

Welke vragenlijsten zijn het meest betrouwbaar, valide en responsief bij patiënten met voorste kruisbandletsel?

Door Samirah Fremouw, eindexamenopdracht Fysiotherapie aan de Hogeschool Utrecht.

Samenvatting

Achtergrond: Er wordt tegenwoordig in de fysiotherapie steeds meer gebruik gemaakt van evaluatieve meetinstrumenten. Ook voor knieklachten bestaan er diverse vragenlijsten. Veel gebruikte vragenlijsten bij voorste kruisbandletsel zijn bijvoorbeeld de Lysholm scale, de Tegner activity rating scale, de Cincinnati knee rating system en de International knee documentation committee. Het doel van deze literatuurstudie is om meer duidelijkheid te verkrijgen met betrekking tot de betrouwbaarheid, validiteit en responsiviteit van de meetinstrumenten.

Kwaliteitsbeoordeling studies: Literatuur search bestaande uit onder andere systematic reviews, cohort studies en clinical trials.

Methodiek: Literatuurstudie, waarin gezocht is naar artikelen in verschillende databanken, zoals PEDro, Medline Pubmed, CINAHL en Cochrane Library. Er is gezocht met behulp van onderstaande key-words.

Resultaten: De Lysholm blijkt een voldoende betrouwbaarheid te hebben met een ICC van 0.95 en een SRM van 0.9 en hij heeft een goede correlatie met de SF-36. De Cincinnati heeft een ICC waarde van 0.88 en ook dit instrument heeft een significant positieve relatie met de Lysholm. De responsiviteit is berekend aan de hand van de ES=1.18-3.49 en de SRM=1.07-2.48. Ten slotte blijkt de IKDC een voldoende betrouwbaarheid te hebben met een ICC waarde van 0.96 en hij correleert significant positief met de SF-36 met een waarde van 0.71.

Conclusie: Alle beschreven meetinstrumenten behalve de IKDC blijken betrouwbaar, valide en responsief te zijn bij patiënten met VKB-letsel. Nader onderzoek is aan te bevelen om de betrouwbaarheid, validiteit en responsiviteit van de Lysholm, Tegner, SARS en IKDC in de Nederlandse vertaling te bepalen.

Key-words: anterior cruciate ligament, questionnaire, rating scale, Lysholm knee score, Tegner activity rating scale, Sports activity rating scale (SARS), Cincinnati sports activity scale (CSAS), International Knee Documentation Committee (IKDC), knee, knee injury, reliability, validity, responsiveness.

Inleiding

Kruisbandletsel leidt vaak tot een vroegtijdig einde van een carrière in de sport. Kruisbandletsel, en vooral voorste kruisbandletsel, is vooral een bekend fenomeen bij voetbal. Ruim 43% van alle kruisbandletsels ontstaan tijdens voetbal. Tevens blijkt dat kruisbandletsel vaker voorkomt bij vrouwen dan bij mannen (Joanna Kvist, 2004).

In het kniegewricht bevinden zich twee kruisbanden. Beide hebben als belangrijkste functie te zorgen voor stabiliteit in het gewricht. De voorste kruisband zorgt voor een remming van de translatie van de tibia ten opzichte van het femur naar ventraal, terwijl de achterste kruisband juist een tegenovergestelde functie heeft, namelijk remming van de translatie naar dorsaal. Wanneer er sprake is van een acuut letsel van de voorste kruisband, zorgt dit niet alleen voor instabiliteit van de knie, maar ook voor een vergrote translatie naar anterior en endorotatie. Dit is de richting waarbij het letsel vaak ontstaat en patiënten tevens door hun knie zakken (giving way).

Afhankelijk van de leeftijd, ernst van het letsel, eventueel bijkomend letsel, sportieve en functionele behoeften van de patiënt en de wil om te revalideren, wordt besloten tot een conservatieve of operatieve behandeling. In beide gevallen staat de hulpvraag van de patiënt centraal en wordt er niet alleen gekeken naar activiteiten van het dagelijks leven, maar ook naar functionele activiteiten, zoals werk en sport. Na revalidatie van een voorste kruisband reconstructie zal de belastbaarheid van de knie verminderd blijven, maar het geeft de patiënt toch weer een kans om terug te keren naar zijn sportniveau.

Om een goed beeld te kunnen krijgen van iemands deelname aan sportactiviteiten zijn verschillende meetinstrumenten ontwikkeld. Voorbeelden hiervan zijn de Tegner activity scale en de Cincinnati sports activity scale. Beide meetinstrumenten bevatten onderdelen waarin gekeken wordt naar type sport, frequentie waarmee gesport wordt en type bewegingen veel voorkomend tijdens sportactiviteiten.

Afgelopen decennia zijn er veel verschillende meetinstrumenten ontwikkeld en momenteel wordt hier in de praktijk steeds meer gebruik van gemaakt. Meetinstrumenten hebben als voordeel dat klinische verschijnselen objectief in maat en getal uitgedrukt kunnen worden. Een meetinstrument maakt zichtbaar. Voor zorgverleners betekent dit dat het eigen handelen inzichtelijk gemaakt wordt en vervolgens verbeterd kan worden. Het is niet zo dat een meetinstrument de klinische expertise van een zorgverlener kan vervangen, maar het kan zeker een toegevoegde waarde hebben bij iedere vorm van therapie. Het doel van deze studie is om vijf vragenlijsten op het gebied van knieklachten met elkaar te vergelijken om vervolgens een uitspraak te kunnen doen over de kwaliteit van de meetinstrumenten in relatie tot voorste kruisbandletsel.

Methodiek

Artikelen zijn gezocht in verschillende online databanken, zoals PEDro, Medline Pubmed, CINAHL en Cochrane Library. Er is gebruik gemaakt van de volgende key-words: anterior cruciate ligament, questionnaire, rating scale, Lysholm knee score, Tegner activity rating scale, Sports activity rating scale (SARS), Cincinnati sports activity scale (CSAS), International Knee Documentation Committee (IKDC), knee, knee injury, reliability, validity, responsiveness. De volgende inclusiecriteria zijn gebruikt:

- Artikelen waarin onderzoek wordt beschreven bij patiënten met VKB-letsel, zowel pre- als postoperatief
- Artikelen waarbij onderzoek is gedaan bij adolescenten
- Artikelen waarin gebruik wordt gemaakt van de meest gebruikte vragenlijsten met betrekking tot VKB-letsel
- Vragenlijsten die gevalideerd vertaald zijn naar het Nederlands
- Artikelen die uitsluitend in het Engels of Nederlands zijn geschreven

Dit heeft uiteindelijk geresulteerd in 22 full text artikelen. In deze literatuurstudie zijn 12 artikelen opgenomen.

Begripsomschrijving

Om een uitspraak te kunnen doen over de kwaliteit van meetinstrumenten spelen de volgende begrippen een belangrijke rol, betrouwbaarheid, validiteit en responsiviteit.

Er is sprake van een betrouwbaar meetinstrument wanneer deze reproduceerbaar is. Dit houdt in dat er wordt gekeken bij herhaling van de meting (test-hertest) in hoeverre de twee metingen dezelfde resultaten geven (intrabeoordelaarsbetrouwbaarheid) of in hoeverre de resultaten overeenkomen tussen twee of meerdere beoordelaars (interbeoordelaarsbetrouwbaarheid). Dit kan berekend worden aan de hand van de Pearson-correlatiecoëfficiënt of de Spearman-correlatiecoëfficiënt. Deze is voldoende betrouwbaar bij $r=0.70$ en een significantie van $p=0.05$. Significantie kan worden omschreven als de kans dat het gevonden resultaat al dan niet op toeval berust.

Een nadeel van de Pearson-correlatiecoëfficiënt is dat hij geen rekening houdt met systematische verschillen tussen twee metingen. Een methode die dat wel doet en die de r-waarden van meer dan twee beoordelaars kan vergelijken is die volgens de intraclasscorrelatiecoëfficiënt (ICC). Hierbij geldt dat er een redelijke overeenstemming is wanneer de ICC tenminste een waarde heeft van 0.70.

Validiteit gaat over de mate waarin een meetinstrument daadwerkelijk meet wat het behoort te meten. Er zijn verschillende vormen van validiteit te benoemen. Binnen deze literatuurstudie vormt constructvaliditeit één van de belangrijkste. De beste manier om dit te kunnen bepalen is om hypothesen op te stellen over de verwachte correlaties met andere meetinstrumenten of over verwachte verschillen in scores op het meetinstrument tussen groepen (discriminante analyse).

Responsiviteit (longitudinale validiteit) zegt iets over de mate waarin een meetinstrument bij een patiënt klinisch relevante veranderingen in de tijd kan vaststellen. Er zijn verschillende manieren om dit te kunnen aangeven. De eerste manier wordt hetzelfde bepaald als constructvaliditeit. Een tweede manier is te kijken of het meetinstrument relevante veranderingen kan onderscheiden van de ruis (minimum detectable difference-MDD of smallest detectable change-SDC) of dat het meetinstrument in staat moet zijn minimaal belangrijke veranderingen aan te kunnen tonen (minimal important change-MIC).

Vragenlijsten

Lysholm knee score

De Lysholm knee score is een vragenlijst ontwikkeld voor het in kaart brengen van de mate van knie instabiliteit op zowel stoornis- als op beperkingniveau. De vragenlijst is dus uitermate geschikt voor patiënten met een letsel van met name de voorste kruisband resulterend in knie instabiliteit, ook wel giving way genoemd. De Lysholm knee score is een evaluatief klinisch meetinstrument, bestaande uit acht items. Per item wordt op een ordinale meetschaal door de patiënt zelf het stoornis- of beperkingniveau weergegeven. Items waarop gescoord dient te worden zijn bijvoorbeeld instabiliteit, slotverschijnselen en traplopen. De score range verschilt per item. Zo verlopen de stappen bij het item instabiliteit gelijkmatig met 5 en bij zwelling ongelijkmatig met stappen van 2 en 4. Een hoge score op de Lysholm scorelijst komt overeen met een geringe mate van knie instabiliteit (ofwel een betere kniefunctie). De maximale score is 100 punten. De beoordeling van de scores is als volgt:

Range	Kniefunctie
95-100 punten	Uitstekend
84-94 punten	Goed
65-83 punten	Matig
0-64 punten	Slecht

Tegner activity rating scale

De Tegner activiteitschaal bevat één item waarbij de patiënt zelf de zwaarte van zijn werk- en sportactiviteiten aangeeft. De score varieert van 0 tot 10, waarbij een hoge score overeenkomt met een hoog activiteitsniveau. Indien de score binnen de range van 6-10 valt, worden recreatieve competitieve sporten bedreven door de proefpersoon. Niveau 2-5 omvatten sport- en werkactiviteiten en niveau 1 of lager omvatten alleen werkactiviteiten. De Tegner activiteitschaal kan gebruikt worden ter aanvulling op de Lysholm score. Het geeft een meer realistisch beeld van iemands kniefunctie. Iemand kan bijvoorbeeld hoog scoren op het item pijn, maar de vraag is dan: tijdens welke activiteit scoort iemand dit?

Sports activity rating scale

Deze vragenlijst is net als de Tegner activity rating scale bedoeld om het activiteitsniveau van de patiënt in kaart te brengen. Het verschil met de Tegner-score is dat de SARS rekening houdt met het aantal dagen in de week dat er gesport wordt, type bewegingen die veel voorkomen tijdens de sportactiviteiten en het type sport zelf. Het betreft een ordinale score met een totaal van 100 punten.

De lijst wordt onderverdeeld in vier niveaus:

1. 4-7 dagen per week sporten; 90-100 punten
2. 1-3 dagen per week sporten; 75-85 punten
3. 1-3 dagen per maand sporten; 55-65 punten
4. geen sport mogelijk; 0-40 punten

Volgens Vlasveld et al. (1993) geeft de SARS een gedifferentieerder beeld dan de Tegner activity rating scale, doordat het item 'frequentie' aan de lijst is toegevoegd. Om een goed beeld te kunnen krijgen van het sportactiviteitsniveau voor en na een kniebandoperatie moet er dus rekening gehouden worden met het type sport voor de operatie, de intensiteit waarmee de sport werd beoefend en verandering in sportactiviteit.

Cincinnati knee rating system

De Cincinnati knee rating system werd voor het eerst beschreven door Noyes et al. (1983). De oorspronkelijke versie bestond uit een sportactiviteitschaal bestaande uit de volgende zes items: lopen, traplopen, squatten/hurken, rennen, springen/landen op het aangedane been en draaien. Na een aantal aanpassingen van de oorspronkelijke versie is de bestaande sportactiviteiten schaal verkleind naar vier items, namelijk lopen, traplopen, rennen en springen/draaien. Er is een subjectieve schaal bijgekomen, waarin wordt gevraagd naar bepaalde symptomen, zoals pijn, zwelling en stabiliteit. Ook wordt er gevraagd naar functionele beperkingen tijdens sport en dagelijkse activiteiten. De totaalscore van de activiteitschaal wordt bepaald door alle gescoorde punten bij elkaar op te tellen. De score wordt op de volgende manier geïnterpreteerd: <30 slecht, 30-54 matig, 55-79 goed, >80 uitstekend.

International Knee Documentation Committee

Het IKDC formulier kan gebruikt worden voor het meten van subjectieve symptomen, het in kaart brengen van sportactiviteiten en het beoordelen van de kniefunctie voor en na het knieletsel. De vragenlijst is ontwikkeld voor mensen met een verscheidenheid aan knieletsels, zoals voorste kruisbandletsel, patellofemorale klachten, meniscusletsel en verwondingen aan de onderste extremiteit. Het IKDC formulier wordt gescoord door alle punten bij elkaar op te tellen en dit puntentotaal om te zetten naar een schaal van 0-100 punten. De berekende score wordt geïnterpreteerd als een maat voor functie, waarbij een hogere score een hoger niveau van functioneren en een lager niveau van klachten weergeeft.

Resultaten

Lysholm knee score

In 1985 presenteerden Tegner en Lysholm een vernieuwde versie van de oorspronkelijke Lysholm knee score uit 1982. De intrabeoordelaarsbetrouwbaarheid is berekend aan de hand van Pearson-correlatiecoëfficiënt, deze kwam op een waarde van $r=0.97$. De interbeoordelaarsbetrouwbaarheid kwam respectievelijk op een waarde van $r=0.90$. Er wordt niets vermeld over de significantie van de resultaten.

Uit een recenter onderzoek van Bengtsson et al. (1996) blijkt dat de Lysholm knee score een intrabeoordelaarsbetrouwbaarheid heeft van 0.75 tussen dag 1 en 3, 0.69 tussen dag 1 en 14 en 0.68 tussen dag 3 en 14 met een significantie van $p=0.0001$. De correlaties zijn berekend aan de hand van Kendall's correlatiecoëfficiënt.

Marx et al. (2001) hebben tevens onderzoek gedaan naar de betrouwbaarheid, validiteit en responsiviteit van de Lysholm knee score. Hieruit is gebleken dat de Lysholm een ICC waarde heeft van 0.95. De gemiddelde score at baseline ligt op 84.1 punten en na follow up op 84.0 punten. De limits of agreement voor de Lysholm score ligt tussen de 3.8 en 8.0 punten.

Validiteit is getest door de Lysholm score te vergelijken met drie andere meetinstrumenten en verschillende subschalen van de SF-36. De correlaties zijn berekend aan de hand van de Spearman-correlatiecoëfficiënt. De Lysholm correleert het laagst met de fysieke component van de SF-36 ($r=0.68$) en de Lysholm laat een goede correlatie zien met de Cincinnati knee rating scale ($r=0.70$). Beide correlaties zijn significant met een p-waarde van 0.01.

Responsiviteit wordt in ditzelfde onderzoek uitgedrukt in de standardized response mean. Deze is voldoende betrouwbaar bij een waarde variërend van 0.9-1.9 (Marx et al. 2001).

Voor de Lysholm geldt een waarde van 0.9 en dit betekent dat het instrument gevoelig is voor veranderingen in de tijd.

De Lysholm knee score is ontworpen in het Engels. Neeb et al. (1993) heeft van de Lysholm score een Nederlandse vertaling gemaakt en uit onderzoek van Strik et al. (1998) blijkt dat de Lysholm score een intrabeoordelaarsbetrouwbaarheid heeft met een kappawaarde van 0.68 voor de preoperatieve groep en 0.67 voor de postoperatieve groep.

De Lysholm scoort een interbeoordelaarsbetrouwbaarheid met een kappawaarde variërend van 0.54-0.85 voor de preoperatieve groep en 0.53-0.69 voor de postoperatieve groep. De correlaties berekend volgens de Spearman-correlatiecoëfficiënt zijn voor zowel de intra- als interbeoordelaarsbetrouwbaarheid voldoende betrouwbaar met de volgende waarden: $r=0.87$ en $r=0.85$. Alle correlatiecoëfficiënten zijn significant met een p-waarde van <0.01 .

Tegner activity rating scale

Er is geen onderzoek gevonden waarin zowel de betrouwbaarheid als responsiviteit wordt onderzocht van de Tegner activiteitschaal bij patiënten met voorste kruisbandletsel. Er is wel onderzoek naar gedaan bij patiënten met meniscuspathologie- of in combinatie met intra-articulaire pathologie. Op basis van een studie door Lysholm et al. (1985) is er wel onderzoek gedaan naar content-validiteit. 47 patiënten met een voorste kruisband laesie moesten in een vragenlijst aangeven hoe moeilijk bepaalde activiteiten uit te voeren waren. Aan de hand van de resultaten is de Tegner activiteitschaal ontwikkeld.

In het postoperatieve onderzoek van Vlasveld et al.(1993) blijkt er een licht positieve significante correlatie te bestaan tussen de Lysholm score en de Tegner activity rating scale. Deze uitkomst is berekend aan de hand van Kendall's Tau-B correlatie en de waarde bedraagt 0.29 met een significantie van $p=0.04$.

Volgens een onderzoek van Wülker et al (1992) blijkt er een significante relatie te bestaan tussen de Lysholm score en de Tegner activity rating scale voor patiënten met VKB-laesie preoperatief. Daarom moet de Tegner activity rating scale altijd in combinatie gebruikt worden met de Lysholm score. Patiënten kunnen hoog scoren op de Lysholm schaal, maar dit zegt niets over hun activiteitsniveau.

Sports activity rating scale

In het onderzoek van Noyes et al. (1989) wordt door een groep van 59 patiënten de SARS ingevuld. De patiënten hebben chronische knieklachten of hebben een voorste kruisband reconstructie ondergaan en zijn weer aan het sporten. In het artikel wordt niets vermeld over de betrouwbaarheid van de SARS.

Wel is er onderzoek gedaan naar de relatie tussen de SARS en de Lysholm score door Vlasveld et al. (1993). Hieruit is gebleken dat de correlatie tussen de Lysholm score en de SARS 0.47 bedraagt, met een significantie van $p=0.009$. De correlatie is berekend via Kendall's Tau-B.

Strik et al. (1998) hebben onderzoek gedaan naar de betrouwbaarheid van de Nederlandse vertaling van de SARS. Uit dit onderzoek is gebleken dat de SARS een voldoende betrouwbaarheid heeft met een waarde variërend van 0.44-0.86 voor de preoperatieve groep en variërend van 0.66-0.84 voor de postoperatieve groep met een significantie van $p=0.01$. De correlaties zijn berekend aan de hand van Cohens's kappawaarde.

Cincinnati knee rating system

Uit een onderzoek van Marx et al. (2001) blijkt dat de Cincinnati knee rating system een voldoende betrouwbaarheid heeft met een ICC waarde van 0.88. De gemiddelde score tijdens dit onderzoek op de Cincinnati schaal bedraagt 71.6 punten. Na de follow up bedraagt de score op de Cincinnati 71.1 punten. 95% van de verschillen tussen de twee metingen ligt 7.8 tot 22.5 punten hoger of lager dan de gemiddelde score tussen de twee tests (limits of agreement).

De Cincinnati blijkt volgens ditzelfde onderzoek tevens een voldoende face- en content validiteit te hebben. De construct validiteit is getest door de Cincinnati te testen ten opzichte van vier andere knie-rating scales en de fysieke en mentale component van de SF-36. De laagste correlatie volgens Spearman bedraagt 0.68 met de fysieke component van de SF-36. De correlatie tussen de Cincinnati en de Activities of daily living scale bedraagt 0.80. Alle correlaties zijn significant met een p-waarde van <0.01 .

Om de responsiviteit te bepalen is in dit onderzoek gebruik gemaakt van de standardized response mean. Deze is voldoende betrouwbaar bij een waarde variërend van 0.9-1.9 (Marx et al. 2001). Voor de Cincinnati schaal geldt een waarde van 0.8 (Marx et al. 2001).

Uit een ander onderzoek van Barber-Westin et al. (1999) blijkt dat de Cincinnati knie rating system ook een voldoende betrouwbaarheid heeft met een correlatiecoëfficiënt variërend van 0.71 tot 1.0 in de niet aangedane groep en een correlatiecoëfficiënt variërend van 0.75 tot 0.98 in de patiëntpopulatie. De correlatiecoëfficiënten zijn berekend aan de hand van de intraclasscorrelatiecoëfficiënt.

Volgens hetzelfde onderzoek heeft de Cincinnati ook een voldoende content- en constructvaliditeit. Acht van de negen geformuleerde hypothesen worden in dit onderzoek bevestigd.

De responsiviteit is getest aan de hand van de standardized response mean en de effect size statistics. Het meetinstrument laat zien dat hij erg gevoelig is voor veranderingen in de tijd. Volgens de standardized response mean, laten zeven van de acht items een grote verandering zien variërend van 1.07 tot 2.48. Ook laten zeven items volgens de effect-size statistics grote veranderingen zien variërend van 1.18 tot 3.49. Er is sprake van een grote verandering bij een Cohen waarde van >0.8 .

International Knee Documentation Committee

Irrgang et al. (2001) hebben onderzoek gedaan naar de betrouwbaarheid en validiteit van de Engelstalige IKDC. Hieruit is gebleken dat de test-hertest betrouwbaarheid voldoende is met een ICC waarde van 0.94. Hierbij wordt een 95% betrouwbaarheidsinterval vermeld van 0.88-0.97. De standaard error of measurement bedraagt 4.6 met een 95% betrouwbaarheidsinterval van 9.0 punten. Dit betekent dat een IKDC score groter dan 9 punten een klinische verandering laat zien en dat een score minder dan 9 punten binnen een acceptabele meetfout valt of als ruis wordt gezien.

De validiteit is berekend door verschillende items van de IKDC te vergelijken met zowel de fysieke, sociale en mentale subschalen van de SF-36. Tevens hebben de auteurs van tevoren een hypothese opgesteld, namelijk dat de IKDC een sterkere relatie zou hebben met de fysieke en sociale subschaal van de SF-36 dan met de mentale subschaal. Deze hypothese wordt bevestigd in het onderzoek aan de hand van de Pearson-correlatiecoëfficiënt. De waarden bedragen respectievelijk 0.63, 0.47 en 0.16 voor de mentale subschaal. Er wordt niets vermeld over de significantie van de resultaten.

De responsiviteit is tevens getest door Irrgang et al. (2006). Eén van de manieren waarop ze dit gedaan hebben is die volgens de receiver operating characteristic (ROC)-curve. Het resultaat laat een AUC-waarde van 0.78 zien met een 95% betrouwbaarheidsinterval van 0.72-0.85. Ook is gekeken naar de effect size en de standardized response mean van het meetinstrument. Deze waarden bedragen respectievelijk 1.13 en 0.94.

Daniël Haverkamp et al. (2006) hebben onderzoek gedaan naar de betrouwbaarheid en validiteit van de Nederlandstalige versie van de IKDC. Hieruit is gebleken dat het instrument voldoende betrouwbaar is met een ICC van 0.96 en een p-waarde van <0.001 . Het 95% betrouwbaarheidsinterval bedraagt 0.94-0.97.

De validiteit is net als bij de originele versie getest door de items te vergelijken met de verschillende subschalen van de SF-36. Hieruit blijkt tevens dat de IKDC een sterkere relatie heeft met de fysieke- en sociale subschaal dan de mentale subschaal. De waarden bedragen respectievelijk 0.71, 0.42 en 0.21 met een p-waarde van <0.01 . De correlaties zijn berekend aan de hand van de Pearson-correlatiecoëfficiënt.

Discussie

In deze literatuurstudie is geprobeerd een beter beeld te krijgen van de betrouwbaarheid, validiteit en responsiviteit van vijf meetinstrumenten voor de knie. Er wordt niet over alle begrippen bij iedere vragenlijst iets beschreven. Zo wordt er bijvoorbeeld bij de originele versies van de Tegner activity rating scale en de SARS niets vermeld over de betrouwbaarheid, validiteit of responsiviteit van de meetinstrumenten. De betrouwbaarheid wordt wel beschreven in een onderzoek van Strik et al. (1998) waarin een studie is gedaan naar de Nederlandse vertaling van beide meetinstrumenten.

De betrouwbaarheid wordt bij veel meetinstrumenten berekend aan de hand van de Spearman- en Pearson-correlatiecoëfficiënt. Beide resultaten houden geen rekening met systematische verschillen tussen twee metingen. Dit wil zeggen dat als één beoordelaar consequent twee punten hoger scoort dan een ander de uitkomst 1 zal zijn, terwijl de resultaten van de beoordelaars niet exact overeen komen. De intraclasscorrelatiecoëfficiënt houdt hier wel rekening mee. Wel blijft het belangrijk voor zowel de Pearson-correlatiecoëfficiënt als de ICC dat er rekening gehouden moet worden met de samenstelling van de groep.

Wanneer de betrouwbaarheid wordt uitgedrukt in ICC kan de Standard Error of Measurement (SEM) gemakkelijk berekend worden. Dit is de pure meetfout, zowel systematische als niet-systematische meetfouten, bij herhaalde metingen. Dit begrip speelt niet alleen een belangrijke rol bij betrouwbaarheid (limits of agreement), maar ook bij responsiviteit. Wanneer de verandering groter is dan de meetfout (ruis) kunnen veranderingen in de tijd worden aangetoond. Deze belangrijke begrippen worden alleen genoemd bij de uitkomsten waarin de betrouwbaarheid wordt uitgedrukt volgens de ICC.

Aan de hand van het begrip betrouwbaarheid alleen kan geen goede uitspraak worden gedaan over de methodologische kwaliteit van een meetinstrument. Daarvoor dient ook rekening gehouden te worden met de begrippen validiteit en responsiviteit.

Validiteit is onder te verdelen in vier verschillende vormen. De belangrijkste voor deze literatuurstudie is de constructvaliditeit. In de literatuur wordt niets vermeld over de constructvaliditeit van de Tegner activity rating scale en de SARS. Nader onderzoek over de validiteit van deze instrumenten is daarom geïndiceerd.

Eén van de manieren waarop de responsiviteit berekend kan worden is volgens de manier van Guyatt's responsiveness statistic. Hierbij geldt: "als Guyatt's responsiveness statistic groter is dan 1,96 is de SDC kleiner dan de MIC. Kleine relevante veranderingen kunnen dan onderscheiden worden van meetfouten." In de literatuur ziet men echter veel varianten op deze berekening van de responsiviteitsmaat volgens Guyatt, zoals de effect size (ES) of de standardized response mean (SRM). Deze effect sizes zijn vooral een maat voor de grootte van de verandering in de hele groep en zeggen niet zoveel over de responsiviteit van het meetinstrument. Resultaten van de onderzoeken kunnen niet met elkaar vergeleken worden. De responsiviteit die is vastgesteld geldt alleen voor een specifieke patiëntencategorie en dat zegt niets over de responsiviteit van dat meetinstrument bij een andere patiëntencategorie. Responsiviteit van alle meetinstrumenten behalve de IKDC (Engelse versie) is daarom geïndiceerd.

In veel van de gebruikte literatuur wordt niets vermeld over de kwaliteit van de onderzoeken. Bij nader onderzoek dient hiermee rekening gehouden te worden.

Eén van de meetinstrumenten die er als beste uitkomt is de Cincinnati sports activity scale. De resultaten laten een voldoende betrouwbaarheid en validiteit zien. Dit meetinstrument is niet vertaald naar het Nederlands, maar de Nederlandse versie van de IKDC bevat items van de oorspronkelijke CSAS. Dat is één van de redenen waarom dit meetinstrument toch in deze studie is opgenomen. En misschien is nader onderzoek voor een Nederlandse versie van de CSAS daarom toch geïndiceerd.

Veel onderzoeken werken met een patiëntenpopulatie met verschillende knieklachten, maar de groepen worden daar niet naar verdeeld. Er is niet duidelijk wat de betrouwbaarheid, validiteit en responsiviteit is van de individuele subgroep voor patiënten met voorste kruisbandletsel van de Lysholm (Marx et al. 2001) en de IKDC (Irrgang et al. 2001).

In het onderzoek van Marx et al. (2001) hebben patiënten met VKB-lletsel meegedaan. Voor het berekenen van de betrouwbaarheid wordt niet duidelijk vermeld wat er met de patiënten is gedaan tijdens het invullen van de vragenlijst at baseline en na de follow up. Hebben ze therapie gehad en zo ja, wat voor therapie? Bij het meten van de responsiviteit wordt vermeld dat 24 patiënten conservatief zijn behandeld en 9 een VKB-reconstructie hebben ondergaan. Er wordt niets vermeld over het oefenprogramma dat de conservatieve patiënten gevolgd hebben. Ook wordt er niets beschreven over wat voor reconstructie de mensen hebben ondergaan en hoe het verloop er heeft uitgezien na de operatie. Al deze gegevens kunnen van invloed zijn op de gegeven resultaten van de meetinstrumenten na de follow up.

Conclusie

	Lysholm	Tegner	SARS	CSAS	IKDC
Betrouwbaarheid (Engels)	ICC 0.95	-	-	ICC 0.88	ICC 0.94
Betrouwbaarheid (Nederlands)	Intra r=0.87 Inter r=0.85	-	Intra r=0.50 Inter r=0.86	-	ICC 0.96
Validiteit (Engels)	Lysholm-SF36 r=0.68	Lysholm-Tegner 0.29	Lysholm-SARS 0.47	CSAS-SF36 0.68	IKDC-SF36 0.63
Validiteit (Nederlands)	-	-	-	-	IKDC-SF36 0.71
Responsiviteit (Engels)	SRM 0.9	-	-	ES 1.18-3.49 SRM 1.07-2.48	ES 1.13 SRM 0.94 AUC 0.78
Responsiviteit (Nederlands)	-	-	-	-	-

ICC: intraclasscorrelatiecoëfficiënt, intra- en inter: intra- en interbeoordelaarsbetrouwbaarheid, ES: effect size, SRM: standardized response mean, AUC: area under the (ROC)-curve.

Uit deze literatuurstudie is gebleken dat de Lysholm een sterkere relatie heeft met de SARS dan de Tegner activity rating scale. De IKDC blijkt tevens een betrouwbaar en valide meetinstrument te zijn voor patiënten met een verscheidenheid aan knieklachten. De vraag die blijft bestaan is in hoeverre het resultaat uit het onderzoek van Haverkamp et al. (2006) betrekking heeft op patiënten met VKB-lletsel. Op basis van deze gegevens valt te zeggen dat de Lysholm in combinatie met de SARS de voorkeur verleent boven de Tegner activity rating scale, de CSAS en de IKDC bij therapie van patiënten met voorste kruisbandletsel.

Nader onderzoek naar de betrouwbaarheid, validiteit en responsiviteit is geïndiceerd voor de Tegner en de SARS in de oorspronkelijke versie. Nader onderzoek naar de validiteit en responsiviteit in de Nederlandse versie is geïndiceerd voor de Lysholm score, de Tegner, SARS en IKDC.

Literatuurlijst

1. G. Aufdemkampe, J. van den Berg, D.A.W.M. van der Windt, Hoe vind ik het *zoeken, interpreteren en opzetten van fysiotherapeutisch onderzoek*, Bohn Stafleu van Loghum, tweede herziende druk, Houten/Mechelen (2003)
2. S.D. Barber-Westin and F.R. Noyes, Assessment of sports participation levels following knee injuries, *The American Journal of Sports Medicine*, 1999;28(1):1-10
3. S.D. Barber-Westin, F.R. Noyes, MD, and J.W. McCloskey, Phd., Rigorous statistical reliability, validity and responsiveness testing of the Cincinnati knee rating system in 350 subjects with uninjured, injured, or anterior cruciate ligament-reconstructed knees, *The American Journal of Sports Medicine*, 1999;27:402-16
4. J. Bengtsson, J. Möllborg, S. Werner, A study for testing the sensitivity and reliability of the Lysholm knee scoring scale, *Knee Surgery, Sports Traumatol, Arthroscopy*, 1996;4:27-31
5. S. Beurskens, R. van Peppen, E. Stutterheim, R. Swinkels, H. Wittink, *Meten in de praktijk stappenplan voor het gebruik van meetinstrumenten in de gezondheidszorg*, Bohn Stafleu van Loghum, Houten (2008)
6. S. Bollen, and B.B. Seedhom, A comparison of the Lysholm and Cincinnati knee scoring questionnaires, *The American Journal of Sports Medicine*, 1991;19:189-90
7. A. M. Garrat, S. Brealey and W. J. Gillespie, Patient-assessed health instruments for the knee: a structured review, *Rheumatology* 2004;43:1414-1423
8. D.Haverkamp, I.N. Sierevelt, S.J.M. Breugem, K. Lohuis, L. Blankevoort and N. van Dijk, Translation and validation of the Dutch Version of the International Knee Documentation Committee Subjective Knee form, *The American Journal of Sports Medicine*, 2006;34:1680-1684
9. L.D. Higgings, Marcus K. Taylor, Daniel Park, Neil Ghodadra, Milford Marchant, Ricardo Pietrobon, Chad Cook, Reliability and validity of the International Knee Documentation Committee (IKDC) subjective knee form, *Joint Bone Spine* 74 2007:594-599
10. R. Hrubesch, C. Rangger, M. Reichkendler, R.F. Sailer, W. Gloetzer and G. Eibl, Comparison of score evaluations and instrumented measurement after anterior cruciate ligament reconstruction, *The American Journal of Sports Medicine*, 2000;6(28):850-6
11. Irrgang JJ, Anderson AF, Boland AL, Harner CD, Kurosaka M, Neyret P, et al. Development and validation of the International Knee Documentation Committee subjective knee form, *AM J Sports Med*, 2001;29:600-13
12. J.J. Irrgang, A.F. Anderson, A.L. Boland, C.D. Harner, P. Neyret, J.C. Richmond, K. D. Shelbourne and the International Documentation Committee, Responsiveness of the international Knee Documentation Committee subjective knee form, *The American Journal of Sports Medicine*, 2006;34(10):1567-1563
13. J. Kvist, Rehabilitation following anterior cruciate ligament injury *current recommendations for sports participation*, *The American Journal of Sports Medicine*, 2004;34(4):269-280
14. J. Lysholm, MD., and J. Gillquist, MD., Evaluation of knee ligament surgery results with special emphasis on use of a scoring scale, *The American Journal of Sports Medicine*, 1982;10:150-154
15. J. Lysholm and Y. Tegner, Review; Knee injury rating scales, *Acta Orthopaedica* 2007;78(4):445-453
16. R.G. Marx, M. D., M.Sc, F.R.C.S.C., Knee rating scales, *Arthroscopy: The Journal of Arthroscopic and Related Surgery*, 2003;19(10):1103-1108
17. R.G. Marx, MD, Msc, FRCS(C), E.C. Jones, MD, MA, A.A. Allen, MD, D.W. Althchek, MD, S.J. O'Brien, MD, S.A. Rodeo, MD, R.J. Williams, MD, R.F. Warren, MD, and T.L. Wickiewicz, MD, Reliability, Validity, and Responsiveness of four knee outcome scales for athletic patients, *J Bone Joint Surg Am* 2001;83:1459-69
18. T.B. Neeb, G. Aufdemkampe, J.H.D. Wagener, L. Mastenbroek, Assessing anterior cruciate ligament injuries: the association and differential value of questionnaires, clinical tests, and functional tests, *The journal of orthopaedic and sports physical therapy*, 1997;6(26):324-31
19. T.B. Neeb, M.L. Mastenbroek, J.H.D. Wagener en G. Aufdemkampe, Anamnestic score-formulier voor knie-(in)stabiliteit, *Fysiopraxis*, 1993;2(17):19-21

20. M.A. Risberg, I. Holm, H. Steen, B.D. Beynnon, Sensitivity to changes over time for the IKDC form, the Lysholm score, and the Cincinnati knee score *A prospective study of 120 ACL reconstructed patients with a 2-year follow-up*, Knee Surgery, Sports Traumatol, Arthrosc, 1999;3(7):152-159
21. N. Sernert, J. Kartus, K. Köhler, S. Sterner, J. Larsson, B.I. Eriksson, J. Karlsson, Analysis of subjective, objective and functional examination tests after anterior cruciate ligament reconstruction *A follow-up of 527 patients*, Knee Surg, Sports Tramadol, Arthrosc, 1999;7:160-165
22. N.A. Sgaglione, W. Del Pizzo, J.M. Fox, and M.J. Friedman, Critical analysis of knee ligament rating systems, The American Journal of Sports Medicine, 1995;23:660-667
23. G. Strik, G. Aufdemkampe, T.B. Neeb en M.L. Mastenbroek, Betrouwbaarheid van Nederlandstalige vragenlijsten bij knieklachten op basis van een voorste-kruisbandlaesie, Nederlands tijdschrift voor fysiotherapie, 1998;108(1):15-20
24. Y. Tegner, M.D., and J. Lysholm, M.D., Ph.D., Rating Systems in the evaluation of knee ligament injuries, Clinical Orthopaedics and Related Research, 1985;198:43-49

Websites:

1. www.fysiovandertas.nl Website van sport- en manueel fysiotherapeuten P. van der Tas en J.M. Klomp-Jacobs met vragenlijsten van de knie.
2. www.ecmr.nl/dnn/AlgemeneMeetinstrumenten/tabid/63/Default.aspx Site van de Hogeschool Zuyd met informatie over meetinstrumenten bij knieklachten.
3. www.orthopaedicscore.com/scorepages/cincinnati.html Site van twee orthopedisch chirurgen uit Enfield UK met informatie over de Cincinnati knee score.
4. www.suburbantherapy.com/patients/files/subknee.pdf Site van Amerikaanse privé fysiotherapie praktijk met informatie over de Cincinnati knee score.