



www.DirectieveTherapie.nl

Dank voor uw download

U kunt er natuurlijk uit citeren, graag zelfs, maar dan wel met bronvermelding. U mag dit artikel ook ruimhartig verspreiden mits het niet voor commerciële doeleinden is. In die gevallen pas na onze schriftelijke toestemming.

Opname in bloemlezingen en readers moedigen wij aan, maar wel graag eerst even overleggen.

Alle rechten van de artikelen liggen bij
de Stichting Cognitie en Psychose.

Voor alle vragen:
info@gedachtenuitpluizen.nl

Praktische vergelijking van uitkomstinstrumenten in de geestelijke gezondheidszorg

Rosa Schoen en Jan Derksen

SAMENVATTING Als gevolg van verschillende ontwikkelingen in de (geestelijke) gezondheidszorg, heeft uitkomstonderzoek meer en meer aan terrein gewonnen. De Clinical Global Impression (CGI), Korte Klachtenlijst (KKL), Health of the Nation Outcome Scale (HoNOS), Symptom Checklist-90-R (SCL-90-R), Brief Symptom Inventory (BSI), Outcome Questionnaire (OQ-45) en Outcome Rating Scale (ORS), Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM) en Depression Anxiety Stress Scale (DASS) zijn veelgebruikte instrumenten voor het meten van de ernst van klachten en/of het dagelijks functioneren. De Manchester Short Assessment of Quality of Life (MANSA) en World Health Organisation Quality of Life-Bref (WHOQoL-Bref) worden veel gebruikt voor het evalueren van kwaliteit van leven. In deze literatuurstudie is een vergelijkend overzicht gegeven van diverse instrumenten die geschikt zouden zijn voor gebruik voor Routine Outcome Monitoring in de volwassenenzorg.

Inleiding

Ondanks het grote aanbod van uitkomstinstrumenten is er een gebrek aan studies waar zowel een overzicht van als een vergelijking tussen deze instrumenten wordt gemaakt. Dit leidt ertoe dat de afweging en keuze voor een instrument in de praktijk bemoeilijkt wordt.

In lijn met diverse ontwikkelingen in de (geestelijke) gezondheidszorg, zoals een toename in zorgconsumptie, toenemende commercialisering en medische, technologische en farmacologische kennis (TNO, z.d.), is in het afgelopen decennium de aandacht voor het behandel-effect op kwaliteit van leven alsmede de kosteneffectiviteit van

MSC R.A. SCHOEN, basispsycholoog, werkt als junior onderzoeker bij ProCES, Pro Persona. E-mail: rosaschoen@hotmail.com.

PROF. DR. J.J.L. DERKSEN is klinisch psycholoog-psychotherapeut en werkt in de eerste lijn bij psychotherapiepraktijk Derksen & Klein Herenbrink te Bemmelen. Daarnaast doceert hij aan de Radboud Universiteit te Nijmegen en de Vrije Universiteit van Brussel.

behandelprogramma's toegenomen (Geraets et al., 2006). In de klinische praktijk wordt steeds meer belang gehecht aan het uitvoeren van evidence-based praktijken, waaronder met empirisch bewijs ondersteunde behandelingen en (diagnostische) onderzoeksmethoden (Geraets et al., 2006; Havenaar, Heemskerk, & Bisseling, 2002).

Daarnaast wordt er wereldwijd geprobeerd om de kwaliteit van psychotherapie voor iedere individuele cliënt te garanderen door middel van uitkomstmanagement (Lambert et al., 2003; Spits & Schippers, 2008). Hierbij worden de resultaten van iedere individuele behandeling – zoals klachtenreductie, dagelijks functioneren of de toename in kwaliteit van leven – gemeten met behulp van vragenlijsten of interviews (Gilbody et al., 2003; Lambert et al., 2003; Walburg, 2003; Wennink & Walburg, 2007).

Want al is het duidelijk dat psychotherapie over het algemeen substantieel voordeel oplevert over een breed spectrum van stoornissen en voorkomende problemen (Lambert & Ogles, 2004; Smith, Glass, & Miller, 1980), niet iedere persoon die psychotherapie ontvangt heeft daar baat bij. Dit is vooral problematisch, omdat onderzoek suggereert dat klinici over het algemeen niet in staat zijn om de patiënten die na afloop van de behandeling slechter af zijn, accuraat te identificeren (Hannan et al., 2005; Westen & Morrison, 2001), zelfs niet wanneer het behandelproces regelmatig geëvalueerd wordt (Spelmans et al., 2006). Door gebruik te maken van uitkomstinstrumenten kunnen klinici de voortgang van de cliënt volgen en zijn zij beter in staat om de lopende behandeling indien noodzakelijk aan te passen en zo de uitkomst van psychotherapie te verbeteren (Lambert et al., 2003).

In een overzichtartikel concludeerden Andrews en Peters (1994) dat uitkomstinstrumenten om geschikt te zijn verder moeten voldoen aan de volgende eisen: ze moeten kort en goedkoop zijn, het moeten multidimensionale instrumenten zijn die een minimale training voor afname vergen en ze moeten gemakkelijk te scoren en interpreteren zijn. Lambert en Hawkins (2004) stellen dat een uitkomstinstrument om geschikt te zijn eenduidig en niet te moeilijk geformuleerd moet zijn, dat afname, scoring en interpretatie eenvoudig moeten zijn en niet te veel tijd mogen kosten en dat er handleidingen voor afname, scoring en interpretatie beschikbaar zijn en dat die eenvoudig te verkrijgen moeten zijn. Daarnaast kan het een voordeel zijn als het instrument via de computer kan worden afgenomen, gescoord en geïnterpreteerd. Ook is het van belang dat uitkomstinstrumenten voldoende betrouwbaar, valide en sensitief voor verandering zijn (Andrews & Peters, 1994; Lambert & Hawkins, 2004).

Betrouwbaarheid geeft de mate aan waarin een meting met behulp van het instrument consistente gegevens oplevert. Hierbij kan onder-

scheid gemaakt worden tussen test-hertestbetrouwbaarheid, waarbij de consistentie bij herhaaldelijke afname van het instrument wordt onderzocht, tussen interne consistentie betrouwbaarheid, waarbij de consistentie van de verschillende items van het instrument wordt onderzocht, en tussen inter-beoordelaarsbetrouwbaarheid, waarbij de consistentie in de verkregen gegevens wordt bekeken wanneer hetzelfde instrument door meerdere beoordelaars wordt afgenomen (Gregory, 2007). In navolging van Nunnally en Bernstein (1994) wordt in dit onderzoek een betrouwbaarheidscoëfficiënt van .80 of hoger als goed beschouwd, van .70 tot .80 als voldoende en een betrouwbaarheid die lager is dan .70 wordt gezien als onvoldoende.

De validiteit geeft de mate aan waarin een instrument meet wat het beoogt te meten. Ook validiteit kan worden opgedeeld in verschillende typen. Allereerst is er de criteriumvaliditeit, die ondersteund wordt wanneer het instrument erin slaagt om een uitspraak te doen over een in werkelijkheid bestaande eigenschap (een criteriumvariabele). Criteriumvaliditeit kan worden opgedeeld in predictieve validiteit, die aangeeft in hoeverre een instrument een voorspelling over de te meten criteriumvariabele kan doen, en concurrente validiteit, die aangeeft in hoeverre de score op een instrument overeenkomt met een tegelijk beschikbare criteriumvariabele. Vervolgens geeft de constructvaliditeit aan in hoeverre de items van een test representatief zijn voor alle denkbare items die mogelijk zijn om het beoogde te meten. De constructvaliditeit wordt vaak onderzocht door verschillende instrumenten met elkaar te vergelijken en de correlatie tussen beide te berekenen. Door instrumenten die hetzelfde beogen te meten met elkaar te vergelijken, wordt de convergente validiteit onderzocht. De divergente validiteit wordt onderzocht door het berekenen van de correlatie tussen instrumenten die iets anders zouden meten. Ten slotte bestaan er discriminerende validiteit (die aangeeft in hoeverre het instrument in staat is om een onderscheid te maken tussen verschillende groepen, zoals dat theoretisch te verwachten is), gezichtsvaliditeit (geeft aan in hoeverre het instrument op het eerste gezicht valide lijkt) en inhoudsvaliditeit (waarmee wordt aangegeven in hoeverre het instrument het beoogde (theoretische) construct meet, waarbij ook de volledigheid van het instrument beoordeeld wordt) (Gregory, 2007). Voor de beoordeling van validiteit met een goed, voldoende of onvoldoende zijn geen eenduidige grenzen aan te geven (Evers et al., 2000). In dit artikel wordt zo veel mogelijk de indeling in de verschillende typen validiteit door de oorspronkelijke auteurs van de geraadpleegde onderzoeken aangehouden.

De sensitiviteit voor verandering van een instrument geeft de mate aan waarin een instrument in staat is om veranderingen in bijvoorbeeld de ernst van de klachten, het dagelijks functioneren of de kwali-

teit van leven daadwerkelijk te detecteren. Om gebruikt te kunnen worden voor Routine Outcome Monitoring dient een instrument hieraan te voldoen.

Froyd, Lambert en Froyd (1996) vonden 1430 uitkomstinstrumenten die in 334 uitkomstonderzoeken gebruikt werden. Ondanks het grote aanbod van uitkomstinstrumenten is er een gebrek aan studies waar een overzicht van en een vergelijking tussen deze instrumenten wordt gemaakt. Gevolg daarvan is dat de afweging en keuze voor een instrument in de praktijk bemoeilijkt wordt. In dit artikel wordt een overzicht geboden van diverse uitkomstinstrumenten die geschikt kunnen zijn voor gebruik in de ambulante geestelijke gezondheidszorg voor volwassenen. Deze instrumenten worden stuk voor stuk beschreven en op de hiervoor genoemde punten met elkaar vergeleken.

Een selectie van de uitkomstinstrumenten is gemaakt op basis van een studie van Van Wijngaarden en Kok (2007), waarin is onderzocht welke instrumenten in geestelijke gezondheidszorginstellingen in Nederland gebruikt worden voor uitkomstmanagement en benchmarking. In het huidige onderzoek ligt de focus op instrumenten die geschikt zouden kunnen zijn voor het evalueren van de effectiviteit van de behandeling. Om die reden dienden de geselecteerde instrumenten de ernst van de klachten (generiek), het dagelijks functioneren en/of de kwaliteit van leven te meten. Cliënttevredenheid, dat eveneens regelmatig geëvalueerd wordt, is in dit overzichtsartikel buiten beschouwing gebleven. Ook zijn hier slechts de instrumenten meegenomen die geschikt zijn voor routinematig uitkomstonderzoek en voor individueel gebruik.

De beschreven selectiecriteria leidden tot een voorlopige selectie van twaalf instrumenten. Omdat met de Hopkins Symptom Checklist-25 (Derogatis, Lipman, Rickels, & Uhlenhuth, 1974) die als een van de in de GGZ gebruikte instrumenten wordt genoemd, slechts angst- en depressieklachten worden gemeten, is ervoor gekozen om in plaats hiervan de 57 items tellende versie van de Hopkins Symptom Checklist (Derogatis, 1979) op te nemen in dit onderzoek. Deze versie geeft een score voor psychische en somatische klachten, als ook een algemene score.

Vervolgens werd via Psychinfo, Web of Science en Google.Scholar gezocht naar mogelijk goede Nederlandstalige instrumenten die in het onderzoek van Van Wijngaarden en Kok (2007) niet genoemd werden. In eerste instantie is hier slechts één instrument, namelijk de General Health Questionnaire (Goldberg, 1972) gevonden. Later zijn ook de Outcome Rating Scale (Miller, Duncan, & Johnson, 2000), Clinical Outcomes in Routine Evaluation – Outcome Measure (Evans

et al., 2000) en Depression Anxiety Stress Scale (Lovibond & Lovibond, 1995) toegevoegd. Tevens werd van twee instrumenten die wel genoemd waren en voldoen aan de inclusiecriteria van dit onderzoek – te weten de Quality Of Life (QOL; Delespaul & Van Os, 2001) en 4Kaas (4-DKL) – onvoldoende informatie gevonden. Om die reden zijn deze instrumenten niet in het verdere onderzoek meegenomen. Van de MECCA werd onvoldoende informatie gevonden om te kunnen beoordelen of dit instrument aan de inclusiecriteria van dit onderzoek zou voldoen. Van deze laatste twee instrumenten konden ook de auteurs van de instrumenten niet met zekerheid vastgesteld worden. Hoewel zowel de General Health Questionnaire (GHQ-28) als de Hopkins Symptom Checklist-57 (HSCL-57) aan de inclusiecriteria van dit onderzoek voldeden, is besloten om deze instrumenten niet in het onderzoek mee te nemen, omdat deze instrumenten uit de handel genomen bleken te zijn.

Uiteindelijk leidde deze procedure tot een selectie van elf instrumenten, waarvan negen instrumenten die ernst van de klachten meten, namelijk de Clinical Global Impression (CGI; Guy, 1976), Korte Klachtenlijst (KKL; Lange & Appelo, 2007), Health of the Nation Outcome Scale (HoNOS; Wing, Beevor, Curtis, Park, Hadden & Bruns, 1998), Symptom Checklist-90-R (SCL-90-R; Derogatis, 1975), Brief Symptom Inventory (BSI; Derogatis, 1975), Outcome Questionnaire (OQ-45; Lambert et al., 1996), Outcome Rating Scale (ORS; Miller, Duncan, & Johnson, 2000), Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM; Evans et al., 2000) en Depression Anxiety Stress Scale (DASS; Lovibond & Lovibond, 1995), en twee instrumenten die de kwaliteit van leven meten, namelijk de Manchester Short Assessment of Quality of Life (MANSA; Priebe, Huxley, Knight, & Evans, 1999) en World Health Organisation Quality of Life-Bref (WHOQoL-Bref; The WHOQOL-Group, 1996). In Tabel 1 wordt een overzicht van enkele eigenschappen van deze instrumenten gegeven.

Vergelijking van uitkomstinstrumenten

In dit artikel wordt een overzicht gegeven van diverse eigenschappen van uitkomstinstrumenten, op basis waarvan een onderlinge vergelijking en afweging mogelijk is. Onderstaand zal eerst een beschrijving van de psychometrische eigenschappen van de instrumenten volgen, waarna praktische aspecten besproken zullen worden.

Tabel 1. Overzicht van enkele eigenschappen van de onderzochte uitkomstinstrumenten.

Instrument	Meetpretentie	Aantal items	Subschalen	Beantwoording
CGI	Meet de globale ernst van psychopathologie alsmede de verandering daarin over in de tijd.	2	Geen	Behandelaar: door het geven van een cijfer tussen de 1 en 7. Patiënt: m.b.v. een 5-puntsschaal.
KKL	Meet de mate van psychisch lijden.	14	Geen	Behandelaar (interview): m.b.v. een 5-puntsschaal.
HoNOS	Meet de geestelijke gezondheidstoestand en het sociaal functioneren van psychiatrische patiënten.	12	4: gedragsproblemen, beperkingen, symptomatologie en sociale problemen.	Patiënt: m.b.v. een 5-puntsschaal.
SCL-90-R	Meet recent ervaren lichamelijke en psychische klachten (het psychoneurotische- somatische onwelbevinden of psychoneuroticisme).	90	8: agorafobie, angst, depressie, somatische klachten, insufficiënte van denken en handelen, wantrouwen en interpersoonlijke sensitiviteit, hostiliteit en slaapproblemen. 9: somatische klachten, cognitieve problemen, interpersoonlijke gevoeligheid, depressieve stemming, angst, hostiliteit, fobische angst, paranóide gedachten en psychotisme. Daarnaast 2 aanvullende schalen, voor het aantal aanwezige symptomen (AAS) en voor de ernst van de aanwezige symptomen (E/AS).	Patiënt: m.b.v. een 5-puntsschaal.
BSI	Verkorte versie van de SCL-90-R; zie aldaar.	53	3: symptomen, interpersoonlijk functioneren en functioneren in de sociale rol (werk/opleiding). Geen	Patiënt: m.b.v. een 5-puntsschaal.
OQ-45	Ontworpen voor het meten van progressie in therapie. Meet zowel ernst van de klachten als dagelijks functioneren.	45		Patiënt: m.b.v. een 5-puntsschaal.
ORS	Meet het behandelverloop. Meet het dagelijks functioneren in mindere mate de ernst van de klachten.	4		Patiënt: d.m.v. het plaatsen van een kruisje op een visueel-analoge schaal.

Instrument	Meetpretentie	Aantal items	Subschalen	Beantwoording
CORE-OM	Ontwikkeld om de effectiviteit (behandelresultaat) en efficacy (kosten in tijd en geld t.o.v. de opbrengst) van de behandeling te meten. ⁵ Meet zowel de ernst van de klachten als dagelijks functioneren en kwaliteit van leven.	34	4: subjectief welzijn, problemen/symptomen, dagelijks functioneren en risico's t.o.v. zichzelf en anderen.	Patiënt: m.b.v. een 5-puntsschaal.
DASS	Meet de ernst van psychische klachten (depressie, angst en stress).	42	3: depressie, angst en stress.	Patiënt: m.b.v. 4 antwoordcategorieën.
MANSA	Meet tevredenheid met verschillende levensdomeinen.	16	Geen	Behandelaar (interview): 12 items m.b.v. een 7-puntsschaal; 4 items met ja/nee. Er is ook een zelf-involversie. Patiënt: m.b.v. 5-puntsschaal.
WHOqoL-Bref	Meet zowel de kwaliteit van lichamelijk en geestelijk functioneren.	26	4 domeinen: fysieke gezondheid, psychische gezondheid, sociale relaties en omgeving-smogelijkheden. Daarnaast 2 items voor algemene kwaliteit van leven en algemene tevredenheid met de gezondheid.	Patiënt: m.b.v. 5-puntsschaal.

Psychometrische eigenschappen

Ernst van de klachten en dagelijks functioneren

Clinical Global Impression (CGI)

De CGI is een instrument dat de globale ernst van de klachten en de verandering daarin over de tijd meet (Beekman & De Beurs, 2004; Havenaar, Van Os, & Wiersma, 2004; Van Wijngaarden & Kok, 2007). Het instrument blijkt, ondanks zijn eenvoud, over een goede validiteit te beschikken. Dit blijkt onder andere uit de gevonden hoge correlaties met andere instrumenten. Daarnaast werd een goede inter-beoordelaarsbetrouwbaarheid gevonden (Haro et al., 2003; Kadouri, Corruble, & Falissard, 2007; Havenaar, Van Os, & Wiersma, 2004). Havenaar et al. (2004) wijzen op een matige test-hertestbetrouwbaarheid van de CGI. Dit is problematisch wanneer het instrument gebruikt wordt om veranderingen in de toestand, zoals dat bij uitkomstonderzoek gebeurt, te meten. In hun artikel wordt echter geen waarde genoemd waarop de auteurs deze beoordeling baseren.

Korte Klachtenlijst

De KKL meet de mate van psychisch lijden (Evers, Braak, Frima, & Van Vliet-Mulder, 2008). De psychometrische eigenschappen van het instrument zijn, volgens de auteurs van het instrument, voldoende tot goed (Lange & Appelo, 2007). Allereerst vonden de auteurs een interne consistentie betrouwbaarheid (coëfficiënt Cronbachs alfa) van .79 (voldoende) bij de algemene bevolking en van .85 (goed) bij psychiatrische patiënten. Daarnaast is een test-hertestcorrelatie van .70 (voldoende) gevonden. Bij gebruik voor uitkomstonderzoek wordt een test-hertestcorrectie noodzakelijk geacht (Evers et al., 2008). De auteurs vonden verder een goede divergente validiteit. Patiënten en de algemene bevolking, evenals ambulante en niet-ambulante patiënten bleken met behulp van de KKL van elkaar te kunnen worden onderscheiden. Ook bleken scores van patiënten voor en na de behandeling van elkaar te verschillen. Als indicatie voor de constructvaliditeit werd ten slotte de hoge correlatie van .86 tussen de score van de KKL en de totaalscore van de SCL-90 gevonden. De Cotanbeoordeling (Evers et al., 2008) op dit punt is echter voldoende. De betrouwbaarheid wordt door de Cotan eveneens als voldoende beoordeeld en de criteriumvaliditeit als onvoldoende (te weinig onderzoek op een te kleine groep).

Health of the Nation Outcome Scale (HoNOS)

De HoNOS meet de geestelijke gezondheidstoestand en het sociaal functioneren van psychiatrische patiënten (Mulder et al., 2004a). Uit onderzoek naar de Nederlandse versie van de HoNOS werd gevonden dat het instrument een redelijke tot goede betrouwbaarheid heeft. De gevonden interne consistentie betrouwbaarheid (Cronbachs alfa) voor de totale schaal bedroeg .78, wat voldoende is. De Intraclass Correlation Coëfficiënten (ICC's) werden berekend als een maat voor de interbeoordelaarsovereenstemming. Deze was voor de totaalscore .92 (goed). De ICC's van de subschalen gedragsproblemen, symptomatologie en sociale problemen bleken voldoende (respectievelijk .85, .76 en .86). De ICC van de subschaal beperkingen bleek met .53 onvoldoende te zijn (Mulder et al., 2004a).

De criteriumvaliditeit werd als voldoende beoordeeld omdat patronen van itemscores bij diverse stoornissen naar verwachting verschilden en de HoNOS redelijk correleerde met andere schalen (concurrente validiteit). Daarnaast werd gevonden dat de totaalscores op de HoNOS per intensiteit van de behandeling van elkaar verschilden. Deze bevinding biedt ondersteuning aan de constructvaliditeit van het instrument (Mulder et al., 2004a).

Symptom Checklist-90-R (SCL-90-R)

De SCL-90-R meet lichamelijke en psychisch recent ervaren klachten (Arindell & Ettema, 2005; Arindell & Ettema, z.d.; Evers, Braak, Frima, & Van Vliet-Mulder, 2004a). Uit onderzoek naar de psychometrische eigenschappen van de SCL-90-R is gebleken dat de interne consistentiebetrouwbaarheid (Cronbachs alfa) van het instrument over het algemeen goed is. Bij twee studies van Arindell en Ettema (1986; Arindell & Ettema, 2005) met verschillende steekproeven was de laagst gevonden waarde .73, wat voldoende is. De meeste waarden van alfa waren boven de .80.

Uit Nederlands onderzoek (Arindell, 2001; Arindell & Ettema, 2005; Koeter, Ormel, & Van den Brink, 1987) kwam tevens naar voren dat bij een tweede afname van de SCL-90-R ook zonder interventies een daling in de gemiddelde pathologiescores kan worden verwacht. In contrast hiermee vonden Weijnenborg en Ter Kuile (2000) voor geen van de schalen van de SCL-90 aanwijzingen voor het hertesteffect. Om het eventuele hertesteffect bij uitkomstonderzoek te vermijden wordt een correctie hiervoor desondanks noodzakelijk geacht (Hafkenscheid, 1993; Koeter et al., 1987).

Als ondersteuning van de discriminerende validiteit kwam naar voren dat de angstschaal en depressieschaal een verschillend patroon van correlaties met andere variabelen vertonen. Daarnaast bleek dat

personen met een angststoornis significant hoger scoorden op de angstschaal van de SCL-90-R. Op de depressieschaal van deze klachtenlijst bleken personen met een depressie tevens significant hoger te scoren. Dit ondersteunt de concurrente validiteit van het instrument. De auteurs vermelden dat op basis van correlatiepatronen tussen de SCL-90-R en psychiatrische interviews die symptomen of sociale disfuncties meten, geconcludeerd kan worden dat ook de concurrente validiteit van de overige subschalen werd ondersteund (Arindell & Ettema, 2005).

Eveneens werden diverse aanwijzingen voor de predictieve validiteit gevonden. Zo bleken de uitkomsten op de SCL-90-R de uitkomsten van interventies en therapie, werkhervatting na ziekte en de intensiteit van rouwklachten na verlies van een kind te kunnen voorspellen (Arindell & Ettema, 1986, 2003).

De SCL-90-R bleek verder goed te differentiëren tussen de gewone bevolking en personen met psychopathologie, tussen psychiatrische patiënten met verschillende psychopathologie, tussen de gewone bevolking en somatische groepen met relatief veel psychopathologie en tussen somatische groepen die verschillen in de ernst van niet-psychische symptomatologie of type psychopathologie (Arindell & Ettema, 1986, 2003). Deze bevindingen gelden als ondersteuning voor de criteriumvaliditeit van het instrument.

Ten slotte was de samenhang met andere variabelen en met scores op andere instrumenten volgens verwachting. Geconcludeerd werd dat ook de convergente en divergente validiteit ondersteund werden en dat de gevonden resultaten in overeenstemming waren met eerder onderzoek (Arindell & Ettema, 1986, 2003).

De Cotan-beoordeling (Evers et al., 2004a) voor de psychometrische eigenschappen van de SCL-90-R is als volgt: de betrouwbaarheid is goed (de betrouwbaarheid van de schaal hostiliteit is voldoende); de begripsvaliditeit is goed; de criteriumvaliditeit is goed.

Brief Symptom Inventory (BSI)

De BSI meet lichamelijk en psychisch recent ervaren klachten (Rivierduinen & LUMC, 2008; Van Wijngaarden & Kok, 2007). In de handleiding van de BSI beschrijft De Beurs de resultaten van psychometrisch onderzoek uitgevoerd op de Nederlandse bewerking van de BSI. De test-hertestbetrouwbaarheid varieert van 0.71 tot 0.89 voor de verschillende dimensiescores en is daarmee voldoende stabiel te noemen (De Beurs, 2006). Van de negen schalen hebben er zeven een interne consistentie betrouwbaarheid (Cronbachs alfa) van 0.80 of hoger, wat goed is. Alleen de schaal psychoticisme (alfa = .71) en de schaal paranoïde gedachten (alfa = .79) scoren hier lager, maar toch

voldoende. De betrouwbaarheid van de totaalscore is 0.96, wat uitstekend genoemd kan worden.

Daarnaast werd een factoranalyse uitgevoerd om te onderzoeken of de Amerikaanse structuur van de BSI goed behouden was gebleven in de Nederlandse bewerking van het instrument. De gevonden factorstructuur kwam over het algemeen goed overeen met de verwachte dimensies. Er werd echter wel een aantal afwijkingen gevonden. Een belangrijke afwijking was dat van de vijf items uit de psychoticismeschaal er slechts twee op deze factor bleken te laden, wat erop wijst dat de overige items iets anders meten dan beoogd (De Beurs, 2006). Deze bevinding doet afbreuk aan de constructvaliditeit. Daarnaast werd wel een grote samenhang met de SCL-90, met correlaties variërend van 0.70 tot 0.80, gevonden (De Beurs & Zitman, 2006). De correlaties tussen de totaalscores van beide tests bedraagt .82. Deze laatste bevindingen zijn, in tegenstelling tot de eerdere, een indicatie voor een goede constructvaliditeit.

Aan de hand van de samenhang tussen de BSI met andere (sub)schalen werd de convergente validiteit onderzocht. Welke psychopathologie-instrumenten naast de BSI werden afgenomen, was afhankelijk van de DSM-IV diagnose(s) waaraan de patiënt voldeed. De relevante BSI-subschaalscores bleken naar verwachting te correleren met deze instrumenten, wat steun biedt aan de convergente validiteit.

Ook voor de criteriumvaliditeit van de BSI werd ondersteuning gevonden. Zo werden verschillen in scores gevonden tussen de groep patiënten en de algemene bevolking, en tussen patiënten met verschillende vormen van psychopathologie (De Beurs, 2006). In ander onderzoek werd eveneens gevonden dat de subschalen angst, fobie en somatische klachten een goed onderscheid maken tussen patiënten met en zonder angststoornis. Een hoge score op interpersoonlijke gevoeligheid of op de depressieschaal bleek te kunnen duiden op respectievelijk een sociale fobie of een depressie (Rivierduinen & LUMC, 2008).

De Beurs en Zitman (2006) concluderen in hun onderzoek dat de Nederlandse versie van de BSI een volwaardige vervanger is van de SCL-90. De Cotan-beoordeling van de psychometrische eigenschappen van de BSI (Evers, Braak, Frima, & Van Vliet-Mulder, 2007) is echter minder positief dan van de SCL-90-R. De betrouwbaarheid is voldoende; de begripsvaliditeit is onvoldoende (oorspronkelijke Amerikaanse subschaalstructuur wordt gehandhaafd, hoewel geen bevestiging van deze structuur is gevonden); de criteriumvaliditeit is onvoldoende (matige sensitiviteit en specificiteit, geen kruisvalidatie).

Outcome Questionnaire (OQ-45)

De OQ-45 is sinds enige jaren beschikbaar in een Nederlandse vertaling en meet naast de ernst van de klachten ook het dagelijks functioneren. Uit onderzoek naar de psychometrische kwaliteiten blijkt dat betrouwbaarheid en validiteit van deze Nederlandstalige versie voldoende tot goed zijn, en in grote lijnen vergelijkbaar zijn met de originele versie (De Jong & Nugter, 2004; De Beurs, Den Hollander-Gijsman, Buwalda, Trijburg, & Zitman, 2005).

De Beurs et al. (2005) vonden dat de interne consistentie (Cronbachs alfa) van de totaalscore (.95) en van de subschalen interpersoonlijk functioneren (.82) en symptomen (.82) goed was. De interne consistentiebetrouwbaarheid van de subschaal sociale rol was echter onvoldoende (.68). Daarnaast werden voldoende hoge test-hertestcorrelaties gevonden. Correlaties tussen (subschalen van) de OQ-45 en (subschalen van) overige instrumenten bleken naar verwachting, wat steun biedt voor de convergente validiteit. De correlaties met de GAF waren echter laag, wat weer wat afbreuk doet aan de convergente validiteit. Ten slotte werden grote verschillen gevonden tussen de psychiatrische patiënten en de algemene bevolking, wat een suggestie is voor de discriminerende validiteit.

De Beurs et al. (2005) concluderen dat uit het onderzoek goede psychometrische eigenschappen blijken: de betrouwbaarheid van de OQ-45 is voldoende en de validiteit is goed, met name van de totaalscore en de symptoomschaal (De Beurs et al., 2005). De Jong et al. (2007) vonden soortgelijke resultaten.

De COTAN-beoordeling van de OQ-45 is als volgt: de betrouwbaarheid is voldoende (de betrouwbaarheid van de schaal Sociale Rol is 'onvoldoende'); de constructvaliditeit is onvoldoende (constructvaliditeit is niet voor alle schalen aangetoond, over het wel verrichte onderzoek wordt te weinig informatie verschaft en het is te beperkt); de criteriumvaliditeit is onvoldoende (voor de doelstelling het meten en controleren van behandelvoortgang wordt geen onderzoek gerapporteerd) (Evers, Braak, Frima, & Van Vliet-Mulder, 2009a).

Outcome Rating Scale (ORS)

De ORS is speciaal ontworpen voor het evalueren van het verloop van de behandeling. Het instrument meet het dagelijks functioneren en in mindere mate de ernst van de klachten (Evers, Braak, Frima, & Van Vliet-Mulder, 2009b; Expertgroep Volwassenen Kortdurend, 2011; Hafkenscheid, 2008). Over de psychometrische eigenschappen van de Nederlandstalige versie van de ORS bleek vrijwel geen onderzoek beschikbaar te zijn. Van Beljouw en Verhaak (2010) hebben vanwege het ontbreken van onderzoek een kleinschalig onderzoek

onder eerstelijnspsychologen gedaan. Hierbij werden de itemscores en de totaalscore van de ORS vergeleken met de subschaalscores en totaalscore van de SCL-90. Hier werden hoofdzakelijk lage (.20 tot .40) tot matige (.40 tot .70) correlaties gevonden. De correlaties bleken hoger tussen (sub)schalen en items waarbij dat verwacht werd, dan bij schalen waarbij een lagere correlatie verwacht werd. De correlatie tussen de totaalscores van beide instrumenten bedroeg .63. Deze bevindingen wijzen op een matige steun voor de constructvaliditeit van de ORS. In hetzelfde onderzoek werd wel steun gevonden voor de interne consistentiebetrouwbaarheid van het instrument. De gevonden waarde van coëfficiënt alfa bedroeg .82 (goed).

Ook in internationaal onderzoek werden matige steun voor de constructvaliditeit en een goede interne consistentie betrouwbaarheid gevonden (Bringhurst, Watson, Miller, & Duncan, 2006; Miller, Duncan, Brown, Sparks, & Claud, 2003).

De Cotan-beoordeling (Evers et al., 2009b) van de psychometrische eigenschappen van de ORS, maar ook van het instrument in zijn geheel, zijn ronduit negatief. Op elk punt scoort de ORS onvoldoende. De negatieve beoordeling van de begrips- en criteriumvaliditeit wordt veroorzaakt doordat hier geen of te weinig onderzoek naar is verricht. De betrouwbaarheid wordt met een onvoldoende beoordeeld vanwege de kleine steekproeven en onduidelijkheid over de wijze waarop de auteurs rekening hebben gehouden met de invloed van het gebruikte dataverzamelingsdesign op de hoogte van de gevonden waarden (Evers et al., 2009b).

Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM)

De CORE-OM is ontwikkeld om de efficiëntie (behandelresultaat) en de efficacy (kosten in tijd en geld ten opzichte van de opbrengst) van de behandeling te meten (Van Beljouw en Verhaak, 2010). Het instrument meet zowel de ernst van de klachten als het dagelijks functioneren en de kwaliteit van leven. Nederlands onderzoek naar de CORE-OM kon niet worden gevonden.

In buitenlands onderzoek naar de psychometrische eigenschappen van het instrument komt een goede interne consistentiebetrouwbaarheid naar voren. De gevonden waarde van Cronbachs alfa bedraagt .94 voor de totaalscore en varieert van .75 tot .90 voor de subschalen (Barkham et al., 2001; Evans et al., 2002). Tevens werd er een goede test-hertestbetrouwbaarheid gevonden voor de subschalen en de totaalscore, variërend van .87 tot .91. De test-hertestbetrouwbaarheid van de subschaal risico's bleek echter met een correlatie van .64 onvoldoende te zijn (Evans et al., 2002). De con-

structvaliditeit werd eveneens ondersteund. Andere instrumenten die de ernst van de klachten of het functioneren meten, bleken naar verwachting met de CORE-OM te correleren (Barkham et al., 2001; Evans et al., 2002). Ten slotte bleek het instrument in staat te differentiëren tussen een patiëntenpopulatie en een gezonde groep mensen (Evans et al., 2002).

Depression Anxiety Stress Scale (DASS)

De DASS meet, zoals de naam al zegt, de mate van depressie, angst en stress (Evers, Braak, Frima, & Van Vliet-Mulder, 2004c). Uit Nederlands onderzoek blijkt dat de DASS over een goede interne consistentie betrouwbaarheid beschikt: de gevonden waarden van alfa bedroegen .95 voor de depressieschaal, .90 voor de angstschaal en .93 voor de stressschaal. Ook bleek de test-hertestbetrouwbaarheid voldoende tot goed te zijn. De gevonden test-hertestcorrelaties bedroegen respectievelijk .75, .89 en .79 (De Beurs, Van Dyck, Marquenie, Lange, & Blonk, 2001).

In dezelfde studie werd onderzoek gedaan naar de constructvaliditeit van de DASS. Het driefactorenmodel van het instrument werd goed teruggevonden. Echter, ook de samenhang tussen de verschillende subschalen van het instrument bleek hoog te zijn, wat op overlap tussen de constructen wijst. Samenhang met andere angst- en depressieschalen bij afname bij een groep studenten wijst hier eveneens op. Bij afname van de DASS bij een groep psychiatrische patiënten werden voornamelijk matige correlaties met andere angstschalen gevonden. In de meeste gevallen was de correlatie met de angstschaal van de DASS en deze overige schalen het hoogst. Het verschil met de overige DASS-subschalen was echter beperkt (De Beurs et al., 2001).

De Cotan-beoordeling (Evers et al., 2004c) van de psychometrische eigenschappen van de DASS is als volgt: de betrouwbaarheid is goed; de constructvaliditeit is onvoldoende (te weinig onderzoek); de criteriumvaliditeit is onvoldoende (geen onderzoek).

Kwaliteit van leven

Manchester Short Assessment of Quality of Life (MANSA; interviewversie)

Hoewel de MANSA veel wordt gebruikt, is hij nog niet goed op zijn psychometrische kwaliteiten onderzocht. Publicaties over de psychometrische eigenschappen van de Nederlandse versie konden niet

worden gevonden. Van verschillende buitenlandse versies is wel onderzoek beschikbaar (Van Wijngaarden & Kok, 2007).

Priebe et al. (1999) vonden bij de Engelstalige versie van de MANSA een interne consistentiebetrouwbaarheid (Cronbachs alfa) van .74, wat voldoende is. Daarnaast vonden zij een grote correlatie tussen de vragen naar de subjectieve kwaliteit van leven van de MANSA met de (identieke) vragen van de Lancashire Quality of Life Profile (LQoLP), waar de MANSA een verkorte versie van is. Alle correlaties waren .83 of hoger. Op basis van de bevindingen in dit onderzoek concludeerden de onderzoekers dat de concurrente validiteit, constructvaliditeit en gezichtsvaliditeit van de MANSA alledrie hoog zijn. Om die reden wordt de MANSA door de auteurs een geschikt instrument bevonden voor het verkrijgen van inzicht in de subjectieve kwaliteit van leven van de respondent.

Björkman en Svensson (2005) onderzochten de Zweedse versie van de MANSA. Zij vonden een interne consistentiebetrouwbaarheid (Cronbachs alfa) van 0.81. Daarnaast werden positieve correlaties gevonden met de aanwezigheid van een sociaal netwerk, empowerment en sociaal functioneren. Negatieve correlaties werden gevonden met psychopathologie, zorgbehoefte, ervaren discriminatie en afwijzing. Deze gevonden positieve en negatieve correlaties bieden steun voor de criteriumvaliditeit van de MANSA.

*World Health Organisation Quality of Life-Bref
(WHOQoL-Bref)*

Uit diverse onderzoeken is gebleken dat de WHOQoL-Bref een voldoende valide en betrouwbaar instrument is (Van Wijngaarden & Kok, 2007). Zo is gevonden dat domeinscores van de WHOQoL-Bref in hoge mate correleren met de domeinscores van de WHOQoL-100 (Trompenaars, Masthoff, Van Heck, Hodiament, & De Vries, 2005).

Trompenaars et al. (2005) onderzochten de Nederlandse versie van de WHOQoL-Bref. Net als in eerder onderzoek vonden zij dat de items goed over de factoren verdeeld waren en dat de items samen de vier factoren fysieke gezondheid, psychisch functioneren, sociale relaties en omgevingsmogelijkheden vormden. De interne consistentiebetrouwbaarheid (Cronbachs alfa) bleek voor sociale relaties onvoldoende (.66), voor psychisch functioneren en omgevingsmogelijkheden voldoende (respectievelijk .74 en .73) en voor fysieke gezondheid goed (.80). Daarnaast werden negatieve correlaties gevonden met de klachtenlijst SCL-90-R en positieve correlaties met de Perceived Social Support Scale, waarmee de algemene perceptie van sociale steun wordt onderzocht. Deze gevonden correlaties bieden steun voor de constructvaliditeit van de WHOQoL-Bref. Op basis van hun

bevindingen concludeerden de onderzoekers dat de WHOQoL-Bref een adequaat instrument is voor het meten van kwaliteit van leven (Trompenaars et al., 2005).

Vergelijking van de psychometrische eigenschappen

Betrouwbaarheid

In tabel 2 is een overzicht gegeven van de betrouwbaarheden van de onderzochte instrumenten.

Op basis van de gegevens in tabel 2 kan gesteld worden dat van de instrumenten die ernst van de klachten meten de SCL-90-R, BSI en de DASS als goed beoordeeld kunnen worden wat betreft hun betrouwbaarheid. Voor de SCL-90-R dient bij uitkomstonderzoek echter een hertesteffectcorrectie te worden toegepast. De OQ-45, KKL en HoNOS worden als voldoende beoordeeld. Ook voor de KKL dient een hertesteffectcorrectie te worden uitgevoerd, als het instrument voor uitkomstonderzoek gebruikt wordt. De betrouwbaarheid van de CGI is over het algemeen voldoende te noemen. Problematisch voor uitkomstonderzoek is echter de matige test-hertestbetrouwbaarheid van dit instrument. Een kanttekening hierbij is wel dat de auteurs geen waarde noemen en geen toelichting bij de beoordeling van de betrouwbaarheid van de CGI geven. De bevindingen met betrekking tot de betrouwbaarheid van de ORS lijken te wijzen op een goede betrouwbaarheid. Wel dient de onvoldoende beoordeling door het Cotan (Evers et al., 2009b) in ogenschouw te worden genomen. Buitenlands onderzoek naar de CORE-OM wijst ook op een voldoende betrouwbaarheid van dit instrument. Nederlands onderzoek hiernaar ontbreekt nog.

Voor de Nederlandse versie van de MANSA zijn nog geen gegevens bekend. Bevindingen in buitenlands onderzoek zijn echter gunstig. De betrouwbaarheid van de subschalen van de WHOQoL-Bref is voldoende.

Voor de MANSA, WHOQoL-Bref, OQ-45, CGI, KKL, HoNOS, DASS, CORE-OM en ORS geldt ten slotte dat meer onderzoek wenselijk is voordat definitieve conclusies kunnen worden getrokken.

Validiteit

In tabel 3, 4 en 5 staan overzichten van de gevonden verschillende vormen van validiteit van de instrumenten.

Tabel 2. Betrouwbaarheden.

Instrument	Interbeoordelaarsbetrouwbaarheid	Test-hertestbetrouwbaarheid	Interne consistentiebetrouwbaarheid (alfa)	Cotian-beoordeling
CGI	Goed (geen waarde genoemd). n.v.t.	Matig (geen waarde genoemd). .70	n.v.t. Variërend van .79 tot .85.	– Voldoende.
KKL		Correctie voor hertesteffect noodzakelijk bij uitkomstonderzoek.		
HoNOS	ICC's: .92; .85; .79; .86. .53.	–	.78.	–
SCL-90-R	n.v.t.	Correctie voor hertesteffect noodzakelijk bij uitkomstonderzoek.	Laagst gevonden waarde voor subschaal .73.	Goed (subschaal hostiliteit voldoende).
BSI	n.v.t.	Voldoende stabiel. Varieert van .71 tot .89 voor de verschillende dimensiescores.	Meeste waarden boven .8. Totaalscore: .96. Laagste waarde .71. Meeste subschalen > .80.	Voldoende.
OQ-45	n.v.t.	Voldoende hoog (geen waarden genoemd).	Totaalscore: .95. Subschalen variërend van .68 tot .82. 82 ⁵	Voldoende (subschaal sociale rol onvoldoende). Onvoldoende. ¹⁹
ORS	n.v.t.	–	(Buitenlands onderzoek): .64 (subschaal) risico's; verder variërend van .87 tot .91.	–
CORE-OM	n.v.t.	Voldoende tot goed: variërend van .75 tot .89.	Goed: variërend van .75 tot .94. Goed: variërend van .90 tot .95. (Buitenlands onderzoek): .74 en .81.	Goed. –
DASS	n.v.t.	–	–	–
MANSA	n.v.t.	–	–	–
WHOQoL-Bref	n.v.t.	–	Variërend van .66 tot .80.	–

Tabel 3. Criteriumvaliditeit, ook uitgesplitst in predictieve validiteit en concurrente validiteit.

Instrument	Criteriumvaliditeit	Predictieve validiteit	Concurrente validiteit
CGI	—	—	—
KKL	Cotan: onvoldoende.	—	—
HoNOS	Patronen van itemscores verschilden bij diverse stoornissen naar verwachting.	—	Redelijke correlaties gevonden met andere schalen.
SCL-90-R	De SCL-90 differentieert goed tussen diverse groepen. Cotan: goed.	Scores op de SCL-90 kunnen diverse gezondheidsgerelateerde uitkomsten voorspellen.	Patronen bij een angststoornis of depressie naar verwachting. Correlatiepatronen tussen de SCL-90 en psychiatrische interviews die symptomen of sociale disfuncties meten naar verwachting.
BSI	Verschillen in scores werden gevonden tussen diverse patiëntengroepen en de algemene bevolking. Scores op de subschalen kunnen wijzen op en maken onderscheid tussen diverse vormen van psychopathologie. Cotan: onvoldoende.	—	—
OQ-45	Cotan: onvoldoende.	—	—
ORS	Cotan: onvoldoende.	—	—
CORE-OM	—	—	—
DASS	Cotan: onvoldoende.	—	—
MANSA	(Buitenlands onderzoek.)	—	(Buitenlands onderzoek.)
WHOQoL-Bref	Correlaties met criteriumgegevens naar verwachting.	—	Goed.

Tabel 4. Constructvaliditeit, tevens uitgesplitst in convergente validiteit en divergente validiteit.

Instrument	Criteriumvaliditeit	Predictieve validiteit	Concurrerende validiteit
CGI	Hoge correlaties met andere instrumenten gevonden (geen waarden genoemd).	—	—
KKL	Correlatie met de SCL-90: .86. Cotan: voldoende.	—	Goed: diverse groepen kunnen van elkaar worden gescheiden. Ook scores van patiënten voor en na behandeling verschillen van elkaar.
HoNOS SCL-90-R	Totaalscores verschilden per behandelintensiteit. Samenhang met andere variabelen en met scores op andere instrumenten was volgens verwachting. Cotan: goed.	— Ondersteund.	— Ondersteund.
BSI	Correlaties met de (subschaal)score(s) van de SCL-90 tussen de .72 en .82. Cotan: onvoldoende.	—	—
OQ-45	Cotan: onvoldoende.	—	Correlaties met andere instrumenten waren naar verwachting.
ORS	Matig: lage tot matige correlaties met de (subschaal)score(s) van de SCL-90. Cotan: onvoldoende.	—	—
CORE-OM	(Buitenlands onderzoek:) correlaties met andere instrumenten bleken naar verwachting.	—	—

Instrument	Criteriumvaliditeit	Predictieve validiteit	Concurrente validiteit
DASS	In zeer kleine studie werd de factorstructuur goed teruggevonden. Hoge correlaties tussen verschillende subschalen. Ook patronen in samenhang met andere angst- en depressieschalen wijzen op overlap tussen de constructen angst, depressie en stress van de DASS. Cotan: onvoldoende.	–	–
MANSA	(Buitenlands onderzoek.) goed: correlaties van .83 en hoger met items van de L-QoLP.	–	–
WHOQoL-Bref	Correlaties met andere vragenlijsten zijn naar verwachting.	–	–

Tabel 5. Discriminerende validiteit, gezichtsvaardigheid en inhoudsvaardigheid.

Instrument	Discriminerende validiteit	Gezichtsvaardigheid	Inhoudsvaardigheid
CGI	-	-	-
KKL	-	-	-
HoNOS	-	-	-
SCL-90-R	De angst- en depressieschaal vertonen een verschillend patroon van correlaties met andere variabelen.	-	-
BSI	-	-	-
OQ-45	Grote verschillen tussen algemene bevolking en psychiatrische patiënten.	-	-
ORS	-	-	-
CORE-OM	In staat om te differentiëren tussen een groep gezonde personen en psychiatrische patiënten.	-	-
DASS	-	-	-
MANSA	-	(Buitenlands onderzoek: Goed.)	-
WHOQoL-Bref	-	-	-

Uit de gegevens uit tabellen 3, 4 en 5 kan worden afgeleid dat de SCL-90-R een goede validiteit heeft en in dat opzicht het beste instrument lijkt te zijn voor het meten van ernst van de klachten. De tot nu toe in Nederlands onderzoek gevonden resultaten voor de validiteiten van de HoNOS, OQ-45, en WHOQoL-Bref lijken gunstig. Er is echter meer onderzoek noodzakelijk voordat definitieve conclusies getrokken kunnen worden. De MANSa en CORE-OM lijken, op basis van buitenlands onderzoek, eveneens veelbelovend. Onderzoek naar de Nederlandstalige versie van deze instrumenten zal moeten uitwijzen of dit ook voor deze versies geldt.

De KKL is voldoende te noemen. Ook voor dit instrument geldt echter dat meer onderzoek noodzakelijk is. De validiteit van de BSI en de ORS kunnen als matig beoordeeld worden. Over de CGI en de DASS is te weinig informatie beschikbaar voor een beoordeling.

Andere aspecten voor de beoordeling van de kwaliteit van instrumenten

In tabel 6 is een overzicht gegeven van de overige aspecten die voor de beoordeling van een uitkomstinstrument van belang worden geacht. Hier volgt een puntsgewijze beschrijving.

Lengte en afnameduur

Voor het meten van ernst van de klachten is de CGI het kortst gebleken. Een nadeel van dit instrument is echter dat het slechts een algemene score van het psychisch onwelbevinden geeft. Ook de ORS is een zeer kort instrument. Dit instrument geeft een vrij algemene score voor (diverse terreinen van) het dagelijks functioneren. Wat betreft de lengte en afnameduur is de KKL eveneens een zeer geschikt instrument voor het meten van de aard en de ernst van de klachten.

De MANSa is de kortste lijst voor het meten van de kwaliteit van leven. Wanneer het instrument wordt afgenomen in de vorm van een interview zal de afname naar verwachting echter meer tijd kosten dan het invullen van de WHOQoL-Bref. Van de MANSa is echter ook een versie beschikbaar die de cliënt zelf kan invullen.

Kosten

Beide instrumenten voor kwaliteit van leven (MANSa en WHOQoL-Bref) zijn vrij verkrijgbaar. Bij de instrumenten voor ernst van de klachten geldt dit voor de CGI, HoNOS, ORS en DASS. De SCL-90-R is het duurste instrument, gevolgd door de BSI.

Tabel 6. Overzicht van verschillende aspecten die voor beoordeling van de uitkomstinstrumenten van belang worden geacht.

Test	Lengte en tijdsduur	Verkrijgbaarheid en kosten	Multidimensionaliteit	Training?	Scoring en interpretatie	Handleiding beschikbaar?	Computer?
CGI	2 items; 1-2 min.	Gratis verkrijgbaar via internet (o.a. via Tijdschrift voor Psychiatrie).	Nee	Ja	Items reflecteren direct hun betekenis; geen extra tijd nodig. Geen normen beschikbaar.	Over de Nederlandse versie geen informatie gevonden. Ja: € 28,-.	Nee
KKL	14 items; 5 min.	Via uitgever Bohn Stafleu van Loghum: € 28,95 voor set van 15 formulieren.	Nee	Nee	Kost 5 min. Optellen van eerste 13 items geeft de totaalscore (mate van psychisch lijden); score op de items geeft aard en ernst van specifieke klachten aan. 14 ^e item geldt als aanvullende informatie. Normen beschikbaar, maar niet representatief. Bij gebruik voor uitkomstonderzoek is een hertestcorrectie noodzakelijk.	Ja: gratis via internet.	Ja
HoNOS	12 items; 5-15 min.	Gratis verkrijgbaar via internet (o.a. via Tijdschrift voor Psychiatrie).	Ja	Ja: één dagdeel.	Door sommeren van itemscores. De totaalscore geeft de geestelijke gezondheidstoestand en het sociale functioneren weer. Normen zijn beschikbaar.	Ja: gratis via internet.	Nee

Test	Lengte en tijdsduur	Verkrijgbaarheid en kosten	Multidimensionaliteit	Training?	Scoring en interpretatie	Handleiding beschikbaar?	Computer?
SCL-90-R	90 items; 12-20 min.	Via uitgever Pearson Assessment and Information B.V. Pen en papier: € 77,35 voor 50 formulieren en € 121,38 voor set antwoordsleutels. P ₂ O: € 5,70 en STM € 2,80 per afname.	Ja	Nee	Respectievelijk 10 en ca. 5 min. m.b.v. 9 antwoordsleutels. Items dienen te worden gesommeerd. Normen beschikbaar. Bij gebruik voor uitkomstonderzoek is een hertesteffectcorrectie noodzakelijk.	Ja: € 116,60	Ja
BSI	53 items; 8-15 min.	Via de uitgever PITS: handscoreingsset (handleiding en 50 formulieren): € 120,-. PC-afnameset (handleiding en pc-programma): € 215,-.	Ja	Nee	10 minuten. Per dimensie en voor de totaalscore wordt het gemiddelde van desbetreffende itemscores berekend. De score op de AAS-schaal wordt berekend door het aantal items waarop een score hoger dan 0 behaald wordt te tellen. De score op de EAS-schaal wordt berekend door de som van de scores op de 53 items te delen door de score op de AAS-schaal. Normen zijn beschikbaar, echter niet representatief. Nieuwe normen worden spoedig uitgebracht.	Ja, o.a. beschikbaar in set samen met het instrument. Ook los verkrijgbaar.	Ja

Test	Lengte en tijdsduur	Verkrijgbaarheid en kosten	Multidimensionaliteit	Training?	Scoring en interpretatie	Handleiding beschikbaar?	Computer?
OQ-45	45 items; < 10 min.	Online verkrijgbaar via oqmeasures.com: \$1,- per patiënt, papieren versie \$100,- per behandelaar	Ja	Nee	Ca. 10 min. Scoring d.m.v. optellen van de items, waarbij 9 items omgescoord dienen te worden. Totalscore geeft de ernst van de problematiek weer. De subscores geven de aard van de problematiek aan. Normen beschikbaar, maar niet representatief. Afkappunten vastgesteld om te bepalen of sprake is van psychopathologie en het bepalen van significante verbetering/verslechtering.	Ja: \$25,-. Verkrijgbaar wanneer er al een licentie is voor het gebruik van de OQ-45.	Ja
ORS	4 items; ± 5 min.	Na registratie online verkrijgbaar bij www.talkingcure.com ; gratis bij individueel gebruik.	Nee	Nee	5 min; scoring gebeurt handmatig. Optellen van de items geeft het algemene functioneren weer, terwijl de items het functioneren op verschillende gebieden weergeven. Behandelaars kunnen zelf het verloop van de verschillende afnames beoordelen. Er zijn geen normen beschikbaar.	Geen handleiding beschikbaar.	Nee
CORE-OM	34 items; ± 10 min.	Het instrument is digitaal en tegen kosten verkrijgbaar via www.reflectum.nl . Wat de kosten zijn kon niet worden teruggevonden.	Ja	Nee	De subschaalitems en totaalscore worden berekend door het gemiddelde van de betreffende items te vermenigvuldigen met 10; daarbij moet rekening gehouden worden met het omscoren van enkele items. Normen en een indicatie voor de benodigde tijd voor scoring en interpretatie konden niet worden teruggevonden.	Geen informatie over gevonden.	De Nederlandse versie is alleen digitaal beschikbaar. ⁵

Test	Lengte en tijdsduur	Verkrijgbaarheid en kosten	Multidimensionaliteit	Training?	Scoring en interpretatie	Handleiding beschikbaar?	Computer?
DASS	42 items; ± 10 min.	Gratis verkrijgbaar via internet (via psychischenwerk.nl).	Ja	Nee	Subscaalscores en totaalscore worden berekend door het optellen van desbetreffende items. Een scoringsformulier is hiervoor beschikbaar. Duur van scoring en interpretatie ca. 10 min. Normen zijn niet beschikbaar.	Niet beschikbaar.	Nee
MANSA	16 items; onbekend.	Gratis verkrijgbaar via internet (o.a. www.phamous.eu).	Nee	Geen informatie gevonden.	Interpretatie geschiedt aan de hand van itemscores en eventueel de totaalscore. Deze laatste wordt verkregen door het optellen van de itemscores. Geen informatie over normen gevonden.	Geen informatie over gevonden.	Nee
WHOQoL-Bref	26 items; onbekend	Gratis verkrijgbaar via internet (o.a. www.ildcare.eu).	Ja	Nee	Totaalscores en dimensiescores kunnen worden berekend door de gemiddelde score op de desbetreffende items te berekenen of door de items te sommeren. Geen informatie over normen gevonden.	Geen informatie over gevonden.	Nee

Multidimensionaliteit

De HoNOS, SCL-90-R, BSI, OQ-45, CORE-OM en DASS voldoen aan het criterium van multidimensionaliteit. Ook de WHOQoL-Bref geeft – binnen het kader van kwaliteit van leven – meerdere dimensies weer. Omdat er bij de KKL en ORS geen sprake is van verschillende subschalen, zijn deze instrumenten niet multidimensioneel. Echter beslaan de items van deze instrumenten wel elk ieder één domein binnen het terrein van respectievelijk de ernst van de klachten en dagelijks functioneren.

Welk instrument uiteindelijk de voorkeur heeft hangt ervan af op welk(e) gebied(en) men het behandelings-effect wil evalueren. Het kan een voordeel zijn om een instrument te gebruiken dat meerdere domeinen, zoals ernst van de klachten, dagelijks functioneren en kwaliteit van leven, meet. De CORE-OM is hier een geschikt instrument voor. De OQ-45 meet zowel ernst van de klachten als dagelijks functioneren.

Training voor afname

Voor de CGI en HoNOS is het volgen van een training noodzakelijk. Of voor het gebruik van de MANSA een training noodzakelijk is, kon niet worden teruggevonden. Voor de overige instrumenten is een training niet vereist.

Scoring en interpretatie

De scores op de items van de CGI en ORS reflecteren direct hun betekenis, waardoor voor scoring en interpretatie (vrijwel) geen extra tijd nodig is. De CGI geeft echter slechts een globaal beeld van de ernst van de klachten. Dit kan de interpretatie bemoeilijken. Ook de ORS geeft, in mindere mate, een globaal beeld van het dagelijks functioneren. Omdat de CGI ook door een beoordelaar wordt ingevuld, is objectiviteit niet geheel gegarandeerd. Dit laatste geldt ook voor de HoNOS. De noodzakelijke training voor afname van beide instrumenten moet hierin tegemoetkomen. Voor de HoNOS zijn tevens verschillende normgroepen beschikbaar: normen zijn beschikbaar voor itemscores, subschaalscores en de totaalscore. Dit vergemakkelijkt de interpretatie. De totaalscore op de HoNOS geeft de geestelijke gezondheidstoestand en het sociale functioneren van de respondent weer.

Voor de SCL-90-R, BSI, KKL en DASS geldt dat de totaalscore de ernst van de klachten representeert, terwijl de subschaalscores van deze instrumenten de aard van de klachten weergeven. Van elk van deze instrumenten zijn normen beschikbaar. De oude normen van de

BSI worden door de Cotan echter als onvoldoende beoordeeld (Evers, Braak, Frima, & Van Vliet-Mulder, 2007), evenals de normen van de KKL (Evers et al., 2000). Nieuwe normen voor de BSI zijn in aantocht. Van de DASS zijn geen normen beschikbaar.

De OQ-45 en de CORE-OM geven behalve een score voor de ernst van de symptomatologie ook een score voor het (dagelijks) functioneren. De CORE-OM geeft daarnaast ook scores voor welzijn en risico's ten opzichte van zichzelf en anderen. Dat deze instrumenten meerdere domeinen bestrijken kan de interpretatie van de instrumenten bemoeilijken, omdat een hoge score kan wijzen op problemen in verschillende, maar niet noodzakelijkerwijs op alle, gebieden. Tevens bleken voor de CORE-OM geen normen voorhanden te zijn.

Omdat normen van de SCL-90-R, OQ-45 en HoNOS beschikbaar zijn, heeft het gebruik van deze instrumenten de voorkeur boven het gebruik van de KKL, BSI, CGI, ORS, CORE-OM en DASS.

Wanneer echter uitsluitend naar de verschillen in scores bij een cliënt bij verschillende momenten van afname wordt gekeken – zonder verdere interpretatie van de scores zelf, hoeft het ontbreken van (representatieve) normen geen belemmering te zijn. Desondanks is goede normering wel wenselijk.

Van de instrumenten die de kwaliteit van leven meten is de scoring en interpretatie van beide instrumenten weinig tijdrovend. De interpretatie van de MANSA geschiedt aan de hand van de itemscores en eventueel een totaalscore. Doordat de WHOQoL-Bref dimensiescores en een totaalscore oplevert, is de interpretatie van dit instrument in vergelijking tot de MANSA eenvoudiger en eenduidiger.

Handleidingen

De handleiding van de HoNOS is gratis online te verkrijgen. De handleidingen van de KKL, SCL-90-R, BSI en OQ-45 zijn eveneens eenvoudig, maar tegen betaling, te verkrijgen. Van de CGI, ORS en DASS zijn geen handleidingen beschikbaar. Ook van de CORE-OM, MANSA en WHOQoL-Bref konden deze eveneens niet worden gevonden.

Computer

Afname, scoring en interpretatie met behulp van de computer is mogelijk bij de KKL, SCL-90-R, BSI, OQ-45 en CORE-OM.

Algemene conclusies en kritische beschouwing

In dit artikel is een overzicht geboden van diverse uitkomstinstrumenten. Deze zijn op tal van aspecten met elkaar vergeleken. Op basis hiervan is een overzicht gemaakt, waarin de instrumenten zo goed mogelijk in een (globale) rangorde geplaatst zijn (bijlage 1). Op basis hiervan is het mogelijk om een afweging te maken tussen de instrumenten en zo te bepalen welk instrument het meest wenselijk is voor het meten van ernst van de klachten, dagelijks functioneren, kwaliteit van leven of een combinatie van deze domeinen. Bij de afweging tussen de verschillende aspecten kunnen de beschreven criteria op een verschillende manier gewogen worden, afhankelijk van de situatie en het doel waarvoor de instrumenten gebruikt worden.

Een aspect dat in alle gevallen van belang is, is de psychometrische kwaliteit van het instrument. Geconcludeerd kan worden dat de SCL-90 op dit aspect als beste naar voren komt. Dit wordt geconcludeerd op basis van de informatie die in de tabellen gepresenteerd is, dus op basis van de (globale) rangordening. Uit het overzicht in bijlage 1 is echter op te maken dat de meeste instrumenten een voldoende tot goede psychometrische kwaliteit lijken te hebben.

Wanneer ook de praktische aspecten bekeken worden, is te zien dat van de instrumenten voor het meten van ernst van de klachten gemiddeld (bij gelijke weging van elk van de aspecten) de KKL, SCL-90 en OQ-45 als meest gunstig naar voren komen. Zo bekeken lijkt de WHOQoL-Bref de voorkeur boven de MANSA te hebben. Er zijn echter andere wegen mogelijk. Bij de keuze voor een uitkomstinstrument is het in de eerste plaats belangrijk een instrument te kiezen dat past bij de situatie en het doel dat de afnemer met het gebruik van het instrument wil bereiken. Afhankelijk van wat de afnemer wil meten, zal een keuze kunnen worden gemaakt voor een instrument dat geschikt is voor een bepaalde doelgroep, dat generiek dan wel stoornisspecifiek is en dat het domein ernst van de klachten, dagelijks functioneren en/of kwaliteit van leven beslaat. De mogelijkheid bestaat dat de afnemer, ondanks de aanbeveling van Andrews en Peters (1994) niet voor een multidimensioneel, maar voor een unidimensioneel uitkomstinstrument kiest, omdat dit beter past bij hetgeen geëvalueerd dient te worden. Daarnaast zal de keuze in de praktijk afhangen van aspecten als de beschikbare tijd en het beschikbare geld.

Bij de vergelijking van uitkomstinstrumenten dient rekening te worden gehouden met het (nog) beperkte onderzoek naar de Nederlandse versies van veel van de in dit onderzoek besproken instrumenten. Meer onderzoek naar de psychometrische eigenschappen van

de verschillende instrumenten is noodzakelijk. Tevens kon in dit onderzoek geen gebruik gemaakt worden van studies waarin verschillende uitkomstinstrumenten rechtstreeks met elkaar, op de in dit overzicht genoemde punten, vergeleken worden. Dergelijk onderzoek zou ten eerste aan te bevelen zijn, aangezien dit een veel steviger fundament legt waarop de keuze voor een bepaald instrument gebaseerd kan worden.

De in dit artikel gebruikte onderzoeken naar diverse instrumenten of naar het al dan niet effectief zijn van uitkomstonderzoek in het algemeen is veelal door dezelfde onderzoekers uitgevoerd. Replicaties en aanvullend onderzoek door anderen zijn hier gewenst. Daarnaast werd een beperking ervaren doordat sommige praktische informatie niet kon worden gevonden. Een betere en gemakkelijkere verkrijgbaarheid hiervan zou wenselijk zijn.

Bij beschouwing van dit artikel dient ten slotte rekening gehouden te worden met de mogelijkheid dat naast de hier besproken instrumenten nog andere geschikte uitkomstinstrumenten in de Nederlandse taal beschikbaar zijn. Hoewel geprobeerd is om alle geschikte instrumenten in dit onderzoek mee te nemen, was over sommige instrumenten die wel in de praktijk gebruikt worden, geen of onvoldoende informatie te verkrijgen. Omdat in dit onderzoek slechts instrumenten zijn meegenomen die in een Nederlandse vertaling beschikbaar zijn, is het te verwachten dat in het buitenland nog andere veelbelovende instrumenten beschikbaar zijn. Vervolgonderzoek zou dit kunnen uitwijzen.

ABSTRACT Due to several developments in (mental) health care, outcome monitoring has become more and more important. The Clinical Global Impression (CGI), Korte Klachtenlijst (KKL), Health of the Nation Outcome Scale (HoNOS), Symptom Checklist-90-R (SCL-90-R), Brief Symptom Inventory (BSI), Outcome Questionnaire (OQ-45), Outcome Rating Scale (ORS), Clinical Outcomes in Routine Evaluation - Outcome Measure (CORE-OM) and Depression Anxiety Stress Scale (DASS) are widely used monitoring tools in the Netherlands. The Manchester Short Assessment of Quality of Life (MANSA) and World Health Organisation Quality of Life-Bref (WHOQoL-Bref) are widely used for evaluation of quality of life. To make a comparison possible, an overview of these instruments is made.

Referenties

- Andrews, G. & Peters, L. (1994). Measurement of consumer outcome in mental health. In: R. Jacobs, *Investigating patient outcome measures in mental health*. York, UK: Centre for Health Economics, University of York, 2009.
- Arindell, W.A. (2001). Changes in waiting-list patients over time: data on some commonly-used measures. Beware! In: W.A. Arindell & J.H.M. Ettema, *Handleiding bij een multidimensionele psychopathologie-indicator. Symptom Checklist SCL-90*. Enschede: Ipskamp Drukkers, 2005.
- Arindell, W.A. & Ettema, J.H.M. (1986). SCL-90. Handleiding bij een multidimensionele psychopathologie-indicator. In: W.A. Arindell & J.H.M. Ettema, *Handleiding bij een multidimensionele psychopathologie-indicator. Symptom Checklist SCL-90*. Enschede: Ipskamp Drukkers, 2005.
- Arindell, W.A. & Ettema, J.H.M. (2003). SCL-90. Handleiding bij een multidimensionele psychopathologie-indicator. In: W. Arindell, H. Ettema, N. Groenman, F. Brook, I., Janssen, J. Slaets, G. Hekster, J. Derksen, J. van der Ende, H. Land, K. Hofman, & S. Dost, De groeiende inbedding van de Nederlandse SCL-90-R. Psychodiagnostisch gereedschap. *De psycholoog* 2003, 11, 576-582.
- Arindell, W.A. & Ettema, J.H.M. (2005). *Handleiding bij een multidimensionele psychopathologie-indicator. Symptom Checklist SCL-90*. Enschede: Ipskamp Drukkers.
- Arindell, W.A. & Ettema, J.H.M. (z.d.). *SCL-90; Symptom Checklist*. Verkregen op 5 oktober 2009 van <http://www.pearson-nl.com/producten/127-scl-90-symptom-checklist.html>.
- Arindell, W., Ettema, H., Groenman, N., Brook, F., Janssen, I., Slaets, J., Hekster, G., Derksen, J., Ende, J. van der, Land, H., Hofman, K., & Dost, S. (2003). De groeiende inbedding van de Nederlandse SCL-90-R. Psychodiagnostisch gereedschap. *De psycholoog*, 11, 576-582.
- Barkham, M., Margison, F.R., Leach, C., Lucock, M., Mellor-Clark, J., Evans, et al. (2001). Service profiling and outcomes benchmarking using the CORE-OM: toward practice-based evidence in the psychological therapies. Clinical Outcomes in Routine Evaluation-Outcome Measures. In: I.M.J. van Beljouw & P.F.M. Verhaak, *Geschiede uitkomstmaten voor routinematige registratie door eerstelijnspsychologen*. Utrecht: NIVEL, 2010.
- Beekman, A.T.F. & Beurs, E. de (2004). Meetinstrumenten bij aanmelding in de psychiatrie. *Tijdschrift voor psychiatrie*, 10, 653-658.
- Beljouw, I.M.J. van & Verhaak, P.F.M. (2010). *Geschiede uitkomstmaten voor routinematige registratie door eerstelijnspsychologen*. Utrecht: NIVEL.
- Beurs, E. de & Zitman, F.G. (2006). De Brief Symptom Inventory (BSI): De betrouwbaarheid en validiteit van een handzaam alternatief voor de SCL-90. In: B. van Wijngaarden & I. Kok, *Een inventarisatie van potentieel geschikte instrumenten voor de Basisset Prestatie-indicatoren geestelijke gezondheidszorg en verslavingszorg*. Utrecht: Trimbos-instituut, 2007.

- Beurs, E. de (2006). *Brief Symptom Inventory. Handleiding*. Leiden: PITS.
- Beurs, E. de (2009). *Brief Symptom Inventory. Handleiding*. Addendum. Leiden: PITS.
- Beurs, E. de, Dyck, R. van, Marquenie, L.A., Lange, A., & Blonk, R.W.B. (2001). De DASS: een vragenlijst voor het meten van depressie, angst en stress. *Gedragstherapie*, 34, 35-53.
- Beurs, E. de, Hollander-Gijsman, M. den, Buwalda, V., Trijburg, W., & Zitman, F. (2005). De Outcome Questionnaire. Psychodiagnostisch gereedschap. *De psycholoog*, 7-8, 393-399.
- Björkman, T. & Svensson, B. (2005). Quality of life in people with severe mental illness. Reliability and validity of the Manchester Short Assessment of Quality of Life. In: B. van Wijngaarden & I. Kok, *Een inventarisatie van potentieel geschikte instrumenten voor de Basisset Prestatie-indicatoren geestelijke gezondheidszorg en verslavingszorg*. Utrecht: Trimbos-instituut, 2007.
- Bringhurst, D.L., Watson, C.W., Miller, S.D., & Duncan, B.L. (2006). The reliability and validity of the Outcome Rating Scale: a replication study of a brief clinical measure. In: I.M.J. van Beljouw & P.F.M. Verhaak, *Geschiede uitkomstmaten voor routinematige registratie door eerstelijnspsychologen*. Utrecht: NIVEL, 2010.
- Derogatis, L.R. (1975). *Symptom Checklist-90-R (SCL-90-R)*. Dutch translation and adaptation Arindell, W.A., & Ettema, J.H.M. (1986, 2004). Amsterdam: Pearson Assessment & Information B.V.
- Derogatis, L.R. (1975, 2006). *Brief Symptom Inventory (BSI)*. Nederlandse bewerking E. de Beurs (2006). Leiden: PITS B.V.
- Derogatis, L.R. (1979). *Hopkins Symptom Checklist-57 (HSCL-57)*. Nederlandse bewerking F. Luteijn & L.F. Hamel (1984). Lisse: Swets en Zeitlinger B.V.
- Derogatis, L.R., Lipman, R.S., Rickels, K., Uhlenhuth, E.H., & Covi, L. (1974). *Hopkins Symptom Checklist-57 (HSCL-57)*. Lisse: Swets en Zeitlinger B.V.
- Ellsworth, J.R., Lambert, M.J., & Johnson, J. (2006). A comparison of the outcome questionnaire-45 and outcome questionnaire-30 in classification and prediction of treatment outcome. *Clinical Psychology and Psychotherapy*, 13, 380-391.
- Evans, C., Connel, J., Barkam, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *The British Journal of Psychiatry*, 180, 51-60.
- Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., McGrath, G., Connell, J., & Audin, K. (2000). *Clinical Outcomes in Routine Evaluation - Outcome Measure*. Verkregen via www.reflectum.nl.
- Evers, A., Braak, M.S.L., Frima, R.M., & Vliet-Mulder, J.C. van (2004a). *Klachtenlijst, SCL-90-R, 1981-2003*. Verkregen op 17 oktober 2009 van http://www.cotandocumentatie.nl/test_details.php?id=116.

- Evers, A., Braak, M.S.L., Frima, R.M., & Vliet-Mulder, J.C. van (2004b). *Depression Anxiety Stress Scal, DASS, 2001*. Verkregen op 1 april 2011 van http://www.cotandocumentatie.nl/test_details.php?id=119.
- Evers, A., Braak, M.S.L., Frima, R.M., & Vliet-Mulder, J.C. van (2005). *Klachtenlijst, SCL-90-R, 1981-2003. Researchbeschrijving. Documentatie van tests en testresearch in Nederland*. Verkregen op 17 oktober 2009 van http://www.cotandocumentatie.nl/test_details.php?id=116.
- Evers, A., Braak, M.S.L., Frima, R.M., & Vliet-Mulder, J.C. van (2007). *Brief Symptom Inventory, BSI, 2006*. Verkregen op 17 oktober 2009 van http://www.cotandocumentatie.nl/test_details.php?id=154.
- Evers, A., Braak, M.S.L., Frima, R.M., & Vliet-Mulder, J.C. van (2008). *Korte Klachtenlijst, KKL, 2007*. Verkregen op 17 oktober 2009 van http://www.cotandocumentatie.nl/test_details.php?id=155.
- Evers, A., Braak, M.S.L., Frima, R.M., & Vliet-Mulder, J.C. van (2009a). *Outcome Questionnaire, OQ-45.2008*. Verkregen op 1 april 2011 van http://www.cotandocumentatie.nl/test_details.php?id=658.
- Evers, A., Braak, M.S.L., Frima, R.M., & Vliet-Mulder, J.C. van (2009b). *Outcome Rating Scale, ORS, 2000*. Verkregen op 1 april 2011 van http://www.cotandocumentatie.nl/test_details.php?id=661.
- Evers, A., Vliet-Mulder, J.C. van, & Groot, C.J. (2000). *Documentatie van tests en testresearch in Nederland*. Assen: Van Gorcum.
- Expertgroep Volwassenen Kortdurend (2011). *ROM Volwassenen kortdurend. Adviesdocument*. Amersfoort: GGZ Nederland.
- Federaal Wetenschapsbeleid (2007). *Inventarisatie screening en assessment instrumenten voor middelenmisbruik en/of psychopathologie*. Verkregen op 5 oktober, 2009 van http://www.belspo.be/belspo/home/publ/pub_ostc/Drug/rDR26Inventar_nl.pdf.
- Froyd, J.E., Lambert, M.J., & Froyd, J.D. (1996). A review of practices of psychotherapy outcome measurement. In: J.R. Ellsworth, M.J. Lambert, & J. Johnson, A comparison of the outcome questionnaire-45 and outcome questionnaire-30 in classification and prediction of treatment outcome. *Clinical Psychology and Psychotherapy 2006, 13*, 380-391.
- Geraets, J.J.X.R., Goossens, M.E.J.B., Haastregt, J.C.M. van, Groot, I.J.M. de, Bruijn, C.P.C. de, Bie, R.A. de, Kôke, A.J., & Heuvel, W.J.A. van den (2006). Implications of process evaluation for clinical effectiveness and clinical practice in a trial on chronic shoulder complaints. *Patient Education and Counseling, 61*, 117-125.
- Gilbody, S.M., et al. (2003). Outcomes measurement in psychiatry: a critical review of outcomes measurement in psychiatric research and practice. In: R. Jacobs (2009), *Investigating patient outcome measures in mental health*. York, UK: Centre for Health Economics, University of York.
- Goldberg, D.P. (1972). *General Health Questionnaire (GHQ)*. Nederlandse bewerking M.W.J. Koeter & J. Ormel. Lisse: Swets en Zeitlinger B.V.
- Gregory, R.J. (2007). Psychological testing. *History, principles, and application* (International Edition). Illinois: Pearson.

- Guy, W. (1976). *Clinical Global Impression (CGI)*. Verkregen via tijdschrift-voorpsychiatrie.nl.
- Hafkenscheid, A. (1993). Psychometric evaluation of the Symptom Checklist (SCL-90) in psychiatry inpatients. In: A. Evers, M.S.L. Braak, R.M. Frima, & J.C. van Vliet-Mulder (2004). *Klachtenlijst, SCL-90-R, 1981-2003*. Verkregen op 17 oktober 2009 van http://www.cotandocumentatie.nl/test_details.php?id=116.
- Hafkenscheid, A. (2008). Routine Process Monitoring: ervaringen uit de praktijk. *Tijdschrift Cliëntgerichte Psychotherapie*, 46, 327-354.
- Hannan, C., Lambert, M.J., Harmon, C., Nielsen, S.L., Smart, D.W., Shimokawa, K., & Sutton, S.W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. In: K. Slade, M.J. Lambert, S.C. Harmon, D.W. Smart, & R. Bailey. (2008). Improving psychotherapy outcome: the use of immediate electronic feedback and revise clinical support tools. *Clinical Psychology and Psychotherapy*, 15, 287-303.
- Haro, J.M., Kamath, S.A., Ochoa, S., Novick, D., Rele, K., Fargas, A., Rodriguez, M.J., Rele, R., Orta, J., Kharbeng, A., Araya, S., Gervin, M., Alonso, J., Mavreas, V., Lavrentzou, E., Lontos, N., Gregor, K., & Jones, P.B. (2003). The Clinical Global Impression-Schizophrenia scale: a simple instrument to measure the diversity of symptoms present in schizophrenia. In: B. van Wijngaarden & I. Kok. (2007). *Een inventarisatie van potentieel geschikte instrumenten voor de Basisset Prestatie-indicatoren geestelijke gezondheidszorg en verslavingszorg*. Utrecht: Trimbo-instituut.
- Havenaar, J.M., Heemskerk, B., & Bisseling, E.M. (2002). Kwaliteit quo vadis? De opkomst van kwaliteitszorg binnen de psychiatrie. *Tijdschrift voor psychiatrie*, 2, 95-106.
- Havenaar, J.M., Os, J. van, & Wiersma, D. (2004). Algemene meetinstrumenten in de psychiatrische praktijk. *Tijdschrift voor psychiatrie*, 10, 647-651.
- Jacobs, R. (2009). *Investigating patient outcome measures in mental health*. York, UK: Centre for Health Economics, University of York.
- Jong, K. de & Nugter, A. (2004). De Outcome Questionnaire: psychometrische kenmerken van de Nederlandse vertaling. In: A. Smit, M. van Sonsbeek, & L. van Dijk, *Uitkomsten Resultaatmeting de Gelderse Roos 2008*. Wolfheze: Gelderse Roos Instituut voor Professionalisering, 2009.
- Jong, K. de, Nugter, M.A., Polak, M.G., Wagenborg, J.E.A., Spinhoven, P., & Heiser, W.J. (2007). The Outcome Questionnaire (OQ-45) in the Dutch population: A cross-cultural validation. *Clinical Psychology and Psychotherapy*, 14, 288-301.
- Kadouri, A., Corruble, E., & Falissard, B. (2007). The improved Clinical Global Impression Scale (iCGI): development and validation in depression. In: B. van Wijngaarden & I. Kok. (2007). *Een inventarisatie van potentieel geschikte instrumenten voor de Basisset Prestatie-indicatoren geestelijke gezondheidszorg en verslavingszorg*. Utrecht: Trimbo-instituut.
- Koeter, M.W.J., Ormel, J. & Brink, W., van den (1988). Totalscore op de CL-90 als maat voor de ernst van de psychopathologie. In: A. Evers, M.S.L. Braak, R.M. Frima, & J.C. van Vliet-Mulder (2004). *Klachtenlijst*,

- SCL-90-R, 1981-2003*. Verkregen op 17 oktober 2009 van http://www.cotandocumentatie.nl/test_details.php?id=116.
- Koeter, M.W.J., Ormel, J., Brink, W. van den, Dijkstra, W., Schoenmacker, J., & Staal, J. (1987). De waarde van de GHQ-28 als toestandmaat. *Tijdschrift voor Psychiatrie*, 29, 667-679.
- Lagerfeld, S. (2009). *Depressie-Angst-Stress-Schaal (DASS)*. Verkregen op 1 april 2011 van <http://www.psychischenwerk.nl/pw.subarticle.php?id=61&aid=826>.
- Lambert, M.J. & Burlingame, G.M. (1996, 2001). *Outcome Questionnaire (OQ-45)*. Verkregen via oqmeasures.com.
- Lambert, M.J. (2001). Psychotherapy outcome and quality improvement: Introduction to the special section on client-focused research. In: M.J. Lambert, J.L. Whipple, E.J. Hawkins, D.A. Vermeersch, S.L. Nielsen, & D.W. Smart (2003). Is it time for clinicians to routinely track patient outcome? A Meta-Analysis. *Clinical Psychology: Science and Practice*, 3, 288-301.
- Lambert, M.J. & Hawkins, E.J. (2004). Measuring outcome in professional practice: Considerations in selecting and using brief outcome instruments. *Professional Psychology: Research and Practice*, 5, 492-499.
- Lambert, M.J. & Ogles, B.M. (2004). The efficacy and effectiveness of psychotherapy. In G.I. Spielmann, K.S. Masters, & M.J. Lambert. (2006). A comparison of rational versus empirical methods in the prediction of psychotherapy outcome. *Clinical Psychology and Psychotherapy*, 13, 202-214.
- Lambert, M.J., Whipple, J.L., Hawkins, E.J., Vermeersch, D.A., Nielsen, S.L., & Smart, D.W. (2003). Is it time for clinicians to routinely track patient outcome? A Meta-Analysis. *Clinical Psychology: Science and Practice*, 3, 288-301.
- Lange, A. & Appelo, M.T. (2007). *Korte Klachtenlijst (KKL)*. Houten: Bohn Stafleu van Loghum.
- Lange, A. & Appelo, M. (2007). De Korte Klachten Lijst (KKL): Handleiding. In: B. van Wijngaarden & I. Kok (2007). *Een inventarisatie van potentiële geschikte instrumenten voor de Basisset Prestatie-indicatoren geestelijke gezondheidszorg en verslavingszorg*. Utrecht: Trimbo-instituut.
- Lovibond, P.F. & Lovibond, S.H. (1995). *Depression Anxiety Stress Scale (DASS)*. Verkregen via psychischenwerk.nl.
- Miller, S.D., Duncan, B.L., Brown, J., Sparks, J.A., & Claud, D.A. (2003). The Outcome Rating Scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. In: I.M.J. van Beljouw & P.F.M. Verhaak, *Geschiede uitkomstmaten voor routinematige registratie door eerstelijnspsychologen*. Utrecht: NIVEL, 2010.
- Miller, S.D., Duncan, B.L. & Johnson, L. (2000). *Outcome Rating Scale (ORS)*. Verkregen via www.talkingcure.com.
- Mulder, C.L., Staring, A.B.P., Loos, J., Buwalda, V.J.A., Sytema, S., & Wierdsma, A.I. (2004a). De Health of the Nation Outcome Scales (HoNOS) als instrument voor 'routine outcome assessment'. *Tijdschrift voor psychiatrie*, 5, 273-284.

- Mulder, C.L., Staring, A.B.P., Loos, J., Buwalda, V.J.A., Kuijpers, D., Sytema, S., & Wierdsma, A.I. (2004b). *De Health of the Nation Outcome Scales (HoNOS) in Nederlandse Bewerking. Handleiding*. Rotterdam: Onderzoekscentrum GGZ Rijnmond & GGZ Groep Europoort.
- Nunnally, J.C. & Bernstein, I.H. (1994). Psychometric theory. In: A. Evers, J.C. van Vliet-Mulder, & C.J. Groot, *Documentatie van tests en testresearch in Nederland*. Assen: Van Gorcum, 2000.
- Priebe, S., Husley, P., Knight, P., & Evans, S. (1999). *Manchester Short Assessment of Quality of Life (MANSA)*. Verkregen via www.phamous.eu.
- Priebe, S., Huxley, P., Knight, S., & Evans, S. (1999). Application and results of the Manchester Short Assessment of Quality of Life (MANSA). *International Journal of Social Psychiatry*, 1, 7-12.
- Rivierduinen & LUMC (2008). *Routine Outcome Monitoring. Documentatie over de uitkomstinstrumenten. ROM bij stemmings-, angst- en somatoforme stoornissen (SAS). Ambulante volwassenenzorg*. Leiden: Rivierduinen/LUMC.
- Slade, K., Lambert, M.J., Harmon, S.C., Smart, D.W., & Bailey, R. (2008). Improving psychotherapy outcome: the use of immediate electronic feedback and revise clinical support tools. *Clinical Psychology and Psychotherapie*, 15, 287-303.
- Smit, A., Sonsbeek, M. van, & Dijk, L. van (2009). *Uitkomsten Resultaatmeting de Gelderse Roos 2008*. Wolfheze: Gelderse Roos Instituut voor Professionalisering.
- Smith, M., Glass, G., & Miller, T. (1980). In: G.I. Spielmans, K.S. Masters, & M.J. Lambert (2006). A comparison of rational versus empirical methods in the prediction of psychotherapy outcome. *Clinical Psychology and Psychotherapy*, 13, 202-214.
- Spielmans, G.I., Masters, K.S., & Lambert, M.J. (2006). A comparison of rational versus empirical methods in the prediction of psychotherapy outcome. *Clinical Psychology and Psychotherapy*, 13, 202-214.
- Spits, M. & Schippers, G.M. (2008). *Hoe kunnen resultaten van zorg aan verslaafden worden gemeten en gebruikt? Over operationalisatie en toepassing in de verslavingszorg van prestatie-indicatoren 1.2 en 1.4 van de basisset prestatie-indicatoren geestelijke gezondheidszorg en verslavingszorg*. Amersfoort: AIAR/Resultaten Scoren.
- Stichting Benchmark GGZ. (2011). *SBG zorgdomeinen, meedomeinen, meetinstrumenten 20110211*. Addendum bij Minimale Dataset. Bilthoven: Stichting Benchmark GGZ.
- TNO (z.d.). *Trends: groeiende consumptie gezondheidszorg*. Geraadpleegd op 5 oktober 2009: <http://healthcare.monsterboard.nl/sector/trends/index.asp>.
- Trompenaars, F.J., Masthoff, E.D., Heck, G.L. van, Hodiament, P.P., & Vries, J. de (2005). Content validity, construct validity, and reliability of the WHOQoL-Bref in a population of Dutch adult psychiatric outpatients. *Quality of Life Research*, 1, 151-160.
- Walburg, J.A. (2003). Uitkomstenmanagement in de gezondheidszorg. Het opbouwen van lerende teams in de zorgorganisaties. In: M. Spits & G.M.

- Schippers (2008). *Hoe kunnen resultaten van zorg aan verslaafden worden gemeten en gebruikt? Over operationalisatie en toepassing in de verslavingszorg van prestatie-indicatoren 1.2 en 1.4 van de basisset prestatie-indicatoren geestelijke gezondheidszorg en verslavingszorg*. Amersfoort: AIAR/ Resultaten Scoren.
- Weijnenborg, P.T.M. & Kuile, M.M. ter (2000). The effect of a group programme on women with the Mayer-Rokitansky-Küster-Hauser syndrome. In: W.A. Arindell & J.H.M. Ettema (2005). *Handleiding bij een multidimensionale psychopathologie-indicator. Symptom Checklist SCL-90*. Enschede: Ipskamp Drukkers.
- Wennink, H.J. & Walburg, J. (2007). *Kompassen voor uitkomstmanagement in de GGZ*. Utrecht: Trimbos-instituut.
- Westen, D., & Morrison, K. (2001). A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. In: G.I. Spielmans, K.S. Masters, & M.J. Lambert (2006). A comparison of rational versus empirical methods in the prediction of psychotherapy outcome. *Clinical Psychology and Psychotherapy*, 13, 202-214.
- WHOQoL-Group (1996). *World Health Organisation Quality of Life-Bref (WHOQoL-Bref)*. Verkregen via tijdschrift voor psychiatrie.
- Wijngaarden, B. van & Kok, I. (2007). *Een inventarisatie van potentieel geschikte instrumenten voor de Basisset Prestatie-indicatoren geestelijke gezondheidszorg en verslavingszorg*. Utrecht: Trimbos-instituut.
- Wing, J.K., Beevor, A., Curtis, R.H., Park, S.B.G., Hadden, S., & Burns A. (1998). *Health of the Nation Outcome Scale (HoNOS)*. Verkregen via tijdschrift voor psychiatrie.

Bijlage 1 Rangorde kwaliteit van uitkomstinstrumenten voor ernst van de klachten per aspect

Betrouwbaarheid

Rangorde

1. Goed: BSI, SCL-90-R, DASS, ORS*
2. Voldoende: OQ-45, KKL, HoNOS, CGI**, WHOQoL-Bref
3. Gunstige buitenlandse bevindingen: CORE-OM, MANSA

* De bevindingen lijken te wijzen op een goede validiteit, maar worden door het Cotan (Evers et al., 2009b) als onvoldoende beoordeeld.

** In de literatuur zijn geen waarden teruggevonden waarop de betrouwbaarheid van de CGI beoordeeld kon worden.

Validiteit

Rangorde

1. Goed: SCL-90-R
 2. Veelbelovend: HoNOS, OQ-45, WHOQoL-Bref (goed, maar meer onderzoek noodzakelijk)
 3. Gunstige buitenlandse bevindingen: CORE-OM, MANSA
 4. Voldoende: KKL
 5. Matig: BSI, ORS
- CGI en DASS: te weinig informatie voor een beoordeling

Lengte en tijdsduur afname

Rangorde

1. CGI
2. ORS, KKL
3. HoNOS, BSI, OQ-45, CORE-OM, DASS
4. SCL-90-R

Het is onbekend hoeveel tijd er nodig is voor afname van de MANSA en WHOQoL-Bref.

Kosten

Globale indeling: afhankelijk van aantal gewenste afnamen.

Rangorde

1. Gratis: CGI, HoNOS, ORS, DASS, MANSA, WHOQoL-Bref
2. OQ-45
3. KKL
4. BSI
5. SCL-90-R

Over de kosten van de CORE-OM is geen informatie gevonden.

Multidimensionaliteit

Rangorde

1. Ja: HoNOS, SCL-90-R, BSI, OQ-45, CORE-OM, DASS, WHOQoL-Bref
2. Nee: CGI, KKL, ORS, MANSA

Training voor afname

Rangorde

1. Nee: KKL, SCL-90-R, BSI, OQ-45, ORS, CORE-OM, DASS, WHOQoL-Bref
2. Ja: CGI, HoNOS

Over een al dan niet benodigde training voor afname van de MANSA kon geen informatie worden gevonden.

Scoring en interpretatie

Wanneer uitsluitend naar de verschillen in scores bij een cliënt bij verschillende afnames gekeken wordt, zonder verdere interpretatie van de scores zelf, hoeft het ontbreken van (representatieve) normen geen probleem te zijn. Goede normering is echter wel wenselijk.

Rangorde

1. Goed: SCL-90-R, OQ-45, HoNOS
2. Goed, maar huidige normen niet representatief: KKL, BSI
3. Geen normen (gevonden): CGI, ORS, CORE-OM, DASS, MANSA, WHOQoL-Bref

Handleidingen

Rangorde

1. Gratis: HoNOS
2. Tegen kosten: KKL, SCL-90-R, BSI, OQ-45
3. Geen handleiding (gevonden): CGI, ORS, CORE-OM, DASS, MANSA, WHOQoL-Bref

Computerafname mogelijk

Rangorde

1. Ja: KKL, SCL-90-R, BSI, OQ-45, CORE-OM
2. Nee: CGI, HoNOS, ORS, DASS, MANSA, WHOQoL-Bref