



TOIMIA

# Opas toimintakyvyn mittarin arviointiin TOIMIA-verkostossa (1.0)

1.6.2014

Kirjoittajat:

Heli Valkeinen, Erikoistutkija, TtT, Terveyden ja hyvinvoinninlaitos  
Heidi Anttila, Erikoistutkija, TtT, Terveyden ja hyvinvoinninlaitos  
Jaana Paltamaa, Erikoissuunnittelija, TtT, Jyväskylän Ammattikorkeakoulu



TOIMIA

## SISÄLLYS

1 OHJEISTUKSEN TAUSTA .....	3
2 JOHDATUS PSYKOMETRIIKAN ARVIOIMISEEN .....	4
3 VALIDITEETIN ELI PÄTEVYYDEN OSA-ALUEET .....	7
3.1 Ilmivaliditeetti (face validity) .....	8
3.2 Sisältövaliditeetti (content validity) .....	8
3.3 Kriteerivaliditeetti (criterion validity) .....	9
3.3.1 Samanaikainen validiteetti ( <i>concurrent validity</i> ) .....	9
3.3.2 Ennustevaliditeetti ( <i>predictive validity</i> ) .....	10
3.4 Rakennevaliditeetti (construct validity) .....	10
3.4.1 Mittarin rakenteen validiteetti ( <i>structural validity</i> ) .....	11
3.4.2 Hypoteesien testaus .....	12
3.4.3 Kulttuurien välinen validiteetti ( <i>cross-cultural validity</i> ) .....	13
4 RELIABILITEETIN OSA-ALUEET .....	15
4.1 Toistomittausten toistettavuus ja mittausvirhe .....	16
4.1.1 Relatiivisen toistettavuuden tilastolliset menetelmät/tunnusluvut .....	17
4.1.2 Absoluuttisen toistettavuuden tilastolliset menetelmät/tunnusluvut .....	18
4.2 Sisäinen yhtenevyys .....	20
5 MUUTOSHERKKYYS ( <i>responsiveness</i> ) .....	21
5.1 Suositeltavia tilastollisia menetelmiä muutosherkkyden mittaamiseen .....	22
5.2 Ei-suositeltavia tilastollisia menetelmiä muutosherkkyden mittaamiseen, mutta suositeltavia muutoksen määrän tulkintaan .....	23
6 TULOSTEN TULKITTAVUUS ( <i>interpretability</i> ) .....	24
6.1 Poikkileikkaustulosten tulkinta .....	25
6.2 Muutostulosten tulkinta .....	25
6.2.1 Pienin havaittava muutos ( <i>Smallest/Minimal Detectable Change, SDC/MDC</i> ) .....	25
6.2.2 Pienin merkittävä muutos ( <i>Minimal Important Change, MIC; Minimal Clinically Important Difference, MCID</i> ) .....	26
7 KÄYTTÖKELPOISUUS .....	28
8 LÄHTEET .....	29
LIITTEET .....	32

## 1 OHJEISTUKSEN TAUSTA

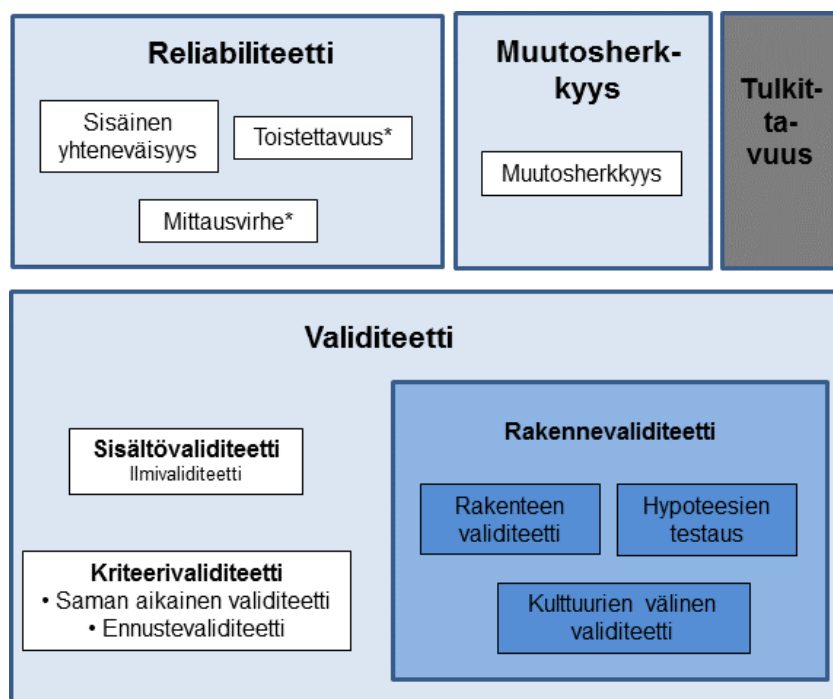
Opas toimintakyvyn mittarin arviointiin TOIMIA-verkostossa on tehty helpottamaan toimintakyvyn mittareiden arviointityötä. Perustiedot mittareista kerätään ”Mittarin perustiedot” -lomakkeelle (liite 1), mutta tämä opas on tarkoitettu erityisesti ”Mittarin psykometriset tiedot” -lomakkeen (liite 2) täyttämistä ajatellen. Mittarin psykometrisilla ominaisuuksilla tarkoitetaan tietoja mittarin validiteetista (pätevyys), reliabiliteetista (toistettavuus tai luotettavuus) ja muutosherkyydestä. Tässä oppaassa käytetään termiä ”psykometriset ominaisuudet” tai ”psykometriikka”, kun tarkoitetaan näitä edellä mainittuja tietoja. Lisäksi mittarin arvioinnissa tarkastellaan mittarin käyttökelpoisuutta. TOIMIA-verkostossa mittarien arviointi tehdään resurssien säästämiseksi usein vain tiettyihin käyttötarkoituksiin. Siksi mittarin **psykometrisiä ominaisuuksia ja käyttökelpoisuutta ei ole tarkoitus arvioida kaikissa mahdollisissa yhteyksissä tai tilanteissa, vaan nämä tiedot tulee koota ja arvioida siitä käyttötarkoituksesta käsin, joka on tarkastelun kohteena. Mittariarvion tekemistä ohjaa siten aina se, millaisessa käyttötarkoituksessa mittaria on tarkoitus käyttää.**

Kappaleessa 2 johdatellaan lyhyesti psykometriikan teorioihin ja yleisiin peruskäsitteisiin. Kappaleessa 3 esitellään validiteetin ja kappaleessa 4 reliabiliteetin osa-alueet. Kappaleessa 5 käydään läpi muutosherkyyden mittaaminen ja kappaleessa 6 tulosten tulkinta. Tulosten tulkinnassa painopiste on muutostulosten tulkinnassa, jotka liittyvät läheisesti mittarin muutosherkyyden analysointiin. Kappaleessa 7 kuvataan lyhyesti kuinka mittarin käyttökelpoisuutta voidaan arvioida.

Psykometriikatietojen poimiminen tutkimusartikkeleista mittarin arviointia varten saattaa olla aikaa vievää salapoliisityötä. Valitettavan usein on niin, että tutkimusartikkelit ovat hieman vaikealukuisia ja tärkeiden tietojen löytäminen saattaa olla joskus haastavaa. Oppaan loppuun on sen vuoksi tehty tarkistuslistoja, jotka voivat olla avuksi mittariarvioita tekeville. Liitteinä (liitteet 3–6) olevat listat eivät kuitenkaan ole kattavia, vaan suuntaa antavia helpottamaan eri asioiden etsimistä ja muistamista.

## 2 JOHDATUS PSYKOMETRIIKAN ARVIOIMISEEN

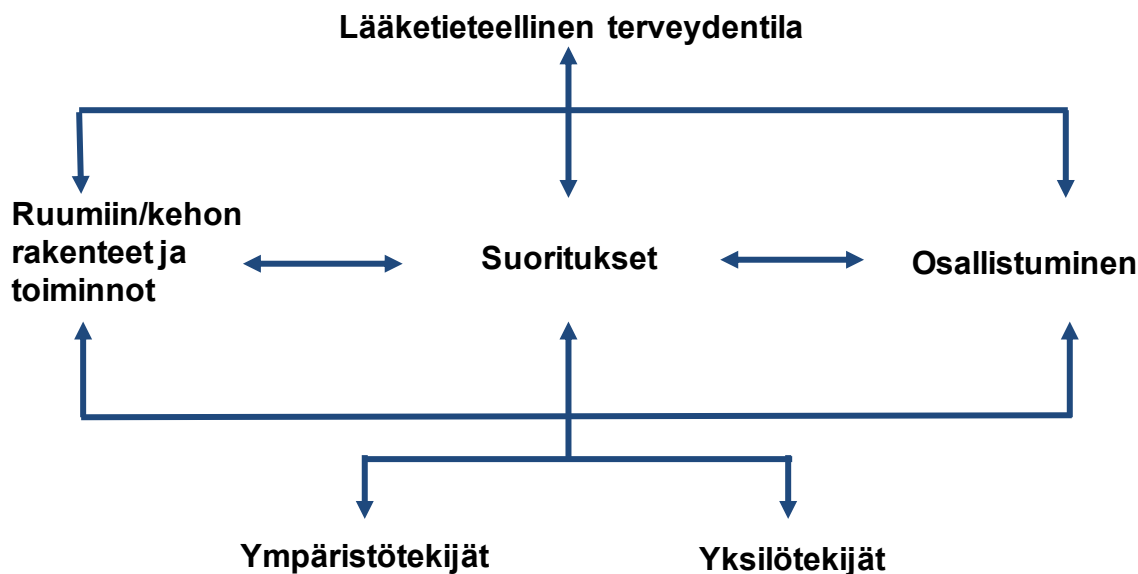
Mittareiden psykometrisilla (*psychometrics*) ominaisuuksilla tarkoitetaan tietoja mittarin pätevyydestä, toistettavuudesta ja muutosherkkyydestä. Lisäksi samassa yhteydessä arvioidaan myös käyttökelpoisuutta. Psykometriikan ”juuret” ovat psykologisessa tutkimuksessa, jossa on kehitetty erilaisia mittaamisen teorioita, esimerkiksi *classical test theory*, *item response theory* (Streiner ja Norman 2008). Eräissä yhteyksissä käytetään myös termiä klinimetria (*clinimetrics*), joka tarkoittaa käytännössä samaa kuin psykometriikka. TOIMIA:ssa on sovittu käytettäväksi termiä ”psykometria” puhuttaessa mittareiden pätevyys-, toistettavuus- ja muutosherkkyysominaisuuksista. Tämä psykometrinen ominaisuuksien jaottelu perustuu kansainvälisen COSMIN-tutkijaryhmän laatimaan jaotteluun ([www.cosmin.nl](http://www.cosmin.nl)) (kuva 1) (de Vet ym. 2011, 4).



**KUVA 1.** Psykometrinen ominaisuuksien jaottelu COSMIN mukaan. \*Testi–uusintatesti, mittaajien välinen ja mittaajan sisäinen. (Mukaeltu Mokkink ym. 2010a)

**Diagnostiset- vs. seuranta- vs. ennustemittarit:** Eri mittarit on kehitetty mittaamaan tiettyä ihmisen terveyteen liittyvää konstruktiota (ominaisuutta). Näitä konstruktiota on tarpeen mitata erilaisissa yhteyksissä, kuten eroteltaessa sairauksia tai sairauden asteita (diagnostiset mittarit), seurattaessa hoidon tai kuntoutuksen vaikutuksia (seurantamittarit) tai haluttaessa ennustaa esimerkiksi jonkin sairauden tai toiminnanvajauden ilmaantumisesta (ennustemittarit). Jotkin mittarit voivat toimia kaikissa edellä mainituissa tilanteissa. Huomioitavaa on, että nykyisin käytetään usein ennustemalleja tai ennustesääntöjä (*prediction rules*) määrittelemään muuttujia, jotka parhaiten kuvaavat sairauden kulkua. Näitä ei tulisi kutsua mittareiksi, sillä ennustemalleissa ja -säännöissä voi olla mukana useita eri konstruktiota ja muuttujia (Steyerberg 2009).

**Mitattava konstruktio vs. mittari:** Mittarin suunniteltu käyttötarkoitus vaikuttaa oleellisesti mittarin kehittämiseen. Siksi mitattavaksi valittu konstruktio on yleensä ”purettava” käsitteisiin mittarin kehittämissivaiheessa. Samalla voidaan tarkastella kattaako mittari kaikki oleelliset mitattavat asiat kyseessä olevasta konstruktiosta. Toimintakykyyn ja toimintarajoitteisiin liittyviä käsitteitä löytyy kattavasti kansainvälisestä ICF-luokituksesta (International Classification of Functional, Disability and Health). Kuvassa 2 on kuvattu ICF-luokituksen osa-alueet ja niiden väliset vuorovaikutussuhteet (WHO 2004, 18).



**KUVA 2.** ICF-luokituksen osa-alueiden vuorovaikutussuhteet (mukaeltu WHO 2004, 18).

Ruumiin/kehon rakenteiden ja toimintojen alueella (kuvan 2 vasen osa) tarvitaan diagnostisia mittareita sairauksien määrittämiseksi, kun suoritusten ja osallistumisen alueella (kuvan 2 oikea osa) tarvitaan seurantamittareita sairauden tai hoidon vaikutusten selvittämiseksi. Ammattilaisten käyttämiä mittareita pidetään usein ns. objektiivisina mittareina ja potilaiden/asiakkaiden itse täyttämiä kyselyitä subjektiivisina mittareina (ns. itsearviointimittarit). Monien objektiivisten mittareiden tulosten tulkinnassa tarvitaan kuitenkin ammattilaisten taitoja ja tietoja. Siten myös objektiivisten mittareiden käyttöön liittyy subjektiivisia piirteitä. Ruumiin/kehon rakenteiden ja toimintojen alueella voidaan käyttää enemmän yksiuotteisia mittareita, mutta suoritusten ja osallistumisen arviointiin ja mittaamiseen tarvitaan huomattavasti moniuotteisempia mittareita. Ruumiin/kehon rakenteiden ja toimintojen alueella mitattavat ominaisuudet ovat tavallisesti havaittavia, mutta osallistumisen alueella mitattavat ominaisuudet ovat usein abstrakteja ja moniuotteisia, esimerkiksi koettu terveys.

**Yksiuotteinen vs. moniuotteinen mittari:** Mittaamisen kohteena oleva konstruktio voi olla esimerkiksi yksittäinen fysiologinen toiminta, kuten käden puristusvoima. Tällöin puristusvoimamittari on yksiuotteinen mittari, koska sillä ei voida mitata mitään muuta ominaisuutta kuin käden puristusvoimaa. Samoin esimerkiksi muistia mittaava testi tai kävelynopeuden mittaaminen on yksiuotteinen mittari. Jos halutaan kuitenkin arvioida alaraajojen toimintaa, niin tällaisen mittarin tulisi sisältää laajasti alaraajojen toimintaa käsittäviä

testejä, kuten kävelynopeus, porraskoulu, lihasvoima, nivelliikkuvuus jne. Tällöin kyseessä olisi moniulotteinen mittari, koska kaikki eri osa-alueet kuvaisivat jollain tavalla alaraajojen toimintaa.

Vaikeasti havaittavia ja mitattavissa olevia asioita, kuten koettu terveys, arvioidaan usein erilaisilla kyselylomakkeilla. Tällöin useilla kysymyksillä pyritään saamaan tietoa yhdestä mitattavasta konstruktiosta (yksiulotteinen mittari). Vaikeasti havaittavissa olevat abstraktit asiat ovat usein niin moniulotteisia, että mittarilla on kartoitettava useita eri osa-alueita, joiden ajatellaan jollain tavoin kuvaavan itse mitattavaa konstruktiota (moniulotteinen mittari). Esimerkiksi elämänlaatua on arvioitava kysymällä kysymyksiä niin terveydestä, ihmissuhteista, fyysisestä liikkumiskyvystä kuin muistamisesta ja omasta itsestä huolehtimisestakin.

**Geneerinen vs. spesifi mittari:** Geneeriset mittarit mittaavat asioita yleisesti ja kokonaisvaltaisesti, esimerkiksi geneeriset elämänlaadun mittarit. Spesifit mittarit ovat taas kehitetty tiettyä sairautta tai tilannetta varten ja näitä ovat esimerkiksi sairauskohtaiset mittarit.

### 3 VALIDITEETIN ELI PÄTEVYYDEN OSA-ALUEET

Pätevyydellä eli validiteetilla (*validity*) tarkoitetaan sitä, mittaako mittari juuri sitä tutkittavan ilmiön konstruktiota, mitä sen on tarkoitus mitata (de Vet ym. 2011, 203). Validiteettia tutkittaessa mittauksia tehdään yleensä poikkileikkausasetelmassa (yksi mittauskerta), mutta joissakin tapauksissa, esimerkiksi kriteerivaliditeettia tutkittaessa, voidaan tarvita pitkittäisasetelmaa. Kun mittaukset tehdään kahdesti ja mitattavassa ominaisuudessa oletetaan tapahtuvan tällä aikavälillä muutosta, on kyseessä muutosherkkyiden tarkastelu. Muutosherkkyys on yksi validiteetin itsenäinen osa-alue, jota tarkastellaan myöhemmin (ks. kappale 5 Muutosherkkyys) (de Vet ym. 2011, 4).

Validiteetin selvittäminen alkaa mittarin sisällön validoinnilla. Ensin tulee tarkastella yleisesti mittarin **ilmivaliditeettia** ja sen jälkeen tarkemmin **sisältövaliditeettia**, esimerkiksi moniulotteisen mittarin kysymysten relevanttiutta ja riittävää laajuutta. Muiden mittareiden osalta tulee tarkastella yleisesti mittarin erityisominaisuuksia, esimerkiksi MRI-kuvantamisessa tulisi selvittää miten hyvin kuvissa näkyvät erilaiset kudokset.

**Kriteerivalidointia käytetään** silloin, kun mittarille on olemassa kultainen standardi (*golden standard*). Tällöin määritellään mittarin yhtenevyys kultaiseen standardiin ja tehdään johtopäätökset, jotka on saatu verrattaessa saatuja tuloksia etukäteen asetettuihin kriteereihin. Tarkoituksena on selvittää onko saatu yhtenevyys riittävä siinä käyttötarkoituksessa, jossa mittaria on tutkittu. Jos mittarille ei ole olemassa kultaista standardia, näyttöä tulee etsiä **rakennevalidoinnilla**. Tällöin edellytyksenä on, että mittarin konstruktio on selkeästi määritelty teorian tai mallin pohjalta. Rakennevalidointia voidaan tehdä yhdistelemällä useampia validointitutkimuksia, joissa kerätään **näyttöä tarkoista hypoteeseista**. Harkittujen ja tarkkojen hypoteesin vahvistuminen validointitutkimuksessa tuottaa enemmän näyttöä kuin heikosti muotoillut hypoteesit. Kun testataan suurta määrää haastavia hypoteeseja, rakennevaliditeettitutkimuksesta tulee vahva työkalu mittarin validoinnissa.

Rakennevalidoinnin erityistapauksia ovat **rakenteen validointi** ja **kulttuurien välinen validointi**. Rakenteen validointia käytetään usein moniulotteisten mittareiden validoinnissa. Kun mittarin rakenne tunnetaan, sen taustalla oleva teoria tiedetään ja halutaan testata onko mittari sen mukainen, käytetään konfirmatorista faktorianalyysiä. Eksploratiivista faktorianalyysiä voidaan käyttää, kun taustalla oleva teoria ei ole tunnettu/tiedossa. Molempien faktorianalyysien ja osio-vaste-mallien (*item response models*; IRT-mallit) avulla voidaan tehdä johtopäätöksiä siitä, toimivatko osiot kuten on suunniteltu. Kulttuurien välistä validointia tarvitaan, kun mittareita käännetään toiselle kielelle.

Validointi on siis jatkuva prosessi. Mittarin validiteettitietoa kerääntyy asteittain ja mittarin validiteetti riippuu siitä millaista tietoa on saatu kerättyä ja miten tuloksia voidaan hyödyntää. Tämä on harvoin täydellinen prosessi. Mittarin validiteettia on aina tutkittava suunnitellussa käyttötarkoituksessa ja niillä tutkittavilla, joilla sitä on tarkoitus käyttää. (Fitzpatrick ym. 1998, Finch ym. 2002, Streiner ja Norman 2008, de Vet ym. 2011).

Validointiin liittyviä keskeisiä asioita ovat (Strauss & Smith, 2009; de Vet ym. 2011):

- Tieto mitattavasta konstruktiosta
- Konstruktion monimuotoisuus (*complexity*)
- Tilannesidonnaisuus
- Validoidaan tuloksia, ei mittausvälineitä
- Tarkkojen hypoteesien muodostaminen

## 3.1 Ilmivaliditeetti (*face validity*)

Ilmivaliditeetti on tärkeä ensisilmäys pätevyyden arvioinnissa. Jos mittari ”ei näytä” sopivalta aiottuun tarkoitukseen, ei sitä kannata ryhtyä validoimaan. Alkuarviota suoritettaessa voidaan esittää kysymys: ”näyttääkö mittari mittaavan sitä, mitä sen on tarkoitus mitata?”. Esimerkki alkuarviosta voisi olla seuraava: jos mittari on alaraajojen toimintakykymittari, niin mittaria tarkasteltaessa odottaa osioiden sisältävän kävelyä, juoksemista, seisomista, porraskävelyä jne. eli alaraajojen toimintakykyä kuvaavia asioita. Jos mittarissa olisi myös yläraajojen toimintaa arvioiva osio tai elämänlaatukysymyksiä, mittarin ilmivaliditeetti olisi kyseenalainen (Finch ym. 2002, Streiner ja Norman 2008, de Vet ym. 2011). Ilmivaliditeettia arvioitaessa ei käytetä tilastollisia menetelmiä.

TOIMIA:n mittariarviointeja tehtäessä ilmivaliditeettia tulee miettiä erityisesti siinä tapauksessa, kun mittaria arvioidaan johonkin muuhun käyttötarkoitukseen kuin alkuperäiseen tarkoitukseen. Esimerkiksi jos mittaria aiotaan käyttää jollain toisella potilasryhmällä tai moniulotteista mittaria aiotaan käyttää eri konstruktion arvioimiseen.

## 3.2 Sisältövaliditeetti (*content validity*)

Sisältövaliditeetilla tarkoitetaan sitä, missä määrin mittarin sisältö on tarkka vastine mitattavasta konstruktiosta (Streiner ja Norman 2008, de Vet ym. 2011). Jos esimerkiksi tavoitteena on mitata alaraajojen toimintakykyä, mittarin osioiden tulisi mitata niitä kaikkia erilaisia toimintoja, joissa alaraajoja käytetään. Jos siis mittarissa on vain 2 osiota: a) asiakkaan 2 min kävelymatka ja b) kyky nousta ylös portaita, se ei ole kokonainen alaraajojen toimintakyvyn mittari, vaan vain osa sitä. Tällöin mittarin sisältövaliditeetti ei ole kunnossa (Finch ym. 2002).

Mittarin sisältövaliditeettia arvioitaessa asiantuntijoiden on oltava selvillä mitattavasta konstruktiosta, mittarin käyttötarkoituksesta ja mittarin rakenteesta ja sen kehittämisprosessista (Streiner ja Norman 2008, de Vet ym. 2011). Mittareiden kehittämisprosessissa mittarin varsinaiset kehittäjät saattavat ”sokeutua” omalle asialleen ja sen vuoksi mittareiden kehittämisessä tulisi käyttää riippumatonta asiantuntijapaneelia, joka on perehtynyt kyseiseen alueeseen syvällisesti ja voi siten arvioida mitattaako mittari koko konstruktiota. Esimerkiksi radiologian asiantuntijat ovat päteviä arvioimaan radiologiaan kehitettäviä mittareita, mutta asiakkaan/potilaan itse täytettävien kyselyiden (ns. itsearviointimittarit eli *patient-reported outcomes*; PRO-mittarit) arvioinnissa parhaita asiantuntijoita ovat potilaat/asiakkaat itse. Sisältövaliditeettia selvitetessä on arvioitava myös sisällön laajuus ja asiaan kuuluvuus suhteessa



konstruktion ja kohderyhmään. Tässä arvioinnin apuna voi käyttää esimerkiksi ICF-luokitusta ja verrata mittarin sisältöä ja kattavuutta siihen (de Vet ym. 2011). Esimerkkinä sisällön validoinnista voidaan mainita WHODAS 2.0., joka on moniulotteinen, useita asteikkoja sisältävä mittari (Üstün ym. 2010). Se on kehitetty aiempien toimintakykymittareiden ja ICF-luokituksen suoritusten ja osallistumisen osa-alueiden pohjalta arvioimaan terveydentilasta johtuvia vaikeuksia, joita ihminen kokee päivittäisessä elämässään.

### 3.3 Kriteerivaliditeetti (*criterion validity*)

Kriteerivaliditeetti jaetaan samanaikaiseen - ja ennustevaliditeettiin ja se tarkoittaa, missä määrin mittarilla saadut tulokset ovat yhtäpitäviä kultaisen standardin kanssa (Finch ym. 2002, de Vet ym. 2011). Kultainen standardi on ulkoinen kriteeri, joka viittaa kunkin aihealueen parhaaseen mahdolliseen käytettävissä olevaan arviointimenetelmään ja joka antaa tarkimman ja luotettavimman tuloksen. Kriteerivaliditeetin mittaamiseksi käytettävän tilastollisen menetelmän valintaan vaikuttaa kultaisen standardin ja tarkasteltavan mittarin muuttujien ominaisuudet (dikotominen, ordinaalinen, jatkuva) sekä mittareiden mittayksiköt (taulukko 1).

**Taulukko 1.** Käytettävät tilastolliset menetelmät kriteerivaliditeetin määrittämiseksi (mukaillen de Vet ym. 2011, 163).

Mitta-asteikko		Sama mittayksikkö	Tilastollinen menetelmä
Kultainen standardi	Tutkittava mittari		
Dikotominen	Dikotominen	Kyllä	Herkkyys ja tarkkuus
	Ordinaalinen	Ei merkitystä	ROC
	Jatkuva	Ei merkitystä	ROC, logistinen regressioanalyysi
Ordinaalinen	Ordinaalinen	Kyllä Ei	Painotettu kappan Spearmanin r / muu korrelaatio
	Jatkuva	Ei merkitystä	ROC; Spearmanin r; regressioanalyysi
Jatkuva	Jatkuva	Kyllä Ei	Bland-Altmanin piirros; ICC Spearmanin r; Pearsonin r

ROC = Receiver Operating Characteristics curve; ICC = Intraclass Correlation Coefficient.

#### 3.3.1 Samanaikainen validiteetti (*concurrent validity*)

Samanaikaisella validiteetilla tarkoitetaan tilannetta, jossa tutkittavaa mittaria verrataan lähes samanaikaisesti kultaiseen standardiin. Esimerkiksi manuaalisesti toteutetun lihasvoimatestin tulosta verrataan dynamometrillä mitattuun tulokseen, jota voidaan tässä tapauksessa pitää kultaisena standardina. Samanaikaista validiteettia tarvitaan validoitaessa erottelevia mittareita ja

seurantamittareita. Samanaikaista validiteettia arvioitaessa käytetään eri tilastomenetelmiä sen mukaan millainen mitta-asteikko ja mittayksikkö ovat tutkittavassa mittarissa ja kultaaisessa standardissa (taulukko 1). Koska kaikkiin mittauksiin sisältyy jonkin verran mittausvirhettä, on mahdollista, että samaa ilmiötä mittaavien mittareiden välinen korrelaatio jää keskinkertaiseksi. Kovin alhainen korrelaatio voi tarkoittaa, että jommankumman vertailtavan mittarin reliabiliteetti on todennäköisesti liian heikko tai että mittarit mittaavat eri ilmiötä (Streiner ja Norman 2008).

### 3.3.2 Ennustevaliditeetti (*predictive validity*)

Ennustevaliditeetilla tarkoitetaan mittarin kykyä ennustaa jokin tuleva tapahtuma. Ennustavia mittareita käytetään laajasti erityisesti diagnostiikassa, kun halutaan arvioida/ennustaa tuleva tapahtuma. Ennustevaliditeettia arvioidaan samoin menetelmin kuin samanaikaista validiteettia, mutta kriteeritapahtuma on tiedossa vasta muutaman päivän tai vuoden päästä. Dikotomisten tulosten tarkastelussa käytetään ristiintaulukointia ja jatkuvia muuttujia arvioidaan vertaamalla niiden välistä korrelaatiota (taulukko 1). Ennustevaliditeettia tutkittaessa on vältettävä aiemman mittauksen tietojen sekoittumista kriteeritapahtuman tulosten arviointiin (*criterion contamination*), jossa mittausten välille rakentuu keinotekoinen korrelaatio. Ennustevaliditeetti olisi aina testattava toisella otoksella kuin millä mittari on kehitetty (Streiner ja Norman 2008).

Esimerkiksi Bergin tasapainoasteikko ennustaa kaatumista kuuden viikon aikana. Kriteeritapahtumana tässä esimerkissä on vastaus kysymykseen: ”kaatuiko asiakas seuraavan kuuden viikon aikana”. Tällaisessa tutkimuksessa on siis mitattu ensin asiakkaan tasapaino käyttäen Bergin tasapainoasteikkoa. Tämän jälkeen on selvitetty kyselyllä, puhelinhaastattelulla tai asiakirjoista kuuden viikon kuluttua, onko asiakas kaatunut vai ei. Verrattaessa varsinaisia tasapainotuloksia ja myöhemmin raportoituja kaatumistietoja saadaan tietoa mittarin kyvystä ennustaa kaatuminen (Finch ym. 2002).

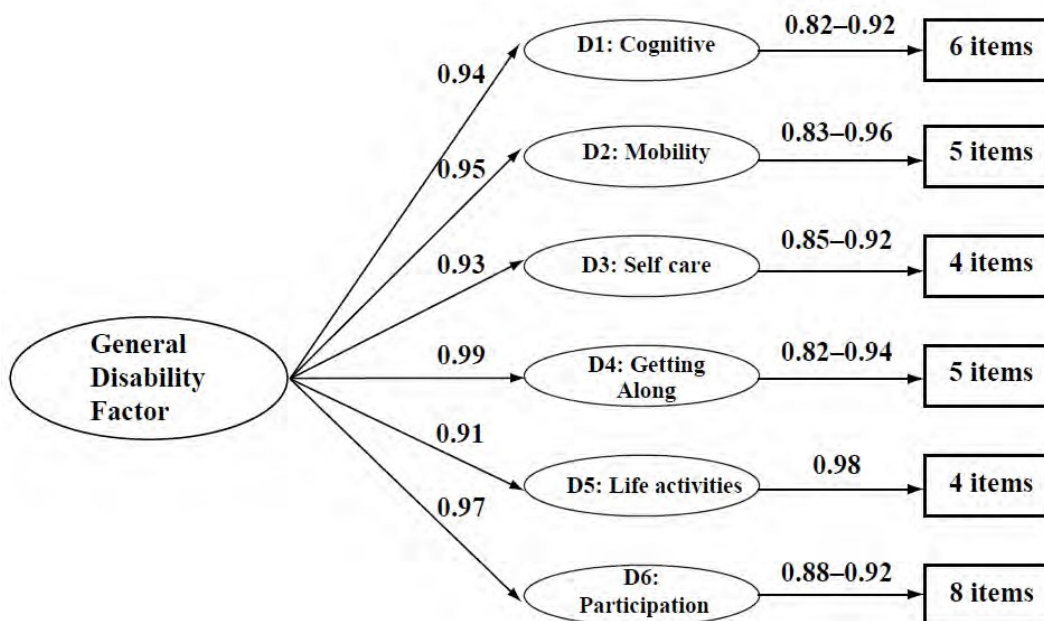
### 3.4 Rakennevaliditeetti (*construct validity*)

Rakennevaliditeetti tarkoittaa sitä, missä määrin mittarin tulokset ovat yhtäpitäviä/johdonmukaisia mitattavan konstruktion kanssa. Rakennevalidointia tarvitaan erityisesti silloin, kun konstruktiolle (esimerkiksi kipu tai elämänlaatu) ei ole olemassa selkeitä kriteereitä tai kultaista standardia. Tällöin tarvitaan teoria tai malli kiinnostuksen kohteena olevasta konstruktiosta ja tutkitaan missä määrin mittari on yhtenevä tämän teorian tai mallin konstruktion kanssa. Jotta yhtenevyyttä voidaan testata, on ensin määriteltävä tarkat hypoteesit oletetusta yhtenevyydestä (ks. 3.4.2) (Mokkink ym. 2010b, de Vet ym. 2011). Rakennevalidointia voi tehdä kolmesta näkökulmasta: rakenteen validiteetti, hypoteesien testaus (konvergentti validiteetti, erotteleva validiteetti ja kyky erotella ryhmiä) ja kulttuurien välinen validiteetti.

### 3.4.1 Mittarin rakenteen validiteetti (*structural validity*)

Rakenteen validiteetti tarkoittaa sitä, missä määrin mittarin tulokset ovat tarkka vastine mitattavan konstruktion ulottuvuuksista (de Vet ym. 2011). Rakenteen validiteettia arvioidaan erityisesti silloin, kun mittari on moniulotteinen. Tavallisimmin rakenteen validiteettia tutkitaan konfirmatorisella faktorianalyysillä tai IRT-malleilla, koska mitattavan konstruktion rakenne on jo etukäteen tiedossa perustuen joko aiempiin tutkimuksiin ja/tai hyvin perusteltuun kehittämisprosessiin. Faktorianalyysin avulla voidaan tehdä johtopäätöksiä toimivatko mittarin osiot odotetusti. Jos mittarin ulottuvuudet eivät kuitenkaan ole tiedossa etukäteen, rakenteen validiteettia voidaan tutkia eksploratiivisella faktorianalyysillä.

Esimerkiksi WHODAS 2.0 -mittarin rakennetta on validoitu faktorianalyysillä (kuva 3) (Üstün ym. 2010). Se osoitti, että mittari mittaa yleisesti toimintakyvyn ongelmia (*general disability factor*) ja että sen osiot jakautuivat kuuteen aiheeseen: ymmärtäminen ja kommunikointi; liikkuminen; itsestä huolehtiminen; ihmissuhteet; arjen toiminnot (kotitaloustehtävät sekä työ/opiskelu); ja yhteisöllinen osallistuminen. Tulokset saadaan siis sekä yksiulotteisesti kaikkien osioiden summapisteinä että moniulotteisesti osa-alueittain, jonka avulla voidaan tarkastella kaikkien asteikkojen muodostamaa profiilia.



KUVA 3. WHODAS 2.0 –mittarin faktorirakenne (Üstün ym. 2010).

### 3.4.2 Hypoteesien testaus

Mittareiden validoinnissa yksi keskeinen asia on selvittää kuinka hyvin mittarin tulokset vastaavat toisella mittarilla mitattuja tuloksia samasta konstruktiosta (*convergent validity*) tai miten mittari eroaa muita konstruktiota mittaavista mittareista (*discriminant validity*). Lisäksi on oleellista tietää pystyykö mittari erottelemaan esim. sairauden asteeltaan erilaisia ryhmiä toisistaan (*known group validity*). Näitä kaikkia kolmea eri validiteetin osa-aluetta tutkitaan hypoteesien testauksella. Ennen varsinaisia mittauksia tutkijan on siis laadittava hypoteesit siitä, miten hän olettaa tutkittavan mittarin vastaavan ja/tai eroavan muista mittareista ja miten hyvin hän olettaa tutkittavan mittarin kykenevän erottelemaan ryhmiä.

Arvioitaessa hypoteesien testaukseen liittyviä validiteetin osa-alueita, tieteellisistä artikkeleista on etsittävä kuvaus validoinnin kohteena olevan mittarin rakenteesta ja sen teoriataustasta (jos tämä on kuvattu artikkelissa). Artikkelissa tulisi olla tarkat tutkimushypoteesit siitä, miten mittarin oletetaan vastaavan tai ei oleteta vastaavan mitattavaa tai jotain muuta konstruktiota; miten tulosten oletetaan eroavan eri ryhmien välillä; ja mitkä oletetaan olevan vastaavuuksien ja/tai erojen suunta ja suuruus. Jotta nämä vertailut voidaan tehdä hyvin, on artikkelissa oltava kuvattuna vertailussa käytetyt mittarit, niiden konstruktiot ja mittausominaisuudet sekä eroteltavat alaryhmät. Artikkelissa on kuvattava tietojen keräys ja tulosten analysointi sekä tuloksissa on arvioitava tulosten yhtäpitävyys artikkelin alussa esitettyihin hypoteeseihin nähden. Lisäksi tutkimuksessa tulisi olla tietoa siitä, missä määrin havaittuja tuloksia saattaa selittää jokin muu teoria tai vaihtoehdot (de Vet ym. 2011).

Hypoteesien testauksessa on kolme erilaista pätevyyden aluetta:

- a) **Rinnakkaisvaliditeetti** (*convergent validity*) (yhtäpitävyys validiteetti; samanaikaisvaliditeetti)

Tämä validiteetin osa-alue arvioi sitä, missä määrin mittarin tulokset ovat yhtäpitäviä toisen mittarin kanssa, jonka uskotaan mittaavan samaa ilmiötä. Esimerkiksi poikittaisasetelmassa käytetyn mittarin tulosten pitäisi korreloida kohtalaisesti tai vahvasti toisen mittarin kanssa, joka on aiemmin validoitu samaan käyttötarkoitukseen. Vastaavasti pitkittäisasetelmassa käytetyllä mittarilla mitatun muutoksen tulokset pitäisi korreloida kohtalaisesti tai vahvasti toisella aiemmin samaan käyttötarkoitukseen validoidulla mittarilla saatuihin tuloksiin. Rinnakkaisvaliditeetin määrittämiseksi käytetään korrelaatiokertoimia (Finch ym. 2002, de Vet ym. 2011). Rinnakkaisvaliditeetin tutkimuksessa käytetään myös *Multi Trait Multi Method* (MTMM) -menetelmää, joka teknisesti on eräänlainen konfirmatorinen faktorianalyysi ja perustuu korrelaatioihin, kuten kaikki faktorianalyysimallit (Campbell ja Fiske 1959).

- b) **Erotteleva validiteetti** (*discriminant validity*)

Erotteleva validiteetti arvioi sitä, missä määrin mittari eroaa muita ilmiöitä mittaavista mittareista. Esimerkiksi alaraajojen toimintakykyä mittaavan mittarin tulokset (tai pitkittäisasetelmassa muutostulokset) korreloivat paremmin geneerisen mittarin fyysisen toimintakykyosan tulosten (tai muutostulosten) kanssa kuin tunne-elämää mittaavan osan tulosten kanssa. Mittarin erottelukykyä tutkitaan korrelaatiokertoimilla (Finch ym. 2002, de Vet ym. 2011).

**c) Kyky erottaa ryhmiä toisistaan (*known group validity*)**

Eryteisesti erottelevaa mittaria validoitaessa arvioidaan sen kykyä erotella kaksi tai useampi ryhmä toisistaan. Ryhmät määritellään hypoteesien asettamisvaiheessa edustamaan eri vaikeustasoja esim. akuutit ja krooniset selkäkipupotilaat. Mittarin tulisi erotella nämä ryhmät toisistaan siten, että mittarin tulokset eroavat selvästi kahdella eri sairauden vaikeustasolla toisistaan. Ryhmien erottelussa usein käytettyjä testejä ovat mm. t-testi ja varianssianalyysit (*analysis of variance, ANOVA; analysis of covariance, ANCOVA*) (Finch ym. 2002, de Vet ym. 2011).

**3.4.3 Kulttuurien välinen validiteetti (*cross-cultural validity*)**

Kulttuurien välistä validiteettia tutkitaan tavallisesti silloin, kun käännetään kysely- ym. lomakkeita toiselle kielelle. Kulttuurien välistä validiteettia tutkittaessa asetetaan ensin hypoteesit, joilla voidaan tutkia esimerkiksi sitä korreloivatko eri kieliversioiden rakenteet tai erottelevatko eri kieliversiot yhtä hyvin relevantteja ryhmiä (de Vet ym. 2011). Nämä ovat kuitenkin voimakkaasti otosriippuvaisia. Sen vuoksi parempi tapa on tutkia mittarin muuttumattomuutta tai vastaavuutta (*measurement invariance*) eli sitä toimiiko mittari, asteikko tai osio täsmälleen samalla tavalla eri populaatioissa. Tällöin mittarilla ei ole havaittavissa eri osioiden välistä erilaista "toimintaa" (*differential item functioning; DIF*). Jos eri osioiden välillä havaitaan eroja, on selvítettävä löytyykö niille käytännön selitys. Usein erot jäävät selittymättä ja silloin osiot voidaan muokata uudelleen tai poistaa. Muokkauksen tai poistojen jälkeen mittaria on tutkittava uudelleen keräämällä uusi data tutkittavasta kohdepopulaatiosta. Kulttuurien välistä validiteettia voidaan tutkia useilla menetelmillä, esimerkiksi konfirmatorisella faktorianalyysillä, logistisella regressioanalyysillä ja IRT-malleilla. Faktori- ja IRT-analyysien etu on se, että mallia voidaan muokata, kun taas logistisessa regressioanalyysissä tätä mahdollisuutta ei ole.

Mittarin kääntämiseksi toiseen kulttuuriin tarvitaan standardoitu kääntämisprosessi, jossa otetaan huomioon kieli, kulttuuri ja maantieteelliset olosuhteet (Beaton ym. 2000). Tämän prosessin tulisi edetä seuraavalla tavalla:

- a) käännös kohdekielille, jonka suorittaa kaksi kohdekielistä kääntäjää
- b) kääntäjät tekevät synteetin ja ulkopuolinen havainnoitsija ratkaisee erilaisuudet
- c) takaisinkäännös, jonka tekee kaksi alkuperäiskielistä kääntäjää
- d) käännösten yhtäpitävyyden tarkastelu, alustavan versio laatiminen (4 kääntäjää, havainnoitsija, menetelmäosaaja, sosiaali- ja terveydenhuollon ammattilaiset, kielen osaaja)
- e) alustavan version testaaminen (kohderyhmä n=15–30)
- f) arvio ja hyväksyntä (alkuperäismittarin kehittäjät)
- g) käännetyn mittarin validointi kohdemaassa voi alkaa

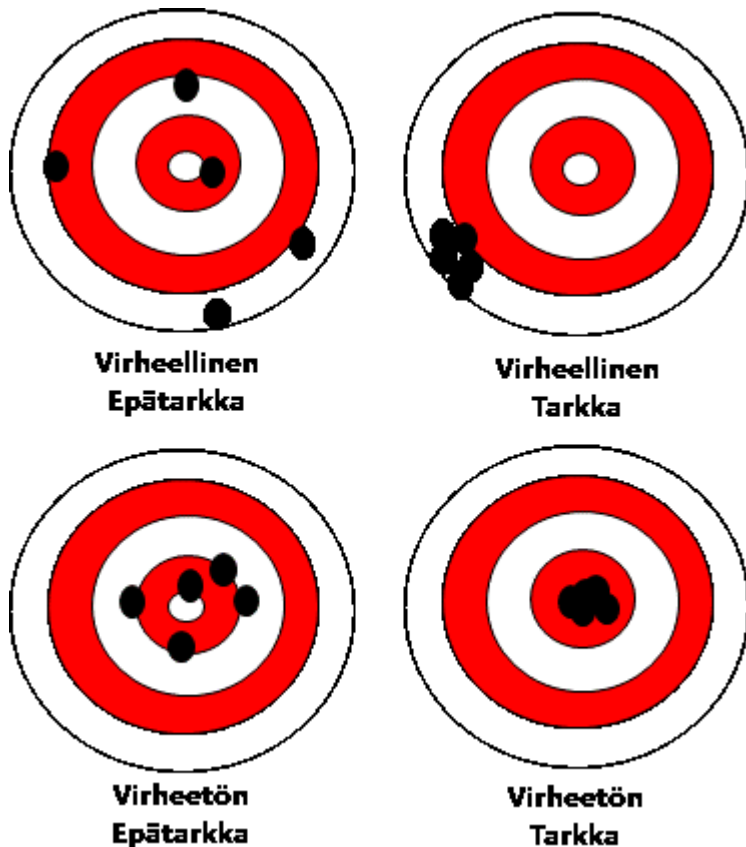
Osa tutkijoista pitää mittarin laadukasta adaptointi- eli soveltamisprosessia tärkeämpänä kuin tarkkaa kääntämistä (Sweine-Verdier ym. 2004), jotta mittari toimisi kohdemaan sosiaalisessa ympäristössä. He ehdottavat takaisinkääntämisprosessin sijaan kahden paneelin menetelmää. Ensimmäinen paneeli koostuu mahdollisimman "tavallisista" kääntäjistä, jotka osaavat hyvin sekä alkuperäiskielen että kohdekielen. He kääntävät mittarin kohdemaan kielelle. Toiseen paneeliin

kootaan 5–7 kohderyhmän edustajaa, eli maallikkoja erilaisilla sosiaali-, sukupuoli- ja koulutustaustoilla. He tarkastelevat pelkästään käännettyä versiota ja arvioivat sen sanamuotoja ja kertovat, mitä ymmärtävät kullakin kysymyksellä. Lisäksi tarvitaan koordinaattori, joka varmistaa tiedonkulun ryhmien välillä, prosessin monitoroinnin ja sen, että pysytään mittarin konstruktiossa ja käyttötarkoituksessa. Tämän jälkeen adaptoidun mittarin validointi kohdemaassa voi alkaa.

Arvioitaessa kulttuurien välistä validiteettia artikkeleista tulisi etsiä käänösprosessin kuvaus ja arvioida, miten hyvin prosessi vastaa jompaakumpaa edellä kuvattua menetelmää. Lisäksi tulisi kuvata, mitä vertailutietoa alkuperäiseen mittariin on saatavilla.

## 4 RELIABILITEETIN OSA-ALUEET

Mittaukset sisältävät usein kahdenlaisia virheitä: systemaattisia virheitä ja satunnaisvirheitä. Systemaattinen virhe syntyy, kun käytetty mittari on esimerkiksi kalibroitu väärin tai mittaustilanne on suunniteltu huonosti. Tällöin mittaukset ovat systemaattisesti virheellisiä, vaikka ne voivatkin olla tarkkoja (kuva 4). Satunnaiset virheet voivat taas johtua monista eri tekijöistä, esimerkiksi olosuhteiden ja vastaajan mielialan ym. satunnaisvirheiden vaikutuksesta. Niitä voidaan yrittää ehkäistä, mutta kokonaan niitä ei voida poistaa.



**KUVA 4.** Mittauksien virheetömyyden/virheellisuuden ja tarkkuuden/epätarkkuuden eri mahdollisuudet.

Reliabiliteetti (*reliability*) ilmaisee, miten luotettavasti ja toistettavasti käytetty mittaus- tai tutkimusmenetelmä mittaa haluttua ilmiötä. Reliabiliteetti-käsite kuuluu yleensä määrälliseen, kvantitatiiviseen tutkimukseen, jossa mittauksen reliabiliteetti tarkoittaa mittaustuloksen toistettavuutta, ei-sattumanvaraisuutta. Kvalitatiivisessa tutkimuksessa reliabiliteetilla tarkoitetaan aineiston käsittelyn ja analyysin luotettavuutta (Golafshani 2003).

Reliabiliteetista on ollut käytössä useita termejä kuten, *objectivity*, *repeatability*, *reproducibility*, *stability*, *agreement*, *association*, *sensitivity* ja *precision* (de Vet ym. 2011). Suomeksi on käytetty termejä tarkkuus, luotettavuus, käyttövarmuus, toimintavarmuus ja johdonmukaisuus. Vain

harvalla näistä termeistä on selkeä määritelmä, joten niiden erot ovat epämääräisiä. *Repeatability* voidaan kuitenkin määritellä samoissa olosuhteissa suoritettujen toistettujen mittausten vaihtelua kuvaavaksi suureksi. *Reproducibility* on taas eri olosuhteissa suoritettujen toistettujen mittausten vaihtelua kuvaava suure. Sillä mitataan esim. eri reagenssien, olosuhteiden, määrittäslaitteiden ja laboratorioden välistä vaihtelua. Hyvä toistettavuus samoissa olosuhteissa on tärkeä edellytys toistettavuudelle eri olosuhteissa. Ilmiön pysyvyyttä tarkoitettaessa käytetään termiä *stability* (Metsämuuronen 2000, de Vet ym. 2011). Kvalitatiivisessa tutkimuksessa reliabiliteettia vastaavina termeinä on käytetty muun muassa seuraavia termejä: *trustworthiness*, *credibility*, *neutrality*, *confirmability*, *consistency*, *dependability*, *applicability* ja *transferability* (Golafshani 2003).

De Vet ym. (2011, 97) mukaisesti TOIMIA:ssa käytetään reliabiliteetin kriteereinä toistettavuutta ja mittausvirhettä sekä mittarin sisäistä yhtenevyyttä (*internal consistency*). Samalla reliabiliteetin käsitteellä on siis määritelty kaksi varsin erilaista mittarin ominaisuutta: mittarin tai menetelmän pysyvyys ajassa (toistettavuus ja mittausvirhe, ks. 4.1) ja mittarin sisäinen yhteneväisyys (ks. 4.2). Vaikka mittari olisi sekä toistettava että sisäisesti yhtenevä, se ei riitä. Mittari voi nimittäin mitata väärääkin asiaa hyvin johdonmukaisesti. Mittarin on siis oltava myös validi. Reliabiliteetti asettaa myös mittarin validiteetille ylimmän mahdollisen rajan siten, että mitä korkeampi reliabiliteetti, sitä korkeampi mahdollinen maksimivaliditeetti mittarilla on (Streiner & Norman 2008).

## 4.1 Toistomittausten toistettavuus ja mittausvirhe

Toistettavuus on erityisen tärkeä silloin, kun halutaan seurata samaa ilmiötä useamman kerran (seurannassa). Toistettavuus ei ole yksi yksittäinen ominaisuus, vaan se täytyy tutkia erikseen kussakin käyttötarkoituksessa, esimerkiksi eri diagnoosien yhteydessä (Metsämuuronen 2000, de Vet ym. 2011). Ei siis ole olemassa toistettavuutta sinänsä, vaan vain testin toistettavuustuloksia erilaisista käyttötilanteista. Kun eri tutkimukset antavat erilaisia tuloksia mittaajien välisistä tai mittauskertojen välisistä toistettavuuksista, on mahdotonta sanoa johtuuko se erilaisista mittausvirheistä eri tilanteissa vai otosten välisistä eroista (Streiner ja Norman 2008).

Toistettavuutta tutkitaan toistomittausten yhtäpitävyydellä. Tällöin pyritään eri tavoin varmistamaan, että mitattava ilmiö ei ole oikeasti muuttunut mittausten välillä. Tutkimusasetelmana on yli ajan tapahtuva testi-uusinta testi -asetelma (*test-retest*), jossa mittaus uusitaan samoille tutkittaville. Testi-uusinta testi -asetelmasta voidaan käyttää myös termiä *intra-rater*, jolloin testataan saman mittaajan tekemien testien toistettavuutta. Toistettavuutta tutkitaan myös eri mittaajien välisenä, jolloin kaikki mittaajat käyttävät samaa mittaria (*inter-rater*). Toistettavuutta voidaan tarkastella relatiivisena (ks. 4.1.1) tai absoluuttisena (ks. 4.1.2).

Suositteluvan tunnusluvun käyttö riippuu muuttujasta (taulukko 2). Streinerin ja Normanin (2008) mukaan Bland-Altmanin -menetelmä (ks. kuva 5) on analoginen ICC:n (Intraclass Correlation Coefficient) mittausvirheen laskemisen kanssa. Bland-Altmanin lisäetuna on, että arvot raportoidaan suoraan ja kuviosta voidaan samalla nähdä muita systemaattisia eroja. Myös kappa-kertoimet ja ICC tuottavat identtisiä tuloksia.



**Taulukko 2.** Erityyppisten muuttujien toistettavuuden ja mittausvirheen laskemisessa käytettäviä keskeisiä tilastollisia menetelmiä (de Vet ym. 2011, 120).

	Jatkuva muuttuja	Luokitteluasteikollinen muuttuja	Nominaaliasteikollinen muuttuja
Toistettavuus	ICC	ICC; painotettu Cohenin kapp	Painottoman Cohenin kapp
Mittausvirhe / Yhtäpitävyys	SEM; LOA	% LOA	% LOA

ICC = Intraclass Correlation Coefficient; SEM = Standard Error of Measurement; LOA = Limits of Agreement - yhtäpitävyysrajat, jotka lasketaan Bland-Altmanin -menetelmällä.

Jos mittaus on toistettava, samat henkilöt saavat samalla mittarilla samanlaisia tuloksia. Ongelmana voi kuitenkin olla se, että itse tutkittava ilmiö voi muuttua mittausten välillä. Myös ihmisillä on taipumus vastata samalla tavalla usealla eri kerralla. Peukalosääntönä voitaisiin sanoa, että aikavälin mittausten välillä tulisi olla noin kaksi viikkoa (Nunnally 1978, Metsämuuronen 2000). Mikäli testi-uusinta testi -asetelman aika on mittausten välillä liian pitkä, testi mittaa todennäköisesti enemmänkin ilmiön stabiiliutta (pysyvyyttä) kuin reliabiliteettiä (Metsämuuronen 2000, de Vet ym. 2011, Kottner ym. 2011). Jos se taas on liian lyhyt, ensimmäisen mittauksen vastausten muistaminen voi vääristää uusintamittauksen vastauksia ja tuloksia.

Parhaiten toistettavuutta voidaan selvittää fysikaalisista mittauksista, joissa mittausten ero eri ajanhetkinä voi aiheutua ainoastaan mittausvirheestä. Esimerkkinä voidaan ajatella kehon painoa. Jos kahden mittauksen välillä ei ole niin pitkää väliä, että paino ehtisi aidosti nousta tai laskea, on kyseessä selvästi mittausvirhe. Jos mittaus ei ole fysikaalinen, vaan esim. jokin kyselylomake, joudutaan olettamaan, että mittaus mahdollisesti vaikuttaa mitattavaan kohteeseen. On esimerkiksi mahdollista, että ihminen muistaa mitä on vastannut aikaisemmin, oppii testin tekemistä jne. Siksi tämän tyyppisissä mittareissa on toiston avulla vaikeampi selvittää sitä, kuinka tarkka mittari on ja kuinka paljon mittaustulokset sisältävät mittausvirhettä.

#### 4.1.1 Relatiivisen toistettavuuden tilastolliset menetelmät/tunnusluvut

Kun tutkitaan sitä miten hyvin toistomittauksissa kukin tutkittava säilyttää asemansa tutkimusjoukossa suhteessa toisiin, käytetään nimitystä relatiivinen reliabiliteetti. Tämän laskemiseen käytetään useimmin Intraclass Correlation Coefficient- (ICC) ja Cohenin kapp - menetelmiä (taulukko 2) (Lohr 2002, de Vet ym. 2011, Kottner ym. 2011).

##### **Intraclass Correlation Coefficient (ICC):**

Jatkuvien muuttujien toistettavuuden laskemiseen käytetään ICC:tä. ICC:n laskemiseen on olemassa useita laskentakaavoja, jotka kaikki perustuvat varianssiin. ICC voi saada arvon, joka vaihtelee välillä 0–1. Tulkinta koostuu ICC:n saamasta arvosta, luottamusvälistä ja tilastollisesta merkitsevyydestä (Shrout ja Fleiss 1979). ICC-tulosten tulkintaan on olemassa useita ohjeita, mutta niitä käytettäessä on huomioitava, että nekin on saatu tutkimuksen tuloksena tietyllä

otoksella. Sen vuoksi arvioitaessa eri tutkimuksissa saatuja ICC-tuloksia on ensin ymmärrettävä mikä on tutkimuksen tarkoitus, millä ja millaisella otoksella tulokset on saatu ja mitä tarkasti ottaen mitattiin. Vasta tämän jälkeen saatuja tuloksia tulisi tarkastella suhteessa ohjearvoihin, mutta pitäen aina mielessä, että kaikki ICC-tulokset ovat riippuvaisia siitä kontekstista, jossa ne on mitattu.

Yllä olevaan viitaten ICC:n tulkintaan on olemassa useita ohjearvoja, joissa tulkinnan rajat vaihtelevat. Yhteistä kaikissa ohjearvoissa on kuitenkin se, että korkean/hyvän ICC:n tulisi olla  $>0.9$  (Vincent 1994, Atkinson ja Nevill 1998, Finch ym. 2002, de Vet ym. 2011). Muut arvot ovat hieman erilaisia lähteistä riippuen, mutta karkeasti voitaisiin todeta, että ICC:n ollessa  $0.7-0.9$  se on keskinkertainen ja alle  $0.7$  olevat arvot heikkoja, joihin tulisi suhtautua varoen (Nunnally ja Bernstein 1994, de Vet ym. 2011). Jos mittari on tarkoitettu kliiniseen käyttöön, ICC:n tulisi olla vähintään  $0.9$ . Tutkimuskäytössä olevan mittarin ICC ei saa olla alle  $0.8$  (Atkinson ja Nevill 1998, Finch ym. 2002, de Vet ym. 2011).

#### **Cohenin kappaa ( $\kappa$ ) tai painotettu Cohenin kappaa:**

Cohenin kappaa on luokkamuuttujien välinen yhdenmukaisuuden mitta kahden tai useamman arvioitsijan tai diagnostisen tavan välillä. Cohenin kappaa mittaa sitä, kuinka paljon todettu yhtäpitävyys ( $p_0$ ) poikkeaa pelkän sattuman perusteella odotettavissa olevasta yhtäpitävyydestä ( $p_e$ ). Tuloksissa suositellaan ristiintaulukoinnin esittämistä. Teoreettisesti on mahdollista saada kapalle negatiivinen arvo, mutta käytännössä kappaa vaihtelee kuitenkin välillä  $0-1$ . Mitä suurempi kappaa arvo on, sitä yhteneväisempiä arvioinnit ovat (Sarna 2007, de Vet ym. 2011). Myös kappaa arvon tulkintaan tulee suhtautua varoen (ks. yllä ICC:n tulkinta). Jos kappaa arvo on alle  $0.40$ , sen voidaan tulkita olevan heikko. Kohtalainen/hyvä kappaa arvo on  $0.41-0.80$  ja erittäin hyvä kappaa arvo on  $0.81-1.00$  (Landis ja Koch 1977).

### **4.1.2 Absoluuttisen toistettavuuden tilastolliset menetelmät/tunnusluvut**

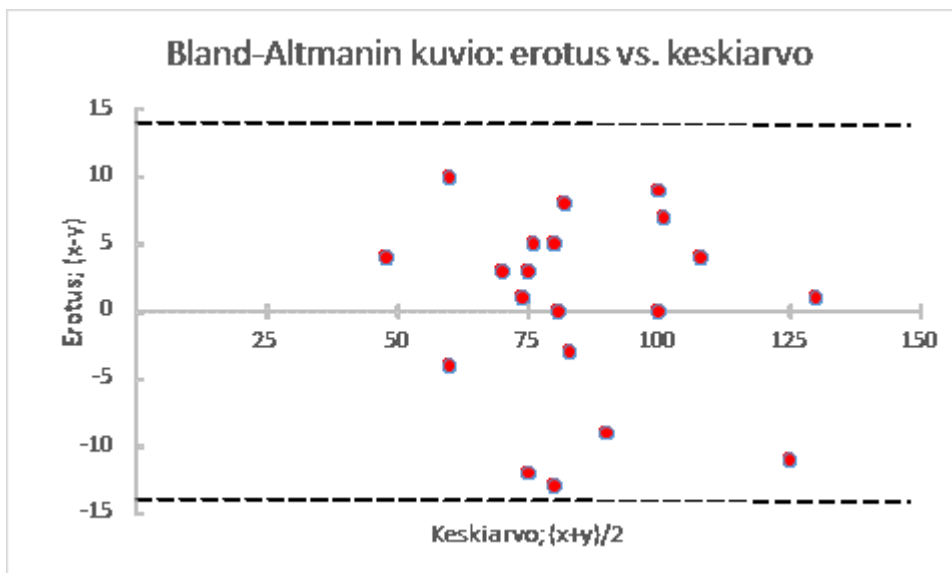
Absoluuttista toistettavuutta tutkittaessa tutkitaan toistettujen mittausten tulosten vaihtelua. Tämän laskemiseen käytetään useimmin mittauksen keskivirhettä, yhtäpitävyysrajoja ja variaatiokerrointa (Lohr 2002, de Vet ym. 2011, Kottner ym. 2011) (Taulukko 2). Absoluuttisen toistettavuuden tulokset ilmoitetaan joko mittarin todellisissa yksiköissä (esimerkiksi kg) tai prosenttiosuutena mitatuista arvoista.

#### **Mittauksen keskivirhe (*Standard Error of Measurement; SEM*):**

SEM ilmoittaa sen, kuinka paljon pelkät mittausvirheet vaikuttavat vaihteluun saaduissa testipistemäärissä (toistettujen mittausten välinen ero). SEM määritellään varianssin ja reliabiliteetin tulona (reliabiliteetin arvona käytetään esim. ICC:tä). Mitä suurempi on varianssi ja mitä pienempi on reliabiliteetti, sitä suurempi on mittauksen keskivirhe. Keskivirheen laskemiseen on olemassa useita tapoja, kuten esimerkiksi:  $SEM = SD \times \sqrt{1 - ICC}$ . Mittauksen keskivirheen yksikkö on sama kuin mittarin yksikkö ja se antaa tietoa siitä, missä rajoissa tutkittavan suoritus vaihtelee pelkän mittausvirheen vaikutuksesta. Mittauksen keskivirheen tulkintaan ei voida antaa yleisiä rajoja. Mitä parempi on reliabiliteetti, sitä pienempi on SEM (de Vet ym. 2011). Siten pienemmät erot mittausten välillä ovat osoitus todellisista eroista mitattavassa ilmiössä.

### Yhtäpitävyysrajat (*Limits of agreement, LOA*) Bland-Altmanin piirroksen avulla (*Bland-Altman's plot*):

Bland-Altmanin piirros on kahden menetelmän tai mittauskerran yhtäpitävyyden (agreement) arvioimiseksi käytetty graafinen menetelmä (Bland ja Altman 1986) (Kuva 5). Piirroksessa x-akselilla on kahdella mittauskerralla saatujen tulosten (x ja y) keskiarvo  $(x+y)/2$  ja y-akselilla niiden erotus  $d=x-y$ . Piirroksessa esitetään keskiarvoa kuvaava viiva ja normaalijakaumaan perustuvat nk. yhtäpitävyysrajat (LOA). Yhtäpitävyysrajat määritellään:  $d \pm 1,96 \cdot SD$ , missä SD on erotusmuuttujan keskihajonta. Jos keskiarvo = 0 ja LOA =  $\pm 8,3$ , mittauksessa ei ole systemaattista virhettä. Tuloksen tulkintaa helpottaa se, että tulos esitetään mittarin yksikköinä. Jos kahden menetelmän tai eri mittauskertojen välinen tulos on yhtäpitävyysrajojen sisäpuolella, tulos on tulkittava mittausvirheeksi (merkitsevyytaso on  $p > 0,05$ ). Jos taas tulos on yhtäpitävyysrajojen ulkopuolella, tulos on tällöin merkitsevä ( $p < 0,05$ ) (Atkinson ja Nevill 1998, Finch ym. 2002, de Vet ym. 2011).



KUVA 5. Bland-Altmanin kuvio.

### Variaatiokerroin (*Coefficient of Variation; CV*):

Variaatiokerroin sopii vain jatkuville suhdeasteikkomuuttujille (eli muuttuja ei saa saada negatiivisia arvoja). Variaatiokertoimen tulkinta helpottuu, jos otos on normaalijakautunut, mutta se ei ole ehdoton edellytys. Variaatiokerroin voidaan laskea seuraavilla tavoilla:

- $CV = (SD / \text{kahden mittauksen keskiarvo})$
- $CV = (SD / \text{kahden mittauksen keskiarvo}) \times 100$

Molemmissa kaavoissa SD tarkoittaa keskihajontaa. Variaatiokerroin ilmaistaan ilman yksikköä (esim. 0.14) tai prosentteina (esim. 14 %). Mitä suurempi variaatiokerroin on, sitä suurempaa on tulosten vaihtelu (Atkinson ja Nevill 1998, Finch ym. 2002, de Vet ym. 2011).

## 4.2 Sisäinen yhtenevyys

Sisäinen yhtenevyys (konsistenssi) määrittelee miten samanlaisia yksittäiset testikysymykset tai tehtävät ovat keskenään, mittaavatko ne samaa konstruktiota samalla tavalla ja voiko ne loogisesti ryhmittää yhteen. Tutkittaessa sisäistä yhtenevyyttä tutkittavalle ryhmälle tehdään yksi mittauskerta samalla mittarilla, saadut tulokset jaetaan keinotekoisesti kahteen osaan ja sopivalla tilastollisella menetelmällä (esim. Cronbachin alfa, 'split halves' tai Kuder-Richardson) analysoidaan tulokset. Nämä testit tuottavat samankaltaisia tuloksia ja kertoimet saadaan helposti, koska analyysiin tarvitaan vain yksi mittauskerta. Poikkileikkausasetelma ei kuitenkaan ota huomioon eri ajankohtien tai eri mittajien aiheuttamaa vaihtelua (Streiner ja Norman 2008). Toinen mahdollinen analyysitapa on faktorianalyysi, jonka avulla etsitään faktorirakenteen luotettavuusmittaa – tätä jälkimmäistä ei kuitenkaan käytetä kovin usein (Metsämuuronen 2000, Finch ym. 2002).

Cronbachin alfa mittaa sisäistä yhdenmukaisuutta (konsistenssia) mm. testeissä, jotka muodostuvat useista eri osioista (*items*). Alfa-arvoja, jotka jäävät alle 0.60, ei pitäisi hyväksyä. Yleisesti hyväksytyt Cronbachin alfa-arvot vaihtelevat 0.70 ja 0.90 välillä ja lisäksi on esitettävä alfan luottamusväli. Korkea alfan arvo, esim. arvo 0.98 ilmaisee, että mittarissa saattaa olla liian paljon samantyyppisiä kysymyksiä. Alfa tuloksesta on huomioitava esim. osiot, jotka ovat varianssiltaan lähellä nollaa, sillä ne eivät erottele tutkittavia toisistaan riittävästi ja laskevat siten mittarin reliabiliteettia. Lisäksi suppea mittarin skaala (esimerkiksi kolmiluokkainen Likert-asteikko) aiheuttaa arvoihin niin vähän vaihtelua, että reliabiliteetti jää matalaksi. Myös osioiden määrä vaikuttaa, sillä pitkä mittari aiheuttaa suurta vaihtelua kokonaismittariin ja se heijastuu reliabiliteetin kasvamisena (Metsämuuronen 2000, de Vet ym. 2011). Jos mittari on moniulotteinen, tai sen sisältämät osiot korreloivat toisiinsa, alfa voi olla negatiivinen, mikä on merkki siitä, että mittarin rakenteessa tai pisteytyksessä on jotain vialla (Streiner ja Norman 2008). Cronbachin alfaa käytetään paljon rutiinimaisesti analysoitaessa moniulotteisten mittareiden sisäistä yhtenäisyyttä. Tilalle on myös esitetty käytettäväksi uutta Tarkkosen menetelmää (Vehkalahti 2000).

Sisällön validoinnilla ja sisäisellä yhtenevyydellä on yhteys. Esimerkiksi, jos tavoitteena on mitata tiettyä piirrettä, vaikeutta tai käyttäytymistä, joka on heterogeeninen, on ymmärrettävää, että mittarilla on heikko sisäinen yhtenevyys. Tällainen tilanne voi olla esimerkiksi joidenkin oirekyselyiden kohdalla, sillä monissa sairauksissa oireet voivat olla kovin erilaisia eikä tällöin ole välttämättä mahdollista rakentaa sisäisesti yhdenmukaista kyselyä. Erilaiset oireet ovat kuitenkin tärkeitä ja oleellisia, joten kyselyn sisäinen validiteetti on tällöin tärkeämpi kuin sisäinen yhtenevyys.

**Tästä jatkuu!!!**

## 5 MUUTOSHERKKYYS (*responsiveness*)

Muutosherkkyys on kohtalaisen uusi käsite, jolla ei ole ollut vakiintunutta tarkastelutapaa. Muutosherkkydestä on käytetty muun muassa käsitteitä *sensitivity to change* ja *responsiveness* (Finch ym. 2002), joista ainakin jälkimmäinen on vakiintunut käyttöön (de Vet ym. 2011). Muutosherkkyys on tärkeä mittarin psykometrinen ominaisuus erityisesti silloin, kun mittarilla arvioidaan muutosta. Muutosherkkyttä on pidetty sekä itsenäisenä psykometrisenä ominaisuutena (Kirsner ja Guyatt 1985) että osana validiteettia (Hays ja Hadorn 1992). Tässä yhteydessä noudatamme De Vetin ym. (2011) esittämää COSMIN-taksonomiaa, jossa muutosherkkyys ymmärretään osaksi validiteettia. Kuitenkin muutosherkkyiden tärkeyden vuoksi se on kuvattu tässä erikseen omissa kappaleissaan.

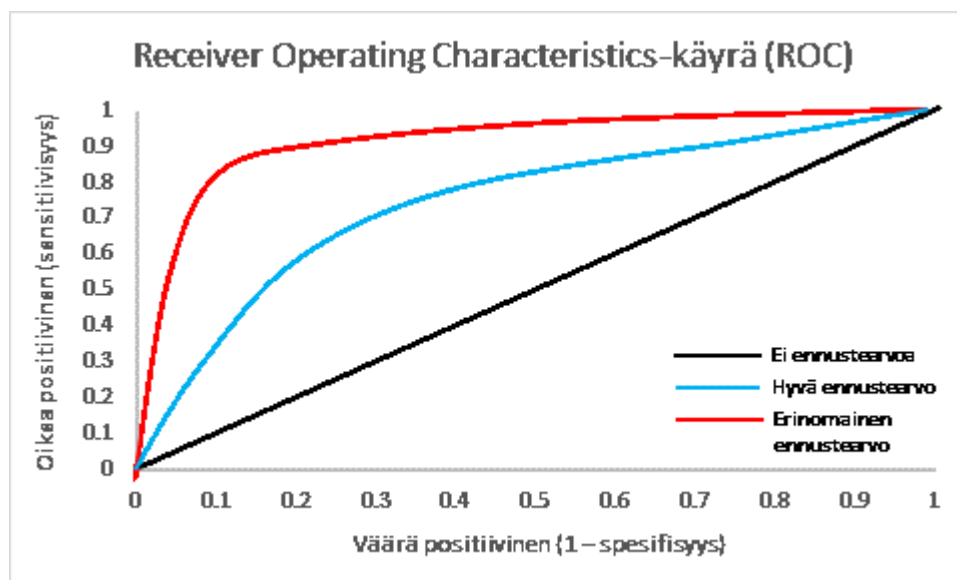
Muutosherkkyys on siis osa mittarin validiteettia, koska määrittelyn mukaan muutosherkkyys on mittarin kyky havaita tarkasteltavassa ominaisuudessa ajassa tapahtuvia muutoksia (vrt. validiteetin määritelmä) (de Vet ym. 2011, 202). Keskeisin ero validiteetin ja muutosherkkyiden tarkastelussa on siten se, että validiteettia tarkasteltaessa kohteena ovat yhden mittauskerran tulokset (*single score*), mutta muutosherkkyttä tarkasteltaessa kohteena on kahden mittauskerran tulosten muutos (*change score*). Siten muutosherkkyiden voidaan ajatella olevan myös ”pitkittäisvaliditeetti”.

Muitakin määritelmiä on ollut käytössä perustuen muunlaisiin näkökulmiin, joista tässä on esillä kaksi. Ensinnäkin mittarin muutosherkkyys on määritelty yleiseksi kyvyksi havaita muutos (*the ability to detect change in genera*). De Vetin ym. (2011, 204) mukaan tämä määritelmä on liian laava, sillä mittarin kyky havaita mitä tahansa muutosta ei ole keskeistä, vaan se miten mittari havaitsee muutoksen juuri mitattavassa konstruktiossa. Toinen erilainen tapa määritellä mittarin muutosherkkyys on määritellä se kyvyksi havaita kliinisesti tärkeä muutos (*the ability to detect clinically important change*). De Vetin ym. (2011, 204) mukaan tämä vaatii määritelmän siitä, mikä on kliinisesti tärkeä muutos. Heidän mielestään tämä on siis enemmän tulosten tulkintaa (ks. kappale 6) kuin puhtaasti mittarin muutosherkkyttä. Tästä syystä ”tarkeys” ei kuulu COSMIN-paneelin määritelmään muutosherkkydestä.

Muutosherkkyiden tutkiminen edellyttää ainakin kahden ryhmän vertailua (stabiili vs. muutos) pitkittäisasetelmassa, jossa on toteutettu vähintään kaksi mittauskertaa (Finch ym. 2002, Lohr 2002, Haley ja Fragala-Pinkham 2006). Aikaväli mittauskertojen välillä voi olla joko lyhyt (muutama viikko) tai pitkä (kuukausia). Aikavälillä ei itsessään ole merkitystä, vaan ainoastaan sillä, onko oletettavaa, että kyseisellä aikavälillä mitattavassa suureessa on mahdollista tapahtua ainakin osalla tutkittavista joko paranemista tai heikkenemistä. Tarkoituksena siis on, että tarkasteltavan mittarin olisi kyettävä havaitsemaan tämä mahdollinen muutos.

## 5.1 Suositeltavia tilastollisia menetelmiä muutosherkkyiden mittaamiseen

Muutosherkkyttä tutkitaan hyvin samalla tavoin kuin validiteettia, koska sitä tarkastellaan joko kriteeri- (*criterion*) tai rakenne (*construct*) -näkökulmasta (vrt. kappaleet 3.3 ja 3.4). Muutosherkkyden kriteerinäkökulma vaatii samanaikaista mittausta tutkittavalla mittarilla ja kultaisella standardilla. Ainoa ero kriteerivaliditeetin mittaamiseen on se, että mittarin muutosherkkyttä tutkittaessa tarkastellaan kahden eri mittauskerran välistä muutosta, ei yksittäisen mittauksen tuloksia. Mittaukset on tehtävä samanaikaisesti sekä tutkittavalla mittarilla että kultaisella standardilla ja huomioitavaa on, että kultaisena standardina käytettävän mittarin muutosherkkyys on oltava tiedossa. Muutosherkkyden tilastollinen tarkastelu tapahtuu käyttäen **samoja tilastollisia menetelmiä kuin tutkittaessa kriteerivaliditeettia** (ks. taulukko 1), esim. ROC-käyrä. Käytettäessä ROC-käyrää (kuva 6; ks. myös kpl 6.2.2) muutosherkkyden tarkastelussa on huomioitava, että tutkittavien joukossa olisi oltava osa sellaisia, joilla tapahtui mittauksien välillä muutos sekä sellaisia, joilla muutosta ei tapahtunut.



KUVA 6. ROC-käyrä (receiver operating characteristics curve).

Jos kultaista standardia ei ole käytettävissä, mittarin muutosherkkyttä tutkitaan **samoilla periaatteilla kuin rakennevaliditeettiäkin**, ts. asettamalla ennalta laadittuja hypoteeseja mittareiden ja/tai ryhmien välisille tuloksille. Hypoteesien tulisi sisältää tulosten odotettu suunta ja suuruus (absoluuttisesti tai suhteellisesti ilmoitettuna). Tulosten tarkastelussa (esim. korrelaatioissa) ei tulisi nojautua p-arvoihin, koska muutosherkkyttä arvioitaessa ei ole oleellista määrittää sitä, eroavatko saadut korrelaatiot merkittävästi nolosta vai eivät. Sen sijaan tulisi keskittyä siihen, ovatko mitatun konstruktion havaittujen korrelaatioiden suunta ja suuruus hypoteeseissa oletettujen korrelaatioiden suuntainen. Tilanne on aivan sama arvioitaessa mittarin kykyä erotella ryhmiä. Muutosherkkyttä tarkasteltaessa ei ole niin oleellista eroavatko ryhmät tilastollisesti merkittävästi toisistaan, vaan ovatko havaitut erot suurempia/pienempiä kuin

oletettiin. Huomattavaa on, että hypoteesien testaus muutosherkkyttä arvioitaessa on harvinaisempaa kuin validiteettia arvioitaessa. Tämän on arvioitu johtuvan siitä, että kirjallisuudessa ja tutkijoiden keskuudessa ei ole ollut selkeää yhteisymmärrystä muutosherkkyden arvioimiseksi. Samoin kuin kriteeri- ja rakennevaliditeetin, myös muutosherkkyden arviointi on kumuloituva prosessi, jossa tietoa kertyy vähitellen.

## 5.2 Ei-suositeltavia tilastollisia menetelmiä muutosherkkyden mittaamiseen, mutta suositeltavia muutoksen määrän tulkintaan

Muutosherkkyttä on joissakin tutkimuksissa mitattu nykytietämyksen valossa menetelmillä, jotka ovat riittämättömiä osoittamaan muutosherkkyttä. Tällaisilla menetelmillä ei saada tietoa muutoksen validiteetista, vaan lähinnä muutoksen määrästä. Tämä tieto on tarpeen tehtäessä tulosten tulkintaa (ks. luku 6). Tällaisia testejä ovat mm. vaikutuksen suuruutta kuvaavat *effect size* (ES) ja *standardized response mean* (SRM) sekä *Guyatt's responsiveness ratio* tai *Guyatt's Change Index* (GCI) ja parittainen t-testi (de Vet ym. 2011).

ES ja SRM antavat rajoittuneen todisteen muutosherkkydestä ilman vertailtavaa kultaista standardia tai vahvasti perusteltuja hypoteeseja odotettavissa olevista muutoksista ja niiden suuruuksista. ES ja SRM ovat kehitetty mittaamaan intervention muutoksen suuruutta (*magnitude of change*), jossa muutoksen suuruus ilmaistaan käyttämällä keskihajontaa (*standard deviation*; SD). Molemmat menetelmät ovat siis vahvasti riippuvaisia keskihajonnasta ja otoksesta. Homogeenisella (ja pienellä) otoksella sekä ES- että SRM-arvot ovat suurempia kuin heterogeenisellä (ja suurella) otoksella (Kazis ym. 1989, Finch ym. 2002, Terwee ym. 2003, Haley ja Fragala-Pinkham 2006). Kun ES:lla lasketun muutoksen suuruus on 0.2, se on määritelty pieneksi (Cohen 1988). Muutoksen suuruus on keskisuuri ES:n arvolla 0.5 ja suuri arvolla 0.8 (Cohen 1988).

Guyattin suhdetta ei myöskään pidetä hyvänä muutosherkkyden laskentatapana, koska se ei anna tietoa muutosarvojen validiteetista (*validity of change scores*). Guyattin suhde lasketaan jakamalla pienin merkittävä muutos (*minimal important change*; MIC) muutoksen keskihajonnalla ( $SD_{\text{change}}$ ). MIC on pienin muutos, minkä tutkittava kokee tärkeäksi ja  $SD_{\text{change}}$  tarkoittaa muutoksen keskihajontaa niillä tutkittavilla, joilla muutosta ei tapahtunut (*standard deviation of change score in stable patients*) (Finch ym. 2002, de Vet ym. 2011, 217). Tämä MIC on muutosarvojen tulkittavuuden mitta, mutta ei muutosarvojen validiteetin mitta. Muutoksen keskihajonta on taas lähinnä arvio mittausvirheestä. Kumpikaan luku ei siis anna tietoa muutosarvojen validiteetista eikä siten kuvaa muutosherkkyttä.

Parittaista t-testiä käytetään verrattaessa alkua- ja lopputilannetta samalla ryhmällä. T-testi antaa tulokseksi p-arvon, joka on muutoksen tilastollisen merkitsevyyden mitta. Tilastollinen merkitsevyys riippuu muutoksen suuruudesta, muutoksen keskihajonnasta ja otoskoosta. Se ei kuitenkaan kerro mitään mittarin muutosherkkyden validiteetista eikä muutosherkkyttä sen vuoksi tulisi analysoida parittaisella t-testillä (de Vet ym. 2011, 216).

## 6 TULOSEN TULKITTAVUUS (*interpretability*)

Tulosten tulkittavuus ei ole mittarin ominaisuus, kuten pätevyys ja toistettavuus, koska se ei kuvaa mittarin laatua. Sen sijaan se kertoo mitä mittauksen tulos tarkoittaa. Kuitenkin se on tärkeä osa mittaamista, sillä tulosten huolellinen tulkitseminen on edellytys hyvin suunnitellulle arviointimenetelmän käytölle kliinisessä työssä ja tutkimuksissa. Sen vuoksi COSMIN ottaa tulosten tulkinnan esille omana erillisenä osana (ks. kuva 1) (de Vet ym. 2011). Tulosten tulkittavuus määritellään mittarilla saatujen määrällisten poikkileikkaustulosten tai muutostulosten laadulliseksi merkitykseksi (de Vet ym. 2011, 228). Lukijan on hyvä huomata, että tulosten tulkinta on oleellinen osa sekä poikkileikkaus- että muutostulosten tarkastelua ja sillä kerätään tässä yhteydessä tietoa mittarin ”toimivuudesta” – ei tutkittavista tai mittauksen kohteena olevasta sairaudesta tai muusta terveydentilasta.

Sekä poikkileikkaus- että muutostulosten tulkinnassa tulisi huomioida seuraavat yleiset asiat:

### **Tulosten jakauma:**

Tulosten jakauman tarkastelu on keskeinen asia tulosten tulkinnassa. Tulosten jakauma on riippuvainen otoksesta, sillä tutkittavien ominaisuudet vaikuttavat mitattavan ominaisuuden jakaumaan (esim. tutkittavien ikä, sairauden aste, sairaus jne.). Tässä yhteydessä jakauman tarkastelun tarkoituksena on kuitenkin oppia käytettävän mittarin ominaisuuksista eikä tutkittavista. Tulosten jakautuminen mittarin asteikolle tarjoaa tietoa itse mittarista, muun muassa ovatko tutkittavien saamat pisteet korkeita vai matalia, ovatko ne jakautuneet tasaisesti koko mittarin asteikolle vai ovatko tutkittavien pisteet kasautuneet johonkin kohdalle asteikkoa. Jos mittari ei voi havaita tulosten hajontaa, se näkyy myös mittarin heikkona tuloksena reliabiliteettianalyseissa.

### **Katto- tai lattiaefektit:**

Kattoefekti tai kattovaikutus tarkoittaa tilannetta, jossa suurin osa tutkittavien tuloksista sijoittuu mittarin asteikon yläpäähän. Vastaavasti lattiaefekti tai lattiavaikutus tarkoittaa tilannetta, jossa tulokset kasautuvat asteikon alapäähän. Näillä on merkitystä etenkin pitkittäistutkimuksissa, joissa katto- tai lattiaefekti voi estää havaitsemasta tutkittavien tilassa tapahtuvia muutoksia. Usein nämä efektit tulevat ongelmallisiksi, kun jo olemassa olevaa mittaria sovelletaan uudelle ryhmälle. Tällöin ei tiedetä miten tulokset jakautuvat mittarin asteikolle ja on mahdollista, että se ei erottele tarpeeksi hyvin tutkittavia juuri mittarin asteikon ylä- ja/tai alapäässä. Huomioitavaa on, että kattoefektiä ei ole mittarilla, jolle ei ole olemassa määriteltyä maksimiarvoa (esim. 10 m:n kävelynopeus).

### **Keskeisten ryhmien poikkileikkaus- ja muutosarvot:**

Keskeisillä ryhmillä tarkoitetaan tässä koko väestöä sekä esimerkiksi eri sairautta sairastavien alaryhmiä, vaikkapa sairauden keston tai vaikeusasteen mukaan. Koko väestöä edustavat mitattavan konstruktion arvot ovat viitearvoja, joihin muiden ryhmien arvoja voidaan verrata. Tulosten tulkintaa ja niiden merkittävyyden arviointia sekä mittarin ”toimivuuden” arviointia helpottavat myös muilta ryhmiltä saatavilla olevat tutkittavan konstruktion poikkileikkaus- ja/tai muutosarvot.



## Onko pienin merkittävä muutos tiedossa?:

Kun mittarilla arvioidaan muutosta kahden mittauskerran välillä, olisi tärkeätä, että pienin merkittävä muutos tutkittavassa konstruktiossa olisi tiedossa ja määritelty (ks. tarkemmin mm. vaikutuksen suuruutta kuvaavat ES ja SRM sekä Guyatt's responsiveness ratio ja parittainen t-testi luvussa 6.2).

## 6.1 Poikkileikkaustulosten tulkinta

Poikkileikkausasetelmassa saatuja tuloksia (tulosten jakaumaa) voidaan tarkastella suhteessa väestöstä mitattuihin **viitearvoihin** ja/tai muiden saatavilla olevien aikaisemmin tutkittujen keskeisten ryhmien arvoihin. Viitearvot voivat perustua mittarin tulosten jakaumaan koko väestössä (vertaa pituus tai paino). Viitearvoiksi voidaan myös ajatella ne arvot, jotka eivät lisää esimerkiksi jonkin sairauden riskiä (vertaa verenpaine) tai ovat hoidettavissa (vertaa verenpaineen lievä nousu). Erottelevien mittareiden **raja-arvot** eri ryhmille vaikkapa sairauden vaikeusasteen tai iän mukaisesti saadaan poikkileikkausasetelmista.

## 6.2 Muutostulosten tulkinta

### 6.2.1 Pienin havaittava muutos (*Smallest/Minimal Detectable Change, SDC/MDC*)

Kaikki mittarin havaitsemat muutokset eivät ole välttämättä todellisia tai oikeita, vaan pienet muutokset voivat olla seurausta mittausvirheestä. Tärkeätä on kuitenkin tietää, mikä on todellinen pienin mittarilla havaittava muutos eli SDC. SDC on yhteydessä mittauksen keskivirheeseen (SEM) ja reliabiliteettiin (Finch ym. 2002, Haley ja Fragala-Pinkham 2006, de Vet ym. 2011) ja se osoittaa kuinka paljon tulokset vaihtelevat tutkittavilla, joilla ei tapahtunut muutosta (eli heidän tuloksensa pysyivät stabiileina). Muutostulokset arvioidaan tietyltä ajanjaksolta, joka kliinisessä työssä tai tutkimuksessa tarkoittaa usein ennen ja jälkeen hoito- tai tutkimusjakson mitattuja arvoja. Tällöin ajanjakso on niin pitkä kuin hoito- tai tutkimusjakso sattuu olemaan.

SDC määritellään **Bland-Altmanin metodilla** (kuva 5), jossa tulosten jääminen 95 %:n yhtäpitävyysrajojen sisäpuolelle voidaan tulkita mittausvirheeksi. Kun tulokset ylittävät yhtäpitävyysrajat, voidaan niiden olettaa olevan mittausvirhettä suurempia todellisia muutoksia. SDC:n yhteys mittauksen keskivirheeseen osoittaa, että käytettäessä mittausvirheeltään pieniä mittareita voidaan suhteellisen pienetkin havaitut muutokset määritellä todellisiksi muutoksiksi. Jos taas mittarin mittausvirhe on suuri, mitattujen muutosten täytyy olla huomattavia ennen kuin voidaan olla varmoja, etteivät ne johdu mittausvirheestä. SDC:n tulkittavuutta helpottaa tuloksen ilmaiseminen käytetyn mittarin mittayksikköinä. Määritettäessä pienintä havaittavaa muutosta tulkinnan on perustuttava testi-uusintatesti-asetelman tuloksiin eikä esimerkiksi Cronbachin  $\alpha$ :n käyttö ole riittävää. Tämä siksi, että Cronbachin  $\alpha$  arvioidaan yhtenä ajanhetkenä eikä se osoita tulosten variaatiota.

## 6.2.2 Pienin merkittävä muutos (*Minimal Important Change, MIC; Minimal Clinically Important Difference, MCID*)

Pienin merkittävä muutos eli MIC tarkoittaa pienintä muutosta, jonka joko tutkittava tai asiantuntija kokee tärkeäksi (*important*). Tällä hetkellä ei ole olemassa yhteistä konsensusta siitä, miten ja millä metodilla MIC tulisi määritellä. Tuloksia laskettaessa tutkittavat jaetaan kahteen ryhmään esim. sen mukaan ovatko tulokset parantuneet (parantunut vs. ennallaan) tai heikentyneet (heikentynyt vs. ennallaan) ulkoisen kriteerin perusteella. Tämän jälkeen tarkastellaan molempien ryhmien sisällä tapahtunutta muutosta (jakaumaa) ja määritellään katkaisuarvo käyttäen menetelmänä esim. ROC-käyrää (kuva 6) (Finch ym. 2002, de Vet ym. 2011). ROC-käyrä on esimerkki ankkureihin perustuvasta menetelmästä (engl. *anchor-based approached*). Tämä edellyttää ulkoista riippumatonta kiinnekohtaa (ankkuria), joka on yksinään tulkittavissa ja joka ainakin kohtalaisesti korreloi arvioinnin kohteena olevaan mittariin. Ulkoinen ankkuri tunnetaan myös kultaisena standardina. Tällainen ankkuri voidaan asettaa tutkittavan tai asiantuntijan näkökulmasta. Tutkittava mittari vaikuttaa siihen, millainen tilastollinen menetelmä tulisi valita (Taulukko 3).

**Taulukko 3.** Pienimmän merkittävän muutoksen (MIC) laskemiseksi käytetyt tilastolliset menetelmät.

Muuttuja	Tilastollinen menetelmä
Dikotominen (muutos vs. ei muutosta)	ROC
Järjestysasteikollinen (esim. Likert)	Cohenin kappa Spearmanin r
Jatkuva	Pearsonin r

ROC = Receiver Operating Characteristics curve

ROC-käyrä on toistomittausasetelmien ja diagnostisten testien yhteydessä käytetty graafinen esitystapa (kuva 6). Se on (x, y) -koordinaatistoon piirretty käyrä, missä x-akselina on väärän negatiivisen tuloksen todennäköisyys (1-spesifisyys) ja y-akselina oikean positiivisen tuloksen todennäköisyys (sensitiivisyys). Sensitiivisyys ja spesifisyys ovat vain yksittäisiä lukuja, mutta ROC-käyrä ottaa ne molemmat huomioon. Optimaalinen katkaisupiste on koordinaatiston vasenta yläkulmaa lähinnä oleva käyrän piste. Mitä enemmän käyrä lähenee tätä vasenta yläkulmaa, sen paremmin mittari erottelee tuloksia. Jos taas käyrä ei eroa suorasta viivasta, mittari ei erottele tuloksia.

ROC-käyrän alle jäävä pinta-ala (*area under curve; AUC*) mittaa testin toimivuutta. AUC voi saada arvon väliltä 0–1, jolloin arvo 1 osoittaa parasta mahdollista toimivuutta. Jos AUC=0.5, mittari ei tällöin pysty erottelemaan tutkittavia, joiden tulos on muuttunut tai pysynyt ennallaan. AUC:n lisäksi tulisi myös ilmoittaa vaihteluväli (*confidence interval; CI*; esim. 95 % CI) ja tilastollinen merkitsevyys (p-arvo). Toistomittausasetelmissä AUC:tä käytetään usein toistomittausten keskiarvon asemesta kuvastamaan hoidosta saatua vastetta tietyllä aikavälillä. ROC-käyrä kuitenkin edellyttää, että otoksessa osalla on tapahtunut muutosta ja osalla tulos on pysynyt ennallaan.

**MIC:n määrittämiseen vaikuttavat useat tekijät** (de Vet ym. 2011, 256). Ensinnäkin MIC riippuu käytetyn ankkurin tyypistä, joka voi olla asetettu potilaan tai asiantuntijan näkökulmasta. Toisekseen, MIC on riippuvainen käytetyn ankkurin pienimmästä tärkeästä muutoksesta (esim. miten potilas itse määrittelee oman tilansa muutoksen). Kolmanneksi, MIC on riippuvainen myös alkutilanteen arvoista ja potilasryhmästä. Jos potilaiden sairauden oireet ovat voimakkaasti haittaavat ja alkutilanteen tulos on siten korkea, muutostulos on tällaisilla potilailla helposti suurempi kuin niillä, joiden alkutilanne oli selvästi parempi. Neljänneksi, MIC on myös riippuvainen muutoksen suunnasta. Muutos voi olla joko paranemista/kehittymistä tai heikentymistä. Suositeltavaa on määritellä sekä kehittymiselle että heikentymiselle pienin merkittävä muutos. Koska MIC on riippuvainen näin monesta edellä mainitusta tekijästä, tutkittavan mittarin tulosten avulla määriteltä MIC ei voi olla kiinteä yksi arvo. Tämän vuoksi MIC:stä tulisi voida esittää vaihteluväli tai MIC:ä tulisi testata erilaisilla menetelmillä ja käyttäen erilaisia ankkureita.

## 7 KÄYTTÖKELPOISUUS

Mittarin käyttökelpoisuus (*feasibility*) ja kliininen käytettävyys (*clinical usefulness*) mittaamisen yhteydessä käytettyinä kuvaavat sitä, onko mittari käyttökelpoinen rutiinikäyttöön ja merkityksellinen kyseisessä käyttötarkoituksessa (Slade ym. 1999). Tiedot käyttökelpoisuudesta voivat perustua sekä kirjallisuuteen että asiantuntijoiden ja mittaria käyttävien henkilöiden kokemuksiin ja arvioihin. Oxfordin englannin kielen sanakirja määrittelee käyttökelpoisuuden seuraavasti: *capable of being done, accomplished or carried out; possible; practicable when applied to 'a design or a project' / capable of being dealt with successfully in any way, either in a material or immaterial sense when applied to 'things in general'*. Kliininen käytettävyys on määritelty seuraavasti (Slade ym. 1999): *"A term used when a measure is appraised in a specific clinical context, for a specific purpose."*

Yksiselitteinen määritelmä ja kriteerit puuttuvat mittareiden käyttökelpoisuuden määrittelyltä. Fitzpatrickin ym. (1998) mukaan mittareiden käyttökelpoisuutta arvioitaessa on selvitettävä

- a) mittarin hyväksyttävyyden potilaan/asiakkaan kannalta (*patient burden and acceptability*):
  - onko mittari tai testi hyväksyttävä ja sopiva tutkittaville?
  - mitkä ovat mahdolliset syyt tutkittavien keskeyttämiselle tai kuinka paljon on puuttuvia tietoja, jos ne on raportoitu?
  - mikä on ajankäyttötarve mittarin tai testin käytössä?
  - mikä on mittarin tai testin saatavuus ja mitkä ovat sen käyttörajoitukset?
  - onko mittari tai testi saatavilla suomenkielellä ja onko se todettu kulttuurista riippumattomaksi?
  
- b) mittarin käyttökelpoisuus mittaajan kannalta (*feasibility*):
  - millainen koulutus tarvitaan mittarin tai testin tekemiseksi?
  - mitä välineitä tarvitaan ja mitä vaatimuksia testin tekeminen asettaa ympäristölle?
  - kuinka yksiselitteiset tulokset ovat ja kuinka helppo tuloksia on tulkita?
  - onko mittarin tai testin tuloksille saatavissa viitearvoja?
  - kuinka turvallinen ja monikäyttöinen (geneerinen) mittari tai testi on, kuinka laajasti sitä käytetään jne.?

## 8 LÄHTEET

Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 1998; 26:217-238.

Beaton DE, Bombardier C, Guillemin F, Bosi-Ferraz M. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186-91.

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307-310.

Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 1959;56:81-105.

Cohen J. *Statistical power analysis for the behavioral sciences*. 2. painos. Lawrence Erlbaum, Hillsdale: NJ, 1988.

Consensus-based Standards for the selection of Health Measurement Instruments (COSMIN). <http://www.cosmin.nl/>

De Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine. Practical guides to biostatistics and epidemiology*. Cambridge: Cambridge University Press, 2011.

Finch E, Brooks D, Stratford PW, Mayo NE. *Physical rehabilitation outcome measures. A guide to enhanced clinical decision making*. 2. painos. Canadian Physiotherapy Association. Ontario: BC Decker Inc., 2002.

Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* 1998;2(14):1-74.

Golafshani N. Understanding Reliability and Validity in Qualitative Research. *The Qualitative Report* 2003;8(4):597-607.

Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther* 2006;86:735-743.

Hays RD, Hadorn D. Responsiveness to change; an aspect of validity, not a separate dimension. *Qual Life Res* 1992;1:73-75.

Kazis LE, Anderson JJ, Meenen RE. Effect sizes for interpreting changes in health status. *Med Care* 1989;27 Suppl 3:S278-279.

Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chron Dis* 1985;38:27-36.

Kottner J, Audige L, Bronson S, Donner A, Gajewski BJ, Hróbjartsson A, Roberts C, Shoukri M, Steiner DL. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64:96-106.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.

Lohr KM. Assessing health status and quality of life instruments: Attributes and review criteria. *Qual Life Res* 2002;11:193-205.

Metsämuuronen J. Mittarin rakentamisen ja testiteorian perusteet. Metodologia-sarja 6. Jaabes, Viro: International Methelp Ky, 2000.

Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010a;19:539-549.

Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes: results of the COSMIN study. *Journal of Clinical Epidemiology* 2010b;63:737-745.

Nunnally JC. *Psychometric theory*. 2nd ed. New York: McGraw-Hill, 1978.

Nunnally JC, Bernstein IH. *Psychometric Theory*. 3. Painos. New York: McGraw-Hill, 1994.

Sarna S. Biostatistiikka lyhyesti –kurssimoniste. 2007. Saatavissa: <http://www.kttl.helsinki.fi/sarna/>

Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psyc Bullet* 1979;86:420-428.

Slade M, Thornicroft G, Glover G. The feasibility of routine outcome measures in mental health. *Soc Psychiatry Psychiatr Epidemiol* 1999;34:243-249.

Strauss ME, Smith GT. Construct validity: advances in theory and methodology. *Ann Rev Clin Psychol* 2009;5:1-25.

Streiner DL, Norman GR. *Health Measurement Scales. A practical guide to their development and use*. 4<sup>th</sup> ed. New York: Oxford University Press, 2008.

Steyerberg, EW. *Clinical prediction models: a practical approach to development, validation and updating*. New York: Springer, 2009.

Swaine-Verdier A, Doward LC, Hagell P, Thorsen H, McKenna SP. Adapting quality of life instruments. Value in Health 2004, 7:S27-S30.

Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PMM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Qual Life Res 2003;12:349-362.

Üstün TB, Chatterji S, Kostanjsek N, Rehm J, Kennedy C, Epping-Jordan J, Saxena S, von Korff M, Pull C in collaboration with WHO/NIH Joint Project. Developing the World Health Organization Disability Assessment Schedule 2.0. Bulletin of the World Health Organization 2010;88:815-823. doi:10.2471/BLT.09.067231. Internet: <http://www.who.int/bulletin/volumes/88/11/09-067231/en/>

Vehkalahti K. Reliability of Measurement Scales: Tarkkonen's general method supersedes Cronbach's alpha. Statistical Research Reports 17. Licensiaattityö. Helsingin yliopisto. Finnish Statistical Society, 2000.

Vincent J. Statistics in kinesiology. Champaign (IL): Human Kinetics Books. 1994.

WHO. Toimintakyvyn, toimintarajoitteiden ja terveyden kansainvälinen luokitus. WHO/Stakes Ohjeita ja luokituksia 2004:4.

Wood-Dauphinee S. Information extraction form for measurement studies. Kirjassa: Finch et. Al. Physical Rehabilitation Outcome Measures, 2002 s. 274-6

## LIITTEET

Liite 1: Lomake 1: Mittarin perustiedot

Liite 2: Lomake 2: Mittarin psykometriset tiedot

Liite 3: Tarkistuslista 1: Arvioitavien artikkeleiden tutkimusaineistoa ja –asetelmaa kuvaavat tiedot

Liite 4: Tarkistuslista 2: Arvioitavien artikkeleiden validiteettitiedot

Liite 5: Tarkistuslista 3: Arvioitavien artikkeleiden reliabiliteettitiedot

Liite 6: Tarkistuslista 4: Arvioitavien artikkeleiden tiedot muutosherkyydestä ja muutostulosten tulkittavuudesta





## LOMAKE 1: MITTARIN PERUSTIEDOT

3.0 (1.6.2014)

1. LOMAKKEEN TÄYTTÄJÄN TIEDOT	
Nimi ja oppiarvo	Pvm
Sähköposti	Puhelin

### 2. MITTARIN IDENTIFIOINTI- JA KÄYTTÖTIEDOT

<b>Käytössä oleva vakiintunut nimi</b>  Suomeksi:  Ruotsiksi:  Englanniksi:
<b>Lyhenne</b>
<b>Mittarin nimen synonyymit</b>
<b>Arvion kohteena oleva versio</b>
<b>Ylläpitotaho</b>

<b>Käyttöoikeus</b> (valitse alla olevista vaihtoehdoista toinen laittamalla rasti ruutuun).	
<input type="checkbox"/>	<b>Rajoittamaton</b>
<input type="checkbox"/>	<b>Rajoitettu</b> (jos käyttöoikeus on rajoitettu, kuvaa alla miten)

<b>Käännösprosessin kuvaus</b>
--------------------------------



## LOMAKE 1: MITTARIN PERUSTIEDOT

3.0 (1.6.2014)

<b>Edellytykset ja välineet mittarin käytölle</b>
Ammattikoulutuksen tarve
Erillisen koulutuksen tai kurssin tarve
Mittarin käyttöön liittyvien lomakkeiden ja ohjeiden saatavuus
Tarvittavien välineiden kuvaus

### 3. MITTARIN KUVAUS

<b>Mittarin alkuperäinen käyttötarkoitus</b> (kuvaa lyhyesti)
---

<b>Tiedonkeruumenetelmät</b> (rastita sopivat vaihtoehdot)	
<input type="checkbox"/>	Haastattelu
<input type="checkbox"/>	Havainnointi
<input type="checkbox"/>	Itse täytettävä kyselylomake
<input type="checkbox"/>	Kliininen tutkimus
<input type="checkbox"/>	Testi tai mittaus

<b>Toimintakyvyn ulottuvuus</b> (rastita sopivat vaihtoehdot)	
<input type="checkbox"/>	Fyysinen toimintakyky
<input type="checkbox"/>	Psyykinen toimintakyky
<input type="checkbox"/>	Sosiaalinen toimintakyky
<input type="checkbox"/>	Kognitiivinen toimintakyky
<input type="checkbox"/>	Yleinen toimintakyky (arkitoiminnot kuten ADL/IADL)
<input type="checkbox"/>	Työkyky
<input type="checkbox"/>	Muu



## LOMAKE 1: MITTARIN PERUSTIEDOT

3.0 (1.6.2014)

<b>Mittarin kuvaamat ICF-luokituksen käsitteet</b>
Kaikki ICF-koodit, joita mittarissa (esim. sen eri osioissa) mitataan
ICF-koodi mittarin tuloksen ollessa yksi lukema (esim. summapistemäärä)

<b>Aikatarve</b>
------------------

<b>Tulkinnan avuksi</b>
<b>Viitearvot</b> (yleiset viitearvot)
<b>Raja-arvot</b> (arvoja, jotka erottelevat eri ryhmiä toisistaan)

<b>Mittari on mukana tutkimuksissa</b> (1-3 keskeistä tutkimusta)
---

<b>Mittari on mukana suosituksissa</b> (1-3 keskeistä suositusta)
---

<b>Mittarin tausta ja kuvaus</b> (yksityiskohtaisesti)
--

**4. MITTARIIN LIITTYVÄT LOMAKKEET**

Merkitse rasti ruutuun, mitkä mittaria koskevat lomakkeet on toimitettu TOIMIA:n toimitukseen.

Lomaketyyppi	Toimitettu TOIMIA:n toimitukseen	Onko lupa tallentaa tietokantaan <sup>1)</sup>
Kyselylomake		
Mittauslomake		
Suoritusohjeet		
Pisteytysohjeet		
Viitearvot		
Muu, mikä		

<sup>1)</sup> Jos mittari on tekijänoikeuksilla suojattu, tarvitaan selvitys (esim. kirjallinen dokumentti) luvasta tallentaa se tietokantaan. Tämä tieto on toimitettava yhdessä mittarin lomakkeiden kanssa TOIMIA:n toimitukseen.

**Jos lomaketta ei voi tekijänoikeus- tms. syiden vuoksi tallentaa TOIMIA-tietokantaan, merkitse tähän tarkka lähde (nettiosoite, artikkeli tms.), jota kautta lomakkeen(t) voi hankkia.**

**5. LÄHTEET**

Tarkista lähdeviitteiden merkintätavat TOIMIA:n käsikirjasta.

<b>Alkuperäinen lähdeviite</b>
<b>Hyödylliset linkit</b>
<b>Muut lähdeviitteet</b>



**6. MITTARIN TIETOJEN HYVÄKSYMINEN**

**Mittarin tiedot hyväksytty asiantuntijaryhmässä**

**Ryhmän nimi:**

**Päivämäärä:**

**Lisätietoja:**



## MITTARIN PERUSTIEDOT –LOMAKKEEN TÄYTTÖOHJEET

Mittarin perustiedot -lomakkeelle kootaan perustiedot toimintakyvyn mittaus- ja arviointimenetelmistä. Täytetty lomake käsitellään ja hyväksytään TOIMIA:n asianomaisessa asiantuntijaryhmässä, minkä jälkeen se lähetetään TOIMIA:n toimitukseen editoitavaksi. Kun lomakkeen tiedot on editoitu ja tarkistettu, se viedään TOIMIA -tietokannan suljetulle puolelle, josta tiedot kerännyt asiantuntija voi vielä ne tarkistaa ennen julkaisemista tietokannassa ([www.thl.fi/toimia/tietokanta/](http://www.thl.fi/toimia/tietokanta/)).

### Yleisiä ohjeita Mittarin perustiedot –lomakkeen täyttöön

- AINA tallentaessasi Mittarin perustiedot –lomaketta merkitse tiedoston nimeen omat nimikirjaimesi ja tallennuspäivämäärä. Jos lomakkeeseen tulee myöhemmin muutoksia, tallenna se aina uudella nimellä, josta näkyy päivämäärä.
- Täytä lomakkeen kaikki kentät. Käytä yleiskieltä ja suomenkielisiä termejä.
- Merkitse kirjallisuusviitteet tekstiin sulkuihin ja kokonaisuudessaan lomakkeen lopussa olevaan lähdeluetteloon. Kirjallisuusviitteiden lähdemerkintään liittyvät ohjeet löydät TOIMIA -käsikirjasta.

### 1. Lomakkeen täyttäjän tiedot

- Nimi ja oppiarvo:  
Merkitse lomakkeen täyttäjän tai täyttäjien nimet "Nimi" –kenttään. Merkitse jokaisen henkilön nimen perään oppiarvo esim. Emilia Esimerkki, TtM. Oppiarvo tulee näkyviin TOIMIA-tietokannassa lomakkeen täyttäjän/täyttäjien nimen perässä.
- Päivämäärä:  
Merkitse "Päivämäärä" –kenttään kaikki ne päivämäärät, jolloin lomakkeen tietoja on muutettu. Merkitse päivämäärän perään tarvittaessa muokkaajan tai muokkaajien nimikirjaimet (esim. 12.3.2013HV, 25.4.2013SS+HV).
- Sähköposti ja puhelin:  
Merkitse yhden (tai tarvittaessa useamman) mittarin tietoja koonneen asiantuntijan sähköpostiosoite ja puhelinnumero. Nämä tiedot eivät tule näkyviin TOIMIA-tietokantaan, vaan ne jäävät ainoastaan TOIMIA:n toimituksen tietoon.

### 2. Mittarin identifointi- ja käyttötiedot

- Käytössä oleva vakiintunut nimi:  
Merkitse mittarin suomen-, ruotsin- ja englanninkieliset nimet, jos ne ovat tiedossa. Jos Suomessa käytetään mittarin englanninkielistä nimeä, merkitse se mittarin "suomenkieliseksi" nimeksi.
- Lyhenne:  
Merkitse mittarin lyhenne, jos se on käytössä.
- Mittarin nimen synonyymit:  
Merkitse mittarista käytössä olevat synonyymit tarvittaessa.
- Versio:  
Jos mittarista on olemassa ja käytössä erilaisia versioita, kuvaa selkeästi se versio, jota tässä lomakkeessa kootut tiedot koskevat. Mittarin muut versiot voidaan myös luetella tässä kohden. Lisäksi tulisi merkitä selvästi, mitä versiota suositellaan käytettäväksi.
- Ylläpitotaho:  
Kuvaa lyhyesti ja selkeästi se taho, joka vastaa mittarin ylläpidosta, kuten esim. päivityksistä. Kirjaa ylläpitotahon tarkat yhteystiedot niiltä osin kuin ne ovat tiedossa: nimi, osoite, puhelinnumero, sähköpostiosoite ja nettisivujen linkki. Jos ylläpitotahoa ei ole tai se ei ole tiedossa, kirjaa myös tämä tieto. Mittarin tietojen kerääjän tulee pyytää kirjallinen lupa käyttäjätaholta mittarin julkaisemisesta (esim. kyselylomake) TOIMIA -tietokannassa ja lupa on toimitettava TOIMIA:n toimitukseen.
- Käyttöoikeus:  
Käyttöoikeudella ilmaistaan se, saako mittaria käyttää vapaasti ilman mitään rajoituksia vai onko sen käyttö rajoitettu. Valitse rajoittamaton käyttöoikeus, jos kukaan ei omista mittarin tekijänoikeuksia (copyright) ja mittaria saa käyttää vapaasti. Valitse rajoitettu käyttöoikeus, kun mittarin käyttöoikeudet (copyright) omistaa jokin taho. Kuvaa tällöin mahdollisimman selkeästi miltä osin käyttöoikeus on rajoitettu, kuka omistaa copyright-oikeudet, tarvitaanko mittarin käyttöön lisenssi, onko mittari maksullinen ja saako mittaria käyttää vain tietyissä tarkoituksissa, esim. tutkimuskäytössä. Toisin sanoen, mittarin käyttö voi olla vapaata esim. tutkimuskäytössä, mutta rajoitettu kaupallisissa tarkoituksissa.
- Käännösprosessin kuvaus:  
Kuvaa mahdollisimman tarkasti, miten mittari on käännetty (esim. suomennettu). Kuka suomennoksen on tehnyt, milloin ja millaisen protokollan mukaan se on tehty. Kirjaa myös tieto siitä, jos suomennosta ei ole tehty. Jos mittarista on saatavilla useita suomennoksia, kirjaa selkeästi mitä suomennosta tulee käyttää.
- Edellytykset ja välineet mittarin käytölle:  
Kuvaa lyhyesti seuraavat asiat:
  - edellyttääkö mittarin käyttö tiettyä ammattikoulutusta
  - onko mittarin käyttöä varten käytävä jokin erillinen koulutus/kurssi



- ovatko kaikki mittarin käyttöön liittyvät lomakkeet ja ohjeet helposti saatavilla ja mistä. Mittarin arvioijien tulee toimittaa kaikki mittarin käyttöön liittyvät lomakkeet ja ohjeet TOIMIA:n toimitukseen (ks. lomakkeen kohta 4 "Mittariin liittyvät lomakkeet").
- Kuvaa lyhyesti mitä välineitä mittauksen tekemisessä tarvitaan. HUOM! Tässä kohden ei kuvata miten mittaus tehdään.

### 3. Mittarin kuvaus

- Mittarin alkuperäinen käyttötarkoitus:  
Kuvaa mihin tarkoitukseen mittari on alun perin kehitetty. Tämä tieto perustuu yleensä alkuperäiseen lähteeseen. Kuvaa alkuperäinen käyttötarkoitus mahdollisimman lyhyesti, mutta selkeästi (vrt. lomakkeen kohta "Mittarin tausta ja kuvaus").
- Tiedonkeruumenetelmät:  
Rasita tiedonkeruumenetelmistä kaikki ne vaihtoehdot, joilla mittaria käyttäen hankitaan tietoa tutkittavasta.
- Toimintakyvyn ulottuvuus:  
Valitse kaikki ne toimintakyvyn ulottuvuudet, joiden osa-alueita mittari mittaa. Jos mittari ei sovi minkään ulottuvuuden alle, valitse kohta "muu".
- Mittarin kuvaamat ICF -luokituksen käsitteet:  
Merkitse mittarin kaikki ne ICF-koodit, joita mittarissa (esim. sen eri osissa) mitataan. Merkitse ICF-koodit mahdollisimman tarkalla tasolla. Jos mittarin tulos on yksi lukema (esim. summapistemäärä), merkitse myös mitä ICF-koodia se vastaa. Jos kyseessä on moniulotteinen mittari, joka koostuu useasta osa-alueesta, merkitään ICF-kuvauskohteet sekä osa-alueille että summaindeksille, jos sellainen on. Ehdotuksesi ICF-koodeista tarkastetaan TOIMIA:n Käsiteryhmän järjestämässä ICF-siltaustyöpajoissa. Voit jättää tämän kohdan myös täyttämättä ja se täydennetään em. työpajassa.
- Aikatarve:  
Merkitse haastatteluun, havainnointiin, kyselylomakkeen täyttöön tai mittauksen keskimäärin kuluva aika. Merkitse aikatarve minuutteina, esim. 3–5 minuuttia. Merkitse lisäksi arvio ajasta, joka kuluu tulosten laskemiseen tai tulkintaan (jos se on tiedossa). Tähän voit merkitä vapaamuotoisesti muista aikatarpeeseen keskeisesti vaikuttavista asioista.
- Tulkinnan avuksi:
  - Viitearvot ovat yleisiä arvoja esim. eri sukupuolille, eri-ikäisille jne. Viitearvoista tulisi laatia erillinen taulukko lähdetietoineen, jos se vain on mahdollista ja toimittaa taulukko TOIMIA:n toimitukseen. Tämä taulukko liitetään tietokantaan erillisenä tiedostona. Jos viitearvot ovat lyhyet ja selkeät kuvata, voit kirjata ne myös tähän "Tulkinnan avuksi" -kohtaan.
  - Raja-arvot erottelevat eri ryhmiä toisistaan. Erottele selkeästi poikkileikkaustutkimuksissa ja pitkittäistutkimuksissa saadut ja vakiintuneessa käytössä olevat raja-arvot, jos mittarista on sellaisia saatavilla.
- Mittari on mukana tutkimuksissa:  
Merkitse vain tärkeimpiä tutkimuksia, joissa mittaria on käytetty. Erityisesti keskeiset Suomessa toteutetut tutkimukset olisi hyvä mainita (1-3 kpl) (lähdeviite ja nettilinkki, jos sellainen on olemassa). Merkitse tarkat lähdetiedot lähdeluetteloon ohjeiden mukaisesti. "Mittari on mukana tutkimuksissa" -kohtaa ei ole pakko täyttää eikä ole myöskään tarkoitus listata kaikkia tutkimuksia, joissa mittaria on käytetty.
- Mittari mukana suosituksissa:  
Merkitse keskeiset ja tärkeimmät kotimaiset ja kansainväliset suositukset, joissa mittari on mukana. Kerro suosituksista lyhyesti, esim. millaisista suosituksista on kyse ja merkitse linkki tai muu lähdetieto.
- Mittarin tausta ja kuvaus:  
Kuvaa mihin tarkoitukseen, kenelle, miten ja milloin mittari on kehitetty. Kuvaa mittarin synty- ja kehityshistoria seikkaperäisemmin kuin "Alkuperäinen käyttötarkoitus" -kohdassa. Jos mittaria käytetään nykyään yleisimmin jossain muussa tarkoituksessa kuin alkuperäisessä, kuvaa tämä uusi käyttötarkoitus selkeästi. Kuvaa myös mittarin sisältö, esim. kyselylomakkeen rakenne ja kysymysten aihepiirit tai testin yleinen kuvaus, asteikko, pisteytys ym. Varsinaisia suoritusohjeita ei tarvitse selittää, jos ne ovat saatavilla erillisenä tiedostona tietokannassa.

### 4. Mittariin liittyvät lomakkeet

- Toimita mittariin liittyvät lomakkeet (esim. kyselylomake, suoritusohje) Word- tai Excel-muodossa (doc, rtf, xls) TOIMIA:n toimitukseen tietokantaan tallentamista varten. TOIMIA:n toimituksessa kaikki lomakkeet tallennetaan TOIMIA:n lomakepohjalle ja niihin liitetään identifikaatiotunnus. Lomakkeisiin kirjataan (tarvittaessa) alkuperäinen lähde, selvitys suomennoksesta ym. tarpeelliseksi arvioituiden tiedot. Jos mittari on tekijänoikeuksilla suojattu, tarvitaan selvitys (esim. kirjallinen dokumentti) luvasta tallentaa se tietokantaan. Tämä tieto on toimitettava yhdessä mittarin lomakkeiden kanssa TOIMIA:n toimitukseen.
- Jos lomaketta ei voi tekijänoikeus- tms. syiden vuoksi tallentaa TOIMIA-tietokantaan, merkitse tarkka lähde (nettiosoite, artikkeli tms.), jota kautta lomakkeen(t) voi hankkia.

### 5. Lähteet

- Tarkista lähdeviitteiden merkintätavat TOIMIA:n käsikirjasta.



- Alkuperäinen lähdeviite:  
Merkitse tähän lähdeviite, jossa mittari on alun perin ensimmäisen kerran kuvattu (mieluiten vain yksi lähde). Lisäksi voidaan merkitä suomenkielinen relevantti lähde, jossa mittari ja/tai sen ominaisuuksia on ensimmäisen kerran kuvattu.
- Hyödylliset linkit:  
Merkitse esim. mittarin kotisivun tai muiden luotettaviksi arvioitujen sivustojen nettiosoitteet, joista mittarista voi saada hyödyllistä tietoa. Tarkista, että linkit ohjaavat suoraan asianomaiseen kohtaan sivustolla.
- Muut lähdeviitteet:  
Merkitse kaikki käyttämiesi artikkeleiden tarkat viitetiedot.

### 6. Mittarin tietojen hyväksyminen

- Merkitse TOIMIA-asiantuntijaryhmän nimi ja päivämäärä milloin mittari on hyväksytty ko. ryhmässä.
- Merkitse tarvittaessa muita huomioitavia asioita mittarin hyväksyntään liittyen "Lisätietoja" –kohtaan.





1. LOMAKKEEN TÄYTTÄJÄN TIEDOT	
Nimi ja oppiarvo	Pvm
Sähköposti	Puhelin

2. ARVIOITAVA MITTARI	
Mittarin nimi	Lyhenne
Käyttötarkoitus	
Mittarin perustiedot (lomakkeen 1 tiedot)	
	Tiedot on kerätty (ei tehdä muutoksia/lisäyksiä)
	Tiedot on kerätty (tehdään muutoksia/lisäyksiä)
	Tietoja ei ole vielä kerätty (tiedot kerätään samassa yhteydessä)

### 3. KÄYTETTYJEN ARTIKKELEIDEN TUTKIMUSASETELMIEN JA –AINEISTOJEN KUVAUS

Kuvaa lyhyesti keskeiset tutkimusasetelmaan ja –aineistoon liittyvät tiedot käytetyistä artikkeleista.

**4. MITTARIN PÄTEVYYS (VALIDITEETTI)**

Kirjaa tulokset kyseisen käyttötarkoituksen kannalta oleellisilta pätevyyden osa-alueilta eri väliotsikoiden alle alla olevan jaottelun mukaisesti.

**Ilmivaliditeetti** (face validity)

**Yhteenveto**

**Sisältövaliditeetti** (content validity)

**Yhteenveto**

**Kriteerivaliditeetti** (criterion validity; jakaantuu samanaikaiseen ja ennustevaliditeettiin - kirjaa tiedot kyseisten otsikoiden alle)

**Samanaikainen validiteetti** (concurrent validity)

**Yhteenveto**

**Ennustevaliditeetti** (predictive validity)

**Yhteenveto**

**Rakennevaliditeetti** (construct validity; jakaantuu rakenteen-, yhtäpitävyys-, erotteleva-, ryhmien erottelu- sekä käännetyn mittarin validiteettiin – kirjaa tiedot kyseisten otsikoiden alle)

**Rakenteen validiteetti** (structural validity)

**Yhteenveto**

**Yhtäpitävä validiteetti** (convergent validity)

**Yhteenveto**



<b>Erotteleva validiteetti</b> (discriminant validity)	
<b>Yhteenveto</b>	
<b>Ryhmiin erotteluvaliditeetti</b> (known group validity)	
<b>Yhteenveto</b>	
<b>Kulttuurien välinen validiteetti</b> (cross-cultural validity)	
<b>Yhteenveto</b>	

**5. MITTARIN TOISTETTAVUUS (RELIABILITEETTI)**

<b>Kirjaa tulokset kyseisen käyttötarkoituksen kannalta oleellisilta toistettavuuden osa-alueilta eri väliotsikoiden alle alla olevan jaottelun mukaisesti.</b>	
<b>Toistettavuus saman mittajaan mittaamana</b> (test–retest; intra-rater)	
<b>Yhteenveto</b>	
<b>Mittaajien välinen toistettavuus</b> (inter-rater)	
<b>Yhteenveto</b>	
<b>Sisäinen yhdenmukaisuus</b> (internal consistency)	
<b>Yhteenveto</b>	

**6. MITTARIN MUUTOSHERKKYYS**

<b>Tietoja mittarin muutosherkkyydestä</b>	
<b>Kriteerivaliditeetti pitkittäisasetelmassa</b>	
<b>Yhteenveto</b>	
<b>Rakennevaliditeetti pitkittäisasetelmassa</b>	
<b>Yhteenveto</b>	
<b>Tietoja mittarin muutostulosten tulkinnasta</b>	
<b>Lattia- ja kattoefekti</b>	
<b>Yhteenveto</b>	
<b>Pienin havaittava muutos (<i>Smallest/Minimal Detectable Change; SDC/MDC</i>)</b>	
<b>Yhteenveto</b>	
<b>Pienin merkittävä muutos (<i>Minimal Important Change, MIC; Minimal Clinically Important Difference, MCID</i>)</b>	
<b>Yhteenveto</b>	

**7. MITTARIN KÄYTTÖKELPOISUUS**

<b>Tietoja mittarin käyttökelpoisuudesta ja käyttökokemuksista</b>
--

**8. LÄHTEET**

Tarkista lähdeviitteiden merkintätavat TOIMIA:n käsikirjasta.

**Alkuperäinen lähdeviite****Hyödylliset linkit****Muut lähdeviitteet****9. MITTARIN PSYKOMETRISTEN TIETOJEN HYVÄKSYMINEEN****Mittarin psykometriset tiedot hyväksytty asiantuntijaryhmässä****Ryhmän nimi:****Päivämäärä:****Lisätietoja**

## MITTARIN PSYKOMETRISET TIEDOT –LOMAKKEEN TÄYTTÖOHJEET

"Mittarin psykometriset tiedot" -lomakkeelle kootaan arvioinnin kohteena olevan mittarin validiteetti-, reliabiliteetti-, muutosherkkyys- ja käyttökelpoisuustiedot pohjautuen tieteelliseen tutkittuun tietoon. Mittarin perustiedot ilmoitetaan – ellei niitä ole aikaisemmin kerätty – lomakkeella 1 "Mittarin perustiedot". Täytetty lomake käsitellään ja hyväksytään TOIMIA:n asianomaisessa asiantuntijaryhmässä, minkä jälkeen se lähetetään TOIMIA:n toimitukseen editoitavaksi. Kun lomakkeen tiedot on editoitu ja tarkistettu, ne viedään TOIMIA -tietokannan suljetulle puolelle, josta tiedot kerännyt asiantuntija voi vielä ne tarkistaa ennen julkaisemista tietokannassa ([www.thl.fi/toimia/tietokanta/](http://www.thl.fi/toimia/tietokanta/)).

### Yleisiä ohjeita Mittarin psykometriset tiedot –lomakkeen täyttöön

- AINA tallentaessasi Mittarin psykometriset tiedot –lomaketta merkitse tiedoston nimeen omat nimikirjaimesi ja tallennuspäivämäärä. Jos lomakkeeseen tulee myöhemmin muutoksia, tallenna se aina uudella nimellä, josta näkyy päivämäärä.
- Täytä lomakkeen kaikki kentät. Käytä yleiskieltä ja suomenkielisiä termejä.
- Merkitse kirjallisuusviitteet tekstiin sulkuihin ja kokonaisuudessaan lomakkeen lopussa olevaan lähdeluetteloon. Kirjallisuusviitteiden lähdemerkintään liittyvät ohjeet löydät TOIMIA -käsikirjasta.

### 1. Lomakkeen täyttäjän tiedot

- Nimi ja oppiarvo:  
Merkitse lomakkeen täyttäjän tai täyttäjien nimet "Nimi" –kenttään. Merkitse jokaisen henkilön nimen perään oppiarvo esim. Emilia Esimerkki, TtM. Oppiarvo tulee näkyviin TOIMIA-tietokannassa lomakkeen täyttäjän/täyttäjien nimen perässä.
- Päivämäärä:  
Merkitse "Päivämäärä" –kenttään kaikki ne päivämäärät, jolloin lomakkeen tietoja on muutettu. Merkitse päivämäärän perään tarvittaessa muokkaajan tai muokkaajien nimikirjaimet (esim. 12.3.2013HV, 25.4.2013SS+HV).
- Sähköposti ja puhelin:  
Merkitse yhden (tai tarvittaessa useamman) mittarin tietoja koonneen asiantuntijan sähköpostiosoite ja puhelinnumero. Nämä tiedot eivät tule näkyviin TOIMIA-tietokantaan, vaan ne jäävät ainoastaan TOIMIA:n toimituksen tietoon.

### 2. Arvioitava mittari

- Mittarin nimi:  
Merkitse arvioinnin kohteena olevan mittarin suomenkielinen ja englanninkielinen nimi.
- Lyhenne:  
Merkitse mittarin lyhenne, jos se on käytössä.
- Käyttötarkoitus:  
Selitä lyhyesti, mihin käyttötarkoitukseen liittyen mittaria ollaan arvioimassa.
- Mittarin perustiedot:  
Mittarin perustiedot kerätään lomakkeelle 1 "Mittarin perustiedot". Merkitse rastilla, onko perustiedot kerätty eikä niihin tarvitse tehdä muutoksia; onko perustiedot kerätty, mutta niihin tehdään muutoksia; tai perustietoja ei vielä ole kerätty ja ne kerätään tämän mittarin osalta tässä samassa yhteydessä.

### 3. Käytettyjen artikkeleiden tutkimusasetelmien ja –aineistojen kuvaus

- Kuvaa lyhyesti keskeiset tutkimusasetelmaan ja –aineistoon liittyvät tiedot kaikista niistä artikkeleista, joiden pätevyys-, toistettavuus- ja muutosherkkyystuloksia ilmoitetaan seuraavissa kohdissa. Keskeisiä tietoja ovat mm. aineiston koko, tutkittavien keski-ikä (+keskihajonta tai vaihteluväli), sukupuolijakauma, potilasaineistosta keskeiset tiedot sairauteen/sairauksiin liittyen jne. Tarvittaessa kirjaa myös muita tietoja, jotka ovat aiheen kannalta oleellisia.

### 4. Mittarin pätevyys (validiteetti)

- Tarkat ohjeet pätevyystietojen kokoamisesta löydät erillisestä TOIMIA:n oppaasta "Opas toimintakyvyn mittarin arviointiin TOIMIA-verkostossa".
- Kokoa pätevyyttä (validiteettia) koskevat tiedot tarkastelun kohteeksi valitun käyttötarkoituksen näkökulmasta: onko mittari pätevä kyseisessä käyttötarkoituksessa / kontekstissa? Mittaako mittari juuri sitä tutkittavan ilmiön ominaisuutta, mitä sen on tarkoituskin mitata?
- Kirjaa tulokset kyseisen käyttötarkoituksen kannalta oleellisilta pätevyiden osa-alueilta eri väliotikoiden alle lomakkeessa olevan jaottelun mukaisesti. Jos jotain osa-aluetta ei ole tutkittu ja siitä ei ole tietoja, merkitse myös puuttuva tieto. Muista merkitä selvästi mistä artikkelista mikäkin tulos on poimittu (sulkuihin tutkimuksen lähdeviite).
- Kirjoita jokaisen alaotsikon alle loppuun lyhyt parin lauseen yhteenveto kyseisen pätevyiden osa-alueen tuloksista, kun olet ne koennut. Näistä yhteenvetolauseista voit koostaa Soveltuvuusarvio –lomakkeen (lomake 3) tiedot.

### 5. Mittarin toistettavuus (reliabiliteetti)

- Tarkat ohjeet pätevyystietojen kokoamisesta löydät erillisestä TOIMIA:n oppaasta ”Opas toimintakyvyn mittarin arviointiin TOIMIA-verkostossa”.
- Kokoa toistettavuutta (reliabiliteettia) koskevat tiedot tarkastelun kohteeksi valitun käyttötarkoituksen näkökulmasta.
- Kirjaa tulokset kyseisen käyttötarkoituksen kannalta oleellisilta toistettavuuden osa-alueilta eri väliotsikoiden alle lomakkeessa olevan jaottelun mukaisesti. Jos jotain osa-aluetta ei ole tutkittu ja siitä ei ole tietoja, merkitse myös puuttuva tieto. Muista merkitä selvästi mistä artikkelista mikäkin tulos on poimittu (sulkuihin tutkimuksen lähdeviite).
- Kirjoita jokaisen alaotsikon alle loppuun lyhyt parin lauseen yhteenveto kyseisen toistettavuuden osa-alueen tuloksista, kun olet ne koonnut. Näistä yhteenvetolauseista voit koostaa Soveltuvuusarvio –lomakkeen (lomake 3) tiedot.

### 6. Mittarin muutosherkkyys

- Tarkat ohjeet muutosherkkyystietojen kokoamisesta löydät erillisestä TOIMIA:n oppaasta ”Opas toimintakyvyn mittarin arviointiin TOIMIA-verkostossa”.
- Kokoa muutosherkkyttä koskevat tiedot tarkastelun kohteeksi valitun käyttötarkoituksen näkökulmasta.
- Jos muutosherkkyttä ei ole tutkittu ja siitä ei ole tietoja, merkitse puuttuva tieto. Muista merkitä selvästi mistä artikkelista mikäkin tulos on poimittu (sulkuihin tutkimuksen lähdeviite).
- Kokoa myös muutostulosten tulkintaan liittyvät tiedot eli tiedot mahdollisista lattia- ja kattoefektistä, tiedot pienimmästä havaittavasta ja kliinisesti merkittävästä muutoksesta omiin kohtiinsa.
- Kirjoita jokaisen kohdan loppuun lyhyt parin lauseen yhteenveto muutosherkkyden tuloksista, kun olet ne koonnut. Näistä yhteenvetolauseista voit koostaa soveltuvuusarvio –lomakkeen (lomake 3) tiedot.

### 7. Mittarin käyttökelpoisuus

- Tarkat ohjeet käyttökelpoisuustietojen kokoamisesta löydät erillisestä TOIMIA:n oppaasta ”Opas toimintakyvyn mittarin arviointiin TOIMIA-verkostossa”.
- Tiedot käyttökelpoisuudesta voivat perustua sekä kirjallisuuteen että asiantuntijoiden ja mittaria käyttävien henkilöiden kokemuksiin ja arvioihin.
- Käyttökelpoisuuden (feasibility) osalta voidaan tarkastella mm. seuraavia piirteitä: hinta, saatavuus/käyttörajoitukset, saatavuus suomenkielillä, koulutus, välineistö, ympäristövaatimukset, ajankäyttötarve, yksiselitteisyys, tulosten tulkinnan helppous (myös mm. viitearvojen saatavuus), hyväksyttävyyys, turvallisuus, monikäyttöisyys (geneerisyys), levinneisyys, kulttuurista riippumattomuus, jne.

### 8. Lähteet

- Tarkista lähdeviitteiden merkintätavat TOIMIA:n käsikirjasta.
- Alkuperäinen lähdeviite:  
Merkitse tähän lähdeviite, jossa mittari on alun perin ensimmäisen kerran kuvattu (mieluiten vain yksi lähde). Lisäksi voidaan merkitä suomenkielinen relevantti lähde, jossa mittari ja/tai sen ominaisuuksia on ensimmäisen kerran kuvattu.
- Hyödylliset linkit:  
Merkitse esim. mittarin kotisivun tai muiden luotettaviksi arvioitujen sivustojen nettiosoitteet, joista mittarista voi saada hyödyllistä tietoa. Tarkista, että linkit ohjaavat suoraan asianomaiseen kohtaan sivustolla.
- Muut lähdeviitteet:  
Merkitse kaikki käyttämiesi artikkeleiden tarkat viitetiedot.

### 9. Mittarin psykometristen tietojen hyväksyminen

- Merkitse TOIMIA-asiantuntijaryhmän nimi ja päivämäärä milloin mittari on hyväksytty ko. ryhmässä.
- Merkitse tarvittaessa muita huomioitavia asioita mittarin hyväksyntään liittyen ”Lisätietoja” –kohtaan.

## Tarkistuslista 1: Arvioitavien artikkeleiden tutkimusaineistoa ja -asetelmaa kuvaavat tiedot

Kaikista niistä artikkeleista, joista valideetti-, reliabiliteetti- ja muutosherkkyytuloksia kerätään, tulee esittää lyhyesti tutkimusaineistoa ja -asetelmaa kuvaavat tiedot. Ne merkitään ”Mittarin psykometriset tiedot” -lomakkeelle kohtaan ”Käytettyjen artikkeleiden tutkimusasetelmien ja -aineistojen kuvaus”.

### 1. Selvitä ensin itsellesi.

- Mitä psykometriikan osa-aluetta/alueita artikkelissa tutkitaan?
- Mikä on tutkimusasetelma: poikittais- vai pitkittäistutkimus?
- Tutkitaanko mittarista erottelevia, ennustavia ominaisuuksia vai onko kyseessä seurantamittari?

### 2. Kokoa tutkimuksista seuraavia perustietoja.

- Tutkittavat (keskeiset sisäänotto- ja poissulkukriteerit):
  - Tutkittavien määrä
  - Tutkittavien ikä (keskiarvo ja/tai keskihajonta ja/tai vaihteluväli)
  - Diagnoosi/toimintakyvyn ongelmat
  - Sairauden kesto ja/tai vaikeusaste
  - Mistä tutkittavat saatiin (esim. kotihoidosta, sairaalasta, kuntoutuslaitoksesta, vapaaehtoisia)
  - Muita keskeisiä tietoja
- Ketkä tekivät mittaukset:
  - esim. asiakkaat täyttivät lomakkeen itse, terapeutit, hoitajat
- Missä maassa tutkimus on tehty, jos kulttuurilla voidaan olettaa olevan vaikutuksia tuloksiin?
- Muita arvioitavan mittarin kannalta oleellisia tietoja, joilla voi olla merkitystä tutkimuksessa saatuihin psykometrisiin tuloksiin?



## Tarkistuslista 2: Arvioitavien artikkeleiden validiteettitiedot

### 1. Mitä pätevyyden osa-aluetta tai osa-alueita tutkittiin?

### 2. Ilmivaliditeetti

- ”Näyttävätkö” mittarin osiot mittaavan sitä, mitä niiden pitäisi mitata?

### 3. Sisältövaliditeetti

- Kuvaako mittarin sisältö kattavasti mitattavaa konstruktioita?
- Onko mittarin kehittämisessä käytetty riippumatonta asiantuntijapaneelia (itsearviointimittareissa potilaita)?

### 4. Kriteerivaliditeetti:

- Samanaikainen validiteetti
  - Onko tutkimuksessa tarkoituksena validoida yhtäpitäviä/erottelevia mittareita ja/tai seurantamittareita?
  - Onko tutkimuksessa vertailumittarina ns. kultainen standardi?
  - Millaiset ovat kultaisen standardin ja tutkittavan mittarin mitta-asteikot ja mittayksiköt?
  - Miten yhtäpitävät/eroavat vertailun tulokset ovat?
- Ennustevaliditeetti
  - Ennustaako mittarin tulos jotain tulevaa tapahtumaa esim. kognitiivisen toimintakyvyn alenemista, masennusta, tasapainon tai lihasvoiman heikkenemistä?

### 5. Rakennevaliditeetti:

- Rakenteen validiteetti
  - Jos analyysissa on käytetty eksploratiivista (teoria mittarin toimivuudesta ei ole tiedossa etukäteen) tai konfirmatorista faktorianalyysia (teoria siitä, miten mittarin tulisi toimia, on tiedossa etukäteen), niin mittaako mittari moniulotteista asiaa (konstruktioita)?
  - Jos on tehty Rasch-analyysi, kuvastavatko osiot samaa latenttia piirrettä? Millainen on osion toimivuus?
- Hypoteesien testaus
  - Konvergentti validiteetti
    - Verrataanko tutkimuksessa tutkittavaa mittaria toiseen samaa ilmiötä mittaavaan mittariin (mittari ei kuitenkaan ole kultainen standardi)?
    - Onko tutkimuksessa selvitetty käytettävien vertailumittareiden kehittäminen ja niiden konstruktioita?
    - Onko tutkimuksessa asetettu etukäteen hypoteesit tutkittavien ryhmien oletettujen tulosten suunnalle ja suuruudelle?
    - Onko mittareiden tuloksia verrattu korrelaatiokertoimen avulla ja korreloivatko tulokset eli ovatko ne yhtäpitävät (konvergentit)?

- Erotteleva validiteetti
  - Onko tutkittavan mittarin tuloksia verrattu saman vai eri konstruktioita mittaavan mittarin tuloksiin (erotteleeko mittari eri konstruktiot)?
  - Onko tutkimuksessa asetettu etukäteen hypoteesit tutkittavien mittareiden oletettujen tulosten suunnalle ja suuruudelle?
  - Kuvattiinko vertailumittareiden psykometriset ominaisuudet?
  - Onko käytetty korrelaatiokerrointa tilastollisena menetelmänä?
- Kyky erotella ryhmiä
  - Onko mittarin tarkoitus erotella eri sairauden asteita tai eri potilasryhmiä toisistaan?
  - Onko tutkimuksessa mukana vähintään kaksi ryhmää, joiden tuloksien oletetaan eroavan toisistaan hypoteesien perusteella?
  - Asetettiin hypoteesien testauksessa suunnat ja suuruudet etukäteen?
  - Onko ryhmien tulosten eroja tutkittu t-testillä tai varianssianalyysillä?
  - Mitkä olivat tutkittujen ryhmien tulokset ja miten ne erosivat toisistaan?
- Kulttuurien välinen validiteetti
  - Millaisella protokollalla käänösprosessi on toteutettu?

### Tarkistuslista 3: Arvioitavien artikkeleiden reliabiliteettitiedot

- 1. Onko tutkimuksessa tehty vähintään kaksi mittausta (testi-uusinta-testi) samoille tutkittaville?**
  - Oliko molemmilla mittauskerroilla sama (intra-rater) vai eri mittaajat (inter-rater)?
  - Kuinka monta mittauskertaa oli ja millä aikavälillä?
  
- 2. Onko kyseessä relatiivinen toistettavuus eli miten tutkittava säilyttää asemansa tutkimusjoukossa?**
  - Onko jatkuvan muuttujan toistettavuuden tarkastelussa käytetty ICC:tä?
  - Onko luokitteluasteikollisen muuttujan toistettavuuden tarkastelussa käytetty ICC:tä tai Cohenin kappakerrointa?
  - Onko nominaaliasteikollisen muuttujan toistettavuuden tarkastelussa käytetty Cohenin kappakerrointa?
  - Mitkä ovat ICC:n tai kappan tulokset?
  - Ovatko tulosten luottamusvälit ilmoitettu?
  
- 3. Onko kyseessä absoluuttinen toistettavuus eli toistettujen mittausten tulosten vaihtelu?**
  - Onko tutkittava muuttuja
    - jatkuva?
    - luokitteluasteikollinen
    - nominaaliasteikollinen?
  - Onko toistettavuuden laskemiseksi käytetty
    - mittauksen keskivirhettä (SEM)?
    - Bland-Altmanin piirrosta?
    - variaatiokerrointa (CV)?
    - Cronbachin alfaa ( $\alpha$ )?
  - Mitkä ovat em. menetelmien tulokset?
  - Onko tulosten luottamusvälit ilmoitettu?
  
- 4. Onko tutkittu mittarin sisäistä yhteneväisyyttä (internal consistency)?**
  - Muodostuuko tutkittava asteikko useista osioista?
  - Onko ryhmälle tutkittavia tehty yksi mittauskerta?
  - Onko tulokset analysoitu käyttäen Cronbachin alfa ( $\alpha$ )?

## Tarkistuslista 4: Arvioitavien artikkeleiden tiedot muutosherkyydestä ja muutostulosten tulkittavuudesta

1. **Onko tutkimuksessa ollut vähintään kaksi mittauskertaa?**
2. **Onko tutkimuksessa ollut mukana vähintään kaksi eri ryhmää?**
  - Onko toisilla tutkittavilla tapahtunut muutosta ja toisilla muutosta ei ole havaittavissa?
3. **Onko muutosta mitattu samalla kerralla sekä tutkittavalla mittarilla että kultaisella standardilla?**
  - Tässä tapauksessa tulisi käyttää samoja tilastollisia menetelmiä kuin kriteerivaliditeetin arvioinnissa, esim. ROC-käyrä.
4. **Jos kultaista standardia ei ole käytetty, niin onko tutkimuksessa laadittu ennalta testattavat hypoteesit esim. eri ryhmien tai eri mittareiden välisille tuloksille?**
  - Vrt. rakennevaliditeetti.
5. **Onko tutkimuksessa käytetty joitain ei-suositeltavista toistettavuuden tilastollisista menetelmistä/tunnusluvuista?**
  - Jos on, niin muista merkitä ne selvästi erilleen lomakkeelle – ja mieti, tarvitseeko niitä ilmoittaa lainkaan, jos on muita tunnuslukuja riittävästi.
6. **Millainen on saatujen tulosten jakauma (pisteet korkeita vai matalia; tasainen vai epätasainen jakauma mittarin asteikolle; pisteiden kasautuminen jollekin kohdalle jne.)?**
  - Onko mittarissa havaittavissa tulosten suhteen katto- ja/tai lattiaefektiä?
  - Onko olemassa viite- ja/tai raja-arvoja kyseisen konstruktion suhteen?
7. **Onko muutostuloksista laskettu pienin havaittava muutos (SDC tai MDC)?**
  - Onko tutkittu Bland-Altmanin piirroksella?
  - Miltä ajanjaksolta ja miltä ryhmiltä muutostulokset on laskettu?
  - Mikä oli tulos?
8. **Onko muutostuloksista laskettu pienin merkittävä muutos (MIC tai MCID)?**
  - Kuka on määritellyt pienimmän merkittävän muutoksen: potilas itse, asiantuntija, tutkija vai onko se määritelty jonkin muun tekijän suhteen?
  - Onko MIC määritelty ROC-käyrän, kaplan vai korrelaatioiden avulla?
  - Miltä ajanjaksolta ja miltä ryhmiltä muutostulokset on laskettu?
  - Mikä oli tulos?