

**S T A D I A**

HELSINGIN AMMATTIKORKEAKOULU

# **Toistettavuus- ja validiteettitutkimus sekä seated flexion testi**

Osteopatian koulutusohjelma,  
Osteopatia  
Opinnäytetyö  
27.11.2006

---

Miia Kurvinen  
Lasse Maijala  
Vertti Pöllänen



Koulutusohjelma		Suuntautumisvaihtoehto	
Osteopatian koulutusohjelma		Osteopatia	
Tekijä/Tekijät			
Miia Kurvinen, Lasse Maijala ja Vertti Pöllänen			
Työn nimi			
Toistettavuus- ja validiteettitutkimus sekä seated flexion testi			
Työn laji	Aika	Sivumäärä	
Opinnäytetyö	Marraskuu 2006	35+ 6 liitettä	
<p>TIIVISTELMÄ</p> <p>Tämän opinnäytetyön tarkoituksena on tehdä toistettavuus- ja validiteettitutkimuksien tekeminen helpommaksi manuaalisen lääketieteen ihmisille. Manuaalisessa lääketieteessä käytettyjä testejä ja hoitomenetelmiä ei ole tutkittu tarpeeksi eikä niiden käytettävyyttä ole tieteellisesti todennettu. Työtä varten on käännetty FIMM:n (Kansainvälinen manuaalisen lääketieteen yhdistys) toistettavuus- ja validiteettitutkimusprotokolla. Sen avulla voi jokainen koulutettu manuaalisen hoitoalan ammattilainen tehdä toistettavuus- ja validiteettitutkimuksia kansainvälisesti julkaistaviksi. Protokollan mukaisesti tehdyt tutkimukset ovat vertailukelpoisia ja helposti toistettavia. Käännöksen tarkoituksena on tuoda tutkimusprotokolla tunnetuksi Suomessa. Tutkimusprotokollan myötä toivotaan manuaalisten hoitoalojen ammattilaisten tekevän kasvavassa määrin vastaavanlaisia tutkimuksia ammatillistumisen lisäämiseksi.</p> <p>FIMM:n protokollan mukaisesti toteutettiin kappatutkimus SI-nivelen liikkuvuutta testaavasta seated flexion testistä. Testistä ei ole aikaisemmin tehty toistettavuus- ja validiteettitutkimusta. Opinnäytetyössä esitellään seated flexion testin suorittamisen eri tapoja. Aikarajoituksista johtuen tutkimuksen kaikkia vaiheita ei suoritettu protokollan edellyttämällä tasolla. Testivaiheessa kappaa-arvo jäi alle kliinisesti merkittävän 0,4 arvon. Alhaisen kappaa-arvon takia ei voida osoittaa testin toistettavuutta ja validiteettia.</p> <p>Johtopäätöksenä tutkimuksesta voidaan todeta, että seated flexion testistä tarvitaan lisää vastaavanlaisia kappatutkimuksia.</p>			
Avainsanat			
Kappaa, toistettavuus, validiteetti, seated flexion testi, osteopatia			



Degree Programme in		Degree
Degree programme of Osteopathy		Osteopath AMK
Author/Authors		
Miia Kurvinen, Lasse Maijala and Vertti Pöllänen		
Title		
Reproducibility and Validity Studies and Seated Flexion Test		
Type of Work	Date	Pages
Final Project	November 2006	35 + 6 appendices
<p>ABSTRACT</p> <p>Purpose of this Final Project is to make reproducibility and validity studies easier for professionals of Manual Medicine to produce. Tests and treatment procedures used in Manual Medicine have not been studied enough to scientifically confirm their usefulness. For this project we have translated the FIMM (International Federation for Manual and Musculoskeletal Medicine) reproducibility and validity study protocol into Finnish. With this protocol the professionals of Manual Medicine in Finland can make reproducibility and validity studies to be published internationally. Research studies done by this protocol are comparable and easy to reproduce. The goal of this project is to make the protocol more familiar in Finland. We hope that with this protocol it will become more popular for professionals of manual medicine to produce this kind of studies for the purpose of professionalism.</p> <p>We produced a kappa study of the seated flexion test according to FIMM protocol. There are no earlier reproducibility and validity studies done of this SI joint mobility test. In the Final Project we have introduced different ways to carry through the seated flexion test. Because of limited time all the stages or golden rules of the protocol were not completed at the required level. At the test stage the kappa value remained under the clinically meaningful value of 0,4. Because of lack of time and the fact that we reached only a poor kappa value we cannot make good conclusions of the reproducibility and validity of the seated flexion test.</p> <p>As an only conclusion we can state that more this kind of studies are needed.</p>		
Keywords		
kappa, reproducibility, validity, seated flexion test, osteopathy		

## SISÄLLYS

1 JOHDANTO	1
2 KAPPA- TUTKIMUS	4
2.1 Kappa-arvon laskeminen	5
2.1.1 Kappa-arvon yhtälö	5
2.1.2 Yhtäpitävyyden (Po) vaikutus kappa-arvoon	6
2.1.3 Oletetun sattumayhtäpitävyyden laskeminen (Pc)	7
2.1.4 Kontingenssitaulukko	7
2.2 Esiintyvyyden vaikutus kappa-arvoon	8
2.3 Erilaiset kappa- tutkimusmenetelmät	10
2.3.1 Tavallinen kappa	10
2.3.2 Yleinen kappa	11
2.3.3 Painotettu kappa	11
3 AIKAISEMPIA TUTKIMUKSIA	11
3.1 Aikaisempia SI-nivelen tutkimuksia	12
3.2 Aikaisempia kappa- tutkimuksia	14
4 SEATED FLEXION- TESTI	15
4.1 Testattavan valmistelu (liite 3)	16
4.2 Tutkijan asento ja SI-nivelen palpaatio (liite 4)	16
4.3 Testattavan suorittama liike	17
4.4 Testin tulkinta	18
4.5 Käytetty testaustapa ja tuloksen tulkitseminen	19
5 KAPPA- TUTKIMUS SEATED FLEXION- TESTISTÄ	20
5.1 Tutkimusolosuhteet, kirjanpito	20
5.2 Harjoitteluvaihe	22
5.3 Yhtäpitävyysvaihe	24
5.4 Testivaihe	25
5.5 Statiikka ja johtopäätökset	26
6 POHDINTA	28
7 LÄHTEET	33
LIITTEET	
1. FIMM toistettavuus- ja validiteettitutkimusprotokollan käännös	
2. Tutkimustilan asettelu (valokuvat)	
3. Testattavan valmistelu (valokuvat)	
4. Tutkijan käsien asettelu (valokuvat)	
5. Testattaville annettava lomake	
6. Tutkijoiden täyttämä lomake	

## 1 JOHDANTO

Tämän opinnäytetyön tarkoituksena on tehdä toistettavuus- ja validiteettitutkimuksien tekeminen helpommaksi ja tuoda tämän tyyppiset tutkimukset tunnetuiksi myös Suomessa. Työtä varten on käännetty FIMM:n (Kansainvälinen manuaalisen lääketieteen yhdistys) toistettavuus- ja validiteettitutkimusprotokolla (liite 1). Kyseisen protokollan mukaisesti toteutettiin kappi- tutkimus SI-nivelen (art. sacroiliaca) liikkuvuutta testavasta seated flexion- testistä. Käännöksen myötä toivotaan manuaalisten hoitoalojen ammattilaisten tekevän kasvavassa määrin vastaavia tutkimuksia ammatillistumisen lisäämiseksi. (FIMM 2004: 7-8; Kuchera 2006.)

Manuaalisessa lääketieteessä käytetään testejä ja hoitotekniikoita arvioimaan kehon tuki- ja liikuntaelimestön toimintoja. Perinteisesti manuaalisten hoitomuotojen ammattilaisille on riittänyt hoitojen kliininen näyttö niiden toimivuudesta. Tieteellistä tutkimusta on aina tehty, mutta selkeät ja järjestelmälliset tutkimusmenetelmät ovat puuttuneet. Nykypäivänä on ymmärretty tieteellisesti todistettavan näytön tärkeys myös manuaalisessa lääketieteessä. FIMM:n määrittelemän toistettavuus- ja validiteettitutkimusprotokollan mukaisesti tehtyjen tutkimusten avulla voidaan parantaa huomattavasti eri koulukuntien välistä kommunikaatiota ja manuaalisten alojen ammatillistumista. Tutkimuksilla voidaan myös tehostaa hoitojen vaikutusta ja toimivuutta tuomalla esille selkeästi näyttöön perustuvia testejä ja hoitotekniikoita. (FIMM 2004: 8.)

FIMM on laatinut kattavan ja laajan toistettavuus- ja validiteettitutkimusprotokollan, jota on helppo käyttää pienilläkin klinikoilla. Sen avulla voi jokainen koulutettu manuaalisen hoitoalan ammattilainen tai opiskelija tehdä toistettavuus- ja validiteettitutkimuksia. Protokollan mukaisesti tehdyt tutkimukset ovat vertailukelpoisia ja helposti toistettavia. Tutkimukset eivät vaadi materiaalisia resursseja ja tämän vuoksi ne ovat edullisia suorittaa. (FIMM 2004: 7-8.)

SI-nivel säätelee painonjakautumista alaraajojen ja yläkehon välillä sekä toimii iskunvaimentimena alaraajoista tuleville voimille. SI-nivel on tärkeä osa ihmisen kykyä kävellä pystyasennossa. (Kapandji 1997: 56.) SI-nivel on ollut kautta aikojen manuaalisten hoitoalojen kiistellyimpiä kohteita sen moninaisten oireiden ja vaikean tutkittavuuden

den sekä hoidon johdosta. SI-nivel heijastaa kipua alaselkään ja alaraajoihin. SI-nivelille tyypillistä on myös toisin kuin muille nivelille, että liikkuvampi puoli oireilee usein enemmän kuin aliliikkuva puoli. (Kuchera 2006.)

Manuaalisilla hoitoaloilla on kehitetty monenlaisia SI-nivelen kipuprovokaatio- ja mobiliteettitestejä, joiden toimivuutta, tarkkuutta ja käytettävyyttä on kritisoitu vähäisen tieteellisen tutkimusaineiston perusteella. Joidenkin lähteiden mukaan SI-nivelen liikkuvuutta on mahdoton havaita manuaalisesti. (Hestboek - Leboeuf- Yde 2000: 258-263.) Kuitenkaan ilman hyvin tehtyjä validiteetti- ja toistettavuustutkimuksia ei voida vetää selkeitä johtopäätöksiä testien toimivuudesta tai toimimattomuudesta. Ongelmana on ollut myös SI-nivelen kohdalla, kuten manuaalisissa hoitomuodoissa laajemmaltikin, selkeästi määriteltyjen tutkimusprotokollien puutteellisuus.

SI-niveltä on ollut vaikea tutkia siinä esiintyvien anatomisten poikkeavuuksien takia. Osalla ihmisistä sen on todettu luutuvan keski-ään jälkeen ja toisilla olevan hyvinkin liikkuva aina vanhuusikään asti. (Sizer - Phelps - Thompsen 2002.) Monen testin on oletettu testaavan SI-nivelen toimintaa, vaikkei testeille ole edes määritelty toistettavuus- ja validiteettiarvoa. Kappa-arvo on yhdenlainen näitä ominaisuuksia mittaava arvo, joka ei ole prosentuaalinen vaan määräytyy tietyn laskentakaavan mukaisesti. Laskentakaavalla saadaan prosentuaalisesta tuloksesta vähennettyä sattuman vaikutus lopulliseen tulokseen. (Metsämuuronen 1997.)

Seated flexion- testi valittiin tutkimuskohteeksi pääasiassa kahdesta syystä. Ensimmäinen kyseisestä testistä ei löytynyt aikaisempia kappa- tutkimuksia. Toisaalta seated flexion- testin suorittamisella ei ole oletettavasti suurta vaikutusta SI-nivelen liikelaajuuteen. Näin oletetaan, koska SI-nivel on valmiiksi vähäliikkeinen ja vahvasti ligamenteilla tuettu nivel verrattuna liikkuvampiin niveliin. Testin suorittamisessa ei myöskään käytetä selkeää ulkopuolista niveltä mobilisoivaa voimaa. Testin aiheuttama liikelaajuuden normalisoituminen voi vaikuttaa testitulosta heikentävästi (FIMM 2004:19).

Kappa- tutkimus on helppo ja yksinkertainen kansainvälisesti standardoitu tapa tehdä toistettavuus- ja validiteettitutkimuksia. Michael Kuchera D.O. FAAO (Research department of Philadelphia College of osteopathic medicine) oli mukana valitsemassa sopivaa tutkimusmenetelmää. Hän toimii tällä hetkellä FIMM:n varapääjohtajana ja on

luennoinut manuaalisen lääketieteen ammattilaisille myös Suomessa. Kuchera on ollut suureksi avuksi kaikissa vastaan tulleissa ongelmissa työn edistyessä.

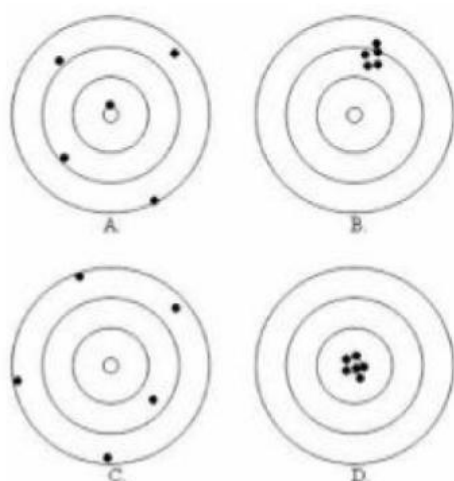
Käännöstyön pohjalta on mahdollista hyödyntää tutkimusprotokollaa tarkemmin. Työssä esiintyy jonkin verran englanninkielisiä termejä ja testien nimiä puuttuvan nimikkeistön takia.

## 2 KAPPA- TUTKIMUS

Manuaalisessa lääketieteessä käytetään diagnostisia testejä ja hoitoja, joista kaikkia ei ole vielä validioitu. Testien reliabiliteetin määrittäminen olisi erittäin olennaista, jotta erilaisten hoitomuotojen vaikuttavuutta voitaisiin vertailla. Reliabiliteetin määrittämisessä käytetään nykyään yleisesti kappastatistiikkaa. (Sim – Wright 2005: 258.)

Reliabiliteetti muodostuu sekä toistettavuudesta että validiteetista (Viera – Garrett 2005: 360). Manuaalisessa lääketieteessä testin toistettavuus kuvastaa löytävätkö kaksi tutkijaa samat löydökset samasta tutkimusjoukosta, tai löytääkö yksittäinen tutkija samat löydökset kahtena eri ajankohtana samasta tutkimusjoukosta. Eli tutkijan sisäinen yksimielisyys (intra-observer agreement) tarkoittaa saman tutkijan kahtena eri ajankohtana tekemien havaintojen välistä toistettavuutta. Tutkijoiden välinen yksimielisyys (inter-observer agreement) tarkoittaa kahden tutkijan tekemien havaintojen toistettavuutta samana ajankohtana. (FIMM 2004: 9.)

Validiteetti kuvaa, miten hyvin testi mittaa asiaa, jota sen oletetaan mittaavan (kultainen standardi: golden standard, kts. liite 1) (FIMM 2004: 9). Alla on esitetty kuvio 1 havainnollistamaan termejä. Keskipiste kuvastaa kultaista standardia. Jos osutaan keskipisteeseen, on tuloksella validiteetti. Jos kaikki osumat menevät samaan kohtaan, on tuloksella toistettavuus. On kuitenkin mahdollista, että keskipisteeseen osutaan vain sattumalta. Kohdassa A yksi osuma on validi (keskipiste), mutta osumilla ei ole keskenään toistettavuutta. Kohdassa B on osumilla selvästi toistettavuus, mutta validiteetti puuttuu. Kohdassa C ei ole kumpaakaan ominaisuutta. Kohdassa D osumilla on sekä hyvä toistettavuus että validiteetti. (Viera – Garrett 2005: 360.)



KUVIO 1. Toistettavuus ja validiteetti.

A kohta: huono toistettavuus, alhainen validiteetti

B kohta: hyvä toistettavuus, huono validiteetti

C kohta: huono toistettavuus ja validiteetti

D kohta: hyvä toistettavuus ja validiteetti

(Viera – Garrett 2005: 361.)



Tutkimuksissa kaksi tutkijaa saattavat antaa sattumalta saman arvion. Jotta sattuman vaikutus testituloksiin otetaan huomioon, on tutkimuksia varten kehitetty erilaisia korjauskertoimia. Tämän alueen tunnetuin mitta on Cohenin kappaa. (Metsämuuronen 1997.) Cohenin kappaa on tilastollinen mitta tutkijoiden välisestä luotettavuudesta (inter-rater reliability). Eli Cohenin kappaa mittaa yhtäpitävyyttä kahden tutkijan välillä. Kappaa-arvoa pidetään vankempaa mittana kuin yksinkertaista prosentuaalista yhtäpitävyyden laskemista, koska kappaa-arvo ottaa huomioon myös sattuman. Jos tutkijoita on enemmän kuin kaksi, pitää käyttää Fleissin kappaa. (Wikipedia 2006.) Sattumakorjatun yksimielisyyskertoimen ideana on se, että halutaan minimoida sattumalta syntyneen yksimielisyyden aiheuttama harha (Metsämuuronen 1997).

## 2.1 Kappaa-arvon laskeminen

Kappaa-arvon laskeminen perustuu eroon yhtäpitävyyden (overall agreement: kuinka suuri yhtäpitävyys on tutkimuksessa) ja oletetun sattumayhtäpitävyyden välillä (expected change agreement: kuinka suuri yhtäpitävyys esiintyisi pelkällä sattumalla). Yhtäpitävyys on siis prosenttiosuus koko tutkimuksen tutkittavista, joista molemmat tutkijat ovat yhtämieltä. (Viera – Garrett 2005: 361.)

Kappaa-arvon laskemiseen on internetistä saatavilla monenlaisia taulukkolaskentaohjelmia, joissa kappaa-arvon laskemiseen vaadittavat kaavat ovat valmiina (Chuang 2002). Tällä tavalla vain perustiedot lisätään kaavaan ja ohjelma laskee automaattisesti kappaa-arvon.

Seuraavissa kappaleissa (2.1.1- 2.1.4) käydään läpi kappaa-arvon yhtälöön vaadittavien arvojen laskentakaaviot. Kaaviot helpottavat ymmärtämään, mistä kappaa-arvo todellisuudessa koostuu.

### 2.1.1 Kappaa-arvon yhtälö

Kaavassa 1 esitetään kappaa-arvon laskentakaava, jossa  $P_o$  on yhtäpitävyys ja  $P_c$  on oletettu sattuma yhtäpitävyys (FIMM 2004: 34). Eli  $P_o$  on osuus, joista tutkijat ovat yhtämieltä.  $P_c$  on osuus, joka voidaan saavuttaa sattumalta (Sim – Wright 2005: 258).

$$\text{Kappa} = \frac{p_o - p_c}{1 - p_c}$$

KAAVA 1. Kappa-arvon kaava (FIMM 2004: 35).

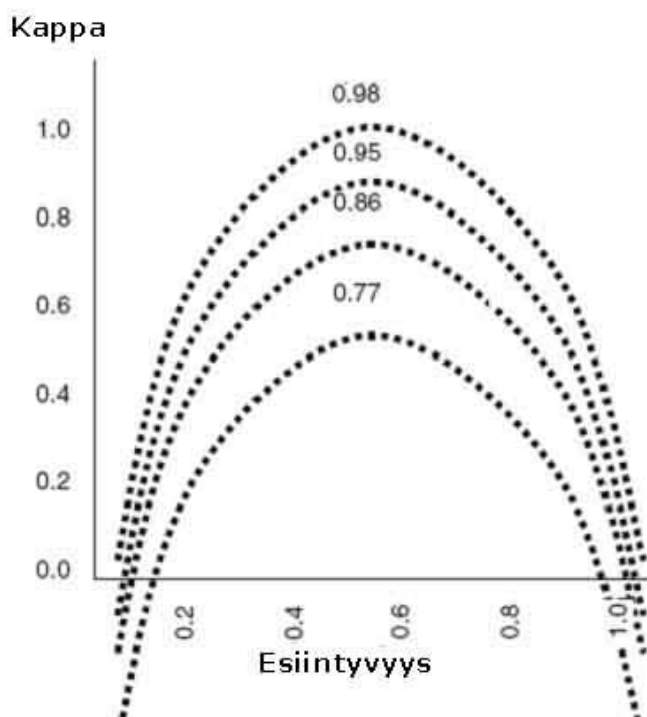
### 2.1.2 Yhtäpitävyyden ( $P_o$ ) vaikutus kappa-arvoon

Yhtäpitävyys lasketaan alla esitetyn kaava 2:n mukaisesti. Kyseiset arvot ( $a$ ,  $d$ ,  $n$ ) saadaan kontingenssitaulukossa, joka on esitetty kuviossa 3 sivulla 7.

Yhtäpitävyys: 
$$P_o = \frac{[a + d]}{n}$$

KAAVA 2. Yhtäpitävyyden kaava (FIMM 2004: 11).

Kappa-arvon taso on riippuvainen tutkijoiden yhtäpitävyydestä. Mitä alhaisempi yhtäpitävyys toistettavuustutkimuksessa on, sitä alhaisemmat ovat suurin ja pienin kappa-arvo. Tämä yhteys esitetään kuviossa 2. Kuvioista käy ilmi, että esiintyvyyss-/kappakäyrässä alhaisella yhtäpitävyyden arvolla (0.86 ja 0.77) alhaisin kappa-arvo on negatiivinen. (FIMM 2004: 22.)



KUVIO 2. Yhteys eri esiintyvyyss-/kappakäyrien ja yhtäpitävyyden arvojen 0.77-0.98 välillä (FIMM 2004: 22).

Kuviossa 2 osoitetaan, että korkealla yhtäpitävyyden (0.98 kuviossa) arvolla suurin kappale-arvo on 1.0 ja alhaisin kappale-arvo on lähes 0 (FIMM 2004: 22).

### 2.1.3 Oletetun sattumayhtäpitävyyden laskeminen (Pc)

Alla esitetyssä kaavassa 3 käydään läpi, miten oletettu sattumayhtäpitävyys lasketaan. Seuraavassa kappaleessa (2.1.4) esitetystä 2 x 2 kontingenssitaulukosta (kuvio 3) voidaan nähdä parhaiten, mitä kaavioon vaadittavat arvot ovat.

Oletettu sattumayhtäpitävyys: 
$$P_c = \frac{a+b}{n} \times \frac{a+c}{n} + \frac{c+d}{n} \times \frac{b+d}{n}$$

KAAVA 3. Oletetun sattumayhtäpitävyyden kaava (FIMM 2004: 35).

### 2.1.4 Kontingenssitaulukko

Kappatulokset esitetään usein 2 x 2 kontingenssitaulukon avulla. Taulukon yhteydessä olisi myös erittäin tärkeää ilmaista esiintyvyys ja yhtäpitävyys. (FIMM 2004: 29.) Alla olevassa 2 x 2 kontingenssitaulukossa (kuvio 3) on esimerkki kahden tutkijan (A ja B) toistettavuustutkimuksen tuloksista.

		Tutkija B		Yhteensä:
		Kyllä	Ei	
Tutkija A	Kyllä	a	b	a+b
	Ei	c	d	c+d
Yhteensä:		a+c	b+d	n

KUVIO 3. 2 x 2 kontingenssitaulukko (FIMM 2004: 10).

a= Tutkija A ja B yhtämieltä (kyllä)

b= Tutkija A ja B erimieltä (A-kyllä, tutkija B-ei)

c= Tutkija A ja B erimieltä (A-ei, tutkija B-kyllä)

d= Tutkija A ja B yhtämieltä (ei)

n= Testattavien kokonaismäärä

Taulukossa ruuduista a ja d nähdään tutkittavat, joista molemmat tutkijat ovat yhtämielisiä (positiivinen/negatiivinen). Ruuduista b ja c nähdään tutkittavat, joista tutkijat ovat erimieltä. Tutkija B:n arvioimat tutkittavat näkyvät alareunassa ( $a+c=kyllä$ ,  $b+d=ei$ ) ja tutkija A:n arvioimat tutkittavat näkyvät oikeassa reunassa ( $a+b=kyllä$ ,  $c+d=ei$ ). (Sim – Wright 2005: 259.)

Kappa-arvot voivat olla negatiivisia tai positiivisia ja ne voivat vaihdella välillä -1 ja +1. Kappa-arvojen merkitsevyyttä arvioidaan yleensä Koch ja Landisin tekemän asteikon mukaisesti, joka esitetään alapuolella (taulukko 1). (Viera - Garrett 2005: 362.)

Kappa-arvo	Yhtäpitävyyden voimakkuus
-1- 0	ei yhtäpitävyyttä
0.0- 0.19	olematon yhtäpitävyys
0.20- 0.39	heikko yhtäpitävyys
0.40- 0.59	kohtalainen yhtäpitävyys
0.60- 0.79	hyvä yhtäpitävyys
0.80- 1.00	erittäin hyvä yhtäpitävyys

TAULUKKO 1: Yhtäpitävyyden tulkitseminen kappa-arvosta (Koch - Landis 1977: 159- 174).

FIMM- protokollassa käytetään kappa-arvoa 0.6 ilmaisemaan hyvää tai erinomaista yhtäpitävyyttä (FIMM 2004: 11). Nolla ilmaisee täysin sattumanvaraista yhtäpitävyyttä eli mikä voidaan saada aikaiseksi pelkällä sattumalla. Negatiivinen arvo viittaa sattumanvaraistakin tulosta huonompaan yhtäpitävyyteen. (Sim – Wright 2005: 259.)

## 2.2 Esiintyvyyden vaikutus kappa-arvoon

Esiintyvyys kuvastaa kuinka paljon jotakin tautia/ kliinistä löydöstä (tunnusmerkki: index condition) on kyseisessä joukossa. Toistettavuustutkimuksissa tunnusmerkin esiintyvyys liittyy ainoastaan tutkimusjoukkoon. (FIMM 2004: 10.) Erilaisissa tutkimuksissa esiintyvyys voi vaihdella erittäin paljon. Tunnusmerkin esiintyvyys voi vaihdella eri joukkojen (ikä, sukupuoli, kansallisuus) välillä. Se voi myös muuttua eri ajan-kohtina. (FIMM 2004: 10.) Tutkimusten mukaan väestö sisältää useita erilaisia alaryhmiä, joilla jokaisella on erilainen esiintyvyys käytettäviin testeihin. Jos tutkimus koos-

tuu pienestä alaryhmästä, voi tutkimukseen liittyä riski, että osalla testeistä on alhainen esiintyvyys. (FIMM 2004: 16.)

Esiintyvyys vaikuttaa suoraan kappa-arvoon, joten julkaistaessa tutkimuksen tuloksia molemmat arvot on erittäin tärkeä mainita. Ilman esiintyvyyden ilmoittamista kappa-arvon luotettavuutta ei voida todeta. (FIMM 2004: 21; Sim – Wright 2005: 262.)

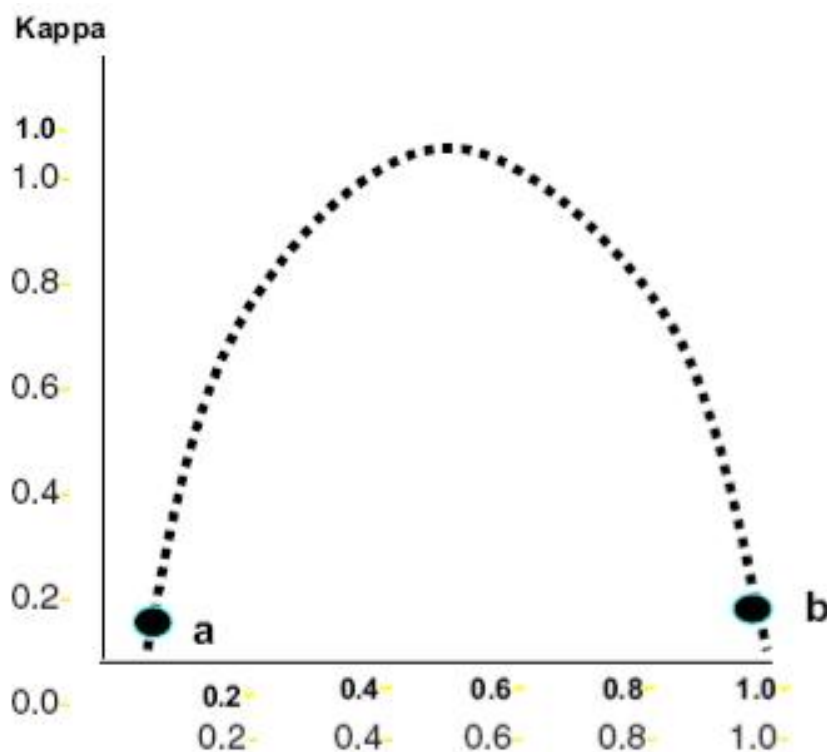
Tutkimuksessa esiintyvyyden olisi hyvä olla 0.50, jolloin se ei vaikuta kappa-arvoon (FIMM 2004: 22). Jos esiintyvyys on erittäin suuri tai pieni, sattumayhtäpitävyys on myös erittäin korkea. Tätä kautta esiintyvyys vaikuttaa kappa-arvoa laskevasti. (Sim – Wright 2005: 261.) Monet toistettavuustutkimukset on todettu kliinisesti merkityksettömiksi alhaisen kappa-arvon takia, mutta tulosten yhteydessä ei ole mainittu esiintyvyyttä. Alhaiset kappa-arvot voivat kuvastaa yhtä hyvin niin suurta kuin pientäkin esiintyvyyttä. (FIMM 2004: 22.)

Esiintyvyys lasketaan kontingenssitaulukosta saatavista arvoista eli kaavaan otetaan arvot ruuduista a, b ja c. Nämä arvot kuvaavat molempien tutkijoiden positiivisiksi arvioituiden tapaukset potilaiden kokonaismäärästä (n). (FIMM 2004: 10.) Eli esiintyvyys (P) lasketaan kaavalla:

$$\text{Esiintyvyys: } P = \frac{a + (b + c)/2}{n}$$

KAAVA 4. Esiintyvyyden kaava (FIMM 2004: 10).

Kuvio 4 esittää kappa-arvon riippuvuuden esiintyvyydestä. Kuviosta voidaan huomioda, että erittäin alhaisella (a) ja korkealla (b) esiintyvyydellä kappa-arvosta tulee erittäin pieni (FIMM 2004: 21).



KUVIO 4. Esiintyvyyden ja kappa-arvon välinen yhteys. Kohdassa a esiintyvyys on erittäin alhainen, kohdassa b esiintyvyys on erittäin korkea. (FIMM 2004: 21.)

### 2.3 Erilaiset kappa- tutkimusmenetelmät

FIMM- protokollan mukaan manuaalisen lääketieteen käytössä on kolme erilaista kappa- tutkimusmenetelmää, joista valitaan jokaiseen tutkimukseen aina soveltuvin vaihtoehto. Nämä erilaiset tutkimusmenetelmät ovat tavallinen kappa, yleinen kappa ja painotettu kappa. (FIMM 2004: 13.)

#### 2.3.1 Tavallinen kappa

Tavallinen kappa valitaan, kun arvioidaan toistettavuutta kahden tutkijan välillä ja käytössä on luokitteluasteikko (kyllä/ei) (FIMM 2004: 13). Yleinen esimerkki tilanteesta, jolloin tutkija haluaa arvioida potilasta luokitteluasteikolla, on määrittää jonkin taudin tai tilan läsnäolo tai puuttuminen. (Sim – Wright 2005: 258.)

### 2.3.2 Yleinen kappa

Yleistä kappa voidaan käyttää, kun tutkijoita on enemmän kuin kaksi. Yleinen kappa antaa arvion eri tutkijoiden välisestä yleisestä toistettavuudesta löytää dysfunktio joltain alueelta, esim. arvioidessa segmentaalisen dysfunktion esiintymistä tietyllä alueella. (FIMM 2004: 13.)

### 2.3.3 Painotettu kappa

Painotettua kappa käytetään, kun testin tulos jakautuu useampaan asteluokkaan luontaisella järjestyksellä, esimerkiksi loppujousto (normaali, elastinen, kova) (FIMM 2004: 14). Painotetussa kapassa voidaan pisteytyksellä vaikuttaa siihen, kuinka suuri eroavaisuus tuloksilla on. Esimerkiksi kipua arvioidessa voidaan käyttää asteikkona ”ei kipua”, ”vähän kipua”, ”kohtuullinen kipu” ja ”vakava kipu”. Tämänlaisessa tilanteessa eroavaisuus yhden kategorian välillä (ei kipua – vähän kipua) ei ole niin suuri kuin kahden kategorian välillä (ei kipua – kohtuullinen kipu). Kun tutkimuksessa käytetään painotettua kappa, saadaan näkyville erot pienten ja suurten erimielisyyksien välillä. (Sim – Wright 2005: 259.)

## 3 AIKAISEMPIA TUTKIMUKSIA

Kappa-arvo on tilastotieteellinen laskennallinen arvo, jota voidaan soveltaa monenlaisen käyttöön. Kappa- tutkimuksia tehdään myös muiden alojen toimesta. Kappaleeseen 3 on kerätty muutamia SI-nivelen testien tutkimuksia ja manuaalisen lääketieteen kappa- tutkimuksia pääasiassa lantion ja alaselän alueelta. Nämä tutkimukset on kerätty työhön esimerkiksi SI-nivelen testien toimivuudesta ja saaduista tuloksista. Tutkimusten johtopäätöksistä voidaan yhteenvetona päätellä, että tutkimuksia pitää tehdä useita samalla tavalla ja samaa testiä/ aluetta koskien, jotta voidaan selvittää kaikki tuloksiin vaikuttavat tekijät ennen lopullisia johtopäätöksiä testien validiteetista.

### 3.1 Aikaisempia SI-nivelen tutkimuksia

Van Deursen ja Patijn tutkivat SI-niveltä kuudella eri testillä ja saivat kohtalaisesta erittäin hyvään olevat kappa-arvot testeille. Johtopäätöksenä tutkimuksesta oli, että SI-niveltä diagnosoitaessa tulisi käyttää kolmea hyväksi todettua testiä. Tutkimuksessa oli mukana kolme tutkijaa. Tutkimuksessa arvioitiin sekä tutkijoiden sisäistä yksimielisyyttä että tutkijoiden välistä yksimielisyyttä. (FIMM 2004: 9,16).

Laslett, Aprill ja McDonald (2005) tutkivat SI-nivelen kipuprovoakaatiotestien diagnostista arvoa ja niiden yhdistelmiä suhteessa hyväksytyyn kriteeristandardiin. Tutkimuksessa fysioterapeutit tutkivat 48 potilasta käyttäen kuutta SI-nivelen kipuprovoakaatiotestiä. Potilaat saivat paikallisen puudutusinjektion SI-niveleen testauksien jälkeen. Testien diagnostista arvoa arvioitiin yksittäin ja usean testin yhdistelminä. Kaikki potilaat, joilla oli positiivinen vaste diagnostiseen injektioon, ilmoittivat kipua myös vähintään yhdesessä SI-nivelen testissä. Sensitiivisyys (sensitivity, kts. liite 1) ja spesifisyys (specificity, kts. liite 1) oli 94 % ja 78 %, kun vähintään kolme SI-nivelen testiä oli positiivisia. Tutkimuksesta pääteltiin, että SI-nivelen provoakaatiotestien yhdistelmät ovat arvokkaita oireellisen SI-nivelen kliinisessä diagnoosissa. Paras ennustusarvo suhteessa intra-artikulaariseen puudutusinjektioon saadaan, kun tehdään kolme tai useampi kuudesta testistä (distraktio, kompressio, reiden kautta tehty SI-nivelen impulssi (thigh trust), SI-nivelen impulssi (SI trust), Gaenslenin testi molemmin puolin) tai mitkä tahansa kaksi neljästä tutkimuksessa valitusta testistä (distraktio, kompressio, reiden kautta tehty SI-nivelen impulssi, SI-nivelen impulssi). Jos mikään kuudesta provoakaatiotestistä ei provosoi kyseistä kipua, voidaan SI-nivel sulkea pois sen hetkisen alaselkävun aiheuttamisesta. (Laslett - Aprill - McDonald - Young 2005: 207-218).

Slipman ym. (1998) tutkivat SI-nivelen provoakaatiotestien luotettavuutta verrattuna kultaiseen standardiin. Kultainen standardi SI-nivel diagnostiikassa on fluoroskooppisesti ohjattu puudutuspistos SI-niveleen. (Huijbregts 2004). 50 potilasta testattiin injektioilla, jos heillä oli ainakin kolmen manuaalisen testin mukaan positiivinen SI-nivelen dysfunktio. Manuaaliset testit sisälsivät aina sekä Patricin testin että sulkus sacralisin ipsilateraalisen arkuuden palpoinnin. Kolmas testi oli mikä tahansa SI-nivelen provoakaatiotesti. Hypoteesina tutkimuksessa oli, että jos injektio auttaisi manuaalisesti löydettyyn dysfunktioon, diagnoosi oli ollut oikea. Tulosten mukaan 30 potilasta 50:stä sai ainakin 80 % avun ja 20 potilasta sai vähemmän kuin 80 % apua. Joten manuaalisilla



testeillä löydettiin 60 %:lla tarkkuudella SI-nivelen dysfunktio. Johtopäätöksenä tutkimuksessa oli, etteivät provokaatiotestit ole tarkkoja SI-nivelen diagnostiikassa. (Chaitow 2003: 16.)

O'Haire ja Gibbonsin (2000) tutkimuksessa osteopaattiopiskelijat osoittivat kolmen yleisen lantion anatomisen maamerkin palpoinnin olevan erittäin vaikeaa. Kymmenen opiskelijaa tutki kymmenen kivutonta testattavaa. Tutkijoiden sisäinen luotettavuus oli ILA:n (Inferior Lateral Angle, Cornua Sacralis) kohdalla pienempi kuin sattuma. SIPS:n (Spina iliaca posterior superior) ja sulcus sacraliksen kohdalla luotettavuus oli keskinertainen. Tutkijoiden välinen luotettavuus oli heikko. Tästä johtopäätöksenä arveltiin, että monien aikaisemmin mainittujen testien luotettavuus on voinut jäädä heikoksi palpaatiokohtien epäspesifisyydestä johtuen. (Chaitow 2003: 16.)

Chaitowin Palpation skills- kirjassa esitetyssä tutkimuksessa todettiin polven nostotestin eli Gilletin testin variaatioineen olevan epäluotettava SI-nivelen tutkimisessa. Testi on kuitenkin ollut yksi käytetyimpiä testejä. Kyseisessä tutkimuksessa todettiin yhden testin käyttämisen olevan epäluotettavaa SI-nivelen liikerajoituksen määrittämiseksi. Monen muunkin tutkimuksen johtopäätöksissä on todettu, ettei manuaalisessa diagnostiikassa voida luottaa vain yhteen testitulokseen. (Chaitow 2003: 17- 19.)

Kiropraktikot suorittivat tutkimuksen alaselkäkivusta 83 kaksosilla. Tutkimus sisälsi useiden eri diagnostisten testien käyttöä, observaatioon perustuvista testeistä liikekivun tutkimiseen ja kipuprovoikaatioon. Heillä oli testien kriteereinä 1. looginen yhteys alaselkäkipuun, 2. korkea sensitiivisyys, 3. korkea spesifisyys, 4. korkea positiivinen ja 5. korkea negatiivinen ennustearvo (predictive value). Tulokseksi saatiin, ettei yksikään testi sellaisenaan täyttänyt kaikkia kriteerejä. Kivuliaista liikkeistä kuitenkin ainakin yksi täytti kaikki viisi kriteeriä. Jokainen lannerangan liikkeistä, lukuun ottamatta vasenta rotaatiota, täyttivät kaikki muut kriteerit paitsi korkean sensitiivisyyden. Johtopäätöksenä oli myös heillä, että vaikei yksittäinen testi ollutkaan tarkka, niin perustuen moniin testeihin pystyttiin erottelemalla tekemään tyydyttävä diagnoosi. (Chaitow 2003: 18.)

### 3.2 Aikaisempia kapp- tutkimuksia

Smith ja Gibbons (1999) osoittivat standing flexion- testin luotettavuuden kyseenalaiseksi SI-nivelen tutkimisessa. Kapp-arvo, joka kertoi tutkijoiden välisen yksimielisyyden, oli 0,052 (42 %) ja tutkijoiden sisäinen yksimielisyys oli 0,46 (68 %). (Chaitow 2003: 16.)

McKenzie ja Taylor (1997) tutkivat 14 fysioterapeutin välistä luotettavuutta rangan segmentaaliossa palpaatiossa. Fysioterapeutit palpoivat viisi tervettä tutkittavaa, joille oli merkattu tietyt nikamat ultraviolettimusteella. Tutkimuksessa myös kolme fysioterapeuttia tutki viisi tervettä tutkittavaa terapeuttien sisäisen luotettavuuden määrittämiseksi. Ensimmäisen kohdan tulos jäi heikoksi ja kapp-arvoksi saatiin vain 0.28, mutta tutkijoiden sisäisessä luotettavuudessa saavutettiin merkityksellisestä lähes täydelliseen olevat tulokset kapp-arvojen vaihtelun olleen 0.61- 0.90 välillä. Heidän päätelmänsä oli, että lisää tutkimustietoa tarvitaan rangan alueen palpaation luotettavuudesta, jos palpaatiota aiotaan terapiatarkoituksessa käyttää. (Chaitow 2003: 17.)

Lewit ja Rosina (1999) analysoivat palpaation avulla tehtäviä SI-nivelen liikkeen diagnostisia tekniikoita. Kirjallisuudesta löytyy palpaation avulla tehtäville SI-nivelen testeille alhaisia tuloksia tutkijan sisäiselle ja tutkijoiden väliselle yhtäpitävyydelle. Tutkimuksessa testattiin 64 yliopistollisen sairaalan potilasta käyttäen uudenlaista tekniikkaa, jota verrattiin muihin tekniikoihin ennen ja jälkeen hoidon. Tutkijoiden väliseksi luotettavuudeksi saatiin kapp-arvo 0.68. He olivat yksimielisiä 82% tapauksista. Johtopäätöksenä he totesivat, että testin helppouden ja virheiden puuttumisen takia uusi testi oli arvokas. Testin mekanismi on kuitenkin heikosti ymmärretty ja se kaipaa jatkotutkimusta. (Lewis – Rosina 1999: 154- 160.)

Freburger ja Riddle (1999) tekivät tutkimuksen SI-nivelen linjauksen mittaamisesta. Tutkimuksen tarkoituksena oli määrittää, voidaanko manuaalisia gonio- ja inklinometrejä käyttää luotettavina mittareina SI-nivelen linjauksen määrittämiseen epäillessä SI-nivelen dysfunktioita. Inklinaatiokulma mitattiin molemmista innominaateista tutkittavan seistessä. Tutkimuksessa saatiin kapp-arvoksi 0.18. Johtopäätöksenä todettiin, että SI-nivelen linjauksen mittarit ovat epäluotettavia. Terapeuttien pitäisi harkita muunlaisia mittareita SI-nivelen arviointiin. (Freburger - Riddle 1999: 1134- 1141.)

Degenhardt ym. osoittivat tutkimuksessaan, että 1-2 tuntia viikoittaista harjoittelua neljän kuukauden aikana parantaa tutkijoiden välistä luotettavuutta lannerangan palpaatiossa. Kappa-arvot parantuivat vähintään kaksinkertaiseksi kaikilla neljällä mitatulla osaluueella. Arkuuden ja kudosuutosten palpaatiossa saatiin harjoittelun avulla kappa-arvot parantumaan yli kliinisesti merkittävän 0.4 raja-arvon. Arkuuden palpaatio sai korkeimman kappa-arvon, joka harjoittelua ennen oli 0.32 ja harjoittelun jälkeen 0.68. Kudosuutosten palpaatiossa saavutettiin suurin edistyminen, kun kappa-arvo nousi 0.12:sta arvoon 0.45. Arkuuden ja kudosuutosten palpaatiossa ylitettiin kliinisesti merkittävä 0.4 raja-arvo. Lannenikamien asennon asymmetrian palpaatiossa kappa-arvo kehittyi 0.17:sta arvoon 0.34. Asymmetrisen liikkeen palpaatiossa kappa-arvo kehittyi 0.10:sta arvoon 0.20. (Degenhardt - Snider - Snider - Johnson 2005.)

#### 4 SEATED FLEXION- TESTI

Seated flexion- testi on yksi monista manuaalisessa lääketieteessä käytetyistä SI-nivelen testeistä. Seated flexion- testillä pyritään arvioimaan sacrumin ja iliumin välistä liikettä, joka tapahtuu selkärangan eteentaivutuksessa. Testin tarkoituksena on löytää puoliero SI-nivelten liikkuvuuksien väliltä. Vaikka testin tarkoituksesta kirjallisuudessa ollaan yksimielisiä, poikkeavat testin suoritustavat lähteestä riippuen. Testin suoritustavan kuvauksista voi löytää päinvastaisiakin yksityiskohtia. Seuraavissa kappaleissa tarkastellaan erilaisia kuvauksia testin suoritustavoista. Suoritustavat on selvyuden vuoksi jaettu seuraavanlaisiin kappaleisiin: testattavan valmistelu, tutkijan asento ja palpaatio, testattavan suorittama liike ja testin tulkinta. Kappaleessa 4.5 kuvaillaan lisäksi kappa- tutkimuksessa käytetty testaus tapa seated flexion- testistä.

SI-niveltä tutkitaan pääasiassa kahdenlaisilla testeillä eli kipuprovoakaatio- ja mobiliteettitesteillä. Kipuprovoakaatiotesteihin kuuluvat fabere- testi eli Patricin testi, SI-nivelen kompressio- ja distraktiotestit, Gaenslenin testi, ligamenttien palpaatiotestit ja piriformis lihaksen palpaatio. (Chaitow 2003: 16- 17; Kuchera 1992: 445; Forst - Wheeler - Fortin - Vilensky 2006: 63.)

Mobiliteettitestejä ovat seated- ja standing flexion testi (Kuchera 1992: 427/439- 440), reiden kautta tehtävä kompressiotesti, SIAS kompressiotesti (Kuchera 1992: 441), sac-

rumin liiketestausta peukaloilla painaen ILA:sta ja sulcus sacralisista (Kuchera 1992: 446), Gilletin testi (Chaitow 2003: 18) ja Sfinksi-testi (Mitchell 2001: 124- 125).

Sulkuksen syvyydestä basis sacralisin määrittämiseksi on yhdentyypinen sacrumin asentotesti (Kuchera 1992: 442- 443). Yksinkertainen testi on Fortin sormitesti. Tässä testissä testattava itse osoittaa sormellaan kivun olevan kohdassa, joka sijaitsee SIPS:n heti mediaali- ja inferioripuolella. (Chaitow 2003: 16- 17.)

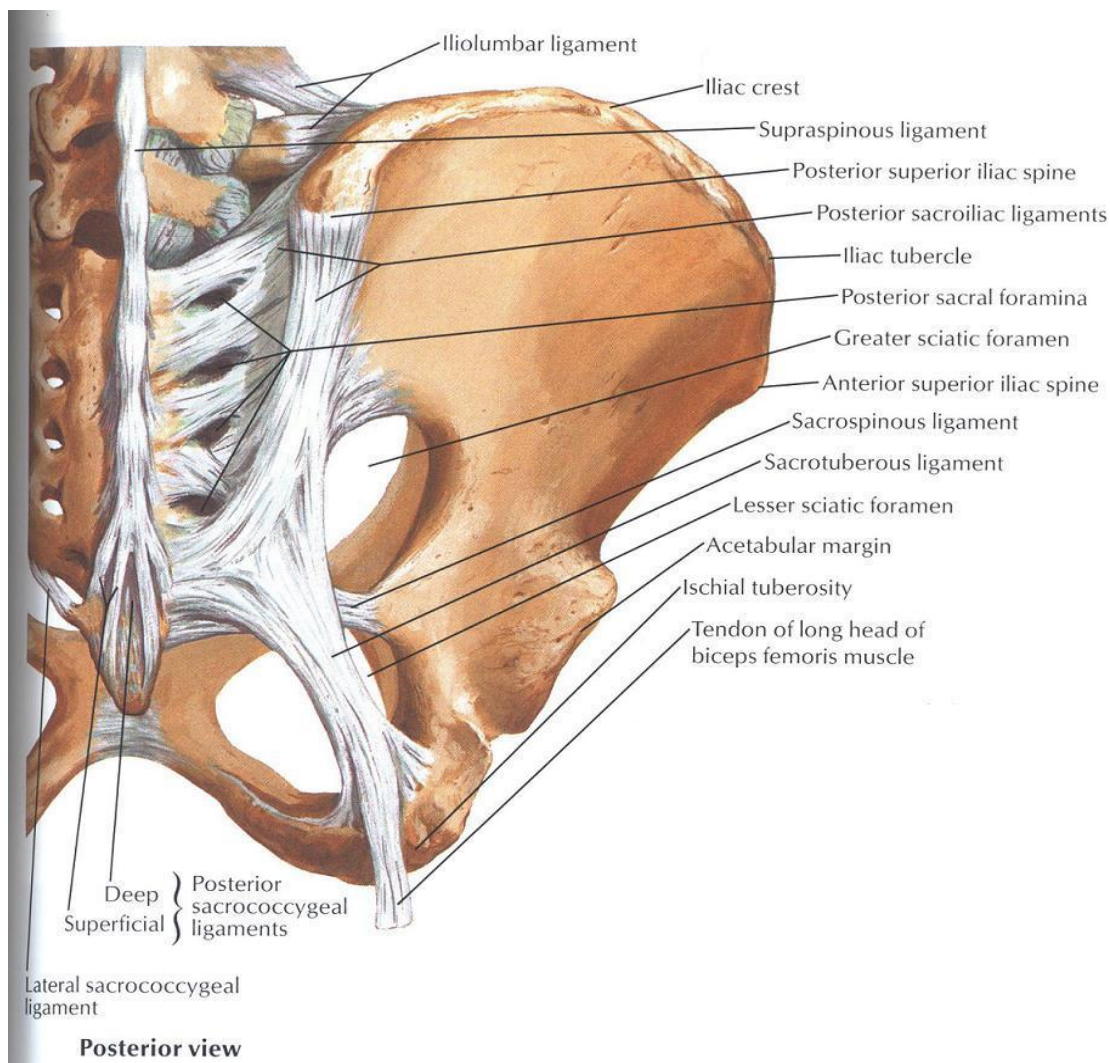
#### 4.1 Testattavan valmistelu (liite 3)

Yleisimmän ohjeen mukaan testattava istuu tasaisella tuolilla jalkapohjat koskettaen alustaan (Petty - Moore 2001: 285; Bandy 2005; Mitchell 2001: 88). Kapplerin tekemässä tutkimuksessa mainitaan kuitenkin, että testin voi suorittaa myös testattavan istuessa hoitopöydällä jalkojen roikkuessa ilmassa (Kappler 2000: 28). Bandy mukaan testattavan polvet ovat kaukana toisistaan (Bandy 2005). Tarkemmaksi polvien väliseksi etäisyydeksi Mitchell mainitsee hartioiden leveyden (Mitchell 2001: 88). Polvien ollessa lähes kiinni toisissaan hankaloituu testattavan vartalon taivuttaminen eteenpäin testin suoritusvaiheessa. Polvien ollessa liian kaukana toisistaan SI-nivelen joustoliike saattaa häiriintyä. (Kuchera 2006.) SIPS tasojen ollessa eri korkeudella voidaan asettaa toisen tuber ischiadicumin alle koroke (DiGiovanna 2005: 315).

#### 4.2 Tutkijan asento ja SI-nivelen palpaatio (liite 4)

Tutkija on testattavan takana joko kyykyssä, polvillaan tai tuolilla istuen. Tutkijan katseen korkeus on hyvä olla lähellä lantion tasoa. (Mitchell 2001: 5; DiGiovanna 2005: 315.) Palpoivien peukaloiden paikaksi mainitaan SIPS (Petty - Moore 2001: 285). Tarkemmin sijainti määritetään SIPS alaosaksi (DiGiovanna 2005: 315; Mitchell 2001: 5). Peukalot asetetaan ensin hivenen kohteen alapuolelle, jolloin peukaloilla voidaan tuoda pehmytkudosta ylöspäin. Tällöin peukalot eivät testiliikkeen aikana liiku SIPS:n yli pehmytkudosten kiristymisen vetämänä. (Kuchera 2006.) Vaihtoehtoiseksi palpaatiokohdaksi mainitaan tuberositas glutealis eli hymykuoppa. Sitä voidaan käyttää, jos SIPS:n palpoinnista osoittautuu vaikeaksi gluteus maximus lihaksesta johtuen. Hymykuoppa sijaitsee S1-tasolla ja SIPS S2-tasolla, ja niiden etäisyys toisistaan on noin 1-2cm. (kts. kuvio 5) (Mitchell 2001: 5.) Lähteistä ainoastaan Chaitow antaa viitettä peukalon

paineen määrästä. Chaitow sanoo peukaloiden kontaktin olevan napakka (Chaitow 2002: 334). Muiden sormien asettaminen crista iliacealle ja lantion sivulle auttavat liikkeen palpatiossa (DiGiovanna 2005: 315; Mitchell 2001: 88; Chaitow 2002: 334).



KUVIO 5. SI-nivelen alueen anatomiset maamerkit ja palpoitavat ligamentit (Netter 2006: 353).

#### 4.3 Testattavan suorittama liike

Testissä testattava taivuttaa vartaloaan eteenpäin (Petty - Moore 2001: 285). Eteentaivutusliike voidaan pyytää suorittamaan vyötäröstä (DiGiovanna 2005: 315). Liikkeen aikana kyynärpäät menevät joko polvien välistä tai ne liukuvat polvien päältä (Kuchera 2006; Mitchell 2001: 88; Greenman 2003: 349). Testattava voi laittaa kädet myös ristiin rinnalle (AOA 2002: 773). Taivutusliike suoritetaan mahdollisimman pitkälle ja hitaasti (Chaitow – DeLany 2002: 334). Taivutusliikkeen alkuvaiheessa tarkkaillaan merkkejä

lantion yläpuolisesta kireydestä eli lanne- ja rintarangan dysrytmiä, sivutaivutusta ja kiertymistä (Greenman 2003: 349). Testattavan eteentaivutuksen viimeiset asteet ovat ratkaisevia testin onnistumiselle (Bandy 2005). Taivutuksen viimeiset asteet ovat kriittisiä ja tuovat esiin pienimmätkin rajoitukset SI-nivelen liikkeessä. Mitchell kehottaa sijoittamaan peukalot uudelleen paikalleen testattavan ollessa kumartuneena. Testattavaa pyydetään tämän jälkeen suoristumaan ylöspäin noin 30cm verran. Testitulos voidaan varmentaa toistamalla liikkeen loppuosa. (Mitchell 2001: 88.) Taivutetussa loppuasennossa voidaan tarkistaa myös ILA symmetria (Greenman 2003: 349).

#### 4.4 Testin tulkinta

SI-nivelen liike on normaali, jos molemmat SIPS:t liikkuvat symmetrisesti superioirisesti vartalon taivutuksen yhteydessä. Jos sacrum ei toisella puolella pääse liikkumaan, se ottaa mukaansa iliumin aiheuttaen asymmetrisen liikkeen puolten välille. Nähtävästi liikkuvampi SIPS on rajoittuneen SI-nivelen puolella (Kuchera – Kuchera 1992: 427; Mitchell 2001: 89.) Testitulos on positiivinen, kun SIPS liikkuu toisella puolella enemmän superiorisesti (Bandy 2005; Greenman 2003: 349; Petty - Moore 2001: 285). SIPS:n liike voi tapahtua myös anteriorisesti (Greenman 2003: 349). Liikkeen ajoituksesta on erilaisia näkemyksiä. DiGiovanna mainitsee aikaisemmin liikkeelle lähtevän puolen, joka liikkuu todennäköisesti myös enemmän ja pidemmälle (DiGiovanna 2005: 315). Mitchell ja Bandy taas painottavat lopussa tapahtuvaa liikettä (Bandy 2005; Mitchell 2001: 88). Mitchell jakaa positiivisen testituloksen neljään luokkaan (0-3). Täydellinen symmetrisyys kuuluu nolla luokkaan. Ensimmäisessä luokassa SIPS liike on 3mm tai alle. Toisessa luokassa 3-9mm. Kolmannessa luokassa yli 9mm. (Mitchell 2001: 88.)

Positiivinen testitulos kertoo vain dysfunktion puolen, ei sen luonnetta tai edes välttämättä paikkaa (AOA 2002). Tämän takia on syytä suorittaa muitakin SI-nivelen testejä (Kuchera – Kuchera 1992: 440). Muita testejä tarvitaan määrittämään onko kyseessä sacrumin fleksio vai torsio, symphysis pubis dysfunktio tai esimerkiksi innominaatin rotaatio. Seated flexion- testin päätarkoituksena Mitchell pitää sacrumin torsio- ja fleksiodysfunktioiden löytämistä tai poissulkemista. Yleisimmäksi SI-nivelen dysfunktioiksi Mitchell mainitsee sacrumin torsion, joista 90 % on tyyppiä ”left on left oblique axis”, eli vasemmalle kiertyminen vasemman vinon akselin ympäri. (Mitchell 2001: 89- 90.)

Testitulos voi myös olla väärä positiivinen tai väärä negatiivinen. Toispuoleinen quadratus lumborum lihaksen kireys voi antaa väärän positiivisen tuloksen. (Chaitow 2002: 334; Greenman 2003: 349.) Lannerangan dysfunktio voi aiheuttaa positiivisen tuloksen seated flexion- testissä (Greenman 2003: 349). Symmetrinen liike voi olla iliosacral dysfunktio (Bandy 2005). Joskus bilateraallinen somaattinen dysfunktio aiheuttaa negatiivisen tuloksen, vaikka SI-nivelissä onkin rajoitus (Kuchera – Kuchera 1992: 440). Bilateraallinen dysfunktio on kuitenkin erittäin harvinainen (Mitchell 2001: 89). Samanaikaisesti esiintyvät muut lantion alueen dysfunktiot saattavat vääristää SI- nivelen dysfunktion löydöstä. Vaatetuksen aiheuttama kiristys, vartalon kiertynyt ja vääntynyt asento, tai takataskussa oleva lompakko voivat vääristää tulosta (DiGiovanna 2005: 315, 318). Testitulos voi olla väärä positiivinen, jos ennen seated flexion- testiä suoritettu standing flexion testi on ollut erittäin positiivinen ja tutkittavalla on iliosacral dysfunktio (DiGiovanna 2005: 318).

#### 4.5 Käytetty testaustapa ja tuloksen tulkitseminen

Kappaleissa 3.1- 3.4 käsitellyistä metodeista valittiin tässä tutkimuksessa käytettyyn testaustapaan parhaiten soveltuvat yksityiskohdat. Tähän päätökseen vaikutti aikaisempi tieto ja kokemus testin suorittamisesta.

Helsingin ammattikorkeakoulu Stadiassa tehdyssä tutkimuksessa testattavat henkilöt istuivat tasaisilla jakkaroilla (kts. liite 2). Jakkaralla istuessa jalkapohjat luonnollisesti tulivat tasaisesti lattialle. Testiin osallistuneille kerrottiin testaustilanteen kulku. Heille myös näytettiin kuinka jalat asetetaan, kuinka suorittaa vartalontaivutusliike ja käsien reitti taivutuksen aikana. Testattavia pyydettiin löysäämään kiristäviä housuja, jotta SIPS:n palpoinnissa mahdollistuisi ja housujen kireys ei vaikuttaisi testiliikkeeseen. Polvien välisen etäisyyden saamiseksi yhtäläiseksi tutkittavia pyydettiin asettamaan polvet kahden nyrkin etäisyydelle toisistaan.

Tutkija paikallisti palpoiden anatomisten maamerkkien (SIPS) sijainnit kyykyssä tai polvillaan testattavan takana. Jos SIPS ei ollut nähtävissä, palpoinnista aloitettiin crista iliaca. SIPS paikallistettiin viemällä sormia crista iliaca pitkin sen posterioriseen päätepisteeseen. Hankalammissa tapauksissa, joissa alueen anatomiset maamerkit ei olleet selkeästi havaittavissa, käytettiin paikallistamiseen poikittaisia sivelyjä. Peukalot

laitettiin hivenen oikean ja vasemman SIPS:n alapuolelle, ja peukalot tuotiin ylöspäin lopulliseen paikkaansa. Ihon ja pehmytkudoksen mukaan tuominen peukaloilla katsottiin pienentävän pehmytkudoksen aiheuttamaa vetoa vartalon taivutusliikkeen aikana. Koska testin tarkoituksena on palpoida SI-nivelen liikettä iliumien kautta, on testituloksen kannalta hyvä vähentää pehmytkudoksista tulevia liikeärsykeitä. Kun peukalot ja muut sormet olivat kohdallaan, testattavat taivuttivat rauhallisesti vartaloon eteenpäin niin pitkälle kuin mahdollista, aikaisemmin esitetyn mallin mukaisesti. Testattavia pyydettiin liu'uttamaan kyynärpäitä polvien päältä. SI-nivelen liikettä seurattiin testiliikkeen aikana niin peukaloilla palpoiden kuin havainnoimalla peukaloiden välistä liikeeroa. Testattavia pyydettiin seuraavaksi suoristumaan takaisin istuma-asentoon, jolloin tutkija irrotti otteen ja testi suoritettiin välittömästi alusta asti uudelleen. Testi suoritettiin testattaville yhtenä ajankohtana enintään kahdesti yhden tutkijan toimesta. Testin toistolla pyritään varmistamaan testitulos.

## 5 KAPPA- TUTKIMUS SEATED FLEXION- TESTISTÄ

Tässä tutkimuksessa käytetään henkilöistä termejä: testattava, tutkittava, tutkija ja tutkimusvastaava. Testattavalla tarkoitetaan henkilöä, kenelle tutkimuksessa suoritettiin seated flexion- testi. Tutkittavat ovat tutkimukseen mukaan valittuja henkilöitä, joista saatuja testituloksia analysoidaan statistiikka ja johtopäätökset kappaleessa. Tutkimusvastaavana toimii yksi tutkimusryhmän jäsen, jonka työnkuvaan kuuluu vastuu tutkimuspäiväkirjasta ja sokkouttamisesta, eikä hän toimi tutkijana.

Testikysymyksenä oli: onko oikea positiivinen? Positiivisiksi testituloksiksi valittiin tapaukset, joissa testi osoitti hypoteettisesti oikean SI-nivelen liikkuvuuden olevan rajoittunut. Tutkimuksessa testattiin yhteensä 109 testattavaa, joista testaushetkellä tutkijat vuorollaan valitsivat yhteensä 40 testattavaa lopulliseen tutkimusjoukkoon.

### 5.1 Tutkimusolosuhteet, kirjanpito

Tutkimuksen toteuttamien tapahtui Helsingin Ammattikorkeakoulu Stadian tiloissa, Vanhan Viertotien toimipisteessä. Testattavina tutkimuksessa oli kaikissa kolmessa tutkimuksen vaiheessa (harjoittelu-, yhtäpitävyys- ja testivaihe) koulun opiskelijoita. Testattavat tulivat 3-5 hengen ryhmissä kerrallaan toimitilan alakertaan sermein eristettyyn



tilaan. Tutkijat eivät olleet missään vaiheessa samanaikaisesti testattavien kanssa samassa tilassa. Kyseisellä tavalla toteutettiin tutkimuksen vaatima sokkouttaminen (kts. liite2). Tutkimustila erotettiin odotustilasta sermein ja täten eivät tutkijat voineet nähdä toisiaan testejä suoritettaessa. Tutkimuksen harjoittelu- ja yhtäpitävyysvaihe suoritettiin kevätlukukaudella 2006 ja testivaihe tehtiin syyslukukaudella 2006.

Ryhmästä valittiin tutkimuksen alussa kaksi henkilöä tutkijoiksi ja kolmas henkilö toimi tutkimusjohtajana ja kirjanpitäjänä. Ensin roolien valinnat oli tarkoitus tehdä siten, että parhaiten yhtämieltä olevat tutkijat olisi valittu tutkijoiksi, mutta tämä järjestely tuntui vievän liian paljon aikaa todelliselta harjoittelujaksolta. Ryhmän sisäisen neuvottelun tuloksena päädyttiin lopulliseen roolijakoon.

Testattaville annettiin täytettäväksi lomake (kts. liite 5), jossa kysyttiin ikä, sukupuoli ja sen hetkinen subjektiivinen kivun tunto alaselän tai lantion alueella sekä toimintahaitta (VAS). VAS-asteikot kerättiin testattavilta mahdollista lisäanalyysia varten. Testattavilta kysyttiin yksinkertaisen lomakkeen avulla alaseläkivun oireellisuutta. Tutkimuksessa käytettiin tähän tarkoitukseen VAS-kyselyä subjektiivisen kivun tunnosta ja sen haittavaikutuksesta. VAS on tehty yksinkertaisella 10cm pitkällä viivalla, johon testattava kipua tuntiessaan merkkää pystyviivalla kohdan, joka kuvaa kivun ja haitan voimakkuutta. Nämä merkinnät voidaan mitata millimetrin tarkkuudella ja saada siitä numeerinen lukuarvo.

Molemmilla tutkijoilla oli omat lomakkeet (kts. liite 6). Lomakkeessa oli merkittynä juoksevat numerot (1-120) ja jokaisen numeron kohdalla y/n (yes ja no) viitaten testin kysymyksen mukaisesti joko: on positiivinen oikealla tai ei ole positiivinen oikealla. Tutkijat merkkasivat jokaisen testattavan kohdalle testin tuloksen ympyröimällä y tai n. Valintavuorossa oleva tutkija valitsi tutkimukseen soveltuvat testattavat ympyröimällä testattavan numeron.

Oikean puolen positiivisuutta päädyttiin kysymään, koska tutkimuksessa apuna ollut Dr. Kuchera kertoi, että tutkimusten mukaan 80 %:lla amerikkalaisesta populaatiosta on oikea SI-nivel aliliikkuvampi verrattuna vasempaan. Tämän takia oli oletuksena, että tutkimusta varten testattaisiin jopa 100 testattavaa, jotta tutkimukseen saataisiin 20 oikealla positiivista ja 20 ei oikealla positiivista testitulosta. Tutkimusta varten selkeimpiä ei oikealla positiivisia ovat vasemmalla positiiviset testitulokset. (Kuchera 2006.)

Tutkimuksen ollessa rakenteeltaan hyvin yksinkertainen, välineiksi tarvittiin penkkejä, sermit, tietokone ja tulostin tutkimustulosten analyysiä ja täytettävien lomakkeiden tekemistä ja tulostamista varten. Kaikki edellä mainitut tarvikkeet saatiin käyttöön Stadion toimipisteestä.

## 5.2 Harjoitteluvaihe

FIMM- protokollassa painotetaan, ettei tutkimukseen kannata ottaa liian montaa testiä samanaikaisesti arvioitavaksi (FIMM 2004: 12). Esimerkiksi, jos samaan niveleen kohdistuu useampi testi, jotka molemmat tutkijat tekevät, voi testitulos muuttua, koska testit saattavat aiheuttaa terapeuttisen vaikutuksen nivelessä. (FIMM 2004: 13.)

Tutkimuksen alussa päädyttiin tutkimaan vain yhden testin toistettavuutta. Harjoitteluvaiheessa oli tarkoitus saada tutkijoiden välinen testin suorittaminen mahdollisimman yhtenäiseksi, käymällä yhdessä läpi kymmenen testattavaa. Testin suoritus perustui useaan lähteeseen. Testin kulku kirjoitettiin läpi yksityiskohtaisesti (kts. kappale 4.5), jotta se olisi mahdollisimman samankaltainen läpi tutkimuksen. Ennen testin aloittamista testattavalle kerrottiin suullisesti testin kulku ja näytettiin myös esimerkki testin suorittamisesta.

Tutkijat eivät saaneet testituloksia heti yhtenäiseksi, joten harjoitteluvaiheen aikana käytiin läpi asioita, joista testitulosten eroavaisuudet saattoivat johtua. Tutkijoiden peukaloiden palpaatiokohta saattoi olla eri kohdilla SI-niveltä, eli testattaessa on erittäin tärkeää käydä yhdessä läpi anatomiset maamerkit, jotta tutkijat löytävät aina saman kohdan SIPS:stä. Tutkijan peukalot saattoivat siirtyä testattavan kumartuessa, eli pehmytkudoksen liikkuminen testin aikana pitää ottaa huomioon.

Testin ohjaus testattavalle saattoi olla huonosti tehty, jolloin testattavan asento vaikutti testitulokseen (lonkan abduktio – SI- nivel kiinni). Testattavan rinta- ja lannerangan liikkuvuus voivat vaikuttaa testitulokseen, esimerkiksi lannerangan aliliikkuvuus saattaa fleksioon mentäessä kallistaa koko lantiota. Pehmytkudosten kireydet voivat vaikuttaa nivelen liikkeeseen/ toispuoleiseen tulokseen.

Harjoitteluvaiheessa käytiin uudestaan läpi kirjallisuutta liikkeen laadun tarkkailemisesta. Tutkijoilla ei syntynyt testituloksen tulkinnasta selkeää yhtenäistä näkemystä. Tämä johtui osaksi tutkijoiden lukemien lähteiden erilaisista testikuvauksista. Kirjallisuudessa käsitellyistä kuvauksista yhdessä mainittiin testissä viimeiseksi liikkuvan puolen olevan merkityksellisin (Bandy 2005).

Tutkimusta tehdessä käytiin paljon keskustelua testin hypoteesista. Seated flexion- testin tarkoitus on testata SI-nivelen liikkuvuutta, mutta esimerkiksi tutkimusryhmän kääntämässä FIMM- protokollassa esitetään toteamus, ettei kokeneinkaan terapeutti voi palpoida SI-nivelen liikkuvuutta (FIMM 2004: 15). Kirjallisuudessa kuitenkin käydään usein seated flexion- testi läpi SI-nivelen testeissä ja myös apuna diagnosointiin. Olisi tärkeää saada monelle testille toistettavuus ja validiteetti, jotta niillä olisi myös diagnostista arvoa. Yhden testin perusteella ei voida vetää johtopäätöstä nivelen liikkeestä. (Van der Wurff - Buijs - Groen 2006.)

Harjoitteluvaiheessa tutustuttiin tarkemmin FIMM- protokollan kapp- tutkimuksen vaihtoehtoihin, ja päädyttiin valitsemaan vaihtoehtoista tavallinen kapp- tutkimus. Tämä sopi parhaiten kyseisistä vaihtoehtoista tutkimuksen asetteluihin ja tutkimuskysymykseen: ”onko seated flexion- testi positiivinen oikean puoleisessa SI-nivelessä?”. Tavallisessa kapp- tutkimuksessa testi määritellään siten, että siihen on selkeä vastaus (kyllä/ei). (FIMM 2004: 9.)

Harjoitteluvaiheessa tutkijat testasivat 20 testattavaa neljän kuukauden aikana (syyskuu-tammikuu). Usean testattavan kohdalla käytiin testi läpi myös molempien tutkijoiden kanssa yksityiskohtaisesti, jotta tutkijat tekisivät testin suorittamisen samalla tavalla. Harjoitteluvaiheessa nousi paljon keskustelua voimasta, jolla peukalot on tarkoitus asettaa SIPS:lle. Kevyempi kosketus tuntui seuraavan enemmän faskiaalista liikettä, kun taas hieman kovemalla paineella liike vaikutti tulevan enemmän nivelestä. Myös peukaloiden paikantamista aina samaan kohtaan SIPS:ä käytiin paljon läpi. Harjoittelun aikana muutama testattavista sanoi, että heistä tuntui tutkijoiden sormien sijainnin olleen keskenään erilainen. Tämän jälkeen tutkijat harjoittelivat neljällä testattavalla tarkemmin. Ensimmäinen tutkija palpoi hänen mielestään hyvän testikohdan, merkkasi kohtaan tussilla merkin ja teki testin. Seuraava etsi kyseiset kohdat silmät sidottuina. Eroavaisuutta oli jonkin verran, joten tutkijat kävivät vielä yhdessä testin läpi ja miten peukaloiden kohta SIPS:llä onnistuisi parhaiten.

### 5.3 Yhtäpitävyysvaihe

FIMM- protokollan mukaan yhtäpitävyysvaiheessa on suositeltavaa saavuttaa yhtäpitävyys, jonka tulisi olla suurempi kuin 80%. Jos osuus on vähemmän kuin 80%, tutkijoiden täytyisi keskustella heidän yksimielisyydestä ja heidän tulisi tehdä harjoitteluvaihe uudelleen (FIMM 2004: 28). FIMM- protokollan mukaan kappa-arvot, jotka ylittävät 0.6, kertovat toistettavuuden olevan hyvästä loistavaan. Arvot välillä 0.4- 0.6 ilmaisevat kohtalaisesta hyvään olevaa toistettavuutta. Alle 0.4 arvot, joihin tutkimuksessa saatu tulos 0.3 kappa-arvolla jäi, ilmaisevat testin yhtäpitävyyden olevan heikko. (FIMM 2004: 11- 12.) Toistettavuus- ja validiteettitutkimusprotokollan mukaan tutkijoiden kuului jatkaa harjoittelua paremman yksimielisyyden saavuttamiseksi ennen lopulliseen testivaiheeseen siirtymistä. Opinnäytetyöhön varattujen tuntien asettamien rajoitteiden takia tutkijat joutuivat kuitenkin siirtymään varsinaiseen testivaiheeseen ilman lisäharjoittelua. Tutkijat olivat yksimielisiä 13 tapauksessa 20:stä. Tutkimusprotokollan mukaisen 80% yksimielisyyden saavuttaminen edellyttäisi 16 yksimielistä testitulosta.

Yhtäpitävyysvaihe suoritettiin aikaisemmalla työjaolla, jossa tutkimusvastaava antoi koehenkilöille ohjeet testin suorittamisesta, toimi sokkouttamisen valvojana sekä tarkisti tutkijoiden kirjaamat tulokset. Tutkijat suorittivat seated flexion- testit ja kirjasivat löydökset ylös. Kaikki toimivat koehenkilöiden värväämisessä Helsingin ammattikorkeakoulu Stadian tiloissa.

Yhtäpitävyysvaiheessa valittiin 20 tutkittavaa testattavien joukosta. Tutkimukseen valitussa seated flexion- testissä kysymykseksi oli valittu ”onko oikea positiivinen?”. Molemmat tutkijat suorittivat testin jokaiselle testattavalle. Tutkija A valitsi ensin testattavista kymmenen tutkittavaa, joista viisi testitulosta oli positiivisia ja viisi negatiivisia. Tämän jälkeen tutkija B valitsi samoin kaksi viiden tutkittavan ryhmää. Yhtäpitävyysvaiheessa käytiin läpi 33 testattavaa, kunnes saatiin käytyä läpi 20 tutkittavan joukko.

Jokaisessa toistettavuustutkimuksessa on tärkeää huolehtia sokkouttamisesta. Sokkouttamista haittaavaa kommunikointia ei saa olla tutkittavien ja tutkijan eikä tutkijoiden välillä. Koska tutkijat eivät sanoneet ääneen löydöstään, ei se välittynyt testattaville tai toiselle tutkijalle. Tutkijoilla A ja B oli erilliset lomakkeet tulosten kirjaamiseen. Tutkijat säilyttivät lomakkeita sokkoutettuna toisiltaan. Tutkija A:n tehdessä testejä odotti

tutkija B vuoroaan sermin takana, jottei testin näkeminen vaikuttaisi tutkija B:n tuloksiin. Tutkija B:n vuoron tultua siirtyi tutkija A sermin taakse, jottei hänen ilmeensä tai eleensä voisi vaikuttaa toiseen.

Testattavat valittiin kyseisenä päivänä koulussa olleista Stadian opiskelijoista. Osallistumisaste oli erinomainen, koska testin suorittaminen vei opiskelijoiden aikaa vain viitisen minuuttia. Kaikista osallistuneista oli tarkoituksena valita selkeät tapaukset. Lyhyessä ajassa suuremman henkilökunnan testaaminen havainnollisti ihmisten yksilöllisyyden. SIPS:n paikallistaminen ei yksilöllisyydestä johtuen ollut aivan suoraviivaista. Jos SIPS ei ollut kovin prominentti ja päällä oli enemmän rasvakudosta, saattoi tulla epävarmuutta peukaloiden oikeasta paikasta. Tämä on voinut olla yksi mahdollinen tekijä testitulosten erimielisyyksille. Tutkimuksesta jätettiin pois testattavat, joiden testitulos oli epäselvä.

#### 5.4 Testivaihe

Testivaihe suoritettiin yhtäpitävyyssvaiheen kaltaisesti tuplaten tutkittavien määrä. Testivaiheessa testattavia tutkittiin, kunnes molemmat tutkijat olivat löytäneet 10 tutkittavaa, joilla testi oli positiivinen ja 10 tutkittavaa, joilla testi oli negatiivinen. Tavoitteena tutkimuksessa oli saada negatiivisiksi tuloksiksi positiivisia vasempia. Testivaihe sisälsi 106 tervettä testattavaa, joista valittiin tutkimusprotokollan mukaisesti otokseksi 20 oikeata positiivista ja 20 ei oikeata positiivista testitulosta. Testattavat olivat Helsingin ammattikorkeakoulu Stadian opiskelijoita. Testivaihe suoritettiin huolimatta heikosta yhtäpitävyyssvaiheesta.

Kappa-arvo laskettiin internetsivuilta

<http://www.dmi.columbia.edu/homepages/chuangj/kappa/calculator.htm> löytyvällä laskimella (Chuang 2002). Seuraavassa kappaleessa on kuvattuna testivaiheen tulokset ja kontingenssitaulukko sekä tehty johtopäätökset saaduista tuloksista.

## 5.5 Statiikka ja johtopäätökset

Seuraavassa esitetään testivaiheen kontingenssitaulukko (kuvio 6) ja kappi-arvon laskentaan käytettävät kaavat listattuna kuvion alla. Yleinen kontingenssitaulukko (kuvio 7) pylvaiden vieressä auttaa lukijaa ymmärtämään kuviossa 6 valmiiksi täytetyt testivaiheen tulokset. Kyseinen taulukko pohjautuu FIMM- protokollan 1. liitteeseen.

		<u>Tutkija B</u>		Yhteensä:
		Kyllä	Ei	
<u>Tutkija A</u>	Kyllä	8	11	19
	Ei	6	15	21
Yhteensä:		14	26	40

Tutkittavien määrä =  $n=40$   
Yhtäpitävyys =  $P_0=0,575$

Oletettu sattuma yhtäpitävyys =  $p_c = 0,5075$

$$\text{Kappa} = \frac{p_0 - p_c}{1 - p_c} = 0,137$$

$$\text{Esiintyvyys } P = (a + [b + c] / 2) / n = 0,4125$$

KUVIO 6. Testivaiheen 2 x 2 kontingenssitaulukko. Alla numeroiden selitykset (kts. tarkemmat selitykset kappaleesta 2) (FIMM 2004: 35).

Tutkimuskysymyksenä oli: onko oikea puoli positiivinen. Kontingenssitaulukon ruutu a kohdassa merkittynä olevan lukumäärän mukaisesti tutkijat olivat yhtämieltä testituloksen positiivisuudesta. Ruutu b kuvastaa tapauksia, joissa tutkija A sai testistä positiivisen tuloksen ja tutkija B sai samoista tutkittavista negatiivisen tuloksen. Ruutu c kuvastaa tapauksia joissa tutkija A sai negatiivisen testituloksen ja tutkija B sai positiivisen tuloksen. Ruutu d kuvastaa tuloksia joissa molemmat tutkijat olivat yhtämieltä testituloksen negatiivisuudesta.

Alla olevassa listassa pienet kirjaimet tarkoittavat kontingenssitaulukossa esiintyviä edellä mainittuja lukuja. Isot kirjaimet listan kaavoissa tarkoittavat pylväiden kirjaimia, joiden kohtien laskutoimitus isolla kirjaimella merkittyyn kohtaan tulee sijoittaa.

$$a = 8, b = 11, c = 6, d = 15$$

$$\text{Pylväs E: } a+b+c+d = n = 40$$

$$\text{Pylväs F: } a+b = 19$$

$$\text{Pylväs G: } a+c = 14$$

$$\text{Pylväs H: } c+d = 21$$

$$\text{Pylväs I: } b+d = 26$$

$$\text{Pylväs J: } a+d = 23$$

$$\text{Pylväs K: Esiintyvyyys} = (a+[b+c]/2)/n \\ = A1/E1+B1/2 \times E1+C1/2 \times E1$$

$$\text{Pylväs L: Yhtäpitävyys } P_o = J1/E1$$

$$\text{Pylväs M: } (a+b)/n = F1/E1 = 0,475$$

$$\text{Pylväs N: } (a+c)/n = G1/E1 = 0,35$$

$$\text{Pylväs O: } (c+d)/n = H1/E1 = 0,525$$

$$\text{Pylväs P: } (b+d)/n = I1/E1 = 0,65$$

$$\text{Pylväs Q: } M1 \times N1 = 0,16625$$

$$\text{Pylväs R: } O1 \times P1 = 0,34125$$

$$\text{Pylväs S: Oletettu sattuma yhtäpitävyys}$$

$$P_c = Q1+R1 = 0,5075$$

$$\text{Pylväs T: } L1-S1 = 0,0675$$

$$\text{Pylväs U: } 1-S1 = 0,4925$$

$$\text{Pylväs V: Kappa-arvo } T1/U1 = 0,137$$

(FIMM 2004: 35)

		Tutkija B		Yhteensä:
		Kyllä	Ei	
Tutkija A	Kyllä	a	b	a+b
	Ei	c	d	c+d
Yhteensä:		a+c	b+d	n

Testattavien määrä = n  
Yhtäpitävyys =  $p_o = \frac{a+d}{n}$

Oletettu sattuma yhtäpitävyys =  $p_c = \frac{a+b}{n} \times \frac{a+c}{n} + \frac{c+d}{n} \times \frac{b+d}{n}$

Kappa =  $\frac{p_o - p_c}{1 - p_c}$

Esiintyvyyys P =  $(a + [b + c] / 2) / n$

Kuvio 7. Yleinen kontingenssitaulukko (FIMM 2004: 36).

Testivaiheen kappa-arvo oli 0,137, joten sen merkitsevyys on olematon. Tämän takia tutkimuksen perusteella ei voida osoittaa testin toistettavuutta eikä validiteettia.

## 6 POHDINTA

Tutkimuksen kaikki vaiheet toteutettiin protokollan mukaisesti, vaikkakin aikarajoitteista johtuen yhtäpitävyysvaiheen tavoitteellista arvoa (80% yhtäpitävyys) ei pystytty saavuttamaan. Alhaiseksi jäänyttä kappa-arvoa voivat selittää monet tekijät. Kokemattomuus tutkimuksen tekemisessä heijastui tutkimuksen suunnitteluun ja toteuttamiseen. Tiukka tutkimusaikataulu lisäsi myös osaltaan jo itsessään vaativan työn haasteellisuutta.

Monet asiat voivat johtaa siihen, että testien suorittaminen tapahtuu eri tavalla. Terauttien työskennelyä vuosia omilla praktiikoillaan, ovat heidän testitapansa muokkautuneet erilaisiksi. Myös kirjallisuudessa on testien kuvauksissa poikkeavuuksia. Tämän takia tutkijat kävivät harjoitteluvaiheessa yhdessä läpi kymmenen testattavaa, jotta saatiin tutkijoiden suoritustavat mahdollisimman samanlaisiksi.

Selkeimmässä tilanteessa positiiviseksi testitulokseksi otettiin tapaukset, joissa toinen SIPS lähti aikaisemmin liikkeelle ja jatkoi liikettä toisen puolen liikkeen loputtua. Jos liikkeen määrä olisi vielä Mitchellin määrittämässä kolmannessa luokassa, eli ylittäisi 9mm (Mitchell 2001: 88), saavutettaisiin todennäköisemmin yksimielisempiä havaintoja. Tutkimusjoukkoon on käytännössä valittava myös epäselviä tapauksia, jos tutkimuksen suorittamisen aikataulu on rajallinen. Mikäli testitulos oli tutkimuksessa toisella kertaa erilainen, tapaus jätettiin muutamaa tapausta lukuun ottamatta lopullisen tutkimusjoukon ulkopuolelle. Muutamissa tapauksissa testitulos oli vähemmän selkeä toisella testikerralla. Jos tällaisia tapauksia joudutaan valitsemaan lopulliseen testijoukkoon, kriteerien tulee olla selkeästi sovittu.

Alussa ja lopussa tapahtuvassa liikemäärässä saattoi olla eroa. Tutkimuksessa päädyttiin valitsemaan liikkeen määrä ensisijaiseksi kriteeriksi. Lopussa tapahtuvalla liikkeellä oli suurempi painoarvo kuin alussa tapahtuvalla liikkeellä. Selvää testin suorittamiseen vaadittavaa kriteeriä ei kuitenkaan pystytty muodostamaan, vaan päätös jäi kummankin tutkijan henkilökohtaisen päätöksen varaan. Jos positiivisen testituloksen valintakriteeriä ei sovita tarkasti ennen tutkimusta, kasvavat erimielisyydet testin tulkinnessa erityisesti tällaisissa tapauksissa. Toinen tutkija voi pitää merkityksellisempänä liikkeen ajoitusta, toinen taas liikkeen määrää. Kirjallisuuden antaessa epätarkkoja ja vaihtelevia ohjeita, muodostuu tutkijoille helposti erilainen käsitys testin suorittamisesta. Olisi siis



tärkeää saavuttaa yhtämielisyys testiprotokollasta mahdollisimman aikaisessa vaiheessa.

Yhtäpitävyysvaiheessa tutkimuksessa saatiin yhtäpitävyyden arvoksi 65% verrattuna testivaiheen 58%. Syynä tähän saattoivat olla tutkittavien määrän tuplaantuminen, ja siten virhemarginaalin kasvamiseen liittyen esimerkiksi tutkijoiden havainnointikyvyn herpaantumiseen. Lisäksi yhtäpitävyysvaihe suoritettiin heti harjoitteluvaiheen jälkeen, jolloin testaajien palpaatiotaidot olivat yhtenäisemmät. Testivaihe suoritettiin noin kolme kuukautta yhtäpitävyysvaiheen jälkeen harjoittelematta välillä suunnitelmallisesti.

Kliinisesti merkittävä kappi-arvo on yli 0,4. Tutkimuksessa kappi-arvoksi saatiin 0,137, joten sen merkittävyys on olematon. Tutkimuksen perusteella ei voida todentaa testin toistettavuutta ja validiteettia. Testiä ei voida myöskään todeta toimimattomaksi kyseisen tutkimuksen perusteella johtuen harjoitteluvaiheen heikosta suorittamisesta. Tutkimuksesta voidaan todeta, että vastaavanlaisia tutkimuksia tarvitaan tulevaisuudessa lisää.

Tutkimusprotokollan hienous perustuu siihen, että tutkimusjoukosta saadaan yhdellä kerralla kerättyä myös tutkimusaineisto. Eli tutkimuksessa testattavia ei tarvitse tämän takia testata kahteen eri ajankohtaan, vaan testattavat valitaan samalla kerralla ja siitä muodostuu myös tutkimusaineisto. Tästä syystä tämä protokolla on erittäin näppärä yksityispraktiikalla toteutettavaksi, koska tutkittavan ei tarvitse käydä kuin kerran.

Tutkijan sisäisen yksimielisyyden tutkiminen olisi hyvä lähtökohta tämänkaltaisissa tutkimuksissa. Tutkijoiden välinen yksimielisyys voisi olla parempi, jos tutkijat olisivat ensin saavuttaneet selkeästi merkittävän tutkijan sisäisen yksimielisyyden. Yhtäpitävyyden saavuttamiseksi vaadittaisiin myös enemmän yhteistä harjoittelua. Tässä vaiheessa tulee olennaiseksi tutkijoiden kyky joustaa ja muokata omaa testin suoritustapaa, jos tutkijoiden testaustavoissa ja tulkinnassa osoittautuu erilaisuuksia. Testejä olisi hyvä jakaa mahdollisimman pieniin osa-alueisiin. Tällöin jokainen testin osa-alue tulisi käytyä läpi yhtäpitävyyden saavuttamiseksi.

Tutkimusjoukko ei ollut todennäköisesti kovin soveltuva tutkimukseen. Asymmetrisen liikkeen SIPS:n tasoissa ajatellaan aiheutuvan sacrumin dysfunktiosta. Jos dysfunktiota ei ole, tai se on hyvin vähäinen, ei selvää liikkeen asymmetriaa tule esille. Liikkeen asymmetrian palpaatio on osoittautunut jo lähtökohtana vaikeaksi myös muissa manuaa-

lisen lääketieteen toistettavuustutkimuksissa. On myös mahdollista, että nuorilla henkilöillä, joilla ei ole merkittävää dysfunktioita sacrumissa, saattaa itse testiliikekin olla riittävä mobilisoimaan SI-niveltä. Ensimmäisen tutkijan havaitsema asymmetria saattaa olla tasoittunut toisen tutkijan suorittaessa testiä. Ihanteelliseen tutkimusjoukkoon olisi mahdollisesti parasta valita henkilöitä, joilla olisi todettu kroonisia oireita, jotka viittaisivat SI-nivelen dysfunktioon. Riittävän suuren tutkimusjoukon löytäminen näillä kriteereillä olisi tietenkin vaikeampaa. Apuna voitaisiin käyttää yksinkertaista testiä, esim. Fortinin sormitestiä. Testissä testattava osoittaa kivun paikantuvan SI-nivelen alueelle.

Tutkimuksessa osa tutkittavista teki testin ohjeista poiketen. Polvet olivat kauempana toisistaan, käsiä ei liu'utettu polvia vasten tai liike suoritettiin pyydettyä nopeammin. Nämä erilaiset tutkittavien suoritukset ovat osaltaan saattaneet vaikuttaa myös tutkimustuloksiin.

Lantion anatomisten maamerkkien palpointi on todettu tutkimuksissa hyvin haasteelliseksi. SIPS:n paikallistaminenkin on ollut vain keskinkertaista osteopaattipiskelijoilla (Chaitow 2003: 16). Suurempaan tutkimusjoukkoon tuli luonnollisesti paljon variaatiota lantion anatomisten maamerkkien suhteen. Joillain henkilöillä SIPS saattoi olla helposti nähtävissä, kun toisilla sen palpaatio oli epävarmaa.

Anatomisten maamerkkien piirtäminen tai merkitseminen testattavaan on yksi keino harjoitella palpaatiota. Tällä keinolla voitaisiin tarkistaa sormien paikka sokkona tehdyn palpoinnin jälkeen. Tässä voidaan käyttää myös erikoisväräkynää ja UV-värilampua, jolloin palpointia ei tarvitse tehdä silmät sidottuina. Harjoittelu- ja yhtäpitävyysvaiheessa tehty videointi voisi toimia hyvin oppimisen edistäjänä. Videointi mahdollistaisi tutkijoiden testaustapojen vertailun, ja sitä voisi käyttää apuvälineenä selvittämään erimielisyyksiä.

Palpaation opetuksen apuvälineenä voitaisiin myös käyttää erilaisia painemittareita ja vaakoja, joilla voitaisiin kalibroida palpaatiossa käytettävä paine samanlaiseksi. Erilaisien palpoitavien ominaisuuksien harjoittelemiseksi voitaisiin käyttää erinäisiä laitteita tai harjoitteita, jotka mahdollistaisivat palpaatiolöydöksen tarkistamisen. Esimerkkinä tästä on yksinkertaisesti hius kirjan välissä, jonka tuntemista harjoitellaan useamman sivun lävitse.

Testattavien palaute voi auttaa tulemaan tietoisiksi tutkijoiden erilaisesta peukaloiden paineesta, ja auttaa testaamistavan kalibroinnissa tutkijoiden välillä. Yksittäisten kokemusten perusteella vaikuttaa siltä, että testin tulos voi muuttua oleellisesti peukaloiden paineen muuttuessa. Muuttamalla painetta voi huomio kohdistua eri kudosteroksiin ja syvyyksiin. Pehmytkudos- ja luutasolla voi liikkuvuus olla erilainen, mikä voi vaikuttaa merkittävästi testin tulokseen.

VAS-kyselyllä yritettiin tutkimuksessa kerätä kätevästi tietoa mahdollisesta kivusta liittyen positiiviseen testitulokseen. VAS lukua ei kuitenkaan ollut mielekästä analysoida yhtämielisyyden ollessa heikko ja tutkittavien ollessa muutamaa poikkeusta lukuun ottamatta oireettomia. Jos tutkimuksessa olisi ollut enemmän oireellisia, olisi myös yhtäpitävyys voinut oireellisten kohdalla olla parempi. Tämän hypoteesin avulla yhdessä hyvän kappa-arvon kanssa olisi voitu arvioida testin soveltuvuutta kliiniseen käyttöön.

Kappastatistiikan soveltuvuutta tutkimuksiin on kritisoitu. Uebersaxin artikkelin mukaan kappa ei todellisuudessa ota huomioon sattumaa. Kritiikkiä on esitetty siitä, että kappa ei selitä erimielisyyteen johtavia tekijöitä. Artikkelissa spekuloidiin myös, että yhtäpitävyys voi olla korkea, vaikka kappa-arvo olisikin heikko. Tätä selitettiin sillä, että tutkijat voivat tulkita saman löydöksen eritavalla. (Uebersax 2002.) Tutkimusasetelmassa on kuitenkin otettava huomioon tällaiset seikat.

Kappa- tutkimuksen suorittaminen on aina hyvin opettavaista, saavutetuista tuloksista riippumatta. Kymmenien ihmisten testaaminen yhden päivän aikana on raskasta ja paljon keskittymistä vaativaa, mutta opettaa paljon palpaatiosta ja ihmisten yksilöllisestä pinta-anatomiasta. Yhden testin tarkempi läpikäyminen opettaa myös ymmärtämään kuinka monesta osa-alueesta luotettava testin suorittaminen riippuu. Osa-alueet ovat samankaltaisia muissakin manuaalisissa testeissä, joten tutkimuksella saadaan eräänlainen perusmalli myös muiden testien suorittamiseksi.

Tutkimuksessa käytetty FIMM- protokollan käänös vaikutti helpolta käyttää. Tutkimuksesta huomattiin, että on tärkeää tehdä harjoittelu- ja yhtäpitävyysvaihe huolellisesti, jotta testivaiheen kappatuloksesta voidaan tehdä valideja johtopäätöksiä. Johtopäätöksenä voidaan todeta, että selkeää enemmistöä positiivisen puolen suhteen ei seated flexion- testillä löytynyt. Hypoteesin mukaan oletettiin oikeita positiivisia olevan

enemmän. Tutkijoiden mielestä tällaista ilmiötä ei voitu todeta olevan. Heikon kappaarvon pohjalta tämän varmistaminen on kuitenkin kyseenalaista.

Tehdyn käännöksen validointi olisi myös syytä toteuttaa, mutta sen tulisi suorittaa jokin riippumaton taho. Näin ollen työ avaa haasteita myös tuleville tutkimuksille, jolloin käännöstyö voitaisiin validoida riippumattoman tahon toimesta.

## 7 LÄHTEET

- American Osteopathic Association 2002: Foundations for osteopathic medicine. Second edition. Lippincott & Williams & Wilkins. Philadelphia, USA. 773.
- Bandy, Bill 2005: Pelvis examination. Verkkodokumentti. Päivitetty 7.9.2005. <<http://faculty.uca.edu/~fletcher/Musculoskel/pelvis-examination.htm>>. Luettu 19.8.2006.
- Chaitow, Leon – DeLany, Judith 2002: Clinical application of neuromuscular techniques, volume 2 the lower body. Churchill Livingstone. London. 334.
- Chaitow, Leon 2003: Palpation and Assessment Skills, Second edition. Churchill Livingstone. London. Sivu 16- 19.
- Chuang, Jen-Hsiang 2002: kappa calculator. Verkkodokumentti. Päivitetty 19.12.2002. <<http://www.dmi.columbia.edu/homepages/chuangj/kappa/calculator.htm>>. Luettu 2.10.2006.
- Degenhardt, Brian - Snider, Karen - Snider, Eric - Johnson, Jane 2005: Interobserver Reliability of Osteopathic Palpatory Diagnostic Tests of the Lumbar Spine: Improvements From Consensus Training. JAOA 105 (10). 465- 473.
- Dowling, Dennis J. 2005: An Osteopathic approach to diagnosis and treatment. Third edition. Lippincott Williams & Wilkins. Philadelphia. USA. 304- 322.
- DiGiovanna, Eileen – Schiowitz, Stanley 2005: Osteopathic approach to diagnosis and treatment. Third edition. Lippincott Williams & Wilkins. Philadelphia, USA. 315.
- FIMM 2004: Reproducibility and validity studies. Verkkodokumentti. Päivitetty 8.9.2006. <[http://www.fimm-online.com/pub/en/data/objects/reproducibility\\_validity.pdf](http://www.fimm-online.com/pub/en/data/objects/reproducibility_validity.pdf)>. Luettu 15.9.2006.
- Freburger, Janet - Riddle, Daniel 1999. Measurement of Sacroiliac Joint Dysfunction: A Multicenter Intertester Reliability Study. Physical Therapy 79 (12). 1134- 1141.
- Forst, Stacy L.- Wheeler, Michael T. - Fortin, Joseph D - Vilensky, Joel A 2006: The Sacroiliac Joint: Anatomy, Physiology and Clinical Significance. Pain Physician 9. 61- 68.
- Greenman, Philip E. 2003: Principles of manual medicine. Third edition. Lippincott Williams & Wilkins. Philadelphia, USA. 349.
- Hestboek, Lise - Leboeuf-Yde, Charlotte 2000: Are Chiropractic Tests for the Lumbo-Pelvic Spine Reliable and Valid? A Systematic Critical Literature Review. Journal of Manipulative and Physiological Therapeutics 23 (4). 258- 263.
- Huijbregts, Peter 2004. Sacroiliac joint dysfunction: Evidence-based diagnosis. Verkkodokumentti. Päivitetty 1.6.2004. <<http://www.kalindra.com/SIJD%20EBP%20Dx.pdf>>. Luettu 2.10.2006.

- Maigne, Jean-Yves - Aivaliklis, Alain - Pferer, Fabrice 1996: Results of sacroiliac joint double block and value of pain provocation tests in 54 patients with low back pain. *Spine* 21(16). 1889- 1892.
- Kapandji, I A 1997: *Kinesiologia III, Selkärangan, rintakehän ja lantion nivelten toiminta.* Ylinen, Jari (Suom.). Laukaa. Medirehab. 56.
- Kappler, Robert E. 2000: Seated flexion test, a study questioning the need for patient's feet on the floor. *AAO journal, publication of the American Association of Osteopathy.* 10 (3). Indianapolis. USA. 28- 30.
- Kuchera, Michael 2006. DO, Professor of Philadelphia College of osteopathic medicine. *Henkilökohtaiset keskustelut ja sähköpostit.* Helsinki. 2005- 2006.
- Kuchera, William – Kuchera, Michael 1992: *Osteopathic principles in practice.* Second edition. USA. 427-430, 439-446.
- Landis, JR - Koch, GG 1977: The measurement of observer agreement for categorical data. *Biometrics* 33. 159- 174.
- Laslett, Mark - Aprill, Charles - McDonald, Barry 2005: Diagnosis of Sacroiliac Joint Pain: Validity of individual provocation tests and composites of tests. *ELSEVIER – Manual Therapy* 10. 207- 218.
- Lewit, Karel - Rosina, Alois. 1999. Why yet another diagnostic sign of sacroiliac movement restriction? *Manipulative Physiol. Therapy* 22(3). 154- 160.  
Pubmed verkkodokumentti. Abstrakti.  
<[http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list\\_uids=10220714&dopt=Abstract](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10220714&dopt=Abstract)> . Luettu 2.10.2006.
- Metsämuuronen, Jari 1997: Työelämän muutosten ja koulutustarpeiden ennakkoinnin analysointijärjestelmä. Päivitetty 22.5.1997.  
<<http://www.mol.fi/esf/ennakointi/metodit/jmstabi2.htm>>. Luettu 1.9.2006.
- Mitchell, Fred L. Jr. 2001: *The muscle energy manual.* 3. Evaluation of pelvis and sacrum. Second edition. MET Press. Michigan. USA. 5, 88-91, 124-125.
- Netter, Frank H. 2006: *Atlas of Human Anatomy.* 4<sup>th</sup> edition. Saunders – Elsevier. Pennsylvania. 353.
- Petty, Nicola - Moore, Ann 2001: *Neuromuscular examination and assessment, handbook for therapists.* Second edition. Churchill Livingstone. London. 285.
- Sim, Julius – Wright, Chris C. 2005: The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requires. *Physical Therapy* 85 (3).
- Sizer, Philip - Phelps, Valerie - Thompsen, Kirk 2002. Disorders of the Sacroiliac Joint. *Pain Practice* 2. Verkkodokumentti. Päivitetty 17.3.2002. <<http://www.blackwell-synergy.com/links/doi/10.1046/j.1533-2500.2002.02002.x/abs/>>. Luettu 1.10.2006.

Uebersax, John 2002: Kappa Coefficient. Verkkodokumentti. Päivitetty 20.7.2002.  
<<http://ourworld.compuserve.com/homepages/jsuebersax/kappa.htm>>. Luettu 3.10.2006.

Van der Wurff - Buijs - Groen 2006: A multitest regimen of pain provocation tests as an aid to reduce unnecessary minimally invasive sacroiliac joint procedures. Verkkodokumentti PubMed.  
<[http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list\\_uids=16401431&dopt=Abstract](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=16401431&dopt=Abstract)> . Luettu 27.11.2006.

Viera, Anthony - Garrett, Joanne 2005: Understanding Interobserver Agreement: The Kappa Statistic. Family Medicine 37 (5). 360- 363.

Wikipedia. Verkkotietosanakirja. Päivitetty 5.8.2006.  
<[http://en.wikipedia.org/wiki/Cohen's\\_kappa](http://en.wikipedia.org/wiki/Cohen's_kappa)>. Luettu 2.9.2006.

## **TOISTETTAVUUS- JA VALIDITEETTITUTKIMUKSET**

Manuaalisen lääketieteen diagnostisissa käytännöissä

Protocol formats

THIRD EDITION

FIMM SCIENTIFIC COMMITTEE

Editor: J. Patijn, MD, PhD

Suomenkielen käännös: J. Karppinen, LT; M. Kurvinen, L. Maijala, V. Pöllänen

Copyright © 2004 by FIMM



## SISÄLLYS

I. JOHDANTO – Tieteellisen toimikunnan puheenjohtaja	3
II. TOISTETTAVUUS JA VALIDITEETTI	5
Nimikkeistö	5
II.1. Reliabiliteetti voidaan jakaa toistettavuuteen (precision) ja validiteettiin (accuracy).	5
II.1.1. Toistettavuus (precision, reproducibility)	5
II.1.2. Validiteetti (accuracy, validity)	6
II.2. Tunnusmerkki ja sen esiintyvyys	6
II.3. Yhtäpitävyys (overall agreement)	7
II.4. Sensitiivisyys ja spesifisyys	8
II.4.1. Sensitiivisyys	8
II.4.2. Spesifisyys	8
II.4.3. Positiivinen ja negatiivinen ennustearvo (predictive value)	8
II.5. Kappa-arvo: tulkitseminen	8
III. DIAGNOSTISEN TOISTETTAVUUSPROTOKOLLAN LÄHTÖKOHTIA MANUAALISESSA LÄÄKETIETEESSÄ	9
III.1. Diagnostisten ja tilastollisten metodien luonteesta	9
III.1.1. Kvalitatiiviset ja semi-kvantitatiiviset diagnostiset menettelytavat (luokittelu- ja järjestysasteikko)	9
III.1.2. Kvantitatiivinen diagnostinen menettelytapa	10
III.1.3. Epäsopiva statistiikka kvalitatiivisissa toistettavuustutkimuksissa	10
III.1.3.1. Prosentuaalinen yksimielisyys (percent agreement)	10
III.1.3.2. Korrelaatiokerroin (correlation coefficients)	11
III.1.4. Sopiva statistiikka kvalitatiivisiin ja semi-kvantitatiivisiin toistettavuustutkimuksiin	11
III.1.4.1. Tavallinen kappa (normal kappa)	11
III.1.4.2. Yleinen kappa (overall kappa)	11
III.1.4.3 Painotettu kappa (weighted kappa)	11

III.1.5. Sopiva statistiikka kvantitatiivisiin toistettavuustutkimuksiin	12
III.2. Diagnostisen menettelytavan tavoite	12
III.3. Arvioitavien testien määrä	14
III.4. tutkijoiden määrä	17
III.5. Testin hypoteesi	18
III.6. Sokkouttaminen (blinding procedures)	19
III.7. Testimenetelmä ja testin arvostelu	19
III.8. Valinta ja henkilöiden määrä	19
III.9. Tilastotieto toistettavuustutkimuksessa: kappa-arvo	19
III.9.1. Kapan riippuvuus esiintyvyyteen	20
III.9.2. Kapan riippuvuus yhtäpitävyyteen (Po)	20
III.9.3. Toistettavuustutkimuksen optimointimenetelmät: Yhtäpitävyyden ja 50 % esiintyvyyden saavuttaminen	21
III.10. Kappatutkimuksen esittäminen	24
III.11. Kappakirjallisuuden lähteet	25
IV. TOISTETTAVUUSTUTKIMUKSEN SEITSEMÄN KULTAISTA SÄÄNTÖÄ	27
V. VALIDITEETTI	30
V.1. Kultainen standardi	30
V.2. Sensitiivisyys ja spesifisyys	31
V.3. Positiivinen ja negatiivinen ennustearvo	32
V.4. Todennäköisyysuhde (likelihood ratio)	33

## **I. JOHDANTO – Tieteellisen toimikunnan puheenjohtaja**

### **I.1. Taustatiedot**

Tämä materiaali pohjautuu International Federation for Manual/Muskoloskeletal Medicine (FIMM) järjestön tieteellisen toimikunnan viidenteen julkaistuun protokollaan.

Protokollan tarkoituksena on luoda standardoitu formaatti validiteetti-, sensitiivisyys- ja spesifisyystutkimuksille. Se antaa ohjeet manuaalisen lääketieteen toistettavuustutkimuksien tekemiseen niin tutkijoille kuin alaa harjoittaville klinikoille. Tieteellisen toimikunnan protokollia kehitetään tulevaisuudessa jatkuvasti.

Protokollien kehittämisen syistä on käyty laajaa keskustelua tieteellisen toimikunnan yleiskokoukselle osoittamissa raporteissa, jotka löytyvät FIMM NEWS julkaisuista. Aikaisemmin julkaistut protokollat ovat vähitellen johtaneet toistettavuustutkimuksiin manuaalisessa lääketieteessä eri maissa.

Perussyyt protokollien kehittämiseen ovat edelleen ajankohtaisia. Tämän vuoksi tekstiin on sisällytetty lyhyesti protokollien taustaa ja yleiskuvaus tieteellisen toimikunnan toiminnasta.

Tiivistettynä manuaalisen lääketieteen ongelma on toimikunnan mukaan seuraava:

**Manuaalisessa lääketieteessä on liian monia erilaisia kouluja ympäri maailmaa, joilla on liian monenlaisia diagnostisia menetelmiä ja liian monia erilaisia terapeuttisia lähestymistapoja.**

Edeltävästä lausunnosta johtopäätöksenä:

**I.1.** Suurin osa manuaalisen lääketieteen kouluista ei vielä ole validioinut omia diagnostisia menetelmiään tuki- ja liikuntaelimestön eri alueilta. Tästä johtuen näiltä diagnostisilta menetelmiltä puuttuvat toistettavuus, validiteetti, sensitiivisyys ja spesifisyys.

**I.2.** Validioinnin puuttumisesta johtuen erilaisten diagnostisten menetelmien vertailu on mahdotonta, ja koulukuntien välinen tieteellinen keskustelu ja informaation vaihdos puuttuvat.

**I.3.** Validioitujen diagnostisten menetelmien puuttuminen manuaalisessa lääketieteessä johtaa heterogeenisiin perusjoukkoihin vaikuttavuustutkimuksissa. Tästä johtuen vertailu vaikuttavuustutkimusten välillä on mahdotonta silloinkin, kun tutkimuksissa on käytetty samankaltaisia terapeuttisia menetelmiä (esim. manipulaatiota).

**I.4.** Nykyisten puutteiden jatkuessa manuaalisen lääketieteen ammatillistuminen hidastuu.

**I.5.** Merkittävimmät syyt huonoon näyttöön manuaalisen lääketieteen vaikuttavuudesta ovat validiteetin puuttuminen diagnostisilta menetelmiltä, huonosti määritellyt terapeuttiset lähestymistavat ja alhainen taso tutkimusten suunnittelussa.

Toimikunnan pyrkimyksenä on luoda foorumi tieteellisen informaation jakamiseen ja sen myötä pohja perusteelliseen keskusteluun eri manuaalisen lääketieteen koulukuntien välillä. Tämän informaation vaihdon on perustuttava luotettavaan tieteelliseen työhön.

Perusteellinen keskustelu käynnistyy, kun päästään vertailemaan eri koulujen tekemiä hyviä toistettavuus-, validiteetti-, sensitiivisyys- ja spesifisyystutkimuksia. Tarkoituksena ei ole määrittää millä koulukunnalla on paras diagnostinen testi tietylle alueelle, vaan määrittää joukko validioituja diagnostisia menetelmiä eri koulukuntien käyttöön, jotka siirtyisivät myös tavallisen lääketieteen käytettäviksi.

Tieteellinen toimikunta haluaa antaa FIMM järjestön kansallisille yhteisöille standardoituja tieteellisiä protokollia tuleviin tutkimuksiin.

Tieteellinen toimikunta järjestää vuosittain konferenssin jonkin kansallisen yhteisön avustuksella. Vuonna 2003 konferenssi järjestettiin Tanskan Odensessa, Danish society for manual medicine järjestön avustuksella. Monet tutkijat esittelivät alustavia tutkimustuloksia, ehdotuksia protokollaformaateiksi ja terapeuttisia päättelyketjuja. Osallistujien välillä käytiin hedelmällistä keskustelua tieteelliseltä pohjalta, jota ei häirinnyt eri oppikuntien näkemuserot.

Toistettavuus-, validiteetti-, sensitiivisyys- ja spesifisyystutkimukset ovat helppoja ja halpoja suorittaa, sekä toimivat parhaana perustana keskustelussa eri manuaalisen

lääketieteen koulujen välillä. Tämän vuoksi niiden tekeminen pitäisi olla ensisijalla manuaalisen lääketieteen tutkimusten teossa.

Kansallisten FIMM järjestöjen aktiivinen osallistuminen on korvaamatonta ja ratkaisevan tärkeää tieteellisen toimikunnan tulevalle työlle.

Tieteellinen toimikunta toivoo antavansa merkittävän panoksen manuaalisen lääketieteen ammatillistumiselle toimittamalla tämän kolmannen protokollan kansallisille FIMM järjestöille.

Dr. Jacob Patijn, MD, PhD, neurologi ja tieteellisen toimikunnan puheenjohtaja

## **II. TOISTETTAVUUS JA VALIDITEETTI**

### **Nimikkeistö**

Yksi merkittävä ongelma lääketieteessä ja tutkimuksessa on eri nimitysten käyttäminen samoista määritelmistä. Sen vuoksi tässä tekstissä käytettävät termit määritellään, jotta lukeminen olisi helpompaa.

**II.1. Reliabiliteetti** voidaan jakaa toistettavuuteen (precision) ja validiteettiin (accuracy).

#### **II.1.1. Toistettavuus (precision, reproducibility)**

Saman tutkijan kahtena eri ajankohtana tekemien havaintojen välistä toistettavuutta kutsutaan tutkijan sisäiseksi yksimielisyydeksi (intra-observer variability/ intra-observer agreement).

Kahden tutkijan tekemien havaintojen toistettavuutta samana ajankohtana kutsutaan tutkijoiden väliseksi yksimielisyydeksi (inter-observer variability/ inter-observer agreement).

Tässä protokollassa käytetään termejä **toistettavuus**, **tutkijan sisäinen yksimielisyys** ja **tutkijoiden välinen yksimielisyys**.

Manuaalisessa lääketieteessä diagnostisen testin toistettavuus kertoo löytävätkö kaksi tutkijaa samat löydökset samasta potilasryhmästä, tai löytääkö yksittäinen tutkija samat löydökset kahtena eri ajankohtana samasta potilasryhmästä.

### **II.1.2. Validiteetti (accuracy, validity)**

Validiteetti ilmaisee miten hyvin diagnostinen testi mittaa asiaa, jota sen oletetaan mittaavan. Tarkemmin sanottuna validiteetti määritetään vertaamalla testin tuloksia kultaiseen standardiin.

Lääketieteen ja myös manuaalisen lääketieteen ongelmana on vertaiskohteena toimivan kultaisen standardin määrittäminen. Arviointiperusteena eli kultaisena standardina käytetään ruumiinavausten ja leikkausten yhteydessä tehtyjä löydöksiä sekä radiologisia havaintoja. Normaalisti väestöstä tehty määrällinen tutkimus liikelaajuudesta toimii kultaisena standardina liikelaajuutta arvioitaessa. Usein lääketieteen asiantuntijoiden muodostama konsensus on ainoa tapa määrittää arviointiperuste.

Kultaisia standardeja tarvitaan testin sensitiivisyyden ja spesifisyyden arviointiin (katso V.1).

## **II.2. Tunnusmerkki ja sen esiintyvyys**

### **II.2.1. Tunnusmerkki (index condition)**

Tunnusmerkki on synonyymi potilaan diagnoosille, jonka on pohjauduttava toistettaviin validioituihin menettelytapoihin. Tutkittaessa diagnostisen testin toistettavuutta, positiivista testitulosta kutsutaan tunnusmerkiksi.

### **II.2.2. Tunnusmerkin esiintyvyys (prevalence of index condition)**

Tunnusmerkin esiintyvyys on tunnusmerkin esiintymistiheys tietyssä perusjoukossa tietyllä hetkellä. Toistettavuustutkimuksissa tunnusmerkin esiintyvyys liittyy ainoastaan tutkimusjoukkoon. On tärkeää ymmärtää, että tunnusmerkin esiintyvyys voi vaihdella instituuttien ja maiden välillä. Se voi myös muuttua eri ajankohtina.

Tämän protokollan toistettavuusosiossa käytetään termejä **tunnusmerkki** ja **tunnusmerkin esiintyvyys positiivisiksi havaittujen testimenetelmien yhteydessä**.

Alla olevan kuvion 2 x 2 kontingenssitaulukossa esitetään teoreettinen esimerkki kahden henkilön (A ja B) toistettavuustutkimuksen tuloksista.

		<u>Tutkija B</u>		Yhteensä:
		Kyllä	Ei	
<u>Tutkija A</u>	Kyllä	a	b	a+b
	Ei	c	d	c+d
Yhteensä:		a+c	b+d	n

Kuvio 1. 2 x 2 kontingenssitaulukko

Ruudut a ja b kuvaavat A tutkijan positiivisiksi arvioimien potilaiden ryhmän. Ruudut a ja c ovat B tutkijan positiivisiksi arvioimien potilaiden ryhmän. Ruudut a, b ja c kuvaavat molempien tutkijoiden positiivisiksi arvioidut tapaukset potilaiden kokonaismäärästä(n). Esiintyvyys (P) lasketaan kaavalla:

$$P = \frac{a + (b + c)/2}{n} \quad (\text{kaava 1})$$

### II.3. Yhtäpitävyys (overall agreement)

Yhtäpitävyys on prosenttiosuus potilaista, joista tutkijat ovat yksimielisiä. Kuviossa 1 ruudut a ja d kuvaavat tapaukset, joista A ja B tutkijat ovat yksimielisiä (kyllä tai ei).

Ruudut b ja c ilmaisevat tapaukset, joista tutkijat ovat erimielisiä. Yhtäpitävyys lasketaan kaavalla:

$$P_o = \frac{[a + d]}{n} \quad (\text{kaava 2})$$

## II.4. Sensitiivisyys ja spesifisyys

### II.4.1. Sensitiivisyys

Testin sensitiivisyys on osuus tapauksista, joilla on tunnusmerkki ja jotka testi paikkansapitävästi löytää.

### II.4.2. Spesifisyys

Testin spesifisyys on osuus tapauksista, joilla ei ole tunnusmerkkiä ja jotka testi paikkansapitävästi löytää.

Tässä protokollassa ”nosografinen sensitiivisyys ja spesifisyys” ovat identtisiä termejä sensitiivisyydelle ja spesifisyydelle.

### II.4.3. Positiivinen ja negatiivinen ennustearvo (predictive value)

Päivittäisessä vastaanotossa sensitiivisyys ja spesifisyys kertovat, onko positiivinen testituloks oikea positiivinen vai väärä positiivinen. Tämä ilmaistaan vastaavasti termeillä ”testin positiivinen ja negatiivinen ennustearvo”. Samaa ilmaisevat myös ”diagnostinen sensitiivisyys ja spesifisyys”.

”Diagnostisesta sensitiivisyydestä ja spesifisyydestä” käytetään identtisesti tässä tekstissä termejä ”**positiivinen ja negatiivinen ennustearvo**”.

## II.5. Kappa-arvo: tulkitseminen

Tässä protokollassa kappastatistiikkaa käytetään toistettavuustutkimuksiin.



**Kappa-arvo** on tilastollinen arvo tutkijan sisäisestä ja tutkijoiden välisestä yksimielisyydestä, jossa sattuma on huomioitu. Kappa-arvot voivat olla negatiivisia tai positiivisia. Kappa-arvot ovat välillä -1 ja +1.

Yleisimmin käytetyn kaavan (Koch & Landis) mukaan kappa-arvojen merkitsevyys on seuraavanlainen:

yli 0.6 hyvä/loistava yhtäpitävyys

0.4-0.6 kohtalainen/hyvä yhtäpitävyys

alle 0.4 huono yhtäpitävyys

Bogduk käyttää arvoa 0.4 rajana hyvälle yhtäpitävyys. Tässä protokollassa käytetään konservatiivista 0.6 raja-arvoa ilmaisemaan hyvästä erinomaiseen yhtäpitävyys.

### **III. DIAGNOSTISEN TOISTETTAVUUSPROTOKOLLAN LÄHTÖKOHTIA MANUAALISESSA LÄÄKETIETEESSÄ**

#### **III.1. Diagnostisten ja tilastollisten metodien luonteesta**

Ennen toistettavuustutkimuksen aloittamista on tärkeää olla selvillä, minkälaisia diagnostisia menettelytapoja on olemassa ja minkälaiset tilastolliset menetelmät ovat soveltuvia käytettäväksi. Yleisesti diagnostiset menetelmät jakautuvat kvalitatiivisiin ja kvantitatiivisiin.

##### **III.1.1. Kvalitatiiviset ja semi-kvantitatiiviset diagnostiset menettelytavat (luokittelu- ja järjestysasteikko)**

Manuaalisessa lääketieteessä kvalitatiiviseen tutkimukseen liittyvät tutkijan ja/tai potilaan subjektiiviset havainnot. Tyypillisiä esimerkkejä ovat loppujousto ja kipuprovoakaatio erilaisissa olosuhteissa (tutkijan provosoima, potilaan oman liikkeen provosoima). Jos löydös on tai puuttuu (kyllä/ei), puhutaan luokitteluasteikosta (nominal data). Esimerkki luokitteluasteikosta on kipuprovoakaatiotestien tulokset (on kipua, ei ole kipua).

Luokitteluasteikon käsittelyssä kappastatistiikka on aiheellinen.

Jos voidaan erotella erilaisia kategorioita (ei loppujousto, pehmeä loppujousto, kova loppujousto, hyvin kova loppujousto), on kyseessä järjestysasteikko (ordinal data) ja on aiheellista käyttää painotettua kappa.

Kun mitataan subjektiivisesti kahden erillisen kategorian eroja, on kyseessä semi-kvantitatiivinen diagnostinen menetelmä. Esimerkkinä tästä olisi tutkimus oikean ja vasemman puolen välisestä erosta liikelaajuudessa tai liikkeen rajoittuneisuudessa.

### **III.1.2. Kvantitatiivinen diagnostinen menettelytapa**

Kvantitatiivisissa tutkimuksissa mittaus tapahtuu useimmiten laitteen avulla ja tulokset ilmaistaan asteina, millimetreinä, kiloina jne. Kvantitatiivisessa tutkimuksessa on siis intervalli- tai jatkuva-asteikko. Kvantitatiivista menettelytapaa varten tarvitaan normiarvoja. Ensin tarvitaan tutkimus normaalista väestöstä, joista menetelmän toistettavuutta arvioidaan samalla perusjoukolla kahtena ajankohtana. Ikä ja sukupuoli ovat myös huomioitava, sillä ne voivat vaikuttaa arvoihin. Kvantitatiiviset tutkimukset voivat toimia kultaisena standardina semi-kvantitatiivisille tutkimuksille.

Kaiken tyyppisissä toistettavuustutkimuksissa käytettävän statistisen menetelmän määrittää kerätyn tiedon luonne (luokittelu-, järjestys-, intervalli-, tai jatkuva-asteikko).

### **III.1.3. Epäsopiva statistiikka kvalitatiivisissa toistettavuustutkimuksissa**

Toistettavuuden mittaamiseen käytetään useasti soveltumatonta statistiikkaa. Olennaisin virhe, että yksimielisyys usein sekoitetaan assosiaatioon, joka on ennustettavuus kahden muuttujan välillä. Seuraavassa listataan toistettavuustutkimuksissa esiintyviä statististen metodien ongelmia.

#### **III.1.3.1. Prosentuaalinen yksimielisyys (percent agreement)**

Prosentuaalinen yksimielisyys on suhde kaikkien havaintojen ja tutkijoiden yhtäläisten tulosten välillä. Prosentuaalinen yksimielisyys ei anna pelkästään mainittuna mitään todellista informaatiota toistettavuudesta. Sen suurin ongelma on, ettei se huomioi kuinka suuren samankaltaisten tulosten määrän pelkkä sattuma voi aiheuttaa.

### **III.1.3.2. Korrelaatiokerroin (correlation coefficients)**

Monissa toistettavuustutkimuksissa käytetään korrelaatio- ja assosiaatiomittaa arvioimaan kliinisen datan toistettavuutta. Ongelmana on, ettei kaikilla tutkimusta tekevillä ole kykyä erotella yksimielisyyttä erimielisyydestä (khiin neliö ja phi kerroin). Systeemistä tutkijan harhaa ei myöskään osata ottaa huomioon (Pearsonin tulomomenttikorrelaatio, järjestyskorrelaatio).

### **III.1.4. Sopiva statistiikka kvalitatiivisiin ja semi-kvantitatiivisiin toistettavuustutkimuksiin**

#### **III.1.4.1. Tavallinen kappa (normal kappa)**

Tavallinen kappa valitaan, kun arvioidaan toistettavuutta kahden tutkijan välillä ja käytössä on luokitteluasteikko (kyllä/ei).

#### **III.1.4.2. Yleinen kappa (overall kappa)**

Kun tutkimuksessa on enemmän kuin kaksi tutkijaa, voidaan käyttää yleistä kappa. Esimerkiksi arvioitaessa segmentaalisen dysfunktion esiintyvyyttä tietyllä alueella, antaa yleinen kappa arvion eri tutkijoiden välisestä yleisestä toistettavuudesta havainnoida dysfunktio kyseiseltä alueelta.

#### **III.1.4.3 Painotettu kappa (weighted kappa)**

Arvioinnin jakautuessa useampaan asteluokkaan, joilla on luontainen järjestys esimerkiksi loppujoustossa (normaali, elastinen, kova) käytetään painotettua kappa. Yksityiskohtien selvittämiseksi tutustu kirjallisuuteen aiheesta, tai kysy tilastotieteen asiantuntijalta.

**III.1.4.4. Kappatutkimuksen merkitsevyys ja luottamusväli (confidence intervals)** voidaan laskea, kun kappa-arvot ovat välillä 0.4-0.6. Laskemalla selvitetään onko arvo sattumaa parempi. Yli 0.6 arvoista laskeminen on tarpeetonta.

Seuraavaksi käsitellään yksityiskohtaisesti tavallista kappastatistiikkaa.

### III.1.5. Sopiva statistiikka kvantitatiivisiin toistettavuustutkimuksiin

Havaintoparin lukujen erotuksen t-testiä (paired t-test) käytetään arvioimaan kvantitatiivisen datan (intervalli- tai jatkuva-asteikko) toistuvia mittauksia.

Varianssianalyysi (ANOVA ICC) on statistinen menetelmä intervalli-asteikon (cm, mm, ym.) käsittelyyn. Siinä laskettavan R-faktorin arvo 1 kertoo identtisistä arvoista. Alle 0 arvoja saadaan, kun toistettavuus puuttuu. ICC:n puute on, ettei se anna informaatiota tutkijoiden välisen erimielisyyden suuruudesta.

Statistiikan valitseminen toistettavuustutkimuksissa riippuu sekä datan luonteesta (luokittelu-, järjestys- tai intervalliasteikko) että tutkimuksen löydöksistä tehtävän kliinisen johtopäätöksen luonteesta.

Esimerkiksi tutkittaessa tarvitaanko kantakoroketta tasaamaan jalkojen pituuseroa, käytetään ANOVA ICC statistiikkaa intervalliasteikon käsittelemiseen. Mitattaessa jalkojen pituuseroa lantion asennon korjaamiseksi asteikon luonne on oikea, vasen tai tasan eli on aiheellista käyttää luokitteluasteikkoa. Sama soveltuu myös semi-kvantitatiiviseen asteikkoon, kuten liikkuvuudeltaan rajoittuneen puolen löytämiseen, jolloin vastauksena on kyllä tai ei.

### III.2. Diagnostisen menettelytavan tavoite

Testattaessa diagnostisten menettelytapojen toistettavuutta on tärkeää olla selvillä testien tavoitteista. On välttämätöntä ymmärtää erot diagnoosin, syndrooman ja diagnostisen testin välillä. Aidossa diagnoosissa tiedetään etiologia ja prognoosi. Syndroomissa esiintyy merkkejä ja oireita yhtäaikaaisesti ja tietyssä perusjoukossa niiden esiintymistiheys voi olla suuri. Etiologia on syndroomissa tuntematon.

Diagnostisia testejä tarvitaan niin diagnooseissa kuin syndroomissa. Diagnostinen testi on klinikon suorittama menettelytapa, joka havainnollistaa kliinisen löydöksen kvalitatiivisella tavalla. Esimerkiksi yhdistelmä sensorista ja motorista puutosta ja positiivinen Laseque voivat olla tyypillinen löydös radikulaari syndroomassa. Etiologia voisi olla niin välilevyn protruusio kuin kasvain hermojuuren aukossa. Molemmat aiheuttaisivat kompressiota hermojuureen.

Manuaalisen lääketieteen eri koulutusjärjestelmissä opetetaan monia testejä. Kaularangan passiivisen liikelaajuuden tutkiminen on yksi esimerkki. Opiskelijat oppivat, kuinka testi kokonaisuudessaan suoritetaan (käsien sijoittaminen, voiman suuruus jne.). Testissä havaittua liikerajoitusta voi kuitenkin selittää monet tekijät, joten testi ei anna mitään informaatiota diagnoosin tekemiseksi.

Ensimmäinen prioriteetti testien tutkimisessa on saada ne toistettaviksi. Toiseksi on löydettävä kultainen standardi testin validioimiseen. Esimerkiksi havaittu liikerajoitus olisi vahvistettava kvantitatiivisella metodilla ja tarkoitukseen suunnitellulla laitteella, joka mittaisi liikelaajuuksia asteissa eri-ikäisiltä ja eri sukupuolta olevilta henkilöiltä.

Tämän jälkeen validioituja testejä voidaan käyttää määrittämään syndroomia manuaalisessa lääketieteessä. Lopuksi on validioitava itse kultainen standardi, joka on useasti erittäin vaikeaa.

**III.2.1.** Yksittäisen diagnostisen testin arviointi antaa informaatiota vain testin menettelytavan toistettavuudesta. Yksittäinen testi ei siis yleensä anna mitään informaatiota diagnoosin tekemistä varten, eikä se näin ollen indikoi mitään hoitotoimenpidettä. Yksittäinen testi harvoin erottaa normaalia sairaasta. Yleisesti ottaen yksittäiseen toistettavaan testiin pohjautuvat sensitiivisyys- ja spesifisyystutkimukset ovat hyödyttömiä, jos kultainen standardi puuttuu.

**III.2.2.** Toistettavien testien yhdistelmäkään eivät anna diagnostista arvoa ilman kultaista standardia. Yhdistelmä positiivisia testejä voi siis esiintyä niin spesifisessä diagnoosissa kuin epäspesifisessä syndroomassa sekä myös normaaleilla henkilöillä. Useamman testin yhtäaikaista testaamista toistettavuustutkimuksessa ei ole suositeltavaa, koska testit saattavat olla riippuvaisia toisistaan.

**III.2.3.** Testin sensitiivisyyttä ja spesifisyyttä voidaan arvioida toistamalla testi väliajan jälkeen samalla potilaalla. Yhdistämällä tällaiset testit muuhun kliiniseen tietoon kasvaa mahdollisuus erottaa potilaat normaaleista verrokeista. Suurimmassa osassa tapauksia tämä ei kuitenkaan johda lisäinformaation saamiseen diagnoosin tekemistä varten. Yleisesti ottaen vain kultaisen standardin määrittämisen jälkeen on hyödyllistä tehdä validioitujen testien yhdistelmään perustuvia sensitiivisyys- ja spesifisyystutkimuksia.

### III.3. Arvioitavien testien määrä

Yleisellä tasolla olevassa toistettavuustutkimuksessa, esimerkiksi alaselkäkkiputkimuksessa, arvioidaan joskus useampaa testiä samanaikaisesti. Tämän kaltaisissa tutkimuksissa monet testit näyttävät alhaista kappa-arvoa ja siten tuomitaan tutkimuksen tekijän puolelta kliinisesti merkityksettömiksi. Koska esiintyvyys ja yhtäpitävyys usein puuttuvat, selvää johtopäätöstä ei voida vetää testien toistettavuudesta, joilla on alhainen kappa-arvo. Heterogeenisten tutkimusten mukaan väestö sisältää useita erilaisia alaryhmiä, joilla jokaisella on erilainen esiintyvyys arvioitavaan testiin. Alaryhmän ollessa pieni tutkimukseen voi liittyä riski, että tutkimuksessa osalla positiivisista testeistä on alhainen esiintyvyys.

Tutkimuksen perusjoukon tulee olla soveltuva arvioitavaan testiin. Esimerkiksi ei kannata tutkia toistettavuutta useilla radikulaarioireiden provokaatiotesteillä alaselkäkkipupotilailla, joilla ei ole merkkejä iskiasoireista, koska positiivinen tulos on erittäin harvinainen kyseisellä joukolla.

Tutkittaessa usean radikulaarioireen provokaatiotestin toistettavuutta samaan aikaan iskiasoireisilla potilailla, voidaan päättää pienin tarvittava määrä positiivisia tuloksia diagnosoimaan radikulaari-syndrooma. Haittana tehtäessä diagnoosia useammalla testillä (esimerkiksi radikulaari-syndrooma, SI-dysfunktio) on, että keskinäiset riippuvuussuhteet muuttuvat (mutual dependency).

Esimerkiksi manuaalisessa lääketieteessä monen SI-testin on tarkoitus testata SI-nivelen dysfunktioita tai hypomobilitteettia. Tämä keskinäinen riippuvuussuhde tutkittiin kuuden SI-nivelen testin toistettavuustutkimuksessa samanaikaisesti (Deursen van, Patijn). Kyseisessä tutkimuksessa kolmen tutkijan (A, B, C) oli tarkoitus käyttää kuutta eri SI-testiä (I-VI) lopulliseen SI-nivelen diagnoosiin (katso kuvio 2 ja 3). Tutkijan testien keskinäistä riippuvuussuhdetta arvioitiin laskemalla kappa-arvo 15 mahdollista kombinaatioparia kuudesta eri SI-testistä.

Testi:	Tutkija:	I	II	III	IV	V	VI
	A					<u>Kappa-arvot</u>	
I	B						
	C						
II	A	-0,09					
	B	0,02					
	C	0,36					
III	A	0,25	-0,01				
	B	0,34	0,17				
	C	0,36	0,22				
IV	A	0,34	-0,29	0,25			
	B	0,06	-0,05	0,15			
	C	0,22	-0,01	0,36			
V	A	<b>0,61</b>	-0,12	0,28	<b>0,43</b>		
	B	0,33	0,39	0,34	0,01		
	C	0,1	0,21	0,21	0,32		
VI	A	<b>0,61</b>	-0,22	0,18	<b>0,43</b>	<b>0,89</b>	
	B	0,23	0,19	0,21	-0,15	<b>0,52</b>	
	C	0,21	0,32	0,24	0,27	<b>0,84</b>	

Kuvio 2. Kuuden SI-testin keskinäinen riippuvuussuhde (I-VI) kolmella tutkijalla (A, B, C). Tummennetut kappa-arvot  $>0.40$  kuvastavat keskinäistä riippuvuussuhdetta.

Kuvio 2 esittää kaikkien kolmen tutkijan eri testiparien arvot, jotka ovat  $>0.40$  tummennettuna käytettäessä kappa-arvoa 0.40 alhaisimpana hyväksyttävänä arvona. Etenkin V ja VI-testien välillä kaikilla tutkijilla esiintyi korkea kappa-arvo ( $+0.89, +0.52, +0.84$ ) kuvastaen keskinäistä riippuvuussuhdetta testien V ja VI välillä. Tämä tarkoittaa, että kaikki kolme tutkijaa tiedostamattomasti päättelivät VI- testin positiiviseksi, jos V testi oli positiivinen. Kyseisessä tutkimuksessa testit II, III verrattuna I, IV ja VI eivät osoittaneet keskinäistä riippuvuussuhdetta (toinen ja kolmas palkki).

Tämä näkökulma keskinäisestä riippuvuussuhteesta on myös erittäin tärkeää toistettavuustutkimuksissa, joissa valitaan testejä samaan kliiniseen diagnoosiin. Kappa-tutkimuksessa voidaan arvioida testin toistettavuuden lisäksi myös tutkijoiden välistä yhtenäisyyttä lopullisesta diagnoosista. Aiemmin mainitussa tutkimuksessa selvisi, että

monesta testistä lopulliseen tutkijoiden SI-diagnoosiin vaikutti vain muutama. Tämä osoitetaan kuviossa 3, jossa on laskettu jokaiselle testille (I-VI) keskinäinen kappo-arvo ja kaikkien kolmen tutkijan (A, B, C) lopullinen diagnoosi.

Testi:	Tutkija:	I	II	III	IV	V	VI	SI- diagnoosi Kappa
I	A							-0,61
	B							0,23
	C							0,21
II	A	-0,09						-0,22
	B	0,02						0,19
	C	0,36						0,32
III	A	0,25	-0,01					0,18
	B	0,34	0,17					0,21
	C	0,36	0,22					0,24
IV	A	0,34	-0,29	0,25				<b>0,43</b>
	B	0,06	-0,05	0,15				-0,15
	C	0,22	-0,01	0,36				0,22
V	A	<b>0,61</b>	-0,12	0,28	<b>0,43</b>			<b>0,89</b>
	B	0,33	0,39	0,34	0,01			<b>0,52</b>
	C	0,1	0,21	0,21	0,32			<b>0,84</b>
VI	A	<b>0,61</b>	-0,22	0,18	<b>0,43</b>	<b>0,89</b>		<b>1</b>
	B	0,23	0,19	0,21	-0,15	<b>0,52</b>		<b>1</b>
	C	0,21	0,32	0,24	0,27	<b>0,84</b>		<b>1</b>

Kuvio 3. Kuuden testin (I-VI) keskinäinen riippuvuus jokaisen tutkijan (A, B, C) lopulliseen diagnoosiin. Tummennetut kappo-arvot  $>0.40$  osoittavat keskinäistä riippuvuutta.

Huomioi ylhäällä oikealla ”SI-diagnoosi” kohdassa, että kaikki kolme tutkijaa käyttivät vain V ja VI testejä lopulliseen SI-diagnoosiin. Kenelläkään tutkijalla (A, B, C) kaikki SI-testit (I-VI) eivät vaikuttaneet lopulliseen diagnoosiin.

Yleisesti on hyvä arvioida samaa kliinistä ominaisuutta enintään kolmella testillä. Olisi hyvä valita täysin erilaisia testimenetelmiä eikä pelkästään yhteen niveleen liittyviä.



### III.4. Tutkijoiden määrä

Toistettavuustutkimusta tehdessä ei ole mitään tilastollista syytä, miksi tarvittaisiin enemmän kuin kaksi tutkijaa. Joissain tutkimuksissa tutkijoita on enemmän arvioitaessa heidän välistä yksimielisyyttä. Ongelma kokeneilla tutkijilla on, että he ovat todennäköisesti kehittäneet omat suoritustavat ja tulkinnat testeille. Useimmat näistä tutkimuksista ovat vailla kunnollista harjoitteluvaihetta testin suorittamisen ja tulkinnan standardointia varten. Tämän kaltaisten tutkimusten tulokset informoivat enemmän tutkijoiden taidoista ja/tai koulutuksellisesta laadusta kuin arvioitavan testin toistettavuudesta. Sama pätee toistettavuustutkimuksissa, joissa testien kapp-arvot on tehty ns. in-vivo tilanteessa, jossa testiprosessin standardoimista ei ole toteutettu (tarkoittaen päivittäistä testin harjoittelua).

Ainoa tapaus, jossa useampi tutkija voi osallistua kapp-tutkimukseen on, kun arvioidaan säännöllisen harjoittelun vaikutusta kapp-arvoon. Samat tutkijat harjoittelevat toistuvasti testin diagnostista suorittamista ja jokaisen harjoitteluvaiheen jälkeen uusi kapp-arvo kuvastaa suuretuessaan harjoituksen onnistumista.

Periaatteessa käytettäessä alla ehdotettua tutkimusformaattia, se antaa meille potentiaalisen toistettavuuden testimenetelmälle. Jos testimenetelmän toistettavuutta ei saavuteta, voidaan toinen tutkimus muodostaa arvioimalla tutkijoiden ominaisuuksien vaikutusta toistettavuuteen.

Toinen virhe käytettäessä liian montaa tutkijaa toistettavuustutkimuksessa on, että testimenetelmällä saattaa olla terapeutista vaikutusta. Jos eri tutkijat toistavat yhdellä potilaalla passiivisesti useaan otteeseen testin (esim. kaularangan liikelaajuus), terapeutinen vaikutus saattaa vaikuttaa liikelaajuuteen ja siten viimeisen tutkijan tulokseen.

Yleensä käytettäessä kyseistä protokollaa, kaksi tutkijaa riittää arvioimaan testin toistettavuutta.

### III.5. Testin hypoteesi

Testin toistettavuutta tutkittaessa on tärkeää keskustella ja analysoida, mitä kyseisen testin on tarkoitus testata. Liikkeenlaajuutta testatessa ei ole ongelmaa, mutta mobiliteettia tutkittaessa, esimerkiksi SI-nivelen hypomobilitteettia, tulee ongelmia. Monissa SI-nivelen toistettavuustutkimuksissa hypoteesina oli, että testeillä oletettiin tutkittavan SI-nivelen liikkuvuutta. Vaikka SI-nivelen liikkuvuus on todistettu useissa ruumiinavaustutkimuksissa, on sen liikkuvuuden testaaminen mahdotonta edes kokeneimmalle tutkijalle. Tämä virheellinen luulo on todennäköisesti syy SI-nivelen testien alhaisiin kappi-arvoihin kirjallisuudessa. Tarkasteltaessa kriittisesti useita huomattavan erilaisia menetelmiä SI-nivelen liikkuvuuden tutkimiseen, on kyseenalaistettava voivatko kaikki menetelmät testata SI-nivelen hypomobilitteettia. Toistettavuustutkimuksissa tutkijan on unohdettava hänelle opetettu testin hypoteesi ja keskityttävä testimenetelmän kaikkiin erilaisiin näkökulmiin ja yksityiskohtiin.

Esimerkiksi kirjallisuuden mukaan Patrickin testin oletetaan testaavan SI-nivelen liikkuvuutta. Tarkasteltaessa kriittisesti testimenetelmää, tutkijat voivat päätellä, että Patrickin testi, joka tutkii loppujoustoa ja liikerajoitusta testaa vain lisääntyntä lihaskireyttä tietyssä lonkanivelen lihasryhmissä.

Hypoteesien vaikutusta SI-testien toistettavuuteen tutkittiin kahdessa tutkimuksessa (Patijn 2000). Ensimmäisessä tutkimuksessa, jossa tutkittiin kuutta SI-nivelen liikkuvuutta arvioivaa testiä, lopputulokseksi tulivat erittäin alhaiset kappi-arvot. Toisessa tutkimuksessa kolmen testin oli tarkoitus testata lihasten hypertoniaa ja siitä johtuvaa liikerajoitusta eri lihasryhmissä lumbosakraalialueella, tuli kappi-arvoksi 0.7. Minkä tahansa testin toistettavuutta tutkittaessa pitää testimenetelmä tutkia vaiheittain ja sopia, mitä testi todellisuudessa testaa. Tähän sopimukseen perustuen tutkijat voivat määrittää vartenotettavamman hypoteesin testille, joka saattaa olla täysin vastakkainen kirjallisuudessa esitettäviin hypoteeseihin.

Parempiin toistettavuustutkimuksen tuloksiin voidaan päästä, kun tutkijat hyväksyvät useita vartenotettavia hypoteeseja testille. Toistettavuustutkimuksessa nämä asiat ovat olennaisia harjoitteluvaiheen testiformaatissa (katso kuvio 10, sivu 27).

### **III.6. Sokkouttaminen (blinding procedures)**

Jokaisessa toistettavuustutkimuksessa on tärkeää huolehtia sokkouttamisesta, ei pelkästään potilas/ tutkija tilanteesta, vaan myös molempien tutkijoiden välillä. Varmista, ettei testiä tehdessä ole kommunikointia tutkijoiden välillä ja käytä tutkijilla tulosten kirjaamiseen erillisiä lomakkeita. Tarvittaessa varmista, etteivät tutkijat ja potilaat kommunikoi keskenään.

### **III.7. Testimenetelmä ja testin arvostelu**

Kuten jo kappaleessa 5 mainittiin, tutkijoiden pitää standardoida koko testimenetelmä ja tapa, jolla he arvostelevat testitulokset. Seuraavassa keskusteltavasta protokollaformaattista (katso kuvio 10, sivu 27) selviää harjoitteluvaiheen merkityksellisyys toistettavuustutkimuksen standardoimiseksi. Lopullisessa julkaisussa pitää keskustella testimenetelmän määrittelystä ja merkityksestä. Ennaltaehkäisemään tutkijoiden ”henkilökohtaista tulkintaa” testin aikana suosittelemme tulostamaan lomakkeeseen testin standardoidun menetelmän ja arvioinnin.

### **III.8. Valinta ja henkilöiden määrä**

Toistettavuustutkimuksen lopullisessa julkaisussa on mainittava primaarilähde, mistä perusjoukon henkilöt on valittu tutkimukseen. Valintamenetelmän pitää olla erittäin selkeä. Yksinkertaisessa toistettavuustutkimuksessa riittää yleensä 40 henkilöä, mikä tekee tutkimuksen toteuttamisesta helpon ja halvan, eikä se edellytä suurta instituuttia.

### **III.9. Tilastotieto toistettavuustutkimuksessa: kappa-arvo**

Kahden tutkijan tekemässä toistettavuustutkimuksessa, joka perustuu luokitteluasteikkoon (kyllä/ei), valittuna menetelmänä on kappa-arvojen arviointi (katso alla).

### III.9.1. Kapan riippuvuus esiintyvyyteen

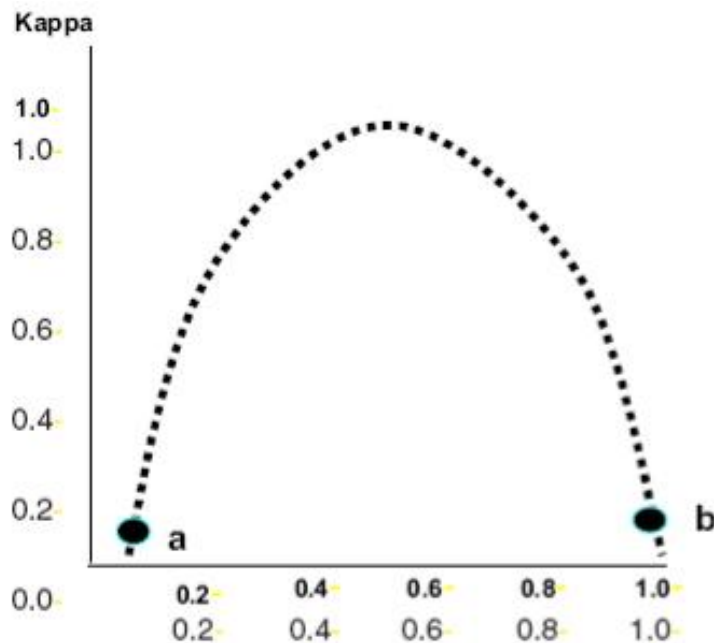
Kirjallisuudessa monet toistettavuustutkimukset diagnostisista testeistä, joiden kappa-arvo on alle 0.6, tuomitaan kliinisesti merkityksettömiksi. Kuitenkin suurimmassa osassa toistettavuustutkimuksista ei ole esitetty mitään tietoa tunnusmerkin esiintyvyydestä eikä yhtäpitävyydestä. Tämä on olennaista, koska kappa-arvo on riippuvainen esiintyvyydestä ja yhtäpitävyydestä.

Julkaistut toistettavuustutkimukset, joissa arvioidaan alhaista kappa-arvoa kliinisesti merkityksettömäksi tai vähäistä merkitystä olevaksi ovat harhaanjohtavia, ellei samalla mainita esiintyvyyttä ja yhtäpitävyyttä.

#### Alhaiset kappa-arvot voivat kuvastaa yhtä hyvin niin suurta kuin pientäkin esiintyvyyttä!

Kuvio 4 esittää kappa-arvon riippuvuuden esiintyvyyteen.

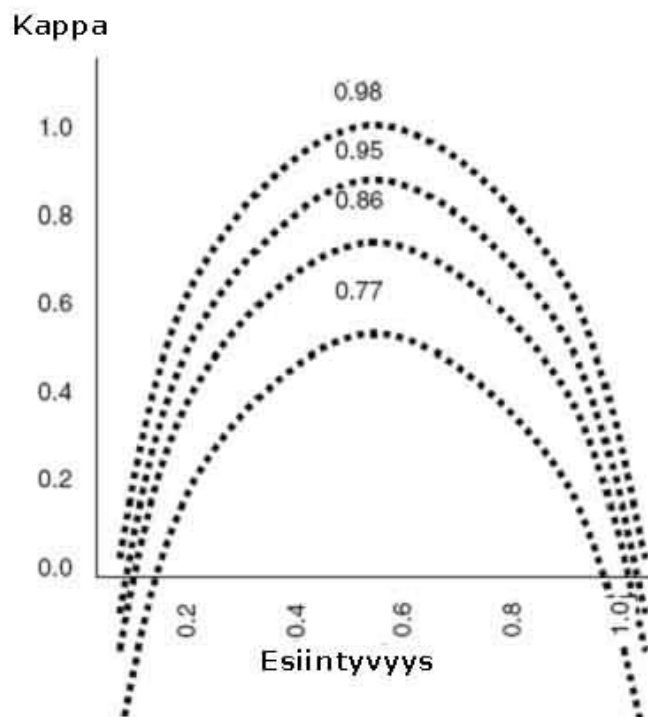
Huomioi, että erittäin alhaisella (a) ja korkealla (b) esiintyvyydellä kappa-arvosta tulee erittäin pieni.



Kuvio 4. Esiintyvyyden ja kappa-arvon välinen yhteys.

### III.9.2. Kapan riippuvuus yhtäpitävyyteen (Po)

Kuviossa 5 todistetaan, että korkealla yhtäpitävyyden (0.98 kuviossa) arvolla suurin kappo- arvo on 1.0 ja pienin kappo-arvo on lähellä 0.



Kuvio 5. Yhteys eri kappa-/esiintyvyysskäyrien ja yhtäpitävyyden arvojen 0.77-0.98 välillä.

Kappo-arvon taso riippuu kahden tutkijan yhtäpitävyydestä. Mitä alhaisempi yhtäpitävyys on toistettavuustutkimuksessa, sitä alhaisemmat ovat suurin ja pienin kappo-arvo. Tämä yhteys esitetään kuviossa 5.

Huomioi, että esiintyvyyss-/kappokäyrässä alhaisella yhtäpitävyyden arvolla (0.86 ja 0.77) pienin kappo-arvo on negatiivinen.

Kappo-arvon riippuvuus sekä esiintyvyyteen että yhtäpitävyyteen todistaa faktan, että kappo-arvo pitää esittää lopullisessa julkaisussa näiden muuttujien kanssa.

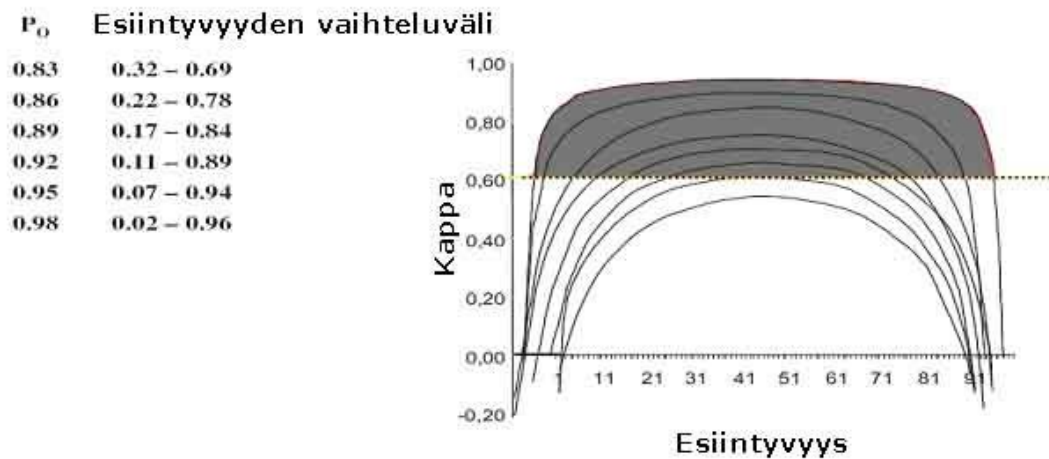
### III.9.3. Toistettavuustutkimuksen optimointimenetelmät: Yhtäpitävyyden ja 50 % esiintyvyyden saavuttaminen

Kun muodostetaan toistettavuustutkimus, lopputuloksena voi olla alhainen kappo-arvo kahdesta altistavasta tekijästä johtuen eli siihen vaikuttavat esiintyvyys ja yhtäpitävyys.

Ensinnäkin, jos yhtäpitävyys on alhaisempi kuin 0.80, on riski pieneen kappa-arvoon. Siitä johtuen harjoitteluvaihe on olennainen, jotta tutkijat yrittävät saavuttaa yhtäpitävyyden arvon suuremmaksi kuin 0.80. Tällä tavalla sen vaikutus lopulliseen kappa-arvoon ei ole merkittävä.

Toiseksi, kuten edellä on mainittu, erittäin alhainen tai korkea esiintyvyys vaikuttavat kappa-arvon pienenemiseen. Tämän takia on kehitetty teoreettinen metodi, jolla vaikutetaan tunnusmerkin esiintyvyyden kasvuun.

Kuviossa 6 esiintyvyys-/kappakäyrät esittävät yhtäpitävyyden vaihteluvälillä 0.83:sta 0.98:een.



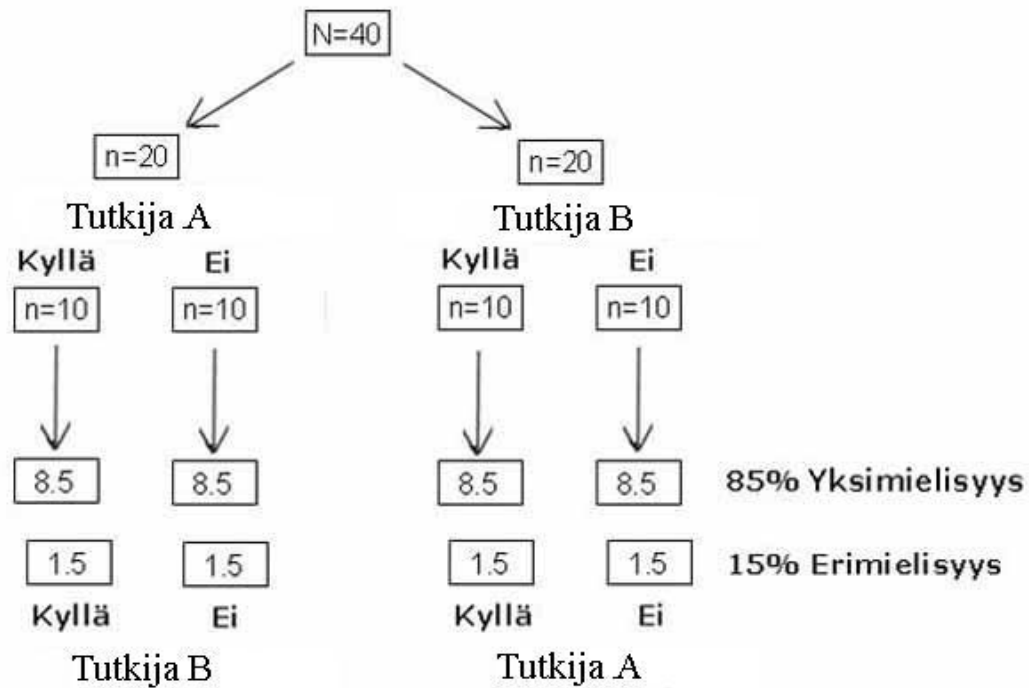
Kuvio 6. Kappa-/ esiintyvyyssäyrät eri yhtäpitävyyden arvoilla. Linja kappa-arvossa 0.60 on raja-arvona hyväksyttävälle kappa-alueelle tämän linjan yläpuolella. Huomioi, että kaksi alinta käyrää ( $P_o$  0.83 ja 0.86) sijaitsevat kappa-arvon 0.60 alapuolella. Käyrät, joiden arvo on  $P_o > 0.90$ , ovat olennainen alue (harmaa) kappa-linjan 0.60 yläpuolella.

Liian korkeasta tai alhaisesta esiintyvyydestä johtuvien odottamattoman alhaisten kappa-arvojen välttämiseksi suosittelemme, että tunnusmerkin esiintyvyys on lähellä 0.50.

Esiintyvyyden 0.50 kappa-arvo sijaitsee aina käyrän huipulla.

Oletetaan, että harjoitteluvaiheella saavutetaan yhtäpitävyyden arvoksi  $P_o$  0.85 (katso kuvio 10, sivulla 27). Toistettavuustutkimuksessa on testattavana 40 potilasta. Molemmat tutkijat (A ja B) valitsevat 20 potilasta ja lähettävät ne toisilleen testattavaksi. Tutkijat lähettävät 10 potilasta, joilla tulos on positiivinen ja 10 potilasta, joilla tulos on negatiivinen. Perustuen yhtäpitävyyden arvoon 0.85, molemmat tutkijat ovat yksimielisiä 85 % positiivisista ja

negatiivisista tuloksista ja erimielisiä 15 % tuloksista. Tämä koeasetelma on esitetty kuviossa 7.



Kuvio 7. Koeasetelma esittää 40 henkilöä yhtäpitävyyden ollessa 0.85, yritettäessä saada tunnusmerkin esiintyvyydeksi 0.50 (positiivinen testi).

Kappa-arvo voidaan laskea henkilöiden määrästä, joista tutkijat ovat yksi-/erimielisiä (kuvio 7). Tulokset esitetään kuviossa 8 2 x 2 kontingenssitaulukossa. Esiintyvyys on 0.50, yhtäpitävyyden arvo on 0.85, joten kappa-arvoksi muodostuu 0.70.

		<u>Tutkija B</u>	
		Kyllä	Ei
<u>Tutkija A</u>	Kyllä	<b>17</b>	<b>3</b>
	Ei	<b>3</b>	<b>17</b>

**Esiintyvyys P: 0.51**

**Yhtäpitävyys Po: 0.85**

**Kappa-arvo: 0.7**

Kuvio 8. 2 x 2 kontingenssitaulukko perustuen kuvion 7 tuloksiin.

Muodostaessa yhtäpitävyysvaihetta toistettavuustutkimuksessa, jossa yhtäpitävyys on yli 0.80 ja sen jälkeen muodostetaan kuvion 7 kaltainen menettelytapa, voidaan vaikuttaa esiintyvyyteen johtaen olennaisesti testimenettelytavan kappa-arvoon. Tämä oletettu teoreettinen rakenne todistettiin oikeaksi käytännössä testaten vuonna 2003 tehdyssä tutkimuksessa (Patijn 2003).

Helpoin tapa laskea kappa-arvo on käyttää taulukkolaskentaohjelmaa, jossa kaava on valmiina. Tällä tavalla vain perustiedot lisätään kaavaan ja se laskee automaattisesti kappa-arvon (katso liite 1). FIMM:n internetsivuilta on mahdollisuus ladata kyseinen ohjelma.

### **III.10. Kappatutkimuksen esittäminen**

Julkaistaessa toistettavuustutkimuksen tuloksia kaikki asiakohdat, jotka käydään läpi kappaleissa 1-8 pitää olla esitettyinä. Lisäksi olennaisia asioita julkaisussa ovat 2 x 2 kontingenssitaulukko, yhtäpitävyys ja esiintyvyys. Tällä tavalla julkaisua lukeva voi



helposti päätellä, mistä aineiston johtopäätökset ovat peräisin. Kuviossa 9 esitetään esimerkki 2 x 2 kontingenssitaulukosta. Kappa-arvot on myös esitetty.

		<u>Tutkija B</u>	
		Kyllä	Ei
<u>Tutkija A</u>	Kyllä	<b>38</b>	<b>0</b>
	Ei	<b>1</b>	<b>1</b>

**Esiintyvyys P: 0.96**

**Yhtäpitävyys Po: 0.98**

**Kappa-arvo: 0.7**

Kuvio9. Toistettavuustutkimuksen 2 x 2 kontingenssitaulukko 40 henkilöstä.

### III.11. Kappakirjallisuuden lähteet

Barlo W, Lay M I, Asen P, A comparison of methods for calculation a stratified kappa, *Statistics in Medicine*, 1991; 10(9): 1465-1472

Cohen J, A coefficient of agreement for nominal scales. *Educ Psychol Measurement*, 1960;20:37-46

Cohen J, Weighted Kappa: Nominal Scale agreement with provision for scaled disagreement of partial credit, *Psychol Bulletin*, 1968:70(4):213-220

Cook R, Kappa, In: *Encyclopedia of Biostatistics* Eds. Armitage P, Colton T, Publishers Wiley, NY, 1998: 2160-2165

Deursen L L J M, Patijn J, Ochhuysen A L, Vortman B J, The value of different clinical tests of the sacroiliac joint, *Ninth International Congress in Manual Medicine*, London 1989, Abstr. 16

Deursen L L J M, Patijn J, Ochhuysen A L, Vortman B J, The value of different clinical tests of the sacroiliac joint, *J Manual Medicine* 1990 (5): 96-99

Deursen L L J M, Patijn J, Ochhuysen A L, Vortman B J, Die Wertigkeit einiger klinischer Funktionstests des iliosakralgelenks, *Manuelle Medizin* 1992 vol 30(6): 43-46

Gjorup T, The Kappa Coefficient and the Prevalence of a Diagnosis, *Meth Inform Med*, 1988;27(4):184-186

Landis R J, Koch G G, The measurement of observer agreement for categorical data, *Biometrics*, 1977;33:159-174

Lantz C, Nebenzahl E, Behaviour and Interpretation of K Statistics: Resolution of the two paradoxes, *J Clin Epidemiology*, 1996;49(4):431-434

Patijn J, Stevens A, Deursen L L J M, Van roy J, Neurological Notions on the Sacroiliac Joint, In *Progress in Vertebral Column Research, First International Symposium on the sacroiliac Joint: Its Role in Posture and Locomotion* Editors A Vleeming, C J Snijders, R Stoeckart, Maastricht, 1991: 128-138

Patijn J, Brouwer R, Lennep van L, Deursen L, The diagnostic value of sacroiliac test in patients with non-specific low back pain, *J Orth Med*, 2000;22(1):10-15

#### IV. TOISTETTAVUUSTUTKIMUKSEN SEITSEMÄN KULTAISTA SÄÄNTÖÄ

Seuraavassa taulukossa on kuvattuna toistettavuustutkimuksen suoritus- ja työvaiheet, joihin kultaiset säännöt perustuvat.

0. Tutkimus olosuhteet, kirjanpito, talous	
1. Ei liian monia testejä 2. Yksimielisyys testin suorittamisesta 3. Yksimielisyys testi-tuloksista 4. Yksimielisyys tutkimus hypoteesista 5. Kaksi tutkijaa 6. Kaikkiaan 10 potilasta	Harjoitteluvaihe
7. Kaksi tutkijaa 8. Kaikkiaan 20 potilasta 9. Sokkouttaminen 10. Potilaiden valitseminen	Yhtäpitävyysvaihe
11. Kaksi tutkijaa 12. Kaikkiaan 40 potilasta 13. Sokkouttaminen 14. Potilaiden valitseminen	Testivaihe
15. Aineisto	
16. Kappa	Statistiikka ja julkaiseminen

Kuvio 10: Toistettavuustutkimuksen suunnitelma

Toistettavuustutkimukset ovat helppoja suorittaa, eivätkä ne ole sidottuina suuriin instituutteihin, kuten korkeakouluihin. Yksityiset praktiikat tai paikat, joissa on vähintään kaksi manuaalisen lääketieteen harjoittajaa, ovat erittäin hyvin soveltuvia tämän tyyppisiin tutkimuksiin.

#### SÄÄNTÖ 1 LUO TOISTETTAVUUSTUTKIMUSTA VARTEN SELVÄ KIRJANPITO JA VASTUURAKENNE

Tutkimuksessa yhden henkilön tulee olla vastuussa koko tutkimusprosessista. Tämä henkilö on vastuussa tutkimuksen lokikirjasta. Lokikirjaan kirjoitetaan kaikki yksimielisyydet ja erimielisyydet ja sitä voidaan käyttää lähteenä ryhmäkeskusteluissa. Tämä henkilö on vastuussa protokollan päivitetystä formaatista.

## **SÄÄNTÖ 2 LUO HARJOITTELUVAIHE ENNEN VARSINAISEN TUTKIMUKSEN ALOITTAMISTA**

Harjoitteluvaiheessa keskustellaan ja päätetään tutkittavat testit ja niiden lukumäärä. Testien määrä riippuu tutkimuksen tarkoituksesta. Harjoitteluvaiheessa tutkijoiden täytyy olla yksimielisiä testien suorittamisen yksityiskohdista, joita he aikovat käyttää tutkimuksessa. 20 potilasta voidaan käyttää keskusteluun tarkkojen menetelmien ja asetelmien luomiseksi. Lopuksi heidän tulee olla yksimielisiä tarkasta testin suorittamisesta ja varmistaa, että jokainen tutkimuksen tekijä tuntee testin yleisesti hyväksytyn määritelmän. On suositeltavaa olla rajoittamatta testin suoritustapaa vain tutkimuspotilaisiin, vaan käyttää kyseistä käytäntöä kaikkiin klinikalla käyviin potilaisiin, harjoittaakseen taitoaan. tutkijoiden täytyy olla yksimielisiä siitä, miten tutkimuksessa käytettävien testien tulokset määritellään. tutkijoiden täytyy suorittaa testi samoille 20 potilaalle ja keskustella testin yksityiskohtaisista johtopäätöksistä. Lopuksi heidän täytyy olla yksimielisiä testin lopputuloksesta ja varmistaa, että kaikki tutkijat tuntevat testituloksen yleisesti hyväksytyn määritelmän. Jokaisen uuden päätöksenteon jälkeen lokikirja on päivitettävä. Jos tutkittavana on useampia testejä, määritellään positiivisten testien vähimmäismäärä lopulliseen positiiviseen testitulokseen. tutkijoiden tulee olla yksimielisiä tutkimuksessa käytettävien testien hypoteettisyydestä. Valittuaan minkä tahansa testin tutkimukseen, tutkijoiden täytyy tarkastella kohta kohdalta koko testin suorittaminen ja olla yksimielisiä siitä, mitä testillä testataan heidän päivittäisessä praktiikassaan.

## **SÄÄNTÖ 3 LUO YHTÄPITÄVYYSVAIHE ENNEN TOISTETTAVUUSTUTKIMUKSEN SUORITTAMISTA**

Tässä vaiheessa on tärkeää saavuttaa yhtäpitävyys, jonka on oltava  $> 0.80$ . Jos tämä on vähemmän kuin 0.80, tutkijoiden täytyy keskustella heidän yksimielisyydestä ja heidän tulee läpäistä harjoitteluvaihe uudelleen.

#### **SÄÄNTÖ 4 KÄYTÄ SOKKOUTTAMISTA TOISTETTAVUUSTUTKIMUKSESSA**

Protokollassa on tultava selville se, miten sokkouttaminen suoritetaan niin tutkijoilta kuin potilailtakin. Ellei kyseessä ole tekijöitä kuten kipu, sokkouttaminen varmistetaan useimmissa protokollissa, kun tietoa ei vaihdeta tutkijoiden ja potilaiden eikä tutkijoiden toistensa välillä. Käytä erillisiä lomakkeita tutkijoiden löydösten kirjaamiseksi.

#### **SÄÄNTÖ 5 MÄÄRITTELE PERUSJOUKKO, JOSTA POTILAAT VALITAAN**

Tutkimuksen lopputuloksessa on tultava esille, miten potilaiden valinta on tehty ja ettei siinä ole käytetty vilppiä.

#### **SÄÄNTÖ 6 MAINITSE PERUSJOUKKON MÄÄRITELMÄ, VALINTAMETODI, SOKKOUTTAMINEN, TESTIMENETELMÄN MÄÄRITELMÄ JA TESTITULOKSET MATERIAALEISSA JA METODEISSA TOISTETTAVUUSTUTKIMUSTA JULKAISTAESSA.**

#### **SÄÄNTÖ 7 NÄYTÄ 2 X 2 KONTINGENSITAUUKKO ESIINTYVYYDESTÄ JA YHTÄPITÄVYYDESTÄ LOPPUTULOKSISSA TOISTETTAVUUSTUTKIMUSTA JULKAISTAESSA**

## V. VALIDITEETTI

### V.1. Kultainen standardi

Kun on saavutettu hyvä toistettavuus testiprosessille, täytyy testin validiteetti varmistaa. Validiteetilla tarkoitetaan sitä, kuinka paljon testi mittaa juuri sitä, mitä sen ajatellaan mittaavan. Tarkemmin sanottuna validiteetti määräytyy sen mukaan, kuinka hyvin testi vastaa kultaista standardia. Tämä on merkittävä ongelma sekä koulu- että manuaalisessa lääketieteessä. Manuaalisessa lääketieteessä monien tyypillisten diagnostisten menetelmien, esim. loppujouaston testaaminen passiivisessa liiketestauksessa, oletetaan testaavan anatomisten rakenteiden liikkuvuutta. Suurimmassa osassa diagnostisista menetelmistä on tiedossa vain hypoteesi. Monille testeille manuaalisessa lääketieteessä standardit ovat vielä luomatta.

Kultaisia standardeja voidaan erotella kahdenlaisia. Ensinnäkin on kultainen standardi testimenetelmälle. Esimerkiksi, jos testi mittaa nivelen liikelaaajuutta tai loppujoustoja, täytyy liikelaaajuudelle saada myös kvantitatiivinen määre (asteina tai N/cm<sup>2</sup>). Kvantitatiivisen metodin arviointiin liittyy myös testaus-/uudelleen testausmenetelmä, jolla nähdään näyttääkö menetelmä saman tuloksen kahdella eri kerralla samalla potilaalla. Toisessa tapauksessa voidaan verrata molempia eli sekä kliinistä testimenetelmää että kvantitatiivista metodologiaa. Toinen kultainen standardi testille liittyy tämän testin hypoteesiin, jota yleisesti opetetaan (SI-mobiliteetti). Tämä on suurin ongelma diagnostiikassa sekä koulu- että manuaalisessa lääketieteessä.

Kultainen standardi kliiniselle testille voi olla radiologinen, kirurginen, ruumiinavauksellinen tai perustuen normaalista populaatiosta saatavaan aineistoon. Tähän mennessä kuvantamismenetelmät (RTG, CT- tai MRI-kuvaus) eivät ole vaikuttavia manuaalisessa lääketieteessä, koska myös suurella joukolla normaaleista (oireettomista) henkilöistä on näillä metodeilla havaittu olevan poikkeavuuksia. Joissakin erityisissä tilanteissa kuten SLUMP-testissä, jolla arvioidaan duurapussin ärsytystä esim. postoperatiivisesta lumbaaliadheesiosta, voi MRI gadolinium kontrasti toimia kultaisena standardina. Joillekin kipuprovokaatiotesteille kultainen standardi on paikallispuudutteen

vaikutus kyseiselle alueelle. Tällaisissa tapauksissa on ongelmana se, ettei voida koskaan olla varmoja paikallisuudutteen laajemmista vaikutuksista, tai jos kyseessä on säteilykipu, ei voida olla varmoja siitä, että kipu liittyy anatomiseen rakenteeseen jota haluamme tutkia. Manuaalisessa lääketieteessä on monia testejä nivelen liikkeen loppujouaston arvioimiseksi. Näihin tapauksiin on olemassa kaksi eri käytäntöä. Ensinnäkin voidaan luoda kvantitatiivinen metodi mittaamaan loppujoustoja. Tässä tapauksessa testi validioidaan kliinisesti. Toiseksi voidaan luoda kvantitatiivinen metodi arvioimaan nivelen liikelaajuutta. Tässä tapauksessa kliinisestä testistä arvioidaan sen mobiliteetti puoli ja siten määritellään testin lopullinen hypoteesi.

Edellä esitetty lista on kaukana täydellisestä, mutta esittää kuinka kultainen standardi luodaan. Joskus hyvin määritellyn standardin puuttuessa, asiantuntijoiden konsensus muiden testien perusteella voi toimia kriteeristandardina. Ongelmana tässä on se, että asiantuntijat ovat yksimielisiä vain testimenetelmästä, joka perustuu hypoteesiin ja täten todellinen validiteetti jää edelleen epäselväksi.

Manuaalisessa lääketieteessä on kuitenkin tärkeämpää ensin määrittellä testin toistettavuus kuin käyttää paljon aikaa määrittämään kultainen standardi.

## **V.2. Sensitiivisyys ja spesifisyys**

Toistettavuustutkimuksissa ei ole järkevää arvioida spesifisyyttä, jos ei kultaisia standardeja ole määritelty.

Sensitiivisyys- ja spesifisyystutkimuksissa 100 henkilöä on riittävä. Sama 100 hengen ryhmä testataan sekä kysymyksillä että myös kultaisella standardilla (katso 2 x 2 kontingenssitaulukko). Tapaukset a ja b on oikein ja c ja d ovat väärät positiiviset ja väärät negatiiviset. Hyvällä testillä tulee olla muutama väärä positiivinen ja muutama väärä negatiivinen tulos. Tunnusmerkin esiintyvyys esitetään kaavalla  $(a+c)/n$ . On tärkeää ymmärtää, että tunnusmerkin esiintyvyys voi vaihdella sekä eri instituuteissa ja maissa että ajasta riippuen.

Testin sensitiivisyys määritellään: testi erottaa oikein osan tapauksista, joilla on tunnusmerkki (a+c). Kaavassa:  $a/(a+c)$ .

Testin spesifisyys määritellään: Testi erottaa oikein osan tapauksista, joilla ei ole tunnusmerkkiä (b+d). Kaavassa:  $d/(b+d)$ .

Sekä sensitiivisyys että spesifisyys on tarpeen määrittäessä testin validiteettia ja on aina esitettävä tutkimuksen kirjallisessa osuudessa.

		<u>Kriteeristandardi</u>		
		Positiivinen	Negatiivinen	
<u>Testitulos</u>				
Positiivinen	a	b	a+b	
Negatiivinen	c	d	c+d	
	a+c	b+d	n=a+b+c+d	

### V.3. Positiivinen ja negatiivinen ennustearvo

Jotta statistiikkaluvut sensitiivisyydestä ja spesifisyydestä voidaan soveltaa päivittäiseen praktiikkaan, täytyy hoitajan tietää kyseisen potilaan tapaus, tietääkseen onko testi todellinen positiivinen vai vain väärä positiivinen. Tämä ilmentyy niin sanotusta ”testin positiivisesta ennustearvosta”. Yllä olevassa 2 x 2 kontingenssitaulukossa kaava testin positiiviselle ennustearvolle on  $a/(a+b)$ . Täytyy ymmärtää, että positiivinen ennustearvo riippuu tunnusmerkin esiintyvyydestä  $(a+c)/n$ . Oletetaan että on 1000 henkilöä, joilla sensitiivisyys 0.8 ja spesifisyys 0.7 ja tunnusmerkin esiintyvyys on 10 % (katso 2 x 2 kontingenssitaulukko).

Tämä tarkoittaa, että jos  $n=1000$ , siten  $a+c=0,10 \times 1000=100$ .



Tapauksessa, jossa annettu sensitiivisyys on  $(a/(a+c))$  on 0,8:

$$a/(a+c) = 0.8$$

$$\} \quad a/100 = 0.8 \quad a = 80$$

$$(a+c) = 100 \quad \} \quad (80+c) = 100 \quad c = 20$$

$$(a+c) = 100$$

$$\text{Jos } a+c = 0.10 \times 1000 = 100, n - a+c = b+d = 1000 - 100 = 900$$

Jos annettu spesifisyys on  $(d/(b+d))$  on 0.7:

$$d/(b+d) = 0.7$$

$$\} \quad d/900 = 0.7 \quad d = 630$$

$$(b+d) = 900 \quad \} \quad (630+b) = 900 \quad b = 270$$

$$(b+d) = 900$$

Testin positiivinen ennustearvo tässä tapauksessa on

$$a/(a+b) = 80/(80 + 270) = 0.22$$

Negatiivinen ennustearvo on samalla tavoin:

$$c/(c+d) = 270/(80 + 630) = 0.30$$

Tunnusmerkin esiintyvyyden ollessa suurempi  $(a+c)/n$ , testin positiivinen ennustearvo  $a/(a+b)$  kasvaa myös yhdessä sensitiivisyys- ja spesifisyyslukujen kanssa. Eli täten positiivinen ennustearvo heijastaa vain tunnusmerkin esiintyvyyttä, eikä itse testin ominaisuutta.

#### V.4. Todennäköisyysuhde (likelihood ratio)

On laskettava todennäköisyysuhde testin esiintymisennusteen (predictive power) arvioimiseksi, joka on erillinen tunnusmerkin esiintyvyydestä. Määritelmän mukaan todennäköisyysuhteen kaava on:

$$\text{Todennäköisyysuhde} = \frac{\text{sensitiivisyys}}{1-\text{spesifisyys}}$$

Testit, joiden todennäköisyysuhde on lähellä 1 tai <1 ovat täysin käyttökelvottomia päivittäiseen praktiikkaan.

Ensin muutama mainittava asia todennäköisyysuhteesta ja sen käytöstä laskettaessa diagnostista luottamussuhdetta (diagnostic confidence odds). Normaalisti olemme tottuneet ajattelemaan prosentteja yleisesti hyväksytyinä tai todellisina positiivisina lukuina. Todennäköisyysuhde ei toimi prosenteissa vaan todennäköisyysyksissä perustuen yleisesti hyväksytyyn ja diagnostiseen varmuuteen. Todennäköisyydet ovat tilanteen muutosten puolesta suhteessa kyseisen ajan tilanteen muutokseen. Esim. jos tilalla on 60 %:n esiintyvyys, testin esiintyvyyden todennäköisyydet ovat  $60 : 40 = 3 : 2$ . Nämä voidaan muuttaa edelleen desimaalilukuihin. Jos esiintyvyyden todennäköisyydet ovat  $3 : 2$ , todennäköisyys on  $3/(3+2)=0.6$ .

Matemaattisesti laskettuna diagnostinen luottamussuhde on laskettu kertomalla todennäköisyysuhde ja esiintyvyyden todennäköisyys.

[Esiintyvyyden mahdollisuudet] x [todennäköisyysuhde] = [diagnostiset luottamussuhteet]

Seuraavassa esimerkissä osoitetaan suuren todennäköisyysuhteen tärkeys suhteessa tilanteen esiintyvyyteen. Oletetaan, että tilalla on 60 % esiintyvyys praktiikalla. Toistettavuus- ja validiteettitutkimuksiin perustuen tiedetään, että sensitiivisyys on 0.8 ja spesifisyys 0.98. Perustuen edelliseen kaavaan,

$$\text{Todennäköisyysuhde} = \frac{\text{Sensitiivisyys}}{1 - \text{spesifisyys}}$$

todennäköisyysuhde on 40. Jos potilas tulee klinikalle valittaen vaivaa, jonka esiintyvyys on 40 %, tämän vaivan mahdollisuus on 60 %. Todennäköisyysuhde tälle vaivalle on siten 6: 4. Suhde diagnostiselle luotettavuudelle on  $6/4 \times 40 = 60$ . Diagnostiset luottamussuhteet=  $60 : 1$ . Diagnostinen luottamussuhde on  $60/60+1 = 0,98 = 98 \%$ .

Tämä tarkoittaa, että on parantanut diagnostista luotettavuutta 60 %:sta 98 %:iin. Tämä on hyvä testi.

Laskettaessa samalla esiintyvyydellä, mutta todennäköisyysuhteella 0.6, diagnostinen luottamussuhde on vain 0.47 tai 47 %. Tämä on pienempi kuin mahdollisuus potilaalla tähän vaivaan astuessaan sisään klinikalle. Tämä on huono testi. Julkaistut validiteettitutkimusten tulokset, joilla pyritään auttamaan päivittäistä ammatinharjoittajaa valitsemaan luotettavin testimenetelmä, ja joissa mainitaan ainoastaan yhtäpitävyys, sensitiivisyys- ja spesifisyysluvut ovat täysin käyttökelttomia. Jos tiedetään esiintyvyys tietylle tilalle, voidaan todennäköisyysuhteeseen perustuen laskea diagnostinen luottamussuhde.

## LIITE 1

## Diagnostiikan reliabiliteetti manuaalisessa lääketieteessä

		Tutkija B		Yhteensä:
		Kyllä	Ei	
Tutkija A	Kyllä	a	b	a+b
	Ei	c	d	c+d
Yhteensä:		a+c	b+d	n

Testattavien määrä = n

$$\text{Yhtäpitävyys} = p_0 = \frac{a+d}{n}$$

$$\text{Oletettu sattuma yhtäpitävyys} = p_c = \frac{a+b}{n} \times \frac{a+c}{n} + \frac{c+d}{n} \times \frac{b+d}{n}$$

$$\text{Kappa} = \frac{p_0 - p_c}{1 - p_c}$$

$$\text{Esiintyvyys } P = (a + [b + c] / 2) / n$$

Seuraavat pylväät voidaan määrittellä taulukkolaskentaohjelmalla, katso edellinen kuvio:

Vain data a, b, c, d tulee täyttää.

Pylväs A: data a (katso 2 x 2 kontingenssitaulukko)

Pylväs B: data b (katso 2 x 2 kontingenssitaulukko)

Pylväs C: data c (katso 2 x 2 kontingenssitaulukko)

Pylväs D: data d (katso 2 x 2 kontingenssitaulukko)

Pylväs E: data n                      Kaava = A1+B1+C1+D1

Pylväs F: data a+b                      Kaava = A1+B1

Pylväs G: data a+c                      Kaava = A1+C1

Pylväs H: data c+d                      Kaava = C1+D1

Pylväs I: data b+d                      Kaava = B1+D1

Pylväs E: data a+d                      Kaava = A1+D1

Pylväs K: Esiintyvyys                      Kaava = A1/E1+B1/2 x E1+C1/2 x E1

Pylväs L: Yhtäpitävyys  $p_0$                       Kaava = J1/E1

Pylväs M: (a+b)/n                      Kaava = F1/E1

Pylväs N: (a+c)/n                      Kaava = G1/E1

Pylväs O: (c+d)/n                      Kaava = H1/E1

Pylväs P: (b+d)/n                      Kaava = I1/E1

Pylväs Q: data pylväs  $M \times N$  Kaava  $=M1 \times N1$

Pylväs R: data pylväs  $O \times P$  Kaava  $=O1 \times P1$

Pylväs S: oletettu sattumayhtäpitävyys  $Pc$  Kaava  $=Q1+R1$

Pylväs T:  $Po - Pc$  Kaava  $=L1-S1$

Pylväs U:  $1 - Pc$  Kaava  $=1-S1$

Pylväs V: Kappa-arvo Kaava  $=T1/U1$

Tutkimustilan asettelu



Penkkien ja sermien asettelu

Testattavan valmistelu



Polvien asettelu kahden nyrkin leveydelle



Kämmenet polvien päälle



Käsivarret liukuvat polvia pitkin eteentaivutuksessa



Tutkijan käsien asettelu



Tutkijan sormien paikka ja käsien asento pystyssä



Tutkijan sormien ja käsien asento eteentaivutuksessa

Potilasnumero \_\_\_\_\_

TUTKIMUSLOMAKE

Sukupuoli: N / M

Ikä: \_\_\_\_\_

Jos sinulla esiintyy alaselkä/lantionalueen kipua, kuinka voimakas kipusi on tänään?

Aseta allaolevalle viivalle pystysuuntainen merkki, joka kuvaa tämänhetken kivun voimakkuutta.

Ei kipua \_\_\_\_\_ Erittäin voimakas kipu

Kuinka paljon kipu haittaa päivittäisiä toimiasi?

Ei haittaa \_\_\_\_\_ Haittaa erittäin paljon

1 Y/N	2 Y/N	3 Y/N	4 Y/N	5 Y/N
----------	----------	----------	----------	----------

6 Y/N	7 Y/N	8 Y/N	9 Y/N	10 Y/N
----------	----------	----------	----------	-----------

11 Y/N	12 Y/N	13 Y/N	14 Y/N	15 Y/N
-----------	-----------	-----------	-----------	-----------

16 Y/N	17 Y/N	18 Y/N	19 Y/N	20 Y/N
-----------	-----------	-----------	-----------	-----------

21 Y/N	22 Y/N	23 Y/N	24 Y/N	25 Y/N
-----------	-----------	-----------	-----------	-----------

26 Y/N	27 Y/N	28 Y/N	29 Y/N	30 Y/N
-----------	-----------	-----------	-----------	-----------

31 Y/N	32 Y/N	33 Y/N	34 Y/N	35 Y/N
-----------	-----------	-----------	-----------	-----------

36 Y/N	37 Y/N	38 Y/N	39 Y/N	40 Y/N
-----------	-----------	-----------	-----------	-----------