvenning) ble tilbudt barna i to av studiene (Deighan et al., 2003; Kellis et al., 1999) for å bli kjent med utstyr og testprosedyrer.

Kellis et al. (1999) testet aktive gutter som trente fotball ($7,9 \pm 1,8$ timer per uke). I studien til Deighan et al. (2003) deltok ingen i regelmessig trening. Merlini et al. (1995) testet barn som var fysisk aktive, men hvor ingen drev med konkurranseidrett. De testet den yngste gruppen barn (6-8 år). I to studier (Merlini et al., 1995; Mohtadi et al., 1990) er det ikke gitt noen aktivitetsbeskrivelse av utvalget. Det finnes noen nyere reliabilitetsstudier på barn (Ayalon, Ben-Sira, Hutzler, & Gilad, 2000; Pierce, Lauer, Shewokis, Rubertone, & Orlin, 2006), men disse er gjennomført på barn med cerebral pareses og derfor ikke sammenlignbare med vårt utvalg.

Reliabilitetsstudier på voksne

Studiene på voksne viser til høy relativ reliabilitet (ICC 0,80 -0,99). Studiene er presentert i tabell 3. To studier undersøkte interrater reliabiliteten (Keskula et al., 1995; Hartmann et al., 2009). Begge studiene viser høy reliabilitet, ICC > 0,90. De fleste studiene har oppgitt SEM/CV og to studier har i tillegg oppgitt et mål på minste reelle endring (Hartmann et al., 2009; Sole et al., 2007).

Hovedandelen av studiene på friske testet bare ett ben (Hartmann et al., 2009; Keskula et al., 1995; Lund et al., 2005; Maffiuletti et al., 2007; Sole et al., 2007). Alle testet PT. I tillegg ble Tw, AVG power/work og AVG PT benyttet i ulike studier. Sole et al. (2007) testet eliteutøvere og aktive/mosjonister som alle drev med løpsrelaterte idretter. Maffiuletti et al. (2007) inkluderte mosjonister. En av studiene stilte ikke noe krav til fysisk aktivitet (Hartmann et al., 2009), og en studie oppga ikke noe aktivitetsbeskrivelse av ikke deltagerne (Lund et al., 2005). Hartmann et al. (2009) var eneste studie som testet et utvalg av eldre ($71,2 \pm 5,5$ år).

Lund et al. (2005) undersøkte reliabiliteten og effekten av læring. De fant høyest reliabilitet for PT (ICC 0,97- 0,98) ved 20 min. intervall mellom målingene. Ingen signifikant læringseffekt ble påvist på tvers av målingene. To studier har sett på reliabilitet etter ACL rekonstruksjon (Brosky, Jr., Nitz, Malone, Caborn, & Rayens, 1999; Ross et al., 2002). Høye ICC verdier ble rapportert også her (ICC 0,81- 0,97). Deltagerne i studien til Ross et al. (2002) var fysisk aktive kadetter som var tilbake i militærtjeneste. I studien til Brosky et al. (1999) bestod utvalget av mosjonister (trening 2- 3 ganger per uke).

Reliabilitet av hinketestene

Hinketestene har vist seg å være reliable på voksne friske (Ross et al., 2002; Munro & Herrington, 2011; Bandy et al., 1994; Bolgla & Keskula, 1997) og etter ACL rekonstruksjon (Reid et al., 2007). Studiene jeg har trukket frem er presentert i tabell 4. Basert på disse studiene viser hinketestene høy relativ reliabilitet med ICC 0,80- 0,97 for distansehopp, med noe lavere verdier for 6m. ICC 0,60- 0,92. Alle studiene oppgir i tillegg absolutt reliabilitet, med unntak av Bany et al. (1994). To studier har også registrert minste reelle endring (Munro & Herrington, 2011; Reid et al., 2007).

Ulike intervall ble gitt mellom testrundene i studiene som strakk seg fra 24-48 timer (Bolgla & Keskula, 1997; Reid et al., 2007) til 4 uker (Ross et al., 2002). I to av studiene bestod utvalget kun av menn (Bandy et al., 1994; Ross et al., 2002), mens Bolgla & Keskula (1997) testet flest kvinner. To studier (Munro & Herrington, 2011) aktivtetsbeskrev utvalget sitt som bestod henholdsvis mosjonister og meget fysisk aktive kadetter. Reid et al. (2007) testet voksne 16 uker etter ACL rekonstruksjon. Alle de inkluderte fulgte samme opptreningsprogram ved samme klinikk.

Munro et al. (2011) kartla læringseffekten assosiert med de fire hinketestene, og undersøkte deretter reliabiliteten ut fra de ”stabile” hoppene. De fant ut at prestasjonene for SH og TH stabiliserte seg etter tre forsøk, mens CH trengte 4 forsøk. For 6m trengte menn tre hopp, og kvinner fire hopp før prestasjonene stabiliserte seg. Ingen av studiene har undersøkt interrater reliabilitet. Vi fant ingen studier som har sett på reliabiliteten av hinkester blant barn, hverken med eller uten ACL skade.

Tabell 2: Oversikt over reliabilitetsstudier for isokinetiske muskelstyrkemålinger på barn uten skade.

Studie	Type test	Utvalg	Reliabilitet
Molnar et al. 1979	Cybox II, Lumex Inc. -Variabel: torque generated on max effort -Albue/kne: eks/fleks -Hastighet: ikke oppgitt	Gutter & jenter 7-15 år n= 100 friske + 60 lett psykisk utviklingshemmet (intra-session) n= 80 (inter-session)	Intra-session: ikke-signif. forskj.(F-ratio): 0-8% Inter-session: ikke-signif. forskj.(Kji-kvadrat): 8,4-8,9% Inter-tester: ikke-signif. forskj. (t-test): 4-12%
Mohtadi et al. 1990	Kin Com, Chattex Corp. -Variabel: PT -Kons/eks. kneekstensjon -60°/sek.	Gutter 10-12 år n= 12 n= 5 (reliabilitet)	Intrarater: t-test: ingen signif. forskj. ved test-retest CV 4,0 – 9,9 %
Merlini et al. 1995	Lido Active, Loredan Biomedical -Variabel: PT -Kne ekst/fleks -100°/sek., 0-100° fleksjon -3 test session: 3 dagers mellomrom -Oppv.: 6-8 rep., 2 set a 5 rep. testing	Gutter 6-8 år n= 12	Resultater <i>dominant</i> ben: Intra-session: Pearson's correlation $r = 0,78-0,99$ Inter-seesion: ICC (1,1) 0,85 (fleks) - 0,95 (ekst) CV 8 % (ekst), CV 12 % (fleks)
Kellis et al. 1999	Cybox Norm dynamometer -Variabel: PT -Kons/eks. kneeks/fleks -60-120-180°/sek., 0-90° knefleks -1 prøvesession, 2 tests, 1 ukes intervall -Oppv.: 3 submaks., 5 maks rep. testing	Gutter 13 år n= 13	ICC (1,1): (60°/sek) 0,90-0,98(kons), 0,85-0,92 (eks) ICC (120°/sek) 0,88-0,96(kons), 0,71-0,88 (eks) ICC (180°/sek) 0,89 (kons), 0,76-0,80(eks) LOA (60°/sek): -4,5 til 7,2 (E) og -4,6 til 1,5 (F)
Deighan et al. (2003)	Biodex System-3 (Biodex Corp., Shirley, NY, USA) -Variabel: PT -Kons/ekst kneekst/fleks -30 og 180/sek., ROM 90° (kons) - 75° (eks) -1 prøvesession, 2 tests, 1 ukes intrvall	Gutter 9-10 år n= 20 (kne + albue) n= 10 (kne)	Resultater høyre ben: ICC (30/sek): 0,76-0,83 (kons), 0,63-0,84 (eks) ICC (180/sek): 0,74-0,90 (kons), 0,72-0,80 (eks) CV: 4 – 11 %, LOA (30°/sek): -8 til 11/(17) (E)

PT= peak torque, ICC= intraclass correlation coefficient, CV= variasjonskoeffisient. LOA = limits of agreement, E = ekstensjon, F = fleksjon.

Tabell 3: Oversikt over reliabilitetsstudier for isokinetiske muskelstyrkemålinger på voksne friske og med ACL skade.

Studie	Type test	Utvalg	Reliabilitet
Keskula et al. 1995	Cybox 6000, Division of Lumex -Variabler: PT & Tw -Kons. knefleks/ekst. -90-0° fleksjon, 60°/sek.- 4 max.rep., 180°/sek. - 20 rep. -Oppv.: 4 rep. -4 test sessions, 3-5 dagers intervall, 3 ukers periode	Voksne, friske n= 16 (8 m + 8 k)	<i>Interrater:</i> ICC (2,1) (PT): 0,90-0,96 SEM 8,9 – 13,3 Nm ICC (Tw): 0,90 – 0,95 SEM 11,3 – 16,8 Nm
Brosky et al. 1999	Biodex Isokinetic Dynamometer, Biodex Corp. -Variabe: PT -Kons. knefleks/ekst. -Oppv.: 3 submax rep. -60°/sek.-5 rep, 360°/sek.-15 rep. -4 test sessions, 2 dg, 7 dg intervall	ACLR n= 15 menn	ICC (PT) intra-session: 0,82-0,97 ICC (PT) inter-sessions: 0,81-0,93
Ross et al. 2002	Kinetic Communication, Chattanooga Group -Variabler: PT index -Kons. kneekstensjon -90-0° fleksjon, 60°/sek. -Oppv.: 3 submax rep., test 5 max rep.	ACLR n= 50 (36 m + 14 k) Test-retest: n= 10 (7 m + 3 k)	ICC (2,1): 0,95 SEM 3,76 Nm
Lund et al. 2005	Biodex System 3 PRO (Biodex Medical System, NY) -Variabel: PT -Kne/albue: fleksjon/ekstensjon -0-90° fleksjon, 60°/s – 3 max rep. -test – retest: 2 målinger dag 1 med 20 min. pause, 1 ukes intervall -1 måling dag 2	Voksne friske n= 13 (4 m + 9 k)	ICC (2,1) (Biodex): 0,92-0,94 SEM: 7,0 - 10,6 Nm LOA (dag 1): -16,3-17 Nm
Maffiuletti et al. 2007	Con-Trex MJ; CMV, AG, Dübendorf -Variabel:PT, AVG work/power -Kons/eksentr. kneekstensjon/fleksjon -60-120- 180°/sek., ROM 70° -2 test sessions, 7 dg intervall -Oppv.: 20 submaks, test 3 maks rep.	Voksne friske n= 30 (15+15)	ICC (2,1): PT (60°/sek.) 0,96 -0,99 CV(PT): 2,8 – 3,6 %
Sole et al. 2007	Kinetic Communicator (KinCom) 500H, Chattecx Corp. -Variabler: PT & Tw -Kons./eksentr. kneekst/fleks -0-85° fleksjon -Oppv: 10 submax + 2 max rep. -60°/sek. -3max rep. -2 test sessions, 7 dager intervall	Voksne friske n= 18 (11 m + 7 k)	ICC (3,1) (PT): 0,92 - 0,95 SEM 5,4 -8,2 % SRD % 15,1 -22,7 % ICC (3,1) (Tw): 0,88 - 0,96 SEM % 4,5 – 9,2 % SRD % 12,1 – 25, 4%
Hartmann et al. 2009	Biodex System 3 (Biodex Medical Systems, Shirley, NY, USA) -Variabler: PT, AVG PT/power -Kne- og ankel ekstensjon/fleksjon -60-120°/sek. -3 test sessions, 2 på dag 1, 7 dagers mellomrom, 2 testere -Oppv.: 6-8 submaks., test 4 maks rep.	Eldre voksne n= 24 (6m+18k)	<i>Interrater:</i> ICC (2,1) PT/AVG PT (60°/s): 0,92 -0,96 SEM % 6,3 -9,0 SRD ₉₅ : 10,1 -17,1 ICC (3,1) PT/AVG PT (60°/s): 0,85-0,93 SEM %: 8,0- 12,8 SRD ₉₅ : 13,8 - 19,7

PT= peak torque, Tw= total work, AP = average power, AVG PT = average PT, ACLR = ACL rekonstruksjon, ICC= intraclass correlation coefficient, CV= variasjonskoeffisient, SEM = standard error of measurement, SRD = minste reelle endring.

Tabell 4: Oversikt over reliabilitetsstudier for hinketestene på voksne med og uten ACL skade.

Studie	Type test	Utvalg	Reliabilitet
Bolgia & Keskula 1997	Alle fire hinketester (Noyes et al. 1991)	Ikke-skadet n= 20, (5 menn+15 kvinner)	ICC (2,3): 0,66 (6m) – 0,95 (TH) – 0,96 (SH/CH) SEM 4,56 cm (SH) – 15,44 cm (TH) – 15,95 cm (CH) – 0,13 sek. (6m)
Reid et al. 2007	Alle fire hinketester (Noyes et al. 1991)	ACL n= 42 (23 m+19 k) n= 35 (reliabilitet)	ICC (2,1): 0,82 (6m) – 0,84 (CH) – 0,88 (TH) – 0,92 (SH) – 0,93 (samlet) SEM fra 3,04 % (samlet) - 5,59 % (6m) MDC ₉₀ fra 7,05 % (samlet) - 12,96 % (6m)
Ross et al. 2007	Alle fire hinketester (Noyes et al. 1991)	Ikke-skadet n= 18 menn	ICC (2,3): 0,92 (SH/6m) – 0,93 (CH) – 0,97 (TH) SEM 4,61 cm (SH) – 11,17 cm (TH) - 17,74 cm (CH) – 0,06 sek. (6m)
Munro et al. 2011	Alle fire hinketester (Noyes et al. 1991)	Ikke-skadet n= 22 (11+11)	ICC (3,1) kvinner: 0,80 (SH/TH) – 0,84 (6m) – 0,87 (CH). SEM 7,9 – 23,2 cm & 0,08 sek. SRD ₉₅ : 22 cm(SH), 64,3 cm (TH), 54,7 cm (CH), 0,21 (6m) ICC (3,1) menn: 0,60 (6m) – 0,80 (SH) – 0,86 (CH) – 0,92 (TH). SEM 7,9 – 21,2 cm & 0,08 sek. SRD ₉₅ : 21,8 cm (SH), 47,6 (TH), 58,7 cm (CH), 0,23 (6m)
Bandy et al. 1994	Alle fire hinketester (Noyes et al. 1991) & vertical hop	Ikke-skadet, menn n= 18, SH n= 18, 6m n=16, vertical n=22, TH & CH	ICC (1,1): 0,85 (vertical) - 0,90 (CH) – 0,92 (6m) - 0,93 (SH) – 0,94 (TH) <i>Dominant ben testet over 3 trials.</i>

ACL= korsbånds rekonstruerte, ICC= intraclass correlation coefficient, SEM= standar error of measurement, LSI= limb symmetry index, KI= konfidensintervall, SRD = minste reelle endring, SH= single hop, TH= triple hop, CH= crossover hop, 6m= 6 meter hop.

2.7 **Testing av barn**

En utfordring ved testing av barn er den store variasjonen som kan forekomme blant annet av modenhet og motorisk utvikling. Dette kan føre til store variasjoner innad i hver aldersgruppe (Holm, Fredriksen, Fosdahl, & Vollestad, 2008) i tillegg til store individuelle variasjoner. I en reliabilitetsstudie som har som mål å registrere stabile målinger kan dette være en utfordring som er større ved å teste barn enn voksne.

En kompleks interaksjon mellom modning og vekst ved antropometriske, nevrologiske, hormonelle og kjønnsassosierte forandringer er viktige i endringen av muskelstyrke fra barn til voksen alder (De Ste Croix, Armstrong, Welsman, & Sharpe, 2002). Selv om det er antatt at barn har tilnærmet lik muskelkomposisjon som voksne, så er evnen til å produsere spenning proporsjonalt til tverrsnittsarealet lavere hos barn enn hos voksne. Rundt 80 % aktivering av muskelspoleenhetene er vist ved viljestyrt muskelspenning (på gutter). Ved styrketrening hos barn er det vist at regelmessig stimuli øker evnen til å generere muskelspenning via bedring av nevralt adaptasjonen, heller enn at muskeltverrsnittet øker (Jones & Stratton, 2000). Vi kan ut i fra dette anta at barn vil trenge lengre tid til tilvenning for å prestere stabilt ved en maksimal styrketest.

Noen studier (tverrsnitt-, longitudinelle- og normative studier) rapporterer forskjeller i isokinetisk muskelstyrke mellom gutter og jenter fra alder 14 år (Barber-Westin, Noyes, & Galloway, 2006; De Ste Croix et al., 2002; De Ste Croix, Armstrong, & Welsman, 1999), mens andre henviser til forskjeller fra 11-12 års alderen (Holm et al., 2008; Wiggin, Wilkinson, Habetz, Chorley, & Watson, 2006). Normative data for SH (gutter/jenter 14,5 år) henviser til at distansen som blir hoppet er relatert til muskelstyrke i underekstremitetene som påvirkes av alder og aktivitetsnivå (De Carlo & Sell, 1997). Det er rapportert store variasjoner innad i hver aldersgruppe for påfølgende hink på ett ben (Holm, Tvetter, Fredriksen, & Vollestad, 2009). Totalvariasjonen fra 7-12 år var henholdsvis 64 %, og 36 % for absolutt og relativ hopplengde.

3. Materiale og metode

3.1 *Bakgrunn for valg av deltagere*

Bakgrunn for metodestudien og valg av deltakere var den pågående prospektive kohortstudien av barn med ACL skade som startet opp i 2006. Kohortstudien bestod per mars 2011 av 49 barn, hvorav åtte barn var operert. Av de ikke-opererte barna (n= 41) var 26 (63 %) gutter og 15 (37 %) jenter. Gjennomsnittsalder ved skade var $11,1 \pm 1,5$ år (gj.sn. \pm SD) (min.-maks.: 7,0-13,0 år). Gjennomsnittsalder ved baseline var $12,0 \pm 1,3$ år (9,0-14,5 år). De to største aktivitetene ved skade var alpint/twin tip (51 %) og fotball (17 %). De fleste av barna i kohortstudien kommer i fra Oslo- og Østlandsområdet.

De ikke-opererte ACL-barna med organiserte aktivitetsdata (n =25) rapporterte ved baseline størst deltagelse i gym på skolen (72 %), fotball (56 %), sykling (48 %) og balansetrening (44 %). De fleste (48 %) oppgav en samlet frekvens på over 2-3 ganger per uke for disse aktivitetene, mens resten av kohortbarna rapporterte frekvens på 0-1 gang (24 %), 4-5 ganger (16 %) og ≥ 6 ganger per uke (12 %).

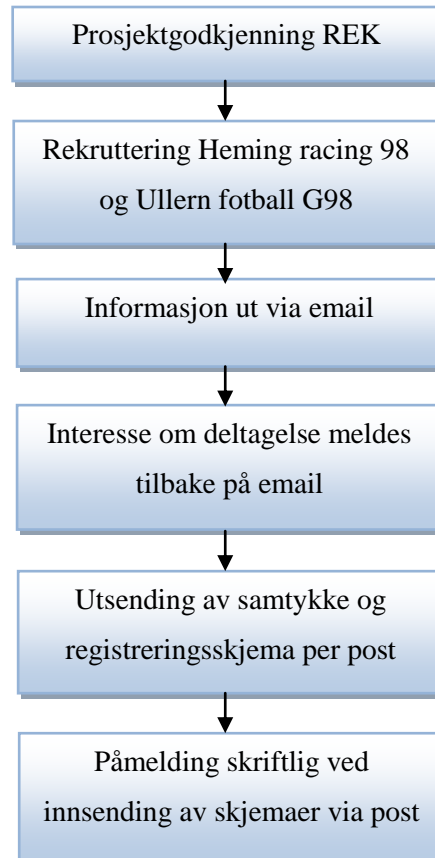
3.2 *Utvalg*

Vi rekrutterte barna til metodestudien fra to idrettslag i Oslo. Deltagerne ble med i studien på bakgrunn av frivillig påmelding. Det var 29 barn som ble inkludert, heretter kalt *metodebarna*. Utvalget bestod av 14 barn fra Ullern fotball akademi og 15 barn fra Heming alpint racing. Alle barna var født i 1998. Det var 23 (79 %) gutter og 6 (21 %) jenter. Barna var alle bosatt i nær avstand til testlokalene på Hjelp24 NIMI på Ullevål stadion.

3.2.1 *Rekrutteringsstrategier*

Vi tok kontakt med Ullern fotballakademi (gutter 98) og Heming racing 98 (gutter og jenter) for å rekruttere barn til metodestudien. Gjennom trener/oppmann for den aktuelle aldersgruppen sendte vi epost til foreldre og foresatte med den første informasjonen om studien. De som var interessert ble bedt om å melde seg på via epost til sin trener/oppmann. Vi mottok en liste med kontaktinformasjon over de interesserte barna og sendte ut et informasjonsbrev og samtykkeerklæring (vedlegg 1) sammen med et

registreringsskjema for aktivitet og skade (vedlegg 2) per brev til barn og foreldre. De som ønsket å delta returnerte samtykkeerklæring og registreringsskjema per post i en ferdig frankert konvolutt til vår forskningskoordinator på NAR (figur 2). Deltagerne fikk skriftlig informasjon om at deltagelsen var frivillig, og at de når som helst hadde rett til å trekke seg fra studiet. Det ble krevd skriftlig samtykke av både barn og foreldre/foresatte før deltagerne ble inkludert i studien.



Figur 1: Flytskjema over rekrutteringsprosedyrer.

Inklusjonskriterier

- Født i 1998.
- Deltagelse i organisert trening innen fotball eller alpint.
- Mulighet for å delta på begge testdager.
- Bosatt i Oslo-området med nærhet til testlokaler.

Eksklusjonskriterier

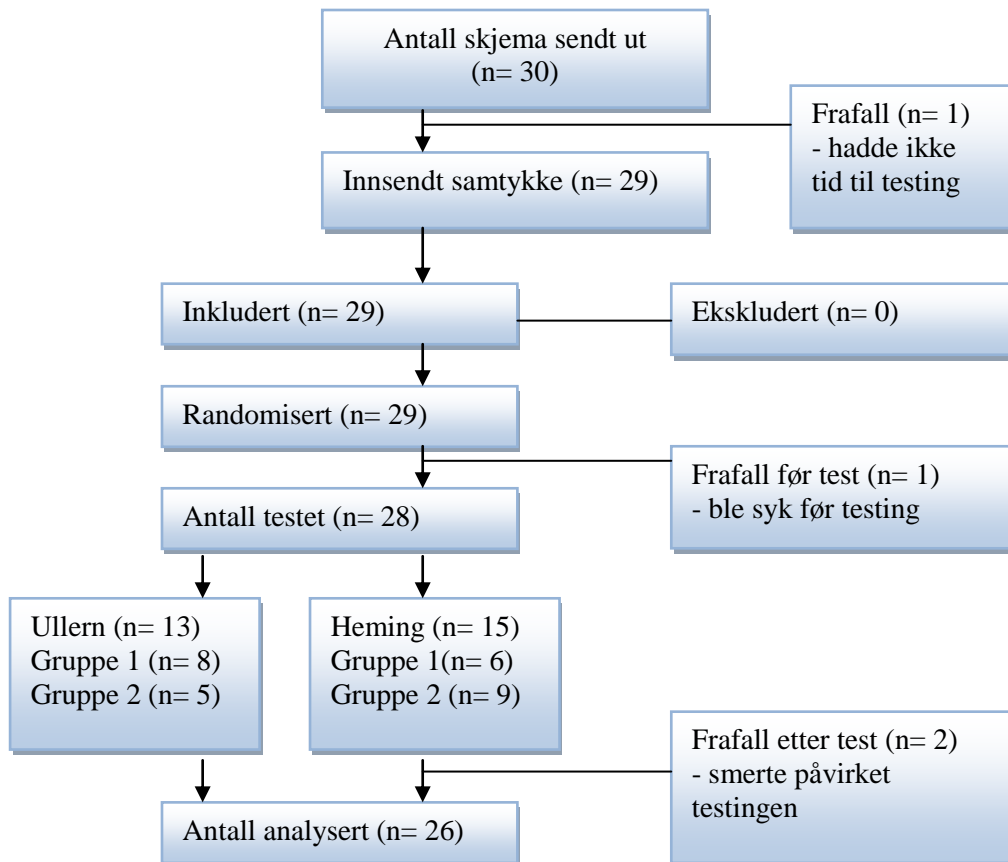
- Kneskade eller annen skade de siste 6 månedene som hindret deltagelse i aktivitet/trening, gav fravær fra trening og/eller ved behov for behandling av helsepersonell.
- Hvis skaden hindret barnet i å delta på trening eller aktiviteter nå.
- Problemer med å hinke.
- Blitt testet tidligere i hinketestene eller isokinetisk muskelstyrke.
- Smerte under testing.

Dersom det oppstod spørsmål om egnethet for deltagelse i prosjektet på bakgrunn av svarene som ble gitt på registreringskjema, ble forelder/foresatt kontaktet via telefon for nærmere undersøkelse av eventuell skadestatus før inklusjon. Vi undersøkte hvilke type skade/plage barnet hadde, når og hvordan plagene oppstod, hvor akutte plagene var nå, eventuell behandling for skaden/plagen, hvordan det påvirket deltagelse i aktiviteter (f.eks. gym på skolen) og om de trente for fullt nå. Hvis det fortsatt var usikkerhet rundt egnethet fikk barnet lov til å komme for å delta, men ble fulgt nøye opp under testing, og trukket ut dersom skaden/plagen ga begrensninger og/eller smerter som påvirket prestasjonen.

3.3 Testere

To fysioterapeuter (HM og IE) var ansvarlige for testingen av barna i metodestudien. Begge har lang erfaring i bruk av de aktuelle testene. En av fysioterapeutene hadde testet siden 2007, mens den andre hadde testet hinketester siden 2003 og isokinetiske muskelstyrketester siden 2006. De hadde jobbet som fysioterapeuter fra henholdsvis 1996 og 1999. Begge har i tillegg til praktisk erfaring, stor teoretisk fagkunnskap innenfor feltet av ACL skader. Den ene testeren disputerte og avsluttet sin doktorgrad (PhD) på voksne med ACL skade i august 2010, mens den andre testeren holder på med sin PhD på barn med ACL skade, som denne metodestudien er en del av.

3.4 Prosedyrer for datainnsamlingen



Figur 2: Flytskjema som viser inklusjon/eksklusjon og frafall ved testing av barn i denne metodestudien.

3.5 Studiedesign

Denne studien er en metodestudie der hinketester og isokinetiske muskelstyrke har blitt reliabilitetstestet på et selektert utvalg av barn uten skade med gitte inklusjons- og eksklusjonskriterier. Studiedesignet for å undersøke intra- og interrater reliabiliteten for de fire hinketestene (SH, CH, TH og 6m) og isokinetisk muskelstyrke (konsentrisk målinger for knefleksjon og ekstensjon ved 60°/sek.) innebar at en gruppe barn ble testet ved tre anledninger i hver av hinke- og styrketestene, fordelt på to ulike testdager med en ukes mellomrom. Barna var blindet for både hinke- og styrketestresultatene ved begge testanledninger.

Barna som hadde gitt sitt samtykke til å delta i denne studien ble innkalt til testing som foregikk på Hjelp24 NIMI Ullevål. Her blir også barna i kohorten testet. Det ble ikke gjennomført noen pilottesting i forkant av reliabilitetsstudiet. Vi brukte en allerede planlagt testdag for ACL-barna i kohortstudien, hvor vi testet åtte barn, for å se på gjennomføringen og rutiner i forhold til testingen. Vi hadde et møte rett i forkant av testingen i metodestudien, hvor protokollen ble gjennomgått sammen med de to testerne. Barna i metodestudien ble ikke tilbudt noen prøvetesting i forkant av selve testdagen for å bli kjent med hinke- og styrketestene. Dette er heller ikke praksis for ACL-barna i den pågående prospektive kohortstudien.

3.5.1 Målemetodene

Isokinetiske muskelstyrketester

Isokinetiske muskelstyrkemålinger av kneekstensjon og fleksjon ble målt med et Biodex 6000 dynamometer (Biodex Medical Systems Inc., Shirley, New York). Vi brukte samme testprotokoll som blir brukt i kohortstudien på barn med ACL skade og som også er beskrevet og brukt i andre studier ved NAR (Eitzen et al., 2010; Moksnes et al., 2008). Kalibrering av instrumentet ble gjort på forhånd før hver testdag i henhold til maskinwarens protokoll. Bevegelsesbanen (ROM) var satt fra 90° knefleksjon til full ekstensjon (0°). Den angulære hastigheten var på 60°/sek. Stolens posisjon ble tilpasset slik at omdreiningsaksen for vektarmen på dynamometeret gikk igjennom fleksjons-/ekstensjonsaksen for kneet. Armen ble stilt i lengde etter testpersonens ben slik at den var ca. 2 cm over helkappen på skoen. Vi korrigerer for effekt av tyngdekraften før test av hvert ben i henhold til prosedyrerene gitt for maskinen. Høyden og dybden av setet ble individuelt tilpasset slik at utgangsstillingen i hoften ca. 100°.

Stolinnstillingene ble lagret ved første testing av hvert barn, slik at samme innstilling ble brukt ved de tre testrundene (t1, t2 og t3) av begge testerne (HM, IE). Testpersonene ble sikret fast til stolen med stropper over brystet, bekken, lår og legger. Det ble utført fire testrepetisjoner etterfulgt av ett minutt pause før fem gjeldende repetisjoner. Kun konsentriske målinger ble utført (figur 3). Mellom testing av hvert ben ble stroppene løsnet på og barnet bedt om å sette seg godt bak i stolen igjen. Barna ble bedt om å krysse armene over brystet under testingen. En utskrift av styrkeresultatene ble lagret manuelt, hvor deretter de aktuelle variablene ble plottet for videre statistiske analyser.

Serie	Hastighet	Muskelgruppe	Forsøk	ROM
Oppvarming	60°/sek.	KE og KF vekselvis	4	90°
Konsentrisk	60°/sek.	KE og KF vekselvis	5	90°

Figur 3: Biodex protokoll. Hvilepause på 1 min. mellom seriene. KE; kneekstensjon, KF; knefleksjon, ROM; bevegelsesutslag.

Hinketester

Vi brukte de fire hinketestene som beskrevet av Noyes et al. (1991) (figur 4).

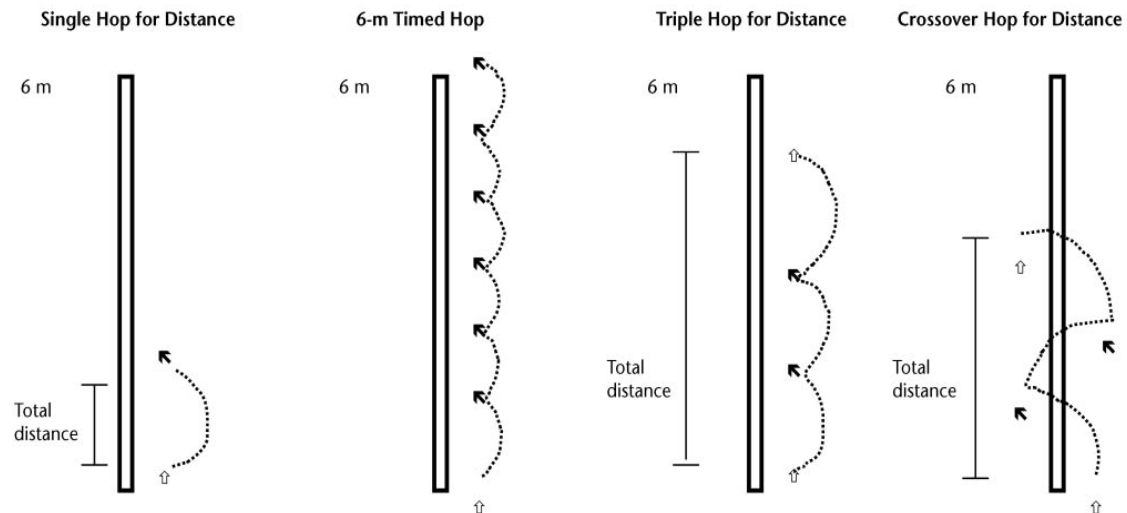
Testprotokollen er den samme som ved testing av barna med ACL skade i den pågående kohortstudien (Moksnes et al., 2008).

En 6 m lang og 15 cm bred tape på gulvet markerte ”testbanen”. Det ble i tillegg lagt et målebånd i lengderetningen for å måle opp hoppelengdene. En stoppeklokke (modell: Rucanor) ble brukt til å måle tiden ved 6 m hoppetest.

SH og TH (distanse): Startposisjon er stående på testbenet med tærne mot startlinjen av tapen. Målingen gjøres fra startlinjen til plasseringen av bakkdelen på helen til testpersonen etter ett eller tre hopp på ett ben. Distansen ble målt til nærmeste cm.

CH (distanse): Startposisjonen er som beskrevet over, stående på samme side av den vertikale tape-linjen som testbenet er. Tre hopp gjøres hvor man krysser linjen så man lander på motsatt side av startposisjonen. Målingen gjøres fra startlinjen til plassering av bakkdelen av helen. Distanse ble målt til nærmeste cm.

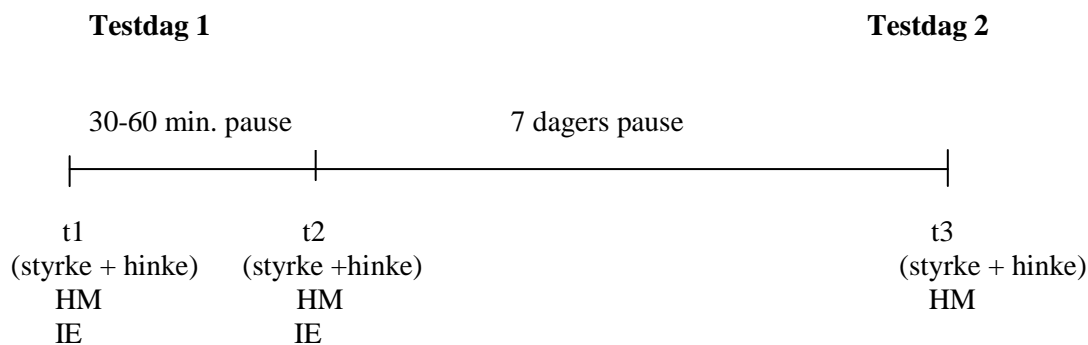
6m (tid): Tapen som er lagt markerer de 6 m som skal hoppes. Startposisjon som beskrevet over. Testeren sier klar-ferdig-gå og starter klokken i det testpersonen starter, og stopper klokken når bakkdelen av helen til testpersonen krysser mållinjen. Tiden ble målt til nærmeste 0,01 sekund.



Figur 4: Presentasjon av hinketestene tatt fra Reid et al. (2007). Testene er opprinnelig beskrevet av Daniel et al. (1982) og Noyes et al. (1991).

3.6 Datainnsamling

Innsamlingen av all data foregikk i lokalene til Hjelp24 NIMI på Ullevål stadion. Testing av de 28 barna ble gjort i to testomganger. Høsten 2010 (oktober-november) testet vi 13 barn fra Ullern fotball akademi, og vinteren 2011 (januar) testet vi 15 barn fra Heming racing. De samme testerne (HM og IE) ble brukt ved begge anledninger. Deltagerne ble testet i hinke- og styrketester i tre testrunder (t1, t2, t3), fordelt på to testdager med en ukes mellomrom (figur 5). Interrater reliabiliteten ble regnet ut i fra målingene til IE og HM fra dag 1, og intrarater reliabiliteten ut i fra målingene til HM fra dag 1 og dag 2.



Figur 5: Tidstabell for den totale undersøkelsesprotokollen med tre testrunder (t1, t2 og t3) og to testere (HM, IE).

Barna utførte to testrunder som inkluderte både hinke- og styrketesting dag 1 med ca. 30-60 min. pause mellom t1 og t2. Hver av målingene for muskelstyrke og hinketestene tok ca. 10-15 min. å gjennomføre. Oppvarming ble gjort før begge testrundene (t1 & t2). Dag 2 var det kun en testrunde (t3) hvor alle barna ble testet av HM. Intervallet mellom testdagene var lik for alle barna. Barna ble testet på samme tid på dagen ved begge anledninger. Det ble ikke gitt noen aktivitetsrestriksjoner for barna mellom dag 1 og 2.

3.6.1 Testrekkefølge

De inkluderte barna til den første testomgangen ble fordelt på to grupper, og deretter i to puljer innad i hver gruppe, etter ønske om testdager og testtider. Det var åtte barn i Gruppe 1 og fem barn i Gruppe 2. De inkluderte barna til den andre testomgangen ble fordelt etter samme prinsipp. Her var det fem barn i Gruppe 1 og ni barn i Gruppe 2 (se figur 3). Rekkefølgen av testing innad i hver pulje ble bestemt tilfeldig og den ble holdt lik for begge testdagene for å sikre lik pause mellom testing for alle barna. Styrketestene ble utført før hinketestene i begge testrundene, i henhold til testprotokollen fra kohortstudien.

Logistikken ved testing på dag 1 var lagt opp slik at for hver testrunde (t1 og t2) ble barna testet først av 'tester 1' i styrke og deretter av 'tester 2' i hinke. Ved neste testrunde (t2) byttet testerne om slik at barna da ble testet først av 'tester 2' i styrke og deretter av 'tester 1' i hinke. Dette ble gjort for å få best mulig flyt i testingen for både barn og testere, og er dessuten slik testingen ofte foregår i praksis i kohortstudien. Hvilke av testerne (HM/IE) som startet med styrke og hinke ble bestemt tilfeldig.

3.6.2 Registreringsskjema

Alle barna som ønsket å delta i studien hadde på forhånd fylt ut et registreringsskjema sammen med foreldre/foresatte, med spørsmål om eventuell skade, funksjon av høyre og venstre ben, om de kunne hinke uten problemer, type aktivitet de hadde drevet med de siste fire ukene og frekvens av disse aktivitetene. Det generelle aktivitetsspørsmålet var lik det kohortebarna blir spurt om i Quesback-spørreskjema hver måned (på email). Registreringsskjemaet ble brukt for å inkludere/ekskludere barn til studien på bakgrunn av oppgitt funksjon og skadestatus, samt for å få en beskrivelse av utvalget. Vi sendte ut

en oppfordring på mail til foreldre/foresatte før begge testdagene, om å melde fra dersom skadestatus til barna hadde endret seg underveis.

3.6.3 Organisert aktivitet & idrett

Den første testdagen svarte barna i tillegg på hvilke aktiviteter/idretter de var med på organisert trening i, og frekvensen av den organisert aktiviteten/treningen (vedlegg 3).

3.6.4 Randomisering av startben

Vi hadde i forkant av testingen randomisert rekkefølgen av hvilket ben som skulle testes først. Vi utførte en pseudorandomisering i SPSS (Statistical Package for Social Science, versjon 18) med en blokkstørrelse på seks (3+3), som sikret lik fordeling av høyre og venstre som startben innen hver blokk. Vi brukte nummererte og forseglete konvolutter hvor det lå en lapp med en randomiseringskode, som svarte til høyre eller venstre ben, og som stemte overens med den fullstendige randomiseringslisten. Randomiseringen ble utført av en forskningsmedarbeider på Hjelp24 NIMI. Hun hadde ingen kjennskap til den fullstendige randomiseringslisten. Hun var heller ikke involvert i selve testingen. Barna ble randomisert med startben fortløpende ettersom de ble inkludert. Ingen av testerne kjente til randomiseringen på forhånd. Randomiseringen ble kun utført en gang, og bestemte dermed startbenet ved testing både for isokinetiske muskelstyrke og hinketester ved de tre testrundene (t1, t2 og t3).

3.6.5 Antropometriske data

Høyde og vekt ble målt av en testmedhjelper på selve testdagen. En digital vekt som stod i treningsrommet på Hjelp24 NIMI og en høydemåler inne på et av behandlingsrommene ble brukt. Barna ble målt til nærmeste cm og kg. Alder ble oppgitt i registreringskjema. Alder ved test ble regnet fra fødselsdato og dato for første testdag.

3.7 Testprosedyrer

3.7.1 Isokinetisk muskelstyrketesting

Barna gjennomførte først en standardisert oppvarming på 5 minutter på ergometersykkel (Technogym) ute i treningsrommet med en testmedhjelper til stede. Deretter utførte de styrketestene i Biodex 6000 maskinen for å måle konsentrisk kneekstensjon og fleksjon. Prinsippene for isokinetisk muskelstyrketesting ble forklart til hver av barna før testingen startet. Barna ble fortalt at: ”Vi skal nå måle hvor sterk du

er. Du skal ta i alt du kan. Det vil si at du skal spark helt opp til det stopper, og bøye helt inn til det stopper.” Det ble gjort fire testrepetisjoner etterfulgt av ett minutt pause før fem gjeldende repetisjonene ble utført. I prøvoforsøkene fikk barna beskjed om å ”ta i det de kunne”. Barna ble gitt standardisert verbal informasjon av begge testere (HM/IE) under selve testen, hvor det ble sagt ”start”, talt antall repetisjoner, og til slutt sagt ”stopp”. Ingen annen verbal eller visuell feedback ble gitt under selve testen. Begge ben ble testet i randomisert rekkefølge.



Figur 6: Isokinetisk muskelstyrke målt med et Biodex 6000 dynamometer. Skriftlig samtykke fra forsøksperson og foresatt er innhentet for bruk av bildet (vedlegg 4).

3.7.2 Hinketester

Etter styrketestingen gikk barna rett opp for å utføre hinketestene som ble utført i en lang gang i annen etasje på Hjelp24 NIMI. Underlaget de hinket på var linoleum. Barna utførte de fire hinketestene SH, CH, TH og 6m i nevnt rekkefølge. Testerne (HM/IE) demonstrerte hvert hink før den aktuelle testen. Barna ble deretter gitt ett prøvehopp. Vi registrerte og noterte to gyldige hopp for hver av hinketestene. De ble gitt inntil fem forsøk (eksklusiv prøvehopp). Et gyldig hopp krevde en stødig landing på testbenet som ble holdt i 2 sek. Re-hopp ble gitt ved ekstra rotasjon/vridning av foten, eller hvis man må støtte seg i bakken/veggen med ben eller overkropp for å holde balansen (Moksnes et al., 2008; Reid et al., 2007). Hinkingen ble utført fortløpende. Det vil si, pause mellom

prøvehopp og test, og hink innad i hver test, var tiden det tok for barnet å gå tilbake til utgangspunktet i tillegg til å konsentrere seg om å hinke på nytt. Pause mellom de fire ulike hinketestene var i tillegg tiden det tok for HM/IE å demonstrere det aktuelle hinket. Det var ingen restriksjoner knyttet til armbevegelser under testing. Begge ben ble testet i randomisert rekkefølge. Registrering av hinkeresultatene ble ført på inn på eget skjema (vedlegg 5). Barna fikk beskjed om å ha på seg samme fottøy begge testdager.

3.8 Databehandling

Resultatene fra metodestudien blir lagret innelåst i manuelle arkiv på Hjelp24 NIMI Ullevål og har blitt lagt inn manuelt i et Excel-ark som lagres på serveren på NIMI og på serveren 'Vilje' ved OUS. Vi aidentifiserte dataene ved elektronisk lagring på pc før de ble lagt inn i SPSS (versjon 18) for statistiske analyser.

3.8.1 Målevariabler

Hovedvariabler for hinketestene er distanse (cm) og tid (sek.). Isokinetisk muskelstyrkemålinger er oppgitt i newton meter (Nm) for alle torque-variabler (PT, TQ30 og AVG PT) og joule (J) for Tw. LSI (%) ble regnet ut og presentert for å illustrere forholdet mellom høyre og venstre ben. Variablene er på kvotenivå og dataene er kontinuerlige.

3.8.2 Utvalgsstørrelse

Vi baserte utvalgsstørrelsen på erfaringsmaterialet fra tidligere reliabilitetsstudier. Studiene som har undersøkt de samme hinketestene som oss har rapportert fra 18 til 35 deltagere (Ross et al., 2002; Reid et al., 2007). Sistnevnte er eneste studie som styrkeberegnet seg frem til sin utvalgsstørrelse, men de har et annet design enn oss. Styrkestudiene for voksne har hatt fra 13 til 30 deltagere (Li, Wu, Maffulli, Chan, & Chan, 1996; Lund et al., 2005) og studiene på barn fra 5-13 deltagere (Kellis et al., 1999; Merlini et al., 1995). Videre ble det diskutert med statistiker på NIH og sett på litteratur angående reliabilitet og utvalgsstørrelse (Donner & Eliasziw, 1987; Walter, Eliasziw, & Donner, 1998; Baumgartner, 2003; Lexell & Downham, 2005).

Med power satt til 80 %, to testere, og forventet reliabilitet på 0,80 (r/ICC), vil man ut i fra grafen (figur 1) i artikkelen til Donner & Eliasziw (1987) kunne avlese at man

trenger ca. $k = 7$ individer. Hvis den estimerte reliabiliteten i steden er forventet å være 0,60 vil man trenge ca. $k = 13-14$ individer, for å med stor sannynlighet kunne oppnå forventet reliabilitet ved kontinuerlige variabler. En god fornemmelse av hva r/ICC bør være, forutsetter at det er gjort en del studier med nettopp denne teknikken fra tidligere. Hypoteseprøving i denne situasjonen er kunstig, da man står overfor et estimeringsproblem, nemlig hvor høy r/ICC er. Dette er grunnen til at man ikke antalls beregner slike studier i samme grad som ved sammenlikning av grupper. Kurvene flater ut for antall personer > 25 til 30 stykker. Det gjør at det ofte velges ut et passende antall deltagere på bakgrunn av tid og ressurser tilgjengelig. Baumgartner et al. (2003) foreslår en utvalgstørrelse på 50-100 individer ved bruk av 90 % KI med $r \geq 0,80$, og to målinger. En utvalgsstørrelse på 15-20 individer blir ofte brukt i reliabilitetsstudier på kontinuerlige data. Større utvalg (30-50) er også blitt foreslått som nødvendig dersom man skal kunne gi praktiske råd for minste relle endring (SRD) (Lexell & Downham, 2005).

3.8.3 Statistiske analyser

De antropometriske variablene (kjønn, alder, høyde, vekt og KMI) er beskrevet som deskriptiv statistikk. Variablene er oppgitt som gjennomsnitt \pm SD. En uavhengig t-test ble brukt for å undersøke eventuelle signifikante forskjeller mellom guttene og jentene i metodestudien. Normalfordeling av dataene ble testet ved Kolmogorov-Smirnov, hvor $p > 0,05$ indikerer at dataene er normalfordelt.

Beste score (av to) ble brukt i analysene for hinketestene (Clark, 2001). For styrkevariablene PT, Tw og TQ30 ble beste score registrert, mens AVG PT er gjennomsnittet av fem repetisjoner ($k = 5$). Resultatene leses direkte ut i fra datautskriften fra styrketesten. LSI er uttrykt i prosent, her beregnet som høyre/venstre $\times 100$. Gjennomsnitt og SD ble kalkulert for de aktuelle variablene. Signifikante forskjeller mellom de to testerne, og mellom testdagene ble undersøkt med en parret t-test. En uavhengig t-test ble brukt for å finne eventuelle signifikante forskjeller mellom kjønn. Signifikante forskjeller mellom høyre og venstre ben ble undersøkt ved bruk av en parret t-test. Signifikansnivået var satt til $p < 0,05$.

Relativ reliabilitet ble uttrykt ved bruk av ICC. En ANOVA *two-way mixed* eller *two-way random* modell ble brukt i beregningen av henholdsvis intrarater (ICC 3,1) og

interrater (ICC 2,1) reliabilitet (Shrout & Fleiss, 1979; McGraw & Wong, 1996). For alle ICC'er ble *agreement* brukt hvor systematiske forskjeller blir tatt hensyn til (McGraw & Wong, 1996; Weir, 2005). For ICC (2,1)/A(3,1) ble *single measure* reliabilitet brukt siden målingene her er representert som beste score. *Average measure* reliabilitet ble brukt for AVG PT, ICC (2,5)/(3,5) (Portney & Watkins, 2000; McGraw & Wong, 1996). Et 95 % KI til ICC ble oppgitt ut i fra SPSS datautskriften. Gjennomsnittsdifferansen til målingene (\bar{d}) og et tilhørende 95 % KI ble oppgitt ut i fra de parrede t-testene.

Absolutt reliabilitet ble estimert ved $SEM = \sqrt{\text{residual mean square error}}$ (RMSE) fra ANOVA (Stratford & Goldsmith, 1997). SEM % ble regnet ut som $SEM/\text{grand mean}$ (fra ANOVA) x 100. (Lexell & Downham, 2005; Flansbjer & Lexell, 2010). LOA ble regnet ut ved $\bar{d} \pm 1,96 \times SD$ (til differansen), hvor \bar{d} mellom målingene til de to testerne ble plottet mot de individuelle gjennomsnittene for hvert sett av målinger og fremstilt i Bland-Altman plott (Atkinson & Nevill, 1998). Graden av heteroskedastisitet ble undersøkt ved bruk av Pearson's korrelasjonskoeffisient, hvor en signifikant r-verdi indikerer heteroskedastisitet (Lund et al., 2005).

Vi beregnet minste reelle endring som; $SRD = 1,96 \times \sqrt{2} \times SEM$, basert på et 95 % KI ($z = 1,96$) og $\sqrt{2}$ (for repeterte målinger) (Copay et al., 2007; Portney & Watkins, 2000). I tillegg regnet vi ut SRD % som representerer endring oppgitt i relative termer, $SRD \% = SRD/\text{grand mean} \times 100$. Systematiske målefeil (bias) tilsvarer \bar{d} , og tilfeldige målefeil (random error) tilsvarere SRD (Lexell & Downham, 2005; Flansbjer & Lexell, 2010).

3.9 Etikk

Måle metodene og testprotokollen vi testet i metodestudien var allerede godkjent gjennom den prospektive pågående kohortstudien på barn med ACL skade. Regional komitè for medisinsk forskningsetikk, Øst-Norge (REK 1) godkjente protokollen 09.11.06 (ref. 684-06288 1.2006.78) (vedlegg 6). En prosjektendring til den opprinnelige studien ble godkjent av Regional komitè for medisinsk og helsefaglig forskningsetikk, Sør- Øst C (REK Sør-Øst C) den 26.08.10 (ref. 2010/1855) (vedlegg 7). Studien ble utført i henhold til de etiske retningslinjene som er gitt ved Helsinkideklarasjonen/ Helseforskningsloven. Data ble oppbevart forsvarlig i henhold til personopplysningsloven.

Testingen ble utført av erfarne fysioterapeuter som tidligere har testet både barn og voksne med og uten ACL skade. Siden vi testet barn (noen under 12 år) var vi avhengige av skriftlig samtykke av forelder/foresatt i tillegg til samtykke fra barnet selv. Vi ga skriftlig og muntlig informasjon til barn og foresatte om studiets formål. Foreldrene fikk tilbud om å være tilstede i testlokalet da barna ble testet. Testene ble gjennomført på Hjelp24 NIMI på Ullevål stadion som er en veletablert institusjon med lang erfaring i behandling og testing av barn og voksne med idrettsskader. Hinketestene er ikke ekstreme av noe slag og inneholder elementer av bevegelse som barn normalt vil utføre i løpet av en uke. Isokinetiske styrketester blir ofte brukt for å undersøke muskelstyrke både hos barn og voksne. Det er ingen kjent risiko forbundet med denne type testing.

4. Resultater

Jeg vil først beskrive utvalget jeg testet med tanke på antropometriske data, funksjon og aktivitetsbeskrivelse. Deretter vil rådata fra styrke- og hinketestene bli presentert som deskriptive data, etterfulgt av intra- og interater reliabiliteten for de ulike målemetodene. Avslutningsvis vil det bli gitt en indikasjon på minste reelle endring til målemetodene.

4.1 Karakteristikk av utvalget

Det var 30 barn som sa seg villig til å delta i metodestudien. Ett barn hadde ikke tid til å bli med på testing, og trakk seg før inklusjon. Av de 29 barna som ble inkludert og randomisert, var det ett barn som ikke stilte til testing pga sykdom. Antropometriske data er oppgitt i tabell 3 for barna som kom til testing (n= 28). Utvalget var homogent med tanke på antropometriske data. Det var størst spredning mellom gutter og jenter for kroppsvekt. Jentene hadde noe større gjennomsnittsverdi for høyde, men det var ingen signifikante forskjeller mellom gutter og jenter for noen av de antropometriske variablene ($p = 0,11-0,82$).

Tabell 5: Antropometriske data for barna i metodestudien (n= 28). Gjennomsnitt \pm SD.

Kjønn	n	Alder (år)	Høyde (cm)	Vekt (kg)	KMI (kg/m ²)
Metodestudien	28	12,4 \pm 0,3	156 \pm 0,1	44 \pm 5,7	18 \pm 1,9
Gutter	22	12,4 \pm 0,3	155 \pm 0,1	44 \pm 6,2	18 \pm 2,0
Jenter	6	12,4 \pm 0,2	158 \pm 0,1	42 \pm 2,5	17 \pm 1,2

KMI = kroppsmasseindeks.

Det var 9 av 28 (32 %) barn som hadde svart *ja* på spørsmålet om de hadde/hadde hatt en skade innen de siste 6 månedene. Av disse svarte 3 av 28 (11 %) barn at de hadde hatt en kneskade (Schlatters og smerter i kneet), og 6 av 28 (21 %) svarte at de hadde hatt en *annen skade* (lysketrek, ryggvondt, strekk kne/legg, brudd i håndledd og helsmerter). Av barna med skade (n= 9) svarte 2 at skaden hindret dem nå, og 4 barn svarte at det hindret dem i å hinke på det aktuelle benet. Alle barna oppga at de trente som vanlig/for fullt. Etter samtale med foreldre ble alle barna i utgangspunktet funnet egnet for studiet (se kap. 3.2.1 for prosedyre).

På spørsmålet om funksjon (min.-maks.; 0-10) av høyre og venstre ben i daglige- og idrettsaktiviteter rapporterte 22 av 28 (79 %) barn at de hadde funksjon 10 (7-10) på høyre ben og 20 av 28 (71 %) barn svarte funksjon 10 (3-10) på venstre ben.

Barna i metodestudien var alle idrettsaktive og med på organisert trening innenfor minst en idrett (n = 28). Det var 7 % som trente i snitt mer enn seks ganger per uke, 79 % trente 4-5 ganger per uke og 14 % som trente 2-3 ganger per uke innenfor sin respektive idretter; alpint eller fotball. På registreringsskjemaet hadde barna svart hvilke aktiviteter (organisert/ikke- organisert) de hadde vært med på de siste fire ukene (n= 28).

Størst deltagelse var det i gym på skolen (100 %), fotball (82 %), styrketrening (61 %) og alpint/telemark (57 %). De fleste av barna (96 %) oppgav en samlet frekvens på over 4-5 ganger per uke for disse aktivitetene.

Etter testingen ble to barn ekskludert pga smerter som påvirket utførelsen av hinke- og styrketestene. Det gjorde at vi totalt inkluderte data for 26 barn som kvalifiserte for statistiske analyser (se fig.2 for frafall). Av disse 26 barna var 13 randomisert med henholdsvis høyre og venstre som startben.

4.2 Deskriptive data til målemetodene

Data fra hinke- og styrketestene var alle normalfordelte. Vi fant ingen signifikante forskjeller mellom gutter og jenter, med unntak av TQ30 knefleksjon høyre ben ($p = 0,04$). Data for gutter og jenter ble derfor slått sammen for videre statistiske analyser. Det var ingen signifikante forskjeller mellom høyre og venstre ben med unntak av TQ30 kneekstensjon ($p = 0,04$ interrater og $p = 0,02$ intrarater). Siden data for intrarater reliabiliteten viste en statistisk signifikant forskjeller for venstre ben ved styrkemålingene (intrarater), ble det besluttet å presentere data for begge ben både for hinke- og isokinetisk muskestyrke. LSI blir presentert i de deskriptive dataene for å belyse forholdet mellom høyre og venstre ben.

4.2.1 Isokinetiske muskelstyrkemålinger

Intrarater

I tabell 4 er data for de konsentriske isokinetiske muskelstyrkemålingene for tester HM presentert. Variabiliteten til målingene (SD) var relativt lik begge dager med unntak av LSI for TQ30. En læringseffekt ble observert for venstre ben ved PTE ($p = 0,01$), TwE ($p = 0,001$), TQ30 ($p = 0,03$) og AVG PT ($p = 0,002$). Gjennomsnittsscore for LSI varierte fra 100,0-142,2 %. Det var 7 barn med forskjell > 50 % ved LSI TQ30. Barna ble ikke ekskludert siden vi ikke hadde bestemt noen ”cut-off” for variasjon. Videre analyser er derfor basert på alle barna ($n = 26$), men en tabell for TQ30 med korrigert n er presentert til slutt under vedlegg (vedlegg 8).

Tabell 6: Gjennomsnitt \pm SD og p -verdier for konsentriske isokinetiske muskelstyrkemålinger ($60^\circ/\text{sek.}$) med hensyn til intrarater reliabiliteten ($n = 26$).

n= 26	Tester HM		p-verdi
	Dag1	Dag 2	
PTE hø (Nm)	104,0 \pm 15,7	106,3 \pm 15,4	0,15
PTE ve	101,9 \pm 17,4	106,7 \pm 15,8	0,01*
LSI (%)	102,8 \pm 9,6	100,0 \pm 8,8	0,09
PTF hø	54,3 \pm 11,0	55,1 \pm 9,2	0,52
PTF ve	52,5 \pm 10,5	53,4 \pm 9,0	0,49
LSI (%)	104,2 \pm 12,3	103,6 \pm 10,9	0,79
TwE hø (J)	480,7 \pm 81,6	492,4 \pm 81,2	0,16
TwE ve	450,4 \pm 77,5	476,0 \pm 77,0	0,001*
LSI (%)	107,2 \pm 9,8	103,7 \pm 7,3	0,03*
TwF hø	280,3 \pm 63,1	274,2 \pm 65,9	0,40
TwF ve	270,3 \pm 60,2	269,1 \pm 64,1	0,87
TwF LSI (%)	104,1 \pm 9,9	102,1 \pm 8,5	0,29
TQ30E hø	61,5 \pm 11,8	60,3 \pm 11,8	0,53
TQ30E ve	51,3 \pm 14,1	56,2 \pm 11,6	0,03*
LSI (%)	144,2 \pm 122,1	109,0 \pm 18,0	0,16
TQ30F hø	49,9 \pm 12,4	49,7 \pm 10,8	0,90
TQ30F ve	47,2 \pm 12,2	49,3 \pm 9,5	0,30
LSI (%)	111,5 \pm 43,6	101,5 \pm 16,1	0,24
AVG PTE hø (Nm)	95,2 \pm 14,7	98,2 \pm 14,6	0,07
AVG PTE ve	92,9 \pm 15,9	97,9 \pm 15,9	0,002*
LSI (%)	103,3 \pm 9,7	100,8 \pm 8,0	0,10
AVG PTF hø	48,9 \pm 9,1	49,5 \pm 8,6	0,56
AVG PTF ve	47,2 \pm 8,9	48,8 \pm 10,6	0,26
LSI (%)	104,0 \pm 7,3	102,8 \pm 11,9	0,60

* $p < 0,05$. E= kneekstensjon, F= knefleksjon, PT= peak torque, Nm= newton meter, LSI = høyre/venstre ben $\times 100$, Tw= total work, J= joule, TQ30= torque ved 30° , AVG PT= average PT.

Interrater

I tabell 5 presenteres data for de isokinetiske muskelstyrkemålingene for begge testere (IE og HM). Det var signifikante forskjeller mellom målingene ved PTE ($p = 0,001-0,002$), PTF venstre ($p = 0,02$), TwE venstre ($p = 0,01$), TQ30E ($p = 0,001-0,01$), AVG PTE ($p = 0,002$) og AVG PTF ($p = 0,01-0,04$). Det indikerer en trend av systematiske målefeil mellom de to testerne. Det ble målt høyere gjennomsnittsverdier av tester HM for alle variabler, unntatt for TwE og TQ30E. LSI varierte fra 102,6-142,2 % .

Tabell 7: Gjennomsnitt \pm SD og p -verdier for konsentriske isokinetiske muskelstyrkemålinger (60°/sek.) med hensyn til interrater reliabiliteten ($n = 26$).

n = 26	Dag 1		p - verdi
	Tester IE	Tester HM	
PTE hø (Nm)	97,8 \pm 16,3	104,0 \pm 15,7	0,001*
PTE ve	95,7 \pm 16,7	101,9 \pm 17,4	0,002*
LSI (%)	102,8 \pm 10,6	102,8 \pm 9,6	0,10
PTF hø	52,6 \pm 8,9	54,3 \pm 11,0	0,19
PTF ve	49,3 \pm 9,5	52,5 \pm 10,5	0,02*
LSI (%)	107,7 \pm 8,3	104,2 \pm 12,3	0,21
TwE hø (J)	492,9 \pm 87,2	480,7 \pm 81,6	0,10
TwE ve	470,2 \pm 83,4	450,4 \pm 77,5	0,01*
LSI (%)	105,3 \pm 9,3	107,2 \pm 9,8	0,40
TwF hø	273,3 \pm 53,7	280,3 \pm 63,1	0,28
TwF ve	257,4 \pm 56,7	270,3 \pm 60,2	0,06
LSI (%)	107,2 \pm 9,9	104,1 \pm 9,9	0,15
TQ30E hø (Nm)	66,9 \pm 15,2	61,5 \pm 11,8	0,01*
TQ30E ve	62,8 \pm 13,4	51,3 \pm 14,1	0,001*
LSI (%)	107,7 \pm 17,1	142,2 \pm 122,1	0,16
TQ30F hø	49,4 \pm 8,3	49,9 \pm 12,4	0,78
TQ30F ve	47,0 \pm 9,6	47,2 \pm 12,2	0,90
LSI (%)	106,6 \pm 12,5	111,5 \pm 43,6	0,57
AVG PTE hø (Nm)	90,7 \pm 15,4	95,2 \pm 14,7	0,002*
AVG PTE ve	88,9 \pm 15,9	92,9 \pm 15,9	0,01*
LSI (%)	102,6 \pm 9,4	103,3 \pm 9,7	0,72
AVG PTF hø	47,1 \pm 8,5	48,9 \pm 9,1	0,04*
AVG PTF ve	43,8 \pm 8,9	47,2 \pm 8,9	0,003*
LSI (%)	108,6 \pm 12,0	104,0 \pm 7,3	0,09

* $p < 0,05$. E= kneekstensjon, F= knefleksjon, PT= peak torque, Nm= newton meter, LSI = høyre/venstre ben x 100, Tw= total work, J= joule, TQ30= torque ved 30°, AVG PT= average PT.

4.2.2 Hinketester

Intrarater

I tabell 6 presenteres data fra hinketestene for tester HM. Det var ingen signifikante forskjeller mellom målingene. Gjennomsnittsverdiene var noe høyere dag 2, med mindre SD enn registrert dag 1. Alle LSI score var > 90 %. Gjennomsnittsscore for LSI varierte fra 98,9 til 102,3 %.

Tabell 8: Gjennomsnitt \pm SD og p-verdier for hinketestene med hensyn til intrarater reliabiliteten ($n = 26$).

n = 26	Tester HM		p-verdi
	Dag 1	Dag 2	
SH hø (cm)	129,5 \pm 17,1	127,9 \pm 15,2	0,42
SH ve	126,9 \pm 15,9	128,0 \pm 14,5	0,54
LSI (%)	102,3 \pm 7,9	101,0 \pm 6,3	0,20
CH hø (cm)	374,2 \pm 57,7	388,5 \pm 53,9	0,06
CH ve	378,5 \pm 51,3	387,4 \pm 49,9	0,21
LSI (%)	98,9 \pm 9,2	100,4 \pm 6,7	0,43
TH, hø (cm)	414,4 \pm 49,7	415,6 \pm 53,8	0,81
TH ve	410,1 \pm 50,2	416,6 \pm 45,2	0,38
LSI (%)	101,4 \pm 8,7	99,8 \pm 6,8	0,24
6m hø (sek.)	1,85 \pm 0,16	1,87 \pm 0,21	0,38
6m ve	1,86 \pm 0,17	1,85 \pm 0,19	0,67
LSI (%)	99,8 \pm 7,0	101,3 \pm 3,8	0,23

* $p < 0,05$. Hø= høyre ben, ve= venstre ben, SH= single hop, LSI= limb symmetry index, CH= crossover hop, TH= triple hop, 6m= 6 meter hop.

Interrater

Tabell 7 viser data fra hinketestene for begge testere (IE og HM). Det var ingen signifikante forskjeller mellom testerne, med unntak for 6m høyre ben ($p = 0,01$). Alle LSI score var > 90 %. Gjennomsnittsscore for LSI varierte fra 98,9 til 102,3 %. SD til målingene var relativt lik for begge testere.

Tabell 9: Gjennomsnitt \pm SD og p-verdier for hinketestene med hensyn til interrater reliabiliteten (n= 26).

n = 26	Dag 1		p - verdi
	Tester IE	Tester HM	
SH hø (cm)	126,7 \pm 16,9	129,5 \pm 17,1	0,10
SH ve	126,2 \pm 14,7	126,9 \pm 15,9	0,70
LSI (%)	100,5 \pm 7,9	102,3 \pm 7,9	0,20
CH hø (cm)	383,7 \pm 50,8	374,2 \pm 57,7	0,23
CH ve	381,3 \pm 50,5	378,5 \pm 51,3	0,64
LSI (%)	101,0 \pm 9,4	98,9 \pm 9,2	0,43
TH hø (cm) ^a	415,0 \pm 51,9	414,6 \pm 50,7	0,92
TH ve ^a	416,2 \pm 46,5	411,4 \pm 50,8	0,43
LSI (%) ^a	99,7 \pm 5,8	101,1 \pm 8,8	0,41
6m hø (sek.)	1,91 \pm 0,20	1,85 \pm 0,16	0,01*
6m ve	1,90 \pm 0,21	1,86 \pm 0,17	0,22
LSI (%)	101,2 \pm 6,0	99,8 \pm 7,0	0,32

*p < 0,05. Hø= høyre ben, ve= venstre ben, SH= single hop, LSI= limb symmetry index, CH= crossover hop, TH= triple hop, 6m= 6 meter hop, ^a= ett barn uten gyldig hopp på TH høyre ben, n= 25 for alle analyser på TH.

4.2.3 Oppsummering av deskriptive data

Det ble generelt registrert høyere score dag 2. Isokinetiske muskelstyrkemålinger viste signifikante forskjeller for venstre ben (intrarater). HM registrerte generelt høyere verdier enn IE for de isokinetiske muskelstyrkemålingene. Signifikante forskjeller ble sett både for høyre og venstre ben mellom testerne (interrater). For hinketestene var det bare 6m som viste en signifikant forskjell (interrater). Alle LSI verdier > 90 %, med unntak av LSI ved TQ30.

4.3 Reliabiliteten til måle metodene

Intra- og interrater reliabilitet er presentert som relativ og absolutt reliabilitet oppgitt i tabell 10–13. En fremstilling av systematiske- og tilfeldige feil, samt øvre og nedre LOA er presentert i Bland-Altman plott (figur 7) for et utvalg av variablene for å belyse de individuelle variasjonene til målingene visuelt. Verdier til Bland-Altman plottene (for alle variabler) kan leses direkte ut i fra reliabilitetstabellene hvor \bar{d} representerer systematiske målefeil (bias) og SRD de tilfeldige målefeilene.

4.3.1 Isokinetisk muskelstyrke

Intrarater reliabilitet

Intrarater reliabiliteten til de isokinetiske muskelstyrkemålingene er presentert i tabell 10. ICC varierte fra 0,56-0,92. De systematiske målefeilene (\bar{d}) varierte i fra -5,0-1,2 Nm og i fra -25,6-6,0 J. Alle \bar{d} verdiene for Nm var lave (≤ 5 Nm) med tilhørende smale 95 % KI. Flertall av negative \bar{d} -verdier viser til høyere målinger dag 2. Et 95 % KI til \bar{d} som ikke går igjennom null indikerer signifikante forskjeller mellom målingene (Batterham & George, 2003). Andelen \bar{d} er lavere enn SRD. De relative verdiene av SEM (SEM %) varierte fra 5,1 %-14,9 % for alle målingene.

Tabell 10: Relativ og absolutt intrarater reliabilitet presentert for de ulike isokinetiske muskelstyrkemålingene ($n=26$).

n= 26	ICC (3,1)	95 % KI	\bar{d}	95 % KI \bar{d}	SEM	SEM %	SRD	SRD %
PTE hø (Nm)	0,87	(0,73 - 0,94)	-2,3	(-5,4; 0,9)	5,5	5,2	15,2	14,4
PTE ve	0,85	(0,62 - 0,94)	-4,8	(-8,0; -1,6)	5,6	5,4	15,5	14,9
PTF hø	0,81	(0,63 - 0,91)	-0,8	(-3,3; 1,7)	4,4	8,1	12,2	22,3
PTF ve	0,77	(0,55 - 0,89)	-0,9	(-3,6; 1,8)	4,7	8,9	13,0	24,6
TwE hø (J)	0,87	(0,74 - 0,94)	-11,7	(-28,1; 4,7)	28,7	5,9	79,5	16,3
TwE ve	0,86	(0,53 - 0,95)	-25,6	(-39,0; -12,3)	23,4	5,1	64,8	14,0
TwF hø	0,84	(0,69 - 0,93)	6,0	(-8,6; 20,6)	25,6	9,2	70,9	25,6
TwF ve	0,84	(0,68 - 0,93)	1,2	(-13,1; 15,6)	25,1	9,3	69,5	25,8
TQ30E hø (Nm)	0,68	(0,40 - 0,84)	1,2	(-2,7; 5,1)	6,7	11,0	18,6	30,6
TQ30E ve	0,60	(0,29 - 0,80)	-4,9	(-9,3; -0,4)	7,8	14,5	21,6	40,2
TQ30F hø	0,71	(0,45 - 0,86)	0,2	(-3,4; 3,9)	6,4	12,8	17,7	35,5
TQ30F ve	0,56	(0,24 - 0,78)	-2,1	(-6,2; 2,0)	7,2	14,9	19,9	41,3
AVG PTE hø (Nm)	0,91 α	(0,80 - 0,96)	-2,9	(-6,1; 0,3)	5,6	5,8	15,5	16,0
AVG PTE ve	0,92 α	(0,73 - 0,97)	-5,0	(-7,9; -2,1)	5,1	5,3	14,1	14,8
AVG PTF hø	0,88 α	(0,74 - 0,95)	-0,7	(-3,0; 1,6)	4,1	8,3	11,4	23,2
AVG PTF ve	0,82 α	(0,61 - 0,92)	-1,7	(-4,7; 1,4)	5,3	11,0	14,7	30,6

ICC= intraklasse korrelasjonskoeffisient, (3,1) two-way mixed, agreement, single measure, α = average measure reliabilitet (ICC (3,5), KI= konfidensintervall, \bar{d} = systematiske målefeil, SEM= standardfeil til målingene= $\sqrt{\text{residual mean square}}$, SEM %= SEM/grand mean x 100, SRD₉₅= minste reelle endring= $1,96 \times \sqrt{2} \times \text{SEM}$, SRD % = SRD/grand mean x 100.

Interrater reliabilitet

Interrater reliabiliteten til de isokinetiske muskelstyrkemålingene er presenter i tabell 11. ICC varierte fra 0,48-0,94. De systematiske målefeilene (\bar{d}) varierte fra -6,3-1,5 Nm og -12,9-19,8 J. Flertallet av negative \bar{d} verdier viser til høyere målinger registrert av HM (tester IE-tester HM). Andelen \bar{d} var lavere enn SRD. De relative verdiene av SEM (SEM %) varierte fra 5,2 %-14,5 % for alle målingene.

Tabell 11: Relativ og absolutt interrater reliabilitet for de ulike isokinetiske muskelstyrkemålingene (n= 26).

n = 26	ICC (2,1)	95 % KI	\bar{d}	95 % KI \bar{d}	SEM	SEM %	SRD	SRD %
PTE hø (Nm)	0,82	(0,42 - 0,91)	-6,3	(-9,5; -3,0)	5,6	5,6	15,5	15,3
PTE ve	0,81	(0,48 - 0,92)	-6,2	(-9,9; -2,5)	5,6	5,6	15,5	15,7
PTF hø	0,81	(0,62 - 0,91)	-1,6	(-4,1; 0,9)	4,4	8,2	12,2	22,8
PTF ve	0,76	(0,49 - 0,89)	-3,3	(-5,9; -0,7)	4,5	8,8	12,5	24,6
TwE hø (J)	0,90	(0,79 - 0,95)	12,2	(0,2; 24,3)	25,8	5,3	71,5	14,7
TwE ve	0,87	(0,69 - 0,95)	19,8	(7,1; 33,2)	25,8	5,6	71,5	15,5
TwF hø	0,85	(0,70 - 0,93)	-7,0	(-19,5; 4,8)	22,6	8,2	62,6	22,6
TwF ve	0,83	(0,64 - 0,92)	-12,9	(-25,0; -0,8)	23,2	8,8	64,3	24,4
TQ30E hø (Nm)	0,65	(0,33 - 0,83)	5,5	(1,2; 9,7)	7,5	11,7	20,8	32,4
TQ30E ve	0,48	(-0,03 - 0,76)	11,5	(6,8; 16,3)	8,3	14,5	23,0	40,3
TQ30F hø	0,60	(0,29 - 0,80)	-0,5	(-4,3; 3,3)	6,7	13,5	18,6	37,4
TQ30F ve	0,66	(0,38 - 0,84)	-0,2	(-3,9; 3,4)	6,4	13,6	17,7	37,6
AVG PTE hø (Nm)	0,92 $\bar{\alpha}$	(0,75 - 0,97)	-4,6	(-7,3; -1,8)	4,8	5,2	13,3	14,3
AVG PT ve	0,94 $\bar{\alpha}$	(0,82 - 0,98)	-3,9	(-6,7; -1,2)	4,8	5,3	13,3	14,6
AVG PTF hø	0,93 $\bar{\alpha}$	(0,83 - 0,97)	-1,8	(-3,6; -0,1)	3,1	6,5	8,6	17,9
AVG PTF ve	0,88 $\bar{\alpha}$	(0,62 - 0,95)	-3,4	(-5,5; -1,3)	3,7	8,1	10,2	22,4

ICC= intraklasse korrelasjonskoeffisient, (2,1) two-way random, agreement, single measure, $\bar{\alpha}$ = average measure reliabilitet (ICC 2,5), KI= konfidensintervall, \bar{d} = systematiske målefeil, SEM= standardfeil til målingene= $\sqrt{\text{residual mean square}}$, SEM %= SEM/grand mean x 100, SRD₉₅= minste reelle endring= $1,96 \times \sqrt{2} \times \text{SEM}$, SRD % = SRD/grand mean x 100.

4.3.2 Hinketestene

Intrarater reliabilitet

Intrarater reliabiliteten for hinketestene er presenter i tabell 12. ICC varierte fra 0,62-0,88 for hinketestene. De systematiske målefeilene (\bar{d}) varierte i fra -14,3-1,5 cm for distansehoppene og -0,02-0,01sek. for hopp på tid. Andelen \bar{d} verdier var lavere enn SRD. Flertall av negative verdier viser til høyere målinger dag 2. De relative verdiene av SEM (SEM %) varierte fra 4,3 %-6,7 % for alle hinketestene.

Tabell 12: Relativ og absolutt intrarater reliabilitet for hinketestene (n=26).

n = 26	ICC (3,1)	95 % KI	\bar{d}	95 % KI \bar{d}	SEM	SEM %	SRD	SRD %
SH, h� (cm)	0,83	(0,65 - 0,92)	1,5	(-2,3; 5,4)	6,8	5,3	18,8	14,6
SH, ve	0,79	(0,59 - 0,90)	-1,2	(-5,2; 2,8)	7,0	5,5	19,4	15,2
CH, h� (cm)	0,77	(0,54 - 0,89)	-14,3	(-29,0; 0,4)	25,7	6,7	71,2	18,7
CH, ve	0,76	(0,53 - 0,88)	-8,9	(-23,1; 5,2)	24,7	6,4	68,4	17,9
TH, h�(cm)	0,88	(0,75 - 0,94)	-1,2	(-11,8; 9,3)	18,4	4,4	51,0	12,3
TH, ve	0,70	(0,44 - 0,85)	-6,5	(-21,4; 8,4)	26,1	6,3	72,3	17,5
6m, h� (sek.)	0,79	(0,58 - 0,90)	-0,02	(-0,07; 0,03)	0,08	4,3	0,22	11,9
6m, ve	0,62	(0,32 - 0,81)	0,01	(-0,05; 0,08)	0,11	5,9	0,30	16,4

ICC= intraklasse korrelasjonskoeffisient, (3,1) two-way mixed, agreement, single measure, KI= konfidensintervall, \bar{d} = systematiske m lefeil, SEM= standardfeil til m lingene= $\sqrt{\text{residual mean square}}$, SEM % = SEM/grand mean x 100, SRD₉₅= minste reelle endring= $1,96 \times \sqrt{2} \times \text{SEM}$, SRD % = SRD/grand mean x 100.

Interrater reliabilitet

Interrater reliabiliteten for hinketestene er presentert i tabell 13. ICC varierte fra 0,72-0,91. De systematiske m lefeilene (\bar{d}) varierte fra -2,8-9,5 cm for distansehoppene og 0,03-0,06 sek. for hopp p  tid. Andelen \bar{d} var lavere enn SRD. Alle 95 % KI til \bar{d} g r igjennom null, unntagen for 6m h yre som viser til signifikant forskjell mellom testerne. De relative verdiene av SEM (SEM %) varierte fra 3,9 %-7,4 % for alle hinketestene.

Tabell 13: Relativ og absolutt interrater reliabilitet for hinketestene (n= 26).

n = 26	ICC (2,1)	95 % KI	\bar{d}	95 % KI \bar{d}	SEM	SEM %	SRD	SRD %
SH h� (cm)	0,87	(0,73 - 0,94)	-2,8	(-6,2; 0,6)	5,9	4,6	16,3	12,8
SH ve	0,84	(0,67 - 0,92)	-0,7	(-4,3; 2,9)	6,3	5,0	17,4	13,8
CH h� (cm)	0,73	(0,49 - 0,87)	9,5	(-6,5; 25,5)	28,0	7,4	77,6	20,5
CH ve	0,82	(0,65 - 0,92)	2,9	(-9,5; 15,2)	21,6	5,7	59,8	15,7
TH h� (cm) ^a	0,91	(0,80 - 0,96)	0,5	(-8,9; 9,9)	16,1	3,9	44,6	10,8
TH ve ^a	0,81	(0,62 - 0,91)	4,8	(-7,6; 17,2)	21,2	5,1	58,7	14,2
6m h�(sek.)	0,77	(0,49 - 0,90)	0,06	(0,02; 0,11)	0,08	4,2	0,22	11,8
6m ve	0,72	(0,47 - 0,87)	0,03	(-0,02; 0,09)	0,10	5,3	0,28	14,7

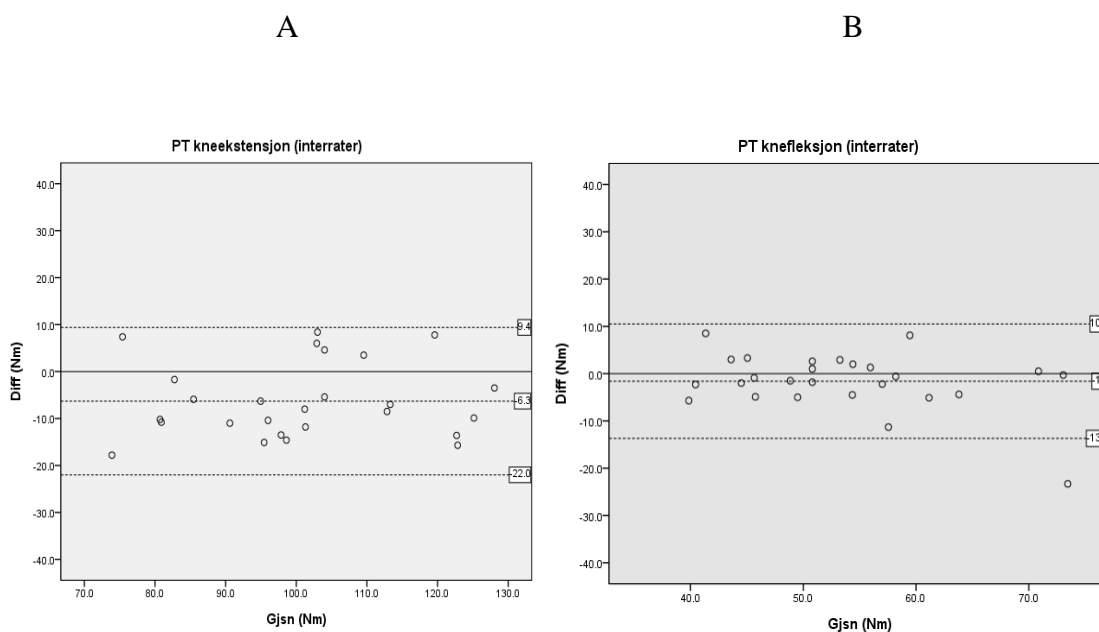
ICC= intraklasse korrelasjonskoeffisient, (2,1) two-way random, agreement, single measure, KI= konfidensintervall, \bar{d} = systematiske m lefeil, SEM= standardfeil til m lingene= $\sqrt{\text{residual mean square}}$, SEM % = SEM/grand mean x 100, SRD₉₅= minste reelle endring= $1,96 \times \sqrt{2} \times \text{SEM}$, SRD % = SRD/grand mean x 100, ^a = ett barn uten gyldig hopp p  TH h yre ben, n= 25 for alle analyser p  TH.

4.3.3 Bland – Altman plott

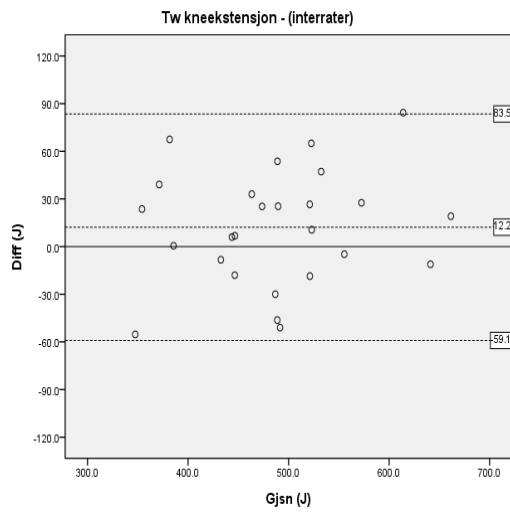
Data som er presentert i figur 7 (A-H) viser for 5 av 8 plott flere verdier under null-linjen enn over, noe som er i tråd med de negative \bar{d} verdiene. Det viser til at HM har registrert høyere verdier enn IE. Andelen av tilfeldige målefeil er større enn \bar{d} .

Heteroskedastisiteten til dataene ble undersøkt ved korrelasjonen, som varierte fra $r = -0,36$ ($p = 0,07$) (PTF) til $r = 0,42$ ($p = 0,03$) (6m). Styrken på korrelasjonen varierte fra liten til medium. Med korrelasjonen for 6m som signifikant. Ved visuell inspeksjon (i samsråd med statistiker) ble dataene allikevel ikke funnet å være heteroskedastiske.

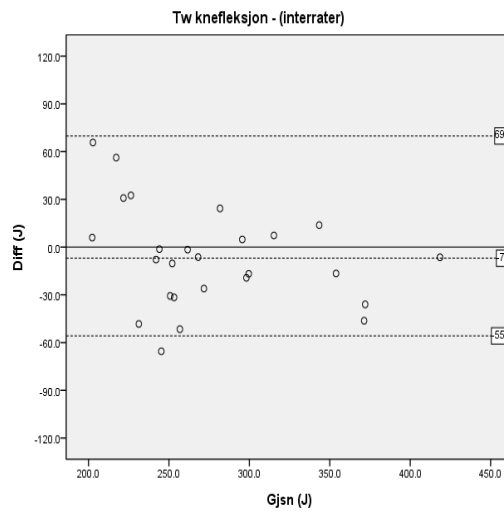
Utvalgstørrelsen var liten ($n = 26$), noe som kan være et bias for vurderingen. Det var få ”uteliggere” registrert. Et smalere 95 % LOA ble registrert for PTF i forhold til PTE, TwF kontra TwE og ved TH kontra CH.



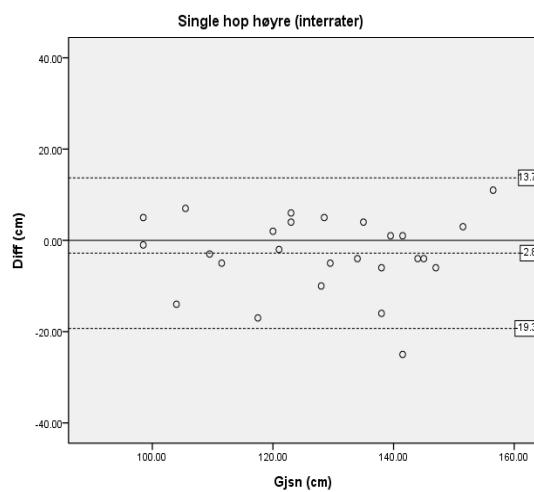
C



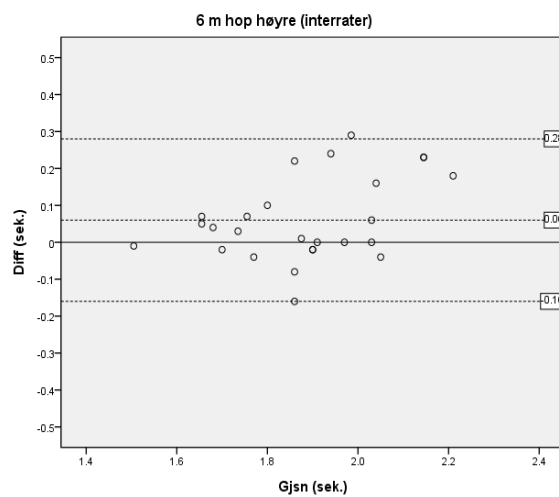
D

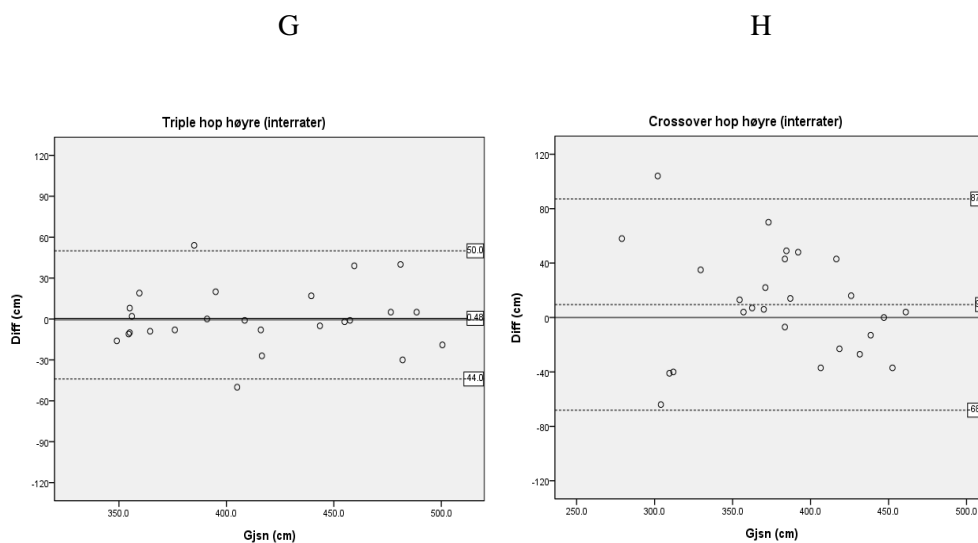


E



F





Figur 7: Bland-Altman plott med hensyn til interrater reliabilitet ($n=26$). PT kneekstensjon (A), PT knefleksjon (B), Tw kneekstensjon (C), Tw knefleksjon (D), SH (E), 6m (F) TH ($n=25$) (G) og CH (H).

4.3.4 Oppsummering av reliabilitet

Isokinetiske muskelstyrkemålinger og hinketester viste ved intra- og interrater moderat til høy relativ reliabilitet (ICC 0,62-0,94). Unntaket var TQ30 som viste lav til moderat reliabilitet (ICC 0,48-0,71). Høyre ben hadde høyere ICC verdier enn venstre ben, og kneekstensjon noe bedre enn knefleksjon. SEM % varierte fra 5,1-14,9 % for styrkevariablene og 3,8-7,4 % for hinketestene. PT og Tw fleksjon viste smalere 95 % LOA enn ekstensjon. CH viste størst 95 % LOA for hinketestene. Andelen av tilfeldige målefeil var større enn de systematiske for alle målinger.

4.4 Responsivness

For de isokinetiske muskelstyrkemålingene ble minste reelle endring (SRD %) registrert for TwE, PTE og AVG PTE med verdier fra 14,0-16,3 % for høyre og venstre ben. TQ30 viste tilsvarende SRD % fra 30-40 %. For hinketestene ble lavest SRD % registrert for TH høyre ben (10,8-12,3 %) og 6m høyre ben (11,8-11,9 %). SRD % varierte fra 14,0-41,3 % for alle styrkevariablene og fra 10,8-20,5 % for alle hinketestene. Det var små forskjeller mellom intra- og interratermålingene.

5. Diskusjon

I metodestudien undersøkte vi intra- og interrater reliabiliteten av hinketester og isokinetiske muskelstyrkemålinger på et utvalg av barn uten skade. Vi beregnet i tillegg et estimat på minste reelle endring til målemetodene ut fra reliabilitetsresultatene.

I diskusjonsdelen oppsummeres først hovedfunnene. Deretter diskuteres resultatene opp i mot funn fra andre reliabilitetsstudier på området. Under diskusjon av metode ser jeg på de ulike faktorene som kan ha påvirket resultatene. Til slutt diskuteres den kliniske verdien av resultatene våre.

5.1 Resultater

Hovedfunn

Vi fant moderat til høy enighet og samsvar mellom målingene gjort av de to testerne i studien vår (interrater), og mellom samme tester (intrarater) fordelt på to ulike testdager. Unntaket var for TQ30 som viste lav til moderat reliabilitet. Hinketestene viste noe bedre interrater enn intrarater reliabilitet. Kneekstensjon viste noe høyere ICC enn knefleksjon ved isokinetisk muskelstyrketesting. Høyre ben viste generelt noe høyere ICC verdier enn venstre ben for begge målemetodene.

Absolutt reliabilitet oppgitt som et mål på de individuelle variasjonene varierte mer for styrkemålingene enn for hinketestene, sett samlet for både intra- og interrater målingene. Lavest absolutt reliabilitet ble sett for PTE, TwE og AVG PTE for styrkemålingene. For de ulike hinketestene var den absolutte reliabiliteten ganske lik, men dårligst verdier ble registrert for CH.

Andelen av tilfeldige målefeil var mye større enn de systematiske målefeilene ved alle målinger. Det betyr at komponenter som biologisk variasjon (barn/testere) og/eller unøyaktigheter i forhold til testprotokoll og prosedyrer har spilt en større rolle enn for eksempel læring ved testingen. Alle gjennomsnittsverdier av LSI > 90 %, med unntak av TQ30, som indikerer en forholdet mellom høyre og venstre ben faller innenfor det som er foreslått som normalt (LSI \geq 85-90 %). Gjennomsnittsdifferansen til målingene varierte i prosent fra 0,3-3,7 % for hinketestene, noe som viser til små individuelle variasjoner for. Styrkemålingene varierte fra 0,4-18,3 %.

Vi registrerte lavest SRD % ved PTE, AVG PTE og TwE for styrkevariablene, hvor en endring $> 16,3$ % kan indikere en reell endring. For hinketestene viste TH og 6m lavest SRD % for høyre ben, hvor endring $> 11,8$ % kan indikere en reell endring.

5.1.1 Reliabilitet av isokinetisk muskelstyrkemålinger

Intrarater reliabilitet

Vi fant fra moderat til høy relativ intrarater reliabiliteten. Størrelsen på SEM verdiene i metodestudien anses som akseptabel sett i forhold til gjennomsnittsverdien for hver av de isokinetiske muskelstyrkemålingene. Vi fant signifikante forskjeller ved PTE, TQ30E, AVG PTE og TwE for venstre ben. Gjennomsnittssifferansene til de signifikante forskjellene (bias) var ≤ 5 Nm for torque-variablene og $\leq 25,6$ J Tw. Det tilsvarer i prosent fra 4,5-8,7 % for de aktuelle målingene. Forskjellen i SEM verdi for de signifikante målingene på venstre ben i forhold til høyre ben var fra 0,1-1,1 Nm for torque-variablene og 5,3 J for Tw. Vi kan ikke se bort i fra den systematiske målefeilen for venstre ben, men basert på differansene i absolutt og prosentverdi, virker resultatene for de isokinetiske muskelstyrkemålingene å være reliable.

Dersom vi hadde kjørt en pilottesting i forkant kunne vi muligens ha fanget opp de systematiske forskjellene og gjort endringer i testprotokoll og prosedyrer som kunne ha minimalisert effekten vi så på venstre ben. Ved intrarater testingen var det signifikante forskjeller både for høyre og venstre ben. Siden forskjellene dag 2 (intrarater) bare gjaldt for venstre ben kan vi anta at prestasjonen på høyre ben hadde stabilisert seg i motsetning til for venstre. Ved å ha innført prøverunde (tilvenning) i forkant av testdagene kan det hende også prestasjonene for venstre ben ville vært mer stabile.

For styrkevariablene så vi størst variasjon ved TQ30. Spesielt ved ekstensjon dag 1 ble det registrert stor differanse mellom målingene for høyre og venstre ben. Resultatene for en av testerne (HM) viste stor variasjon, mens for den andre testeren (IE) viste resultatene mye mindre variasjon. Vi kan derfor stille spørsmål om målingene var valide, siden barna tydelig forstod oppgaven/testen på et annet tidspunkt. Barna som viste stor variabilitet var en blanding av fotball- og alpinbarn, hvorav 5 av 7 ble testet først av HM. Av barna var 4 av 7 randomisert med venstre som startben. Vi kan ikke utelukke at det har forekommet en læringseffekt for noen av disse barna, som tydelig har testet med mindre variasjon gang nummer to (tester IE) og dag 2 (tester HM).

Våre ICC verdier var lavere enn hva Kellis et al. (1999) rapporterte i sin studie på gutter ($13,0 \pm 0,4$ år), hvor ICC (1,1) varierte fra 0,90-0,98 for konsentriske målinger ved $60^\circ/\text{sek}$. To andre studier på barn (Deighan et al., 2003; Merlini et al., 1995) oppga også ICC verdier, men en direkte sammenligning er vanskelig da testhastighet ($30\text{-}100\text{-}180^\circ/\text{sek}$.) og aldersgruppe (6-10 år) er ulik vår studie. Studiene viste imidlertid fra moderat til høy relativ reliabilitet (ICC 0,76-0,95) for konsentriske målinger. Kellis et al. (1999) rapporterte noe lavere ICC verdier for eksentriske målinger, i forhold til konsentrisk, mens Deighan et al. (2003) ikke fant noen spesiell forskjell mellom testmetodene. Tre studier (Deighan et al., 2003; Kellis et al., 1999; Merlini et al., 1995) rapporterte høyere ICC for kneekstensjon enn knefleksjon. Det stemmer også overens med funn fra vår studie. Kellis et al. (1999) oppga et mye smalere 95 % LOA (konsentrisk, $60^\circ/\text{sek}$.) for PT enn vi fant i vår studie, men de fant i likhet med oss et bredere 95 % LOA for PTE enn for PTF. To av studiene på voksne angir de samme resultatene med bredere 95 % LOA for PTE (Hartmann et al., 2009; Sole et al., 2007).

Ingen av de tidligere studiene på barn har oppgitt SEM verdier, men i stedet brukt CV (Deighan et al., 2003; Merlini et al., 1995; Mohtadi et al., 1990). I studiene varierte CV for de konsentriske målingene fra 4,0 – 12,0 % for PT. Våre SEM % verdier varierte tilsvarende for PT fra 5,2-8,9 %. Til sammenligning har studiene på voksne rapportert CV og SEM % verdier fra 2,8-7,0 % (Maffioletti et al., 2007; Sole et al., 2007), mens studien på eldre (Hartmann et al., 2009) angir SEM % for PT fra 8,0-12,8 %.

Vi undersøkte i tillegg reliabiliteten til styrkevariablene Tw og AVG PT. Våre resultater er lik hva Sole et al. (2007) oppga som SEM % for Tw, mens for AVG PT hadde vi mye lavere SEM % verdier enn Hartmann et al. (2009).

Resultatene våre på absolutt reliabilitet er sammenlignbare med de individuelle variasjonene som er vist i de andre barne- og voksenstudiene. Vi hadde et bredt 95 % LOA ved PT enn hva Kellis et al. (1999) anga for sitt utvalg, men Hartmann et al. (2009) som testet eldre viste til større individuelle variasjoner. SEM verdiene våre er også i henhold til de andre studiene, i den grad vi kan sammenligne våre målinger med studien til Sole et al. (2007) som testet et utvalg unge voksne (gj.sn. 21 år) og Hartmann et al. (2009) som testet eldre ($71,2 \pm 5,5$ år). En CV på under 6 % er foreslått å demonstrere god reliabilitet av de isokinetiske muskelstyrkemålingene på den voksne populasjonen (Deighan et al., 2003). I en annen studie blir det nevnt at en CV opp mot

13 % er forventet når man undersøker friske individer over flere ganger (Ayalon et al., 2000). Våre SEM % varierte fra 5,1-14,5 % for alle målinger, med en gjennomsnittlig CV på 6,6-10,5 %.

Gjennomsnittsmålingene for utvalget vårt for PT er i tråd med hva Kellis et al. (1999) fant på sitt utvalg av aktive fotballgutter (n = 13). Deres utvalg var relativt lik metodebarna med tanke på antropometriske mål og aktivitetsbeskrivelse. De rapporterer gjennomsnittsverdier for PT (konsentrisk, 60°/sek.) ved test-retest (begge ben) fra 98,5-100,9 Nm for PTE og 61,1-66,8 Nm for PTF. Våre gjennomsnittsverdier lå på 101,9-106,7 Nm for PTE og 52,5-55,1 Nm for PTF. Kellis et al. (1999) hadde et mindre SD til gjennomsnittsmålingene enn vi registrerte i studien vår. Deres utvalg var således en mer homogen gruppe enn metodebarna med tanke på prestasjon. Til forskjell fra oss ga de en prøverunde før selve testdagen i tillegg til å bruke visuell feedback under testingen. De hadde også lagt inn 10 min. pause mellom testing av hvert ben.

Det finnes normative data på isokinetisk muskelstyrkemålinger for barn (Holm et al., 2008; Wiggin et al., 2006). Wiggins et al. (2006) kritiserer selv i sin studie at de ikke korrigerer for tyngdekraften. De kan dermed ha over-/underestimert PT. En norsk studie testet 376 skolebarn fra Oslo-området i alderen 7-12 år (Holm et al., 2008).

Konsentriske målinger av PT målt ved 60°/sek. (Cybex 6000 dynamometer) ble oppgitt for hvert alderstrinn som absolutt (Nm) og relative verdier (PT/kroppsvikt x 100). Absoluttverdier for aldersgruppen 12 år (gutter/jenter) var 99/97 Nm for PTE og 52/56 Nm for PTF. Gjennomsnittsverdiene kan sammenlignes med våre, men vi testet et mer selektert utvalg av barn og hadde derfor mindre SD enn i normalmaterielet.

Interrater reliabilitet

Estimatene for interrater reliabilitet er i overenstemmelse med våre resultater for intrarater reliabiliteten. Signifikante forskjeller mellom målingene til de to testerne ble observert for både høyre og venstre ben. Dette var ikke et resultat av en ren læringseffekt, siden data for hver tester bestod av en blanding av første- og andregangsresultater. Gjennomsnittlig differanse til de signifikante forskjellene var $\leq 11,5$ Nm for torque-variablene, hvor TQ30 stod for den høyeste differansen, og $\leq 19,8$ J for Tw. Det tilsvarer i prosent fra 2,5-18,5 % for de aktuelle målingene.

Differanseverdiene er noe større for torque-variablene enn registrert ved intrarater

reliabiliteten. Vi kan igjen spørre oss om for eksempel en pilottesting for testerne, og tilvenning for barna, hadde resultert i mindre systematiske målefeil, og bedre reliabiliteten.

To studier har undersøkt interrater reliabilitet på voksne (Hartmann et al., 2009; Keskula et al., 1995). Begge rapporterte høye ICC verdier fra 0,92-0,96 for PT. Keskula et al. (1995) hadde fire testere til å teste et utvalg unge voksne ($n = 16$). Hartmann et al. (2009) brukte to testere for å teste et utvalg av eldre ($n = 24$). Våre SEM verdier er relativt lik de Hartmann et al. (2009) oppga for PT, men vi registrerte lavere SEM verdier for AVG PT og et smalere 95 % LOA. Vi beregnet AVG PT fra gjennomsnittet til fem målinger mens de regnet ut i fra tre målinger, noe som kan spille inn på variasjonen til målingene.

Vi hadde lavere ICC verdier, og høyere SEM verdiene for Tw enn hva Keskula et al. (1995) fant. Dette viser til større individuell variasjon hos barna våre enn utvalget de testet som var unge voksne ($25,4 \pm 4,5$ år). Resultatene til begge interraterstudiene er basert på målinger av dominant ben. Våre resultater er mest sammenlignbare med utvalget av eldre som Hartmann et al. (2009) testet. Både gruppen av barn, som vi testet, og eldre kan tenkes å variere mer i prestasjon enn det unge/voksne gjør (Baumgartner, 2003).

To studier på barn (Deighan et al., 2003; Kellis et al., 1999), og studien på eldre (Hartmann et al., 2009) tillot en prøverunde før selve testdagen for å minimere læringseffekten. Sistnevnte studie fant noen få, men signifikante forskjeller ved intra- og interrater målingene. Kellis et al. (1999) fant bare signifikante forskjeller mellom test-retest for det ikke-dominante (non-preferred leg). På bakgrunn av de systematiske målefeilene vi registrerte både ved intra- og interrater målingene kunne det ha vært interessant og sett utfallet av en prøverunde.

Oppsummering

Vi fant i studien vår moderat til høy reliabilitet for de isokinetiske muskelstyrkemålingene. Våre ICC verdier noe lavere enn hva andre studier rapportert. SEM/SEM % verdiene våre er mest sammenlignbare med utvalget av barn i studien til Kellis et al. (1999) og Hartmann et al. (2009) som testet eldre. TQ30 viste dårligst ICC

og størst SEM av styrkevariablene vi testet. Vi har ingen studier å sammenligne graden av reliabiliteten for TQ30 med. Vi fant ingen store forskjeller i relative eller absolutte reliabilitet mellom intra- eller interrater målingene. Hartmann et al. (2009) viser til noe bedre inter- enn intrarater reliabilitet. Våre absoluttverdier for PT er sammenlignbare med normalmaterial (Holm & Vollestad, 2008). Kneekstensjon viste noe høyere ICC verdier enn knefleksjon, men samtidig ble det registrert lavere absolutte reliabilitet (SEM og LOA) for fleksjon. Vi fant både ved intra- og interrater målingene systematiske forskjeller. Selv om de absolutte differansene til målingene var små, og med lave SEM verdier, er dette allikevel et viktig funn å ta hensyn til videre i en eventuell modifisering av protokoll og testprosedyrer for barn.

5.1.2 Reliabilitet av hinketestene

Intrarater reliabilitet

Den relative reliabiliteten for hinketestene varierte også fra moderat til høy. Våre ICC verdier er allikevel dårligere enn hva andre studier har rapportert for de samme fire hinketestene. Hinketestene har vist høy relativ reliabilitet (ICC 0,80-0,97) testet på voksne, med unntak av 6m som også for dette utvalget viser tendens til lavere reliabilitet (ICC 0,60- 0,77) (Bolglia & Keskula, 1997; Booher, Hench, Worrell, & Stikeleather, 1993; Munro & Herrington, 2011). Ross et al. (2002) fant høy reliabilitet for alle hink (inkludert 6m) med ICC (2,3) > 0,92. Reid et al. (2007) fant tilsvarende testet på et utvalg etter ACL rekonstruksjon (ACLR) med ICC (2,1) fra 0,82-0,92 (basert på LSI).

Størrelsen på SEM vi fant i metodestudien anses som akseptabel med referanse til gjennomsnittsverdien for hver av hinketestene. For 6m er SEM resultatene for metodebarna (0,08-0,11 sek.) sammenlignbare med resultatene på voksne (0,06-0,13) (Bolglia & Keskula, 1997; Munro & Herrington, 2011; Ross et al., 2002). Reid et al. (2007) oppga SEM % på 5,6 % for 6m, hvor vi fant 4,3 % og 5,9 % for henholdsvis høyre og venstre ben. Vi fant absolutte SEM verdier for SH på 6,8 og 7,0 cm (høyre og venstre ben). Dette er noe høyere enn hva andre studier har rapportert, hvor SEM har varierer i fra 4,6-5,9 cm (Bolglia & Keskula, 1997; Booher et al., 1993; Ross et al., 2002). Munro et al. (2011) angir SEM på 7,9 og 7,9 cm (menn og kvinner) som er mer i tråd med våre resultater. Det kan imidlertid være vanskelig å diskutere størrelsen av

våre funn opp i mot studiene på voksne som er utført på utvalg med gjennomsnittsalder ± 20 år. SEM er til en viss grad også avhengig av størrelsen på målingene (Hopkins, 2000). Ross et al. (2002) testet i tillegg et utvalg av veldig aktive kadetter. Det er derfor å forvente at barna våre viser større individuell variasjon enn det som er funnet på de andre utvalgene.

Vi fant størst variasjon for CH som hadde høyest SEM % og det 95 % LOA med størst bredde. Dette er i overensstemmelse med Munro et al. (2011) som også oppga høyest SEM for CH i forhold til gjennomsnittsdistanse som ble hoppet. Reid et al. (2007) rapporterer høyest SEM % for 6m, men også stor variasjon ved CH testet på ACLR pasienter.

Det finnes ingen normative data for hinketestene for aldersgruppen vi testet. Det nærmeste er data for SH basert på målinger av 2508 high-school elever fra USA (gj.snitts alder 14,5 år) (De Carlo & Sell, 1997). Guttene hoppet i gjennomsnitt 154 og 155 cm, og jentene hoppet 123 cm og 120 cm på henholdsvis høyre og venstre ben. Metodebarna hoppet fra 126-130 cm ved SH. Våre gjennomsnittsverdier og SD var sammenlignbare med resultatene til jentene i normmaterialet.

Vi så tegn til at målingene for dag 2 var høyere enn ved dag 1, selv om ingen av disse endringene var signifikante. Det er også rapportert i andre studier at verdiene har steget fra måling til måling (Bolglia & Keskula, 1997; Munro & Herrington, 2011; Ross et al., 2002). Det er usikkerhet i litteraturen rundt hva som er riktig antall prøvoforsøk å inkludere for å minimalisere læringseffekten for hinketestene (Bolglia & Keskula, 1997; Booher et al., 1993; Ross et al., 2002). En studie (Booher et al., 1993) rapporterte signifikante forskjeller mellom målinger fra samme dag, og mellom to testdager (48 timers intervall), ved å gi ett prøvoforsøk og to testhopp (1+2). Bolglia & Keskula (1997) økte antall prøvoforsøk og antall testhopp (3+3) og viste bare signifikante forskjeller mellom 1. og 6. hopp for SH. Ross et al. (2002) hadde samme antall forsøk som (3+3) Bolglia & Keskula. De viste til en bedring av score, men ingen signifikant forskjell fra gang til gang. Munro et al. (2011) målte 6 hopp og fant at 3 hopp var nok for å stabilisere score ved SH og TH, men CH krevde 4 hopp før stabilisering og 6m 3 og 4 hopp (menn/kvinner). Alle studiene viser en økning i score fra forsøk til forsøk, og i så måte en indikasjon på læring underveis, om enn ikke statistisk signifikant. Reid et

al. (2007) brukte også (1+2) forsøk ved testing av ACL pasienter etter rekonstruksjon. De rapporterer signifikante forskjeller mellom 1. og 2.testrunde (48 timers intervall) for kontroll-benet, men ikke mellom 2.- 4. testrunde (6 uker).

Vi ga ett prøveforsøk og deretter inntil 5 forsøk på å klare to gyldige hopp som ble notert. Barna i metodestudien brukte i snitt 3,4 forsøk på SH, 4,0 på CH og 3,8 på TH og 2,2 på 6m. En kritikk til registreringen av hopp i metodestudien kan være at vi ikke konsekvent ga alle barna fem testhopp. Ved å øke antall noterte hopp kunne vi, lik Munro et al. (2011), rapportert når prestasjonen stabiliserte seg. Det kunne ha økt variabiliteten mellom individene våre og dermed resultert i høyere ICC verdier. En motsatt effekt ville være dersom andelen av tilfeldige målefeil (biologiske variasjoner) økte, enten hos barna eller testerne, og dermed reduserte ICC verdiene. En protokoll som inkluderer fem testhopp på hvert ben kan i praksis kanskje være noe anstrengende for pasienter etter ACL skade.

Interrater reliabilitet

Estimatene for interrater reliabiliteten er i overenstemmelse med våre resultater for intrarater reliabilitet. ICC var også her fra moderat til høy. Interrater resultatene viser noe høyere relativ reliabilitet, mens de individuelle variasjonene (SEM/SEM%) var noe høyere ved intrarater reliabiliteten. Vi har ikke funnet studier som har undersøkt interrater reliabilitet for hinketestene som vi kan sammenligne våre resultater med. Det var ingen signifikante forskjeller i målingene til de to testerne, med unntak for 6m hopp på høyre ben. SEM på 0,08 sek. antyder allikevel at variasjonene vil forekomme innenfor et akseptabelt tidsintervall.

Oppsummering

Hinketestene viste fra moderat til høy reliabilitet testet på et utvalg av barn. Det var ingen store forskjeller i relative eller absolutte reliabilitet mellom intra- eller interrater. Størrelsen på SEM verdiene vi fant i metodestudien var akseptabel, sett i forhold til gjennomsnittsverdiene registrert for hver av hinketestene. Vi har ingen reliabilitetsstudier på barn å sammenligne resultatene våre med. Det er derfor vanskelig å få en klar indikasjon på hvor bra eller dårlige våre SEM verdier er. Vi fant lavest SEM for TH høyre og 6m høyre ben. Vi fant ingen systematiske forskjeller mellom målingene, med

unntak av 6m for høyre ben (interrater). CH viste størst variasjon representert ved høyest SEM og bredest 95 % LOA.

5.1.3 Responsivness til målemetodene

Isokinetiske muskelstyrkemålinger

I metodestudien varierte SRD % fra 14,0-40,3 % for alle styrkevariablene . For torque-variablene (PT, TQ30 og AVG PT) registrerte vi laveste SRD for AVG PTE (14,3-14,6 %) som tilsvarte 13,3 Nm for begge ben (interrater). Hartmann et al. (2009) oppga SRD på 10,0 Nm (interrater) for AVG PT, men de anga en lavere absolutt gjennomsnittsverdi enn barna i metodestudien. For TQ30E fant vi en SRD % fra 30,6-40,2 %. Det betyr at for å vurdere om måleforskjellen representerer reell endring så må en re-test overstige 30-40 %, eller >20 Nm. Det er ganske mye når den gjennomsnittlige målingen av TQ30E var på rundt 65 Nm. Våre SRD % verdier for PTE (14,4-14,9%) var lavere enn rapportert i studien til Sole et al. (2007), men vi hadde høyere verdier for PTF. Minste reelle endring for Tw var lik for begge studier, hvor vi fant SRD % på 15 % for TwE og 25 % TwF.

Hinketestene

I studien vår varierte minste reelle endring i prosent fra 10,8 %-20,5 %. Reid et al. (2007) oppga minste reelle endring fra 8,1 %-13,0 %. Vi fant for SH (høyre ben) en SRD på 12,8 % (interrater). Det betyr at forskjeller mindre enn 13 % mest sannsynlig ikke representerer en reell endring for vårt utvalg. For at måleforskjellen skal representere en reell endring må den falle utenfor 95 % KI til SRD, som tilsvarer et 95 % LOA. For SH høyre ben er LOA fra -19,3-13,7 cm (se Bland-Altman plot fig. 7E). Det betyr at dersom et barn reduserer hoppdistansen sin med mindre enn 19,3 cm, eller forbedrer lengden med mindre enn 13,7 cm så vil forskjellen ikke vurderes som en reell endring i hoppdistanse. Forskjellen i distanse må altså falle utenfor LOA, enten i negativ eller positiv retning. Vi ser da at jo smalere intervallet av 95 % LOA er, jo lettere er det å fange opp en potensiell reell endring i funksjon. Som tidligere nevnt er størrelsen på dette intervallet avhengig av størrelsen på SEM/SD som viser til spredningen til målingene.

Munro et al. (2010) oppga minste reelle endring hvor verdiene for menn og kvinner var 22,0 og 21,8 cm for SH. Vi fant tilsvarende SRD (intrarater) for høyre og venstre ben på 18,8 og 19,4 cm for SH. Det kan være vanskelig å sammenligne resultatene direkte da Munro et al. (2011) korrigerer for benlengde i sine resultater, noe ikke vi gjorde.

Oppsummering

Resultatene våre viser at i gjennomsnitt for styrkevariablene PT og AVG PT vil en forskjell større enn 15 % (13 Nm) kunne representere en reell endring. For Tw må måleforskjellen registrert være større enn 15-25 %, eller over 70 J. Våre resultater fra kan sammenlignes med to andre studier på voksne/eldre (Hartmann et al., 2009; Sole et al., 2007). Hinketestene hadde lavere SRD % (10,8-20,5 %) totalt enn hva vi fant for styrkevariablene (14,0-40,3 %). Det betyr generelt at en mindre prosentvis forskjell kan representere en reell endring for hinketestene, enn for styrkevariablene.

5.2 Metode

Jeg vil her diskutere de metodiske forholdene som kan ha påvirket resultatene i denne studien relatert til utvalg, testere, studiedesign, datainnsamling, målemetoder og valg av statistiske analyser.

5.2.1 Utvalget

Barna i studien ble inkludert på bakgrunn av frivillig påmelding. De kom fra et begrenset område i Oslo. Alle var meget fysisk aktive (79 % trente 4-5 ganger i uken) og født det samme året (1998). Barna var homogene med tanke på antropometriske variabler. På bakgrunn av at vi hadde et lite utvalg og en homogen gruppe så er ikke resultatene generaliserbare til populasjonen av barn generelt. Resultatene kan først og fremst generaliseres til et lignende utvalg av barn.

Vi inkluderte, som tidligere nevnt, alle de interesserte barna ($n = 29$). Det er mulig vi kunne vært strengere i forhold til de gitte eksklusjonskriteriene på skade. Ved nærmere undersøkelse var det ingen av barna som falt inn under vår definisjon av skade, men 9 av 28 oppga likevel plager. Kanskje burde vi heller rettet spørsmålene mer mot fysiske plager, uavhengig av hvordan det påvirket tid borte fra trening/kamp, eller behov for medisinsk tilsyn (Fuller et al., 2006). Inklusjon av barn med eventuelle plager kan ha påvirket andelen av målefeil negativt i denne studien.

Maksimal funksjon (score på 10) ved inklusjon ble rapportert å være lavere for venstre ben (71 % mot 79 %) enn for høyre (n = 28). Etter at to barn ble ekskludert på grunn av smerte ved testing ble forholdet at 77 % hadde maksimal funksjon på venstre ben mot 81 % på høyre (n = 26). Det er usikkert om disse små forskjellene i funksjon kan ha medvirket til forskjellene vi så i reliabilitet mellom høyre og venstre ben.

Metodebarna var en homogen gruppe med tanke på antropometri og aktivitet, men det var likevel en stor spredning i målinger. Dette er i tråd med resultater fra tidligere normative studier på barn (Holm et al., 2008; Holm et al., 2009) som viser til stor spredning innen en aldersgruppe. Det kan være en kritikk at vi ikke satte noen grense/”cut off” for hvor stor individuell målevariasjon vi tillot før de ble ekskludert fra analysene, eller før vi eventuelt lot barna testes om igjen. På et så lite utvalg vil ”uteliggere” kunne ha stor betydning på gjennomsnittsverdien og bidra til systematisk eller tilfeldige målefeil som tilsammen vil kunne påvirke reliabiliteten til målingene.

Vi ga ikke noe aktivitetsbegrensninger til barna mellom testdagene. Det kan ha vært en konfunderende faktor til prestasjon, da vi for eksempel vet at noen av barna hadde spilt kamper og vært på treningssamling helgen før første/andre test. Det kan ha påvirket de to testdagene forskjellig ved at barna var mer/mindre sliten, opplagt, motivert, støl i muskulatur osv. Vi oppfordret imidlertid barn og foreldre til å melde fra dersom skadestatus hadde endret seg underveis.

5.2.2 Testere

De to fysioterapeutene som testet i denne studien ble valgt ut som representanter fra et større miljø av forskere og klinikere ved NAR/Hjelp24 NIMI Ullevål. Testing av pasienter og friske individer er en del av den daglige praksisen ved denne institusjonen. Vi valgte å bruke to testere med relativt lik erfaring (praktisk og teoretisk) i metodestudien. Lik erfaring er likevel ingen garanti for at testerne vil vurdere kvalitet og kvantitet av variabelen som skal måles helt likt. Intrarater reliabilitet for hver tester burde ideelt sett etableres før man sammenligner testere (Portney & Watkins, 2000). Vi fulgte en standardisert protokoll og prosedyrer ved testing, men kunne likevel ikke sikre oss at begge testerne opptrådte helt likt. På tross av standardisert verbal informasjon, rutiner for innstillinger og sikring i Biodex dynamometeret, presentasjon og godkjenning av hinketester osv., kan det være andre faktorer som fremtoning,

formidling og humør/motivasjon som kan ha påvirket interaksjonen mellom tester og barn.

Ideelt sett foregår interrater testing slik at begge testerne observerer/tester barna samtidig. For mange variabler som krever interaksjon fra testeren, er ikke dette mulig å gjennomføre. Med denne type målinger kan interrater reliabiliteten bli påvirket av at den "sanne responsen", for eksempel muskelstyrke, forandrer seg underveis fordi forsøkspersonene blir slitne (Portney & Watkins, 2000).

Kritikk til studien er at vi ikke inkluderte en pilottesting i forkant for å undersøke intrarater reliabilitet for hver av testerne. Vi kunne da fått en god indikasjon på hvor stabile testerne var i sine målinger, og oppdaget eventuelle feil og mangler ved protokoll og prosedyrer. På bakgrunn av den høye kompetansen til testerne vi brukte kan det sikkert diskuteres om de kan regnes som "tilfeldig" valgte. Vi tror allikevel at resultatene våre kan være generaliserbare gir en indikasjon på interrater reliabilitet når man bruker to testere med likt kunnskaps- og erfaringsnivå.

5.2.3 Studiedesign

Reliabilitet er som tidligere nevnt en forutsetning, men ingen garanti for at en målemetode er valid (Portney & Watkins, 2000; Batterham & George, 2003). Første steg er derfor å dokumentere reliabiliteten til en målemetode. I vår praksis er ofte flere testere involvert når ACL-barna i kohortstudien kommer til testing. Det var derfor naturlig for oss å undersøke både intra- og interrater reliabiliteten.

Design av metodestudien, og logistikk og gjennomføring av testingen gjenspeiler "praksis som den er" i en testsituasjon for ACL-barna i kohortstudien. Vi kan beskrive designen vår som en *delvis-standardisert* tilnærming (Domholdt, 2005). En *delvis-standardisert* tilnærming beskriver reliabiliteten ved et moderat nivå av standardisering som er oppnåelig i en klinisk setting. En kritikk mot vår design er at vi kunne valgt en "strengere" utforming som kunne gitt oss et bedre estimat av reliabiliteten under ideelle forhold. En slik design ville ha inkludert forsøk på å korrigere for flere av faktorene som kan ha påvirket resultatene våre. Vi kunne da bestemt øvre grense av reliabilitet for en komponent og hatt mer grunnlag for å separere "målestøy" fra variabiliteten hos forsøkspersonene (Domholdt, 2005). I mange metodestudier blir mer enn en komponent

av reliabilitet undersøkt, eller er flettet sammen slik at de ikke klart kan skilles fra hverandre (Domholdt, 2005). Vi så på både intra-, interrater og intrasubjekt reliabilitet. Disse komponentene av reliabilitet kan i mitt design ikke fullstendig separeres fra hverandre. For eksempel vil dag-til-dag reliabiliteten involvere variasjon i målinger som er relatert både til tester og barn.

Det er få studier som har undersøkt interrater reliabilitet som vi kan sammenligne design med. Vi fant to studier på isokinetisk muskelstyrkemålinger (Hartmann et al., 2009; Keskula et al., 1995). I studien til Keskula et al. (1995) brukte de fire testere som testet alle forsøkspersonene, mens Hartmann et al. (2009) brukte to testere, og tre testrunder fordelt på to testdager, slik som oss. Ofte blir det valgt å bruke en tester på grunn av ressurser, og/eller for å eliminere bort interrater variabiliteten (Sole et al., 2007; Maffiuletti et al., 2007).

5.2.4 Datainnsamling og testprosedyrer

Testrekkefølge

Barna ble testet i isokinetisk muskelstyrkemålinger og gikk deretter direkte opp for å utføre hinketestene. Prosedyren ble fulgt for hver av de tre testrundene. Det kan være at rekkefølgen i testing har påvirket prestasjonen ved begge målemetoder. Den store variasjonen vi så ved isokinetisk muskelstyrkemålinger kan ha sammenheng med for lite oppvarming før barna skulle prestere maksimalt, enten ved at den standardiserte oppvarmingen på sykkel (5 min.) ikke var tilstrekkelig, eller at oppvarmings-repetisjonene i Biodex maskinen var for få.

Ved testing etter ACL skade er det anbefalt å teste styrken før hinketestene, for å undersøke styrkeforholdet mellom skadet og kontrollben, og om skadet ben genererer nok styrke til å anbefale videre hinketesting (Brosky, Jr. et al., 1999). For hinketestene kan liten restitusjonstid/pause etter styrkemålingene ha påvirket evnen til å konsentreres seg om å hoppe langt/fort, eller ført til flere forsøk for å oppnå et gyldig hopp. Ross et al. (2002) som både testet isokinetisk muskelstyrke og hinketester oppga 5 min. pause mellom de to målemetodene. Det at hinketestene generelt ga mer stabile målinger kan også være et resultat av at barna hadde vent seg til å være i en testsituasjon og hadde ”varmet opp” med styrketesting og dermed greide å prestere jevnere på hinketestene. Eventuelt har graden av variasjon mer med kompleksiteten til hver målemetode å gjøre.

Som følge av hvordan designen på studien var lagt opp, bestod resultatene/rådata til testerne (IE og HM) fra dag 1 av en blanding av første- og andregangsresultater (t1 og t2) for alle barna. Det betyr at en eventuell motivasjons-, tretthet, og/eller læringseffekt for barna er fordelt likt på begge testerne. De systematiske forskjellene vi ser kan derfor i større grad tilskrives forskjeller mellom testerne selv.

Testintervall

I studien ble pausen mellom testrundene (første dag) ulik for barna. Pausen varierte fra 30-60 min. Ulik restitusjonstid for de som fikk lengst pause (60 min.), og de som fikk kortest pause (minimum 30 min.) kan ha påvirket den andre testrunden. Lund et al. (2005) viste til høyest reliabilitet ved et kort intervall mellom testing (20 min.) for isokinetiske muskelstyrkemålinger ved 60°/sek. (n = 13, gj.sn. alder 32 år). De testet imidlertid kun isokinetisk muskelstyrke, og ikke hinketester. Det er mulig at kombinasjonen av de to målemetodene krever en lenger restitusjonstid mellom testingene.

Ved testing av interrater reliabiliteten lå antallet effektive timer testing på rundt 3,5-4,0 timer for hver av testerne, avhengig av størrelsen på gruppen av barn. Hver test tok ca. 10-15 min. å gjennomføre. Testerne hadde to pauser i løpet av den første testdagen. Et eksempel på testmengde kunne være at åtte barn kom til testing. Hver tester måtte da instruere i isokinetiske muskelstyrkemålinger åtte ganger og demonstrerte de fire ulike hinketestene åtte ganger. Systematiske målefeil kan ha forekommet underveis dersom testerne ble mer eller mindre kritiske til det de observerte/målte (Beyer & Magnusson, 2003; Thomas, Silverman, & Nelson, 2005). Det er en fare også i vår studie at testerne ble lei underveis, og dermed ”snillere” eller ”strengere” i sin bedømmelse og nøyaktighet. Vi testet ettermiddag og kveld etter at begge testerne (IE og HM) hadde jobbet i forkant. Mange av barna kom også rett fra skolen. Både testere og barn kan dermed ha blitt fortere slitne og ukonsentrerte enn om vi hadde foretatt testingen tidligere på dagen.

Vi valgte å ha en ukes intervall mellom de to testdagene. Alle barna ble testet på samme tid begge dager siden det er angitt at fysisk prestasjon kan være påvirket av tid på døgnet (Brosky, Jr. et al., 1999). En studie som testet både isokinetisk muskelstyrke og hinketester oppga testintervall på fem dager (Ross et al., 2002). Studiene som bare har

testet isokinetisk muskelstyrke rapporterte en ukes intervall mellom testing, og for hinketestene ble det rapportert fra 24-48 timer (Reid et al., 2007) til 4 ukers intervall (Ross et al., 2002).

Flere studier har rapportert minst 30 sek. pause mellom hvert hink (Bolglia & Keskula, 1997; Booher et al., 1993; Ross et al., 2002) og opptil 2 min. mellom hver hinketest (Munro & Herrington, 2011; Reid et al., 2007). Vi hadde i protokollen vår liten pause både mellom hvert hink og hver hinketest. Vi kan spørre oss om dette kan ha påvirket prestasjonen slik at barna underveis i testingen ble slitne og dermed presterte dårligere (kortere distanse/lengre tid) og/eller mer variabelt.

Randomisering

Det eksisterer usikkerhet i litteraturen rundt definisjon og effekten av dominant ben (Sadeghi, Allard, Prince, & Labelle, 2000; Clark, 2001). Det er også diskutert om bendominans er oppgavespesifikk, og i tilfelle hvilket ben som kan regnes som dominant for hvilken oppgave (Holm et al., 2009; Parker, Larkin, & Ackland, 1993). På bakgrunn av dette valgte vi å ikke ta hensyn til dominans i denne studien, men testet både høyre og venstre ben i randomisert rekkefølge.

5.2.5 Målemetodene

Vi så at de isokinetiske muskelstyrkemålingene var forbundet med en større andel systematiske forskjeller enn hinketestene. Vi fant en gjennomsnittlig CV fra 6,6-10,5 %, men det var store individuelle forskjeller blant barna (min.-maks.: 1,5-28,5 %). Det kan tenkes at styrketesting ble oppfattet som vanskeligere av barna og var en mer ukjent oppgave å utføre enn hinketestene. Vi har også nevnt tidligere at barns evne til å utvikle maksimal kraft i større grad kan avhenge av læring/tilvenning i form av adaptasjon av nevralt komponenter. En tilvenningsrunde før testing som inneholder både submaksimal og maksimal kontraksjoner er blitt anbefalt (Jones & Stratton, 2000).

For liten tid til tilvenning kan ha gjort det vanskeligere for barna å prestere på et jevnt nivå i en kompleks test som ved maksimale isokinetisk muskelstyrkemålinger. Samtidig er bruken av en slik prøverunde ikke validert i litteraturen, og i daglig praksis kan det være vanskelig å gjennomføre på grunn av tid og ressurser. En slik prøverunde er ikke vanlig å gi ACL-barna i kohortstudien, og som Ross et al. (2002) påpeker i sin studie på

hinketestene, ville vi også undersøke i en reell setting slik testingen foregår hos oss i praksis.

Vi utførte korreksjon for tyngdekraften i henhold til protokollen til Biodex dynamometeret, hvor benet veies statisk ved 10 grader fleksjon. Dette utføres for hvert ben så verdiene kan være forskjellige for høyre og venstre. Det er blitt stilt spørsmål ved bruk av denne metoden for barn, siden barn kan finne det vanskeligere å slappe helt av (Jones & Stratton, 2000). Det vil kunne påvirke nøyaktigheten ved korreksjonen, hvor vi kan komme til å underestimere torque ved kneekstensjon og overestimere for knefleksjon. Den manuelle korreksjonen bidrar også til en variasjonskomponent mellom testerne.

Hinketestene viste mer stabile målinger. Dette var også en mer kjent oppgave å utføre for barna. Vi så tegn til at målingene bedret seg fra gang til gang også der, men eneste signifikante forskjell var mellom testerne (interrater) for 6m. En årsak kan være at vi brukte en manuell stoppeklokke til å ta tiden med og dermed ikke var i stand til å registrerer endringer som ikke oppdages av det blotte øye. Små feil ved manuell tidtaking kan potensielt resultere i store forskjeller når vi måler ned til nærmeste 0,01 sek.

Målemetodene vi brukte ga alle resultater i form av kontinuerlige data. Det sikrer i utgangspunktet en større spredning av score enn ved nominal-/ordinaldata. Vi korrigerer ikke styrke- eller hinkevariablene for kroppsvekt eller høyde/benlengde. I studier hvor man skal se på styrkeutvikling over tid, hvor vekst og modning kan forekomme, eller man skal sammenligne utvalg av ulik størrelse, er det viktig å ha en ”størrelsesfri” styrkevariabel å sammenligne (De Ste Croix et al., 2003). I et reliabilitetsstudie går man ut i fra at ingenting skal endre seg fra gang til gang. Ut i fra testintervallet man velger antar man at høyde og vekt ikke forandrer seg, men at variabiliteten man ser heller er et resultat av andre biologiske variasjoner. Som følge av at vi ikke korrigerer variablene våre er absoluttverdiene våre mindre sammenlignbare med andre studier på barn som ikke har samme antropometri. Samtidig kunne kanskje korreksjon for vekt/høyde ført til mindre spredning av score for utvalget vårt og dermed påvirket spredningen av prestasjon. Det igjen kunne påvirket størrelsen på ICC.

5.3 Valg av statistiske metoder

I metodestudien ble ICC brukt som et mål på relativ reliabilitet. De individuelle variasjonene relatert til hver måling i form av absolutt reliabilitet ble uttrykt ved SEM (SEM %) og LOA. I tillegg benyttet vi SEM til å beregne minste reelle endring uttrykt som SRD.

5.3.1 Relativ reliabilitet

I vår studie tok vi hensyn til de systematiske forskjellene, dvs. forskjeller i gjennomsnittsverdier for intra- og interrater målingene, i beregningen av ICC. Dersom den systematiske målefeilen er stor, vil valg av modell påvirke størrelsen på ICC hvor modell 3 > modell 2 > modell 1 (Dvir, 2003; Portney & Watkins, 2000; Weir, 2005). I våre resultater var komponenten av tilfeldig målefeil mye større enn de systematiske. Da vil valg av modell ha mindre betydning for størrelsen på ICC, og resultatene fra ulike ICC vil være mer sammenlignbare. Generelt vil ICC verdier basert på et gjennomsnittet av flere målinger, f.eks. ICC 2,k/3,k (Shrout & Fleiss, 1979), gi høyere verdier enn for enkeltmålinger, spesielt hvis *average measures* avleses i SPSS. To av reliabilitetsstudiene på hinketestene (Bolglia & Keskula, 1997; Ross et al., 2002) brukte ICC (2,k) modellen som baserte seg på gjennomsnittet av flere målinger. De rapporterer også de høyeste ICC verdiene. For styrkevariabelen AVG PT benyttet vi samme modell (ICC 2,k/3,k), da målingene var gjennomsnittet av fem repetisjoner. Hartmann et al. (2009) brukte samme beregning av ICC for AVG PT i sin studie. Det gjør at sammenligninger mellom for eksempel PT og AVG PT i vår studie egentlig er vanskelig fordi ICC for hver av dem er basert på enten enkeltmålinger (single measure) eller gjennomsnittsmålinger (average) som vil gi ulikt utfall i SPSS.

ICC blir påvirket av heterogeniteten til utvalget som testes, dvs. liten variasjon mellom individene (between-subjects) vil resultere i en lavere ICC verdi enn for en gruppe med større spredning (Weir, 2005). For å undersøke graden av homo-/heterogenitet er man avhengig av å sammenligne med studier som oppgir lignende gjennomsnittsverdier, da SD til en viss grad er avhengig av størrelsen på gjennomsnittsverdien. Barna i metodestudien var en mer heterogen gruppe enn utvalget av aktive fotballgutter i studien til Kellis et al. (1999), men var en mer homogen gruppe enn utvalget fra normmaterialet på isokinetiske muskelstyrkemålinger på samme aldersgruppe (Holm et al., 2008). Årsaken til at vi har dårligere ICC enn rapportert i mange studier kan være et

resultat av flere ting: 1) valg av ICC modell (inkludering av systematiske forskjeller), 2) homogent utvalg og 3) større andel av tilfeldige målefeil enn i andre studier. Utvalget til Kellis et al. (1999) hadde mye mindre individuelle variasjoner (sett ved LOA) enn det metodebarna hadde og rapporterte dermed høyere ICC verdier, selv med et mer homogent utvalg.

Vi kan forbedre reliabiliteten ved å øke variabiliteten mellom subjektene og redusere andelen systematiske og tilfeldige målefeil (Streiner & Norman, 2008). Som tidligere nevnt kunne vi i vårt tilfelle økt antall hopp for hver hinketest. Vi kunne også tilført flere testrunder, økt antall testere, inkludert et mindre homogent utvalg, eller økt lengden og kompleksiteten på testen. For å redusere målefeilene kunne vi gjennomført grundigere opplæring/samkjøring av testerne, tilvenning for barna og eventuelt tilpasset enda bedre protokoll og prosedyrer til utvalget som vi testet.

5.3.2 Absolutt reliabilitet

Variasjonen mellom barna (between-subject) var mye større enn de individuelle variasjonene (within-subject), noe som også er vanlig (Atkinson & Nevill, 1998). Våre SEM verdier var beregnet ut i fra de tilfeldige målefeilene ($\sqrt{\text{RMSE}}$ fra ANOVA), og tar ikke hensyn til de systematiske målefeilene. De fleste av reliabilitetsstudiene jeg så på beregnet SEM ved bruk av SD og ICC, hvor SEM da er knyttet opp til valg av ICC metode og inkludering/eksklusjon av systematiske målefeil (Weir, 2005). Tre studier (Hartmann et al., 2009; Lund et al., 2005; Reid et al., 2007) beregnet som oss, SEM ut i fra error-komponenten fra ANOVA. SEM er da uavhengig av ICC modell som er valgt, variasjonsbredden til målingene, og kan lettere sammenlignes på tvers av studier SEM er mindre påvirket av heterogeniteten til utvalget siden den ikke tar hensyn til variasjonene mellom individer, men heller karakteriserer 'within-subject' variasjonene (Weir, 2005). SEM dekker som nevnt rundt 68 % av variabiliteten til en måling. Ved repeterte målinger er det derfor anbefalt å inkludere et 95 % KI, eller som vi har gjort i denne studien, LOA og/eller SRD som gir 95 % sikkerhet innenfor hvilke verdi/intervall de repeterte målingene vil falle.

Det er blitt hevdet at å presentere både SEM og LOA er u hensiktsmessig fordi de er for nært forbundet med hverandre (Hopkins, 2000). Flere studier har derimot valgt å presentere begge deler (Flansbjerg & Lexell, 2010; Hartmann et al., 2009; Sole et al.,

2007). I denne studien ble det også valgt å ta med LOA for å beskrive de individuelle forskjellene visuelt, og ikke bare i tabell.

5.3.3 Måling av endring i funksjon

SRD er kritisert for å være et konservativt mål for å registrere om en reell endring har skjedd (Beyer & Magnusson, 2003; Flansbjer & Lexell, 2010). Reid et al. (2007) beregnet minste reelle endring med 90 % sikkerhet. De argumenterte for bruken av et mer liberalt mål når man skal tolke klinisk betydning av målinger på individnivå. Vi beregnet minste reelle endring med 95 % sikkerhet ($SRD = SEM \times \sqrt{2} \times 1,96$), noe som fører til et høyere estimat av SRD. Det betyr at størrelsen på forskjellen som kan kalles en reell endring stiger.

Tilsvarende beregning med 95 % sikkerhet er brukt i tre andre studier (Hartmann et al., 2009; Munro & Herrington, 2011; Sole et al., 2007). En alternativ tilnærming til SRD er LOA. Hopkins et al. (2000) argumenterte for at SEM er et bedre alternativ enn LOA. De trekker også fram at et 95 % LOA er for strengt, spesielt hvis man har med idrettsutøvere å gjøre, hvor små endringer kan ha stor betydning. Det er foreslått at en utvalgsstørrelse på 30-50 er nødvendig for å kunne forme praktisk anvendbare 95 % SRD og SRD % (Lexell & Downham, 2005). Utvalgsstørrelsen i studien vår var sånn sett liten ($n = 26$), og vi må ta dette i betraktning når vi vurderer den kliniske betydningen av våre resultater.

5.4 *Kliniske implikasjoner*

Vi har i denne metodestudien undersøkt reliabiliteten til hinketestene og isokinetiske muskelstyrkemålingene på et utvalg av friske barn. Metodestudien har gitt oss muligheten til å undersøke rutinene rundt testing slik de brukes i praksis i vurderingen av knefunksjon på barn med ACL skade. Vi fant, som nevnt, ingen studier som har undersøkt reliabiliteten av de fire hinketestene beskrevet av Noyes et al. (1991) på et utvalg av barn. Det var derfor ekstra viktig å få et estimat på hva vi kan forvente av reliabilitet på et utvalg som er mer aktuelt å sammenligne ACL-barna i kohortstudien med. Det er viktig i enhver studie å reliabilitetsteste målemetodene man benytter under gitte testforhold, på et aktuelt utvalg og med representative testere (Streiner & Norman, 2008). Med reliabilitetsdata på målemetodene vet vi da i større grad hvordan vi kan vektlegge de endringene vi ser ved testing.

Ut i fra våre funn ser det ut som barna varierer mer ved isokinetiske muskelstyrkemålinger enn ved hinketestene. Det kan bety at en form for tilvenning med apparatet og prosedyrer før den isokinetiske testingen kunne vært fordelaktig. Jeg vil tro at dette også gjelder for barna med ACL skade, som vil kunne variere enda mer i prestasjon. De systematiske forskjellen vi fant mellom testerne kan indikere et behov for å gjennomgå testprotokollen og prosedyrerene som brukes på ACL-barna på nytt. Dersom man skal kunne stole på resultatene og minimere variasjonen mellom testerne så er det viktig å gjennomføre en regelmessig gjennomgang og samkjøring av testrutinene.

Styrkevariablene som kom best ut, med lav absolutt reliabilitet og høy relativ reliabilitet var PT, AVG PT og Tw kneekstensjon. TQ30 viste stor variasjonen av score fra måling til måling. Det betyr at det vil være vanskelig å fange opp en reell endring ved bruk av denne variabelen. På et utvalg som vil variere enda mer enn metodebarna er den derfor ikke å anbefale som mål på knefunksjon alene. Hinketestene viste generelt god reliabilitet. CH var den testen som skilte seg ut med størst individuell variasjon hos barna, og dermed vil være minst egnet alene for å vurdere endring i funksjon. For både styrke- og hinketestene er det viktig med innlæring og pauser da vi så forbedring av resultatene fra gang til gang, om enn ikke signifikante for hinketestene.

Selv om vi fant at hinketestene og isokinetiske muskelstyrkemålinger var reliable på et selektert utvalg av barn i denne studien så må vi være forsiktige med å overføre disse resultatene direkte til barn med ACL skade. Som vi nevnte innledningsvis så er en test er ikke reliabel i seg selv, men er nært knyttet opp til utvalget den har blitt testet på og forholdene for testingen (Streiner & Norman, 2008). Variasjonen til målemetodene kan være knyttet opp til alder, kjønn, skadestatus, aktivitetsnivå og funksjon.

Resultatene fra denne studien viste at intra- og interrater reliabiliteten ga omtrent like ICC, SEM og SRD verdier. Det gir oss en trygghet på at dersom vi i praksis bruker ulike testere, med lik erfaring og kjennskap til målemetoden, så vil enigheten testerne imellom være akseptabel.

6. Konklusjon

Formålet med studien var å undersøke intra- og interater reliabiliteten mellom to tester for isokinetiske muskelstyrkemålinger og hinketester på et utvalg av barn uten skade. Vi ønsket også å gi et estimat på minste reelle endring til målemetodene ut i fra reliabilitetsdataene.

Vi fant at hinketestene og de isokinetiske muskelstyrkemålingene var reliable på et utvalg av aktive friske barn. Både intra- og interater målingene viste god reliabilitet med relativt like verdier. Generelt varierte ICC fra moderat til høy. Resultatene viste at målemetodene også ga akseptable SEM verdiene med tanke på gjennomsnittsverdiene til de aktuelle målingene. SEM % viste til større individuelle variasjoner forbundet med isokinetisk muskelstyrkemålinger enn ved hinketestene.

Både hinketestene og de isokinetiske muskelstyrkemålingene viste evnen til å kunne fange opp endringer over tid på utvalget vi testet. Basert på reliabilitetsresultatene indikerte SRD at måleforskjeller på over 20 % for hinketestene og over 25 % for de isokinetiske muskelstyrkemålingene potensielt kan vurderes som en reell endring i knefunksjon.

Referanser

- Atkinson, G. & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.*, 26, 217-238.
- Ayalon, M., Ben-Sira, D., Hutzler, Y., & Gilad, T. (2000). Reliability of isokinetic strength measurements of the knee in children with cerebral palsy. *Dev.Med.Child Neurol.*, 42, 398-402.
- Bahr, R. (2009). *Aktivitetshåndboken: Fysisk aktivitet i forebygging og behandling*. [Oslo]: Helsedirektoratet.
- Bales, C. P., Guettler, J. H., & Moorman, C. T., III (2004). Anterior cruciate ligament injuries in children with open physes: evolving strategies of treatment. *Am.J.Sports Med.*, 32, 1978-1985.
- Bandy, W. D., Rusche, K. R., & Tekulve, F. Y. (1994). Reliability and limb symmetry for five unilateral functional tests of the lower extremities. *Isokinetics and Exercise Science*, 4, 108-111.
- Barber, S. D., Noyes, F. R., Mangine, R. E., McCloskey, J. W., & Hartman, W. (1990). Quantitative assessment of functional limitations in normal and anterior cruciate ligament-deficient knees. *Clin.Orthop.Relat Res.*, 204-214.
- Barber-Westin, S. D., Noyes, F. R., & Galloway, M. (2006). Jump-land characteristics and muscle strength development in young athletes: a gender comparison of 1140 athletes 9 to 17 years of age. *American Journal of Sports Medicine*, 34, 375-384.
- Batterham, A. M. & George, K. P. (2003). Reliability in evidence-based clinical practice: A primer for allied health professionals. *Physical Therapy in Sport*, 4, 122-128.

- Baumgartner, T. A. (2003). *Measurement for evaluation in physical education and exercise science*. (vols. 7th ed.) Boston: McGraw-Hill.
- Bent, N. P. (2009). Selecting outcome measures in sports medicine: a guide for practitioners using the example of anterior cruciate ligament rehabilitation. *British Journal of Sports Medicine*, *43*, 1006-1012.
- Beyer, N. & Magnusson, P. (2003). *Målemetoder i fysioterapi*. København: Munksgaard.
- Bolgia, L. A. & Keskula, D. R. (1997). Reliability of lower extremity functional performance tests. *J.Orthop.Sports Phys.Ther.*, *26*, 138-142.
- Booher, L. D., Hench, K. M., Worrell, T. W., & Stikeleather, J. (1993). Reliability of three single-leg hop tests. *Journal of Sport Rehabilitation*, *2*, 165-170.
- Brosky, J. A., Jr., Nitz, A. J., Malone, T. R., Caborn, D. N., & Rayens, M. K. (1999). Intrarater reliability of selected clinical outcome measures following anterior cruciate ligament reconstruction. *J.Orthop.Sports Phys.Ther.*, *29*, 39-48.
- Caine, D., Caine, C., & Maffulli, N. (2006). Incidence and distribution of pediatric sport-related injuries. *Clin.J.Sport Med.*, *16*, 500-513.
- Caine, D., Maffulli, N., & Caine, C. (2008). Epidemiology of injury in child and adolescent sports: injury rates, risk factors, and prevention. *Clin.Sports Med.*, *27*, 19-50, vii.
- Clark, N. C. (2001). Functional performance testing following knee ligament injury. *Physical Therapy in Sport*, *2*, 91-105.
- Copay, A. G., Subach, B. R., Glassman, S. D., Polly, D. W., Jr., & Schuler, T. C. (2007). Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J.*, *7*, 541-546.
- Daniel, D., Malcolm, L., & Stone, M. L. (1982). Quantification of Knee Stability and Function. *Contemporary orthopaedics*, *5*, 83-91.

- De Carlo, M. S. & Sell, K. E. (1997). Normative data for range of motion and single-leg hop in high school athletes. / Normes pour les tests de l'amplitude de mouvement et le bond sur une jambe chez les sportifs universitaires. *Journal of Sport Rehabilitation*, 6, 246-255.
- De Ste Croix, M. B., Armstrong, N., Welsman, J. R., & Sharpe, P. (2002). Longitudinal changes in isokinetic leg strength in 10-14-year-olds. *Ann.Hum.Biol.*, 29, 50-62.
- De Ste Croix, M. B. A., Armstrong, N., & Welsman, J. R. (1999). Concentric isokinetic leg strength in pre-teen, teenage and adult males and females. *Biology of Sport*, 16, 75-86.
- De Ste Croix, M. B. A., Deighan, M. A., & Armstrong, N. (2003). Assessment and interpretation of isokinetic muscle strength during growth and maturation. *Sports Medicine*, 33, 727-743.
- de Vet, H. C., Terwee, C. B., Ostelo, R. W., Beckerman, H., Knol, D. L., & Bouter, L. M. (2006). Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual.Life Outcomes.*, 4, 54.
- Deighan, M. A., De Ste Croix, M. B. A., & Armstrong, N. (2003). Reliability of isokinetic concentric and eccentric knee and elbow extension and flexion in 9/10 year old boys. *Isokinetics & Exercise Science*, 11, 109-115.
- Denegar, C. R. & Ball, D. W. (1993). Assessing reliability and precision of measurement: an introduction to intraclass correlation and standard error of measurement. *Journal of Sport Rehabilitation*, 2, 35-42.
- Domholdt, E. (2005). *Rehabilitation research: principles and applications*. (vols. 3rd ed.) St. Louis, Miss.: Elsevier Saunders.
- Donner, A. & Eliasziw, M. (1987). Sample size requirements for reliability studies. *Statistics in Medicine*, 6, 441-448.
- Drouin, J. M., Valovich-mcLeod, T. C., Shultz, S. J., Gansneder, B. M., & Perrin, D. H. (2004). Reliability and validity of the Biodex system 3 pro isokinetic

- dynamometer velocity, torque and position measurements. *Eur.J.Appl.Physiol*, 91, 22-29.
- Dvir, Z. (2003). *Isokinetics: muscle testing, interpretation and clinical applications*. (vols. 2nd ed.) Edinburgh: Churchill Livingstone.
- Eastlack, M. E., Axe, M. J., & Snyder-Mackler, L. (1999). Laxity, instability, and functional outcome after ACL injury: copers versus noncopers. *Med.Sci.Sports Exerc.*, 31, 210-215.
- Eitzen, I., Eitzen, T. J., Holm, I., Snyder-Mackler, L., & Risberg, M. A. (2010). Anterior cruciate ligament-deficient potential copers and noncopers reveal different isokinetic quadriceps strength profiles in the early stage after injury. *Am.J.Sports Med.*, 38, 586-593.
- Eitzen, I., Holm, I., & Risberg, M. A. (2009). Preoperative quadriceps strength is a significant predictor of knee function two years after anterior cruciate ligament reconstruction. *Br.J.Sports Med.*, 43, 371-376.
- Eitzen, I., Moksnes, H., Snyder-Mackler, L., Engebretsen, L., & Risberg, M. A. (2010). Functional tests should be accentuated more in the decision for ACL reconstruction. *Knee.Surg.Sports Traumatol.Arthrosc.*
- Fitzgerald, G. K., Axe, M. J., & Snyder-Mackler, L. (2000). A decision-making scheme for returning patients to high-level activity with nonoperative treatment after anterior cruciate ligament rupture. *Knee.Surg.Sports Traumatol.Arthrosc.*, 8, 76-82.
- Flansbjerg, U. B. & Lexell, J. (2010). Reliability of knee extensor and flexor muscle strength measurements in persons with late effects of polio. *J.Rehabil.Med.*, 42, 588-592.
- Fuller, C. W., Ekstrand, J., Junge, A., Andersen, T. E., Bahr, R., Dvorak, J. et al. (2006). Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *Clin.J.Sport Med.*, 16, 97-106.

- Hartmann, A., Knols, R., Murer, K., & De Bruin, E. D. (2009). Reproducibility of an isokinetic strength-testing protocol of the knee and ankle in older adults. *Gerontology, 55*, 259-268.
- Holm, I. (1996). *Quantification of muscle strength by isokinetic performance*. Biomechanics Laboratory/Physiotherapy Dept., National Hospital, Orthopaedic Centre, University of Oslo, Oslo.
- Holm, I., Fredriksen, P., Fosdahl, M., & Vollestad, N. (2008). A normative sample of isotonic and isokinetic muscle strength measurements in children 7 to 12 years of age. *Acta Paediatrica, 97*, 602-607.
- Holm, I., Tveter, A. T., Fredriksen, P. M., & Vollestad, N. (2009). A normative sample of gait and hopping on one leg parameters in children 7-12 years of age. *Gait.Posture., 29*, 317-321.
- Holm, I. & Vollestad, N. (2008). Significant effect of gender on hamstring-to-quadriceps strength ratio and static balance in prepubescent children from 7 to 12 years of age. *Am.J.Sports Med., 36*, 2007-2013.
- Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Med., 30*, 1-15.
- Hopper, D. M., Goh, S. C., Wentworth, L. A., Chan, D. Y. K., Chau, J. H. W., Wootton, G. J. et al. (2002). Test-retest reliability of knee rating scales and functional hop tests one year following anterior cruciate ligament reconstruction. *Physical Therapy in Sport, 3*, 10-18.
- Jones, M. A. & Stratton, G. (2000). Muscle function assessment in children. *Acta Paediatrica, International Journal of Paediatrics, 89*, 753-761.
- Kannus, P. (1994). Isokinetic evaluation of muscular performance: implications for muscle testing and rehabilitation. *Int.J.Sports Med., 15 Suppl 1*, S11-S18.
- Kellis, E., Kellis, S., Gerodimos, V., & Manou, V. (1999). Reliability of isokinetic concentric and eccentric strength in circumpubertal soccer players. / Fiabilite des

- mesures de la force isocinetique concentrique et excentrique chez des joueurs de football autour de la puberte. *Pediatric Exercise Science*, 11, 218-228.
- Keskula, D. R., Dowling, J. S., Davis, V. L., Finley, P. W., & Dell'omo, D. L. (1995). Interrater Reliability of Isokinetic Measures of Knee Flexion and Extension. *J.Athl.Train.*, 30, 167-170.
- Kunnskapssenteret (2010). Ordliste. Hentet 01-03-2011, fra <http://www.kunnskapssenteret.no/Verkt%C3%B8y/2028.cms>
- Lexell, J. E. & Downham, D. Y. (2005). How to assess the reliability of measurements in rehabilitation. *Am.J.Phys.Med.Rehabil.*, 84, 719-723.
- Li, R. C., Wu, Y., Maffulli, N., Chan, K. M., & Chan, J. L. (1996). Eccentric and concentric isokinetic knee flexion and extension: a reliability study using the Cybex 6000 dynamometer. *Br.J Sports Med*, 30, 156-160.
- Lund, H., Sondergaard, K., Zachariassen, T., Christensen, R., Bulow, P., Henriksen, M. et al. (2005). Learning effect of isokinetic measurements in healthy subjects, and reliability and comparability of Biodex and Lido dynamometers. *Clin.Physiol Funct.Imaging*, 25, 75-82.
- Lysholm, J. & Tegner, Y. (2007). Knee injury rating scales. *Acta Orthopaedica*, 78, 445-453.
- Maffiuletti, N. A., Bizzini, M., Desbrosses, K., Babault, N., & Munzinger, U. (2007). Reliability of knee extension and flexion measurements using the Con-Trex isokinetic dynamometer. *Clin.Physiol Funct.Imaging*, 27, 346-353.
- McGraw, K. O. & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. [References]. *Psychological Methods*, .1.
- Merlini, L., Dell'Accio, D., & Granata, C. (1995). Reliability of dynamic strength knee muscle testing in children. *J.Orthop.Sports Phys.Ther.*, 22, 73-76.
- Mohtadi, N. & Grant, J. (2006). Managing anterior cruciate ligament deficiency in the skeletally immature individual: a systematic review of the literature. *Clin.J.Sport Med.*, 16, 457-464.

- Mohtadi, N. G., Kiefer, G. N., Tedford, K., & Watters, S. (1990). Concentric and eccentric quadriceps torque in pre-adolescent males. *Can.J.Sport Sci.*, *15*, 240-243.
- Moksnes, H., Engebretsen, L., & Risberg, M. A. (2008). Performance-based functional outcome for children 12 years or younger following anterior cruciate ligament injury: a two to nine-year follow-up study. *Knee.Surg.Sports Traumatol.Arthrosc.*, *16*, 214-223.
- Moksnes, H. & Risberg, M. A. (2009). Performance-based functional evaluation of non-operative and operative treatment after anterior cruciate ligament injury. *Scand.J.Med.Sci.Sports*, *19*, 345-355.
- Molnar, G. E., Alexander, J., & Gutfeld, N. (1979). Reliability of quantitative strength measurements in children. *Arch.Phys.Med.Rehabil.*, *60*, 218-221.
- Munro, A. G. & Herrington, L. C. (2011). Between-Session Reliability of Four Hop Tests and the Agility T-Test. *J.Strength.Cond.Res.*, *25*, 1470-1477.
- Noyes, F. R., Barber, S. D., & Mangine, R. E. (1991). Abnormal lower limb symmetry determined by function hop tests after anterior cruciate ligament rupture. *Am.J.Sports Med.*, *19*, 513-518.
- Oiestad, B. E., Holm, I., Gunderson, R., Myklebust, G., & Risberg, M. A. (2010). Quadriceps muscle weakness after anterior cruciate ligament reconstruction: a risk factor for knee osteoarthritis? *Arthritis Care Res.(Hoboken.)*, *62*, 1706-1714.
- Parker, H. E., Larkin, D., & Ackland, T. R. (1993). Stability and change in children's skill. *Psychological research*.*55* (2) (pp 182-189), 1993.*Date of Publication: 1993.*, *55*, 1993-189.
- Pierce, S. R., Lauer, R. T., Shewokis, P. A., Rubertone, J. A., & Orlin, M. N. (2006). Test-retest reliability of isokinetic dynamometry for the assessment of spasticity of the knee flexors and knee extensors in children with cerebral palsy. *Arch.Phys.Med.Rehabil.*, *87*, 697-702.

- Pincivero, D. M., Heller, B. M., & Hou, S. (2002). The effects of ACL injury on quadriceps and hamstring torque, work and power. / Effets d ' une blessure au ligament croise anterieur sur les quadriceps et le moment de torsion, le travail et la puissance des muscles de la loge posterieure de la jambe. *Journal of Sports Sciences*, 20, 689-696.
- Portney, L. G. & Watkins, M. P. (2000). *Foundations of clinical research: applications to practice*. (vols. 2nd ed.) Upper Saddle River, N.J.: Prentice Hall.
- Reid, A., Birmingham, T. B., Stratford, P. W., Alcock, G. K., & Giffin, J. R. (2007). Hop testing provides a reliable and valid outcome measure during rehabilitation after anterior cruciate ligament reconstruction.
- Ross, M. D., Irrgang, J. J., Denegar, C. R., McCloy, C. M., & Unangst, E. T. (2002). The relationship between participation restrictions and selected clinical measures following anterior cruciate ligament reconstruction. *Knee.Surg.Sports Traumatol.Arthrosc.*, 10, 10-19.
- Ross, M. D., Langford, B., & Whelan, P. J. (2002). Test-retest reliability of 4 single-leg horizontal hop tests.
- Rousson, V., Gasser, T., & Seifert, B. (2002). Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Stat.Med.*, 21, 3431-3446.
- Sadeghi, H., Allard, P., Prince, F., & Labelle, H. (2000). Symmetry and limb dominance in able-bodied gait: a review. *Gait.Posture.*, 12, 34-45.
- Shanmugam, C. & Maffulli, N. (2008). Sports injuries in children. *Br.Med.Bull.*, 86, 33-57.
- Shaw, T., Chipchase, L. S., & Williams, M. T. (2004). A users guide to outcome measurement following ACL reconstruction. *Physical Therapy in Sport*, 5, 57-67.
- Shea, K. G., Pfeiffer, R., Wang, J. H., Curtin, M., & Apel, P. J. (2004). Anterior cruciate ligament injury in pediatric and adolescent soccer players: an analysis of insurance data. *J.Pediatr.Orthop.*, 24, 623-628.

- Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol.Bull.*, 86, 420-428.
- Sole, G., Hamren, J., Milosavljevic, S., Nicholson, H., & Sullivan, S. J. (2007). Test-retest reliability of isokinetic knee extension and flexion. *Arch.Phys.Med.Rehabil.*, 88, 626-631.
- Stratford, P. W. & Goldsmith, C. H. (1997). Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Phys.Ther.*, 77, 745-750.
- Streiner, D. L. & Norman, G. R. (2008). *Health measurement scales: a practical guide to their development and use.* (vols. 4th ed.) Oxford: Oxford University Press.
- Thomas, J. R., Silverman, S. J., & Nelson, J. K. (2005). *Research methods in physical activity.* (vols. 5th ed.) Champaign, Ill.: Human Kinetics.
- van Mechelen, W., Hlobil, H., & Kemper, H. C. G. (1992). Incidence, severity, aetiology and prevention of sports injuries. A review of concepts. *Sports Medicine*, 14, 82-99.
- Walter, S. D., Eliasziw, M., & Donner, A. (1998). Sample size and optimal designs for reliability studies. *Stat.Med.*, 17, 101-110.
- Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength & Conditioning Research (Allen Press Publishing Services Inc.)*, 19, 231-240.
- WHO (2001). *International classification of functioning, disability and health: ICF.* Geneva: World Health Organization.
- Wiggin, M., Wilkinson, K., Habetz, S., Chorley, J., & Watson, M. (2006). Percentile values of isokinetic peak torque in children six through thirteen years old. *Pediatr.Phys Ther*, 18, 3-18.
- Wright, R. W. (2009). Knee injury outcomes measures. *J.Am.Acad.Orthop.Surg.*, 17, 31-39.

Vedlegg

		Side
Vedlegg 1	Informasjonsbrev & samtykke	95
Vedlegg 2	Registreringsskjema	97
Vedlegg 3	Organisert trening	99
Vedlegg 4	Samtykke foto/video	100
Vedlegg 5	Registrering hinkeresultater	101
Vedlegg 6	Godkjenning REK kohortstudien	102
Vedlegg 7	Godkjenning REK metodestudien	103
Vedlegg 8	Tabell med korrigerte TQ30 resultater	104

Vedlegg 1



Forespørsel om deltagelse i forskningsprosjektet:

”Testing av hinketester og styrketester for barn”

Skade av fremre korsbånd er en alvorlig kneskade som forekommer relativt hyppig hos voksne og har vært økende også for barn. Vi har en studie som pågår ved Norsk forskningssenter for aktiv rehabilitering (NAR) (et forskningssamarbeid mellom Ortopedisk avdeling – Oslo universitetssykehus, Hjelp24 NIMI og Norges idrettshøgskole) hvor vi har fulgt opp barn med fremre korsbåndskade i flere år. Vi tester knefunksjonen til disse barna ved blant annet å bruke hinketester og en styrketest. Nå ønsker vi å teste barn uten korsbåndskade med de samme testene. Resultatene fra testing av barn uten korsbåndskade kan hjelpe oss til å bedre tolke resultatene fra barna med korsbåndskade.

I denne studien ønsker vi å undersøke knefunksjonen hos barn som er 11-12 år (født i 1998) og som ikke har skadet fremre korsbånd, eller har annen skade som kan hindre deltagelse i testene. Det vil være to erfarne fysioterapeuter som tester alle barna ved Hjelp24 NIMI på Ullevål stadion. Barna i dette prosjektet skal gjennomføre fire hinketester og en styrketest tre ganger, fordelt på to ulike dager med en ukes mellomrom. Det tar ca. 1 time å gjennomføre selve testingen. I tillegg kommer noen minutter til informasjon i forkant og etterkant.

Hinketestene inneholder elementer av bevegelse som barn normalt utsetter seg for i løpet av en uke i fritidsaktivitet eller idrett. Styrketesten er mye brukt for å måle muskelstyrke både hos barn og voksne. Det er ingen kjent risiko ved å delta i disse testene.

Resultatene fra testene vil bli lagret innelåst i manuelle arkiv. Du har til enhver tid full rett til innsyn i dataene. Avidentifisering av data skjer ved elektronisk lagring på PC for statistiske analyser (lagres kun med nummer). Ingen av dataene sammenholdes med elektroniske registre. Lagring av data vil foregå i henhold til personopplysningsloven. Regional etisk komité har vurdert prosjektet og godkjent gjennomføringen.

Det er helt frivillig å delta i studien. Du kan når som helst, og uten å oppgi noen grunn, trekke ditt samtykke til å delta i studien. Dersom du ønsker å delta, undertegner du samtykkeerklæringen og fyller ut registreringsskjema som er vedlagt.

Hvis du senere ønsker å trekke deg eller har spørsmål underveis, kan du kontakte fysioterapeut og masterstudent Marianne B. Johnsen 92 43 96 63, eller fysioterapeut Håvard Moksnes 95 05 56 65.

Vedlegg 1



Samtykkeerklæring til deltakelse i prosjektet

Jeg har lest informasjonsskrivet og blitt forklart hva prosjektet innebærer, og sier meg villig til å delta i studien. Jeg har forstått at deltagelsen er frivillig.

Sted

Dato

Underskrift av prosjektdeltager

Underskrift av foreldre/foresatte
(dersom deltager er under 18 år)

Vedlegg 2



Hvordan vil du gradere din nåværende funksjon av høyre ben på en skala fra 0 til 10?

(10 er helt normal funksjon, og 0 er at du ikke kan gjøre noen av dine daglige aktiviteter/idrettsaktiviteter)

Kan ikke 0 1 2 3 4 5 6 7 8 9 10 Ingen
gjøre daglige begrensninger
aktiviteter i aktivitet

Hvordan vil du gradere din nåværende funksjon av venstre ben på en skala fra 0 til 10?

(10 er helt normal funksjon, og 0 er at du ikke kan gjøre noen av dine daglige aktiviteter/idrettsaktiviteter)

Kan ikke 0 1 2 3 4 5 6 7 8 9 10 Ingen
gjøre daglige begrensninger
aktiviteter i aktivitet

Hvilke av aktivitetene under har du drevet med innefor de siste 4 ukene?

(Sett ring rundt de aktuelle aktivitetene - sett gjerne flere ringer)

Gym på skolen – håndball – fotball – basket – innebandy – volleyball – kampsport – turn – ishockey/bandy – tennis/squash/badminton – alpint/telemark – snowboard – langrenn – løping – sykling – svømming – styrketrening – balansetrening – dans/aerobic

Annet:

Hvor mange ganger i uken har du gjennomsnittlig utført disse aktivitetene (til sammen)?

0-1 ganger per uke 2-3ganger 4-5 ganger 6 eller flere

Vedlegg 3



Navn:

Fødselsdato:

1. Hva slags idrett/aktivitet er du med på organisert trening i (høst-vinter-vår-sommer)?

-
2. Hvor mange ganger i uken trener du?

Idrett:

Antall ganger per uke:

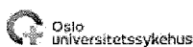
Idrett:

Antall ganger per uke:

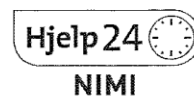
Idrett:

Antall ganger per uke:

Vedlegg 4



Ortopedisk avdeling
FOU-enheten, NAR



Informasjon om fotografering og/eller videoopptak:

Hei, vi ønsker å ta bilder og/eller videoopptak av deg, mens du deltar i prosjektet du er med i. Disse bildene vil bli brukt offentlig i presentasjoner både i inn og utland.

Det er frivillig å delta i fotograferingen og/eller videoopptaket. Dersom du velger å ikke delta, trenger du ikke å oppgi grunn. Om du skulle bestemme deg for ikke å delta, får dette ingen behandlingmessige konsekvenser for deg nå eller i fremtiden. Skulle du på et senere tidspunkt ønske at vi ikke bruker bilde og/eller videoopptak vi har tatt av deg, kan du kontakte prosjektleder, eller forskningskoordinator ved NAR som vil fjerne bildene. De presentasjoner som er utlevert, kan vi imidlertid da ikke tilbakekalle. Bildene blir lagret på en sikker forskningsserver.

Har du spørsmål, kan du kontakte til Kristin Bølstad på telefon 23 26 56 61 eller epost kristin.bolstad@hjelp24.no

Med vennlig hilsen

Kristin Bølstad
Forskningskoordinator
NAR/Ortopedisk avdeling
Oslo universitetssykehus
Hjelp24 NIMI

Tlf: 23265661
E-post: kristin.bolstad@hjelp24.no

Samtykkeerklæring

Jeg har lest og blitt forklart informasjonen på medfølgende informasjonsskriv, og sier meg villig til å delta i fotografering og/eller videoopptak. Jeg har forstått at disse bildene og/eller fotoopptakene kan brukes i offentlige sammenhenger og at deltakelse er frivillig.

Sted Dato Underskrift

Oslo universitetssykehus HF

Telefon:
02770

Besøksadresse:
Kirkeveien 166
0407 OSLO

Bankgiro:
1644 06 05897

Foretaksnr.:
993 467 049

HELSE SØR-ØST

Vedlegg 5



Registrering av hinkeresultater – testomgang 2 (januar 2011)

Tester (fysio):	Testdag (1/2):	Dato:	Klokkeslett:			
Gruppe (1/2):	Pulje (a/b):	Testnr.:	ID nr.:			
Navn:	Fødselsdato:	Hvilket ben foretrekker du å sparke en ball med?:				
Høyde:	Vekt:	Randomisert startben: Hø <input type="checkbox"/> Ve <input type="checkbox"/>				
Testben 1			Testben 2			
Test	1	2	Forsøk	Beste	Gj.sn.	Differanse
Single hopp (cm)						%
Crossover hopp (cm)						
Triple hopp (cm)						
6 m hopp (s)						

Testprosedyre: 1 submaks prøvehopp for hver hinketest. Deretter registreres 2 gyldige hopp målt til nærmeste 0,5 cm. Det gis inn til

5 forsøk (prøvehoppet er utenom disse) for å oppnå minst ett gyldig hopp. Antall forsøk brukt per ben kan noteres.

Gyldig hopp = en stødig landing på testbenet som holdes i 2 sek. Re-hopp ved ekstra rotasjon/vridning av foten, eller hvis man må støtte seg i bakken/veggen med ben eller overkropp for å holde balansen.

Vedlegg 6

REGIONAL KOMITE FOR MEDISINSK FORSKNINGSETIKK Øst-Norge (REK I)

Fysioterapeut Håvard Moksnes
NIMI Ullevål
Sognsveien 75 D
0805 Oslo

Deres ref.:

Vår ref.: 684-06288 1.2006.78

Dato: 20. november 2006


Barn med fremre korsbåndsskade. En prospektiv kohortstudie av barn under 13 år

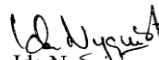
Regional komité for medisinsk forskningsetikk, Øst-Norge, vurderte prosjektet på sitt møte 09.11.06.

Komiteen har ingen innvendinger mot at studien blir gjennomført, men ber om at det utarbeides eget informasjonsskriv for barn over 12 år.

Komiteen ber om å få informasjonsskrivet til barn til orientering.

Med vennlig hilsen


Knut Engedal
professor dr.med.
leder


Ida Nyquist
sekretær

Kopi: Forskningsdirektør Andreas Moan, Ullevål Univeristetsssykehus

Vedlegg 7



UNIVERSITETET I OSLO DET MEDISINSKE FAKULTET

Håvard Moksnes
Sognsveien 75 D
0805 Oslo

Regional komité for medisinsk og helsefaglig
forskningsetikk Sør-Øst C (REK Sør-Øst C)
Postboks 1130 Blindern
NO-0318 Oslo

Telefon: 22 84 46 67

Dato: 26.08.2010
Deres ref.:
Vår ref.: 2010/1855 (oppgis ved henvendelse)

E-post: post@helseforskning.etikkom.no
Nettadresse: <http://helseforskning.etikkom.no>

Barn med fremre korsbåndskade. En prospektiv kohortstudie av barn under 13 år

Vi viser til innsendt prosjektendringskjema for overnevnte studie, mottatt 17.07.2010.

Endringen består i at man ønsker å inkludere en kontrollgruppe av barn uten fremre korsbåndskade for å undersøke reliabiliteten av de fire hinketestene og isokinetiske styrketester som blir brukt i den pågående studien. To erfarne fysioterapeuter vil utføre testingen, og barna rekrutteres fra idrettslag i Oslo-området.

Vedtak:
Komiteen godkjenner prosjektendringen.

Tillatelsen er gitt under forutsetning av at prosjektendringen gjennomføres slik det er beskrevet i prosjektendringsmeldingen og endringsprotokoll, og iht. de bestemmelser som følger av helseforskningsloven med forskrifter.

Forskningsprosjektets data skal oppbevares forsvarlig, se personopplysningsforskriften kapittel 2, og Helseinspektorens veileder for *Personvern og informasjonssikkerhet i forskningsprosjekter innenfor helse- og omsorgssektoren*:
http://www.helseinspektoren.no/samspill/informasjonssikkerhet/norm_for_informasjonssikkerhet_i_helsesektoren_232354

Komiteens vedtak kan påklages til Den nasjonale forskningsetiske komité for medisin og helsefag, jf. Forvaltningslovens § 28 flg. Eventuell klage sendes til REK Sør-Øst. Klagefristen er tre uker fra mottak av dette brevet.

Med vennlig hilsen

Arvid Heiberg (sign.)
Professor dr. med.
Leder

Tor Even Svanes
Seniorrådgiver

Vi ber om at alle henvendelser sendes inn via vår saksportal:
<http://helseforskning.etikkom.no> eller på e-post til: post@helseforskning.etikkom.no. Vennligst oppgi vårt saksnummer/referansenummer i korrespondansen.

Vedlegg 8

Tabell A: Gj.sn. og p-verdier for TQ30 med korrigert n med hensyn til intra- og interrater reliabiliteten.

Testvariabel (Nm)	Tester HM			p -verdi	Testvariabel	Dag 1		
	Dag 1	Dag 2				Tester IE	Tester HM	
TQ30E h� (n= 20)	62,6 ± 13,1	60,9 ± 12,3	0,45	TQ30E h� (n= 21)	69,4 ± 14,4	62,2 ± 12,7	0,08	
TQ30E ve	55,2 ± 10,0	58,7 ± 11,2	0,052	TQ30E ve	63,4 ± 14,1	56,0 ± 9,4	0,053	
TQ30E LSI (%)	114,1 ± 18,9	104,7 ± 15,6	0,03*	TQ30E LSI (%)	110,0 ± 13,9	111,3 ± 15,2	0,93	
TQ30F h� (n= 24)	51,0 ± 11,4	49,9 ± 11,2	0,50	TQ30F h� (n= 24)	49,3 ± 8,6	51,0 ± 11,4	0,32	
TQ30F ve	48,7 ± 10,9	49,5 ± 9,8	0,68	TQ30F ve	47,3 ± 9,7	48,7 ± 10,9	0,43	
TQ30F LSI (%)	105,8 ± 13,2	101,5 ± 16,3	0,28	TQ30F LSI (%)	105,7 ± 12,1	105,8 ± 13,1	0,97	

Tabell B: Relativ og absolutt intra- og interrater reliabilitet for TQ30 med korrigert n.

Testvariabel (Nm)	ICC	95 % KI til ICC	đ	95 % KI đ	SEM	SEM %	SRD	SRD %
TQ30E h� (n = 20)	0,70 ^a	(0,39 - 0,87)	1,7	(-2,9; 6,3)	7,0	11,3	19,4	31,4
TQ30E ve	0,71 ^a	(0,40 - 0,88)	-3,5	(-7,1; 0,04)	5,4	9,4	15,0	26,3
TQ30F h� (n= 24)	0,77 ^a	(0,54 - 0,89)	1,1	(-2,2; 4,3)	5,4	10,8	15,0	29,7
TQ30F ve	0,61 ^a	(0,28 - 0,81)	-0,8	(-4,7; 3,1)	6,5	13,2	18,0	36,7
TQ30E h� (n= 21)	0,65 ^b	(0,20 - 0,85)	7,2	(-1,0; 15,5)	8,0	12,2	22,2	34,0
TQ30E ve	0,58 ^b	(-0,32 - 0,84)	7,4	(-0,10; 14,9)	5,7	9,4	15,8	26,2
TQ30F h� (n= 24)	0,74 ^b	(0,50 - 0,88)	-1,7	(-5,2; 1,8)	5,1	10,1	14,1	28,1
TQ30F ve	0,80 ^b	(0,60 - 0,91)	-1,4	(-5,1; 2,2)	4,5	9,5	12,5	26,1

^a= ICC (3,1) intrarater reliabilitet, ^b= ICC (2,1) interrater reliabilitet

