



Dit rapport is een uitgave van het NIVEL.
De gegevens mogen worden gebruikt met
bronvermelding.

Geschikte uitkomstmaten voor routinematige registratie door eerstelijnspsychologen

I.M.J. van Beljouw
P.F.M. Verhaak

U vindt dit rapport en andere publicaties van het NIVEL in PDF-format op: www.nivel.nl

ISBN 978-90-6905-997-6

<http://www.nivel.nl>

nivel@nivel.nl

Telefoon 030 2 729 700

Fax 030 2 729 729

©2010 NIVEL, Postbus 1568, 3500 BN UTRECHT

Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook zonder voorafgaande schriftelijke toestemming van het NIVEL te Utrecht. Het gebruik van cijfers en/of tekst als toelichting of ondersteuning in artikelen, boeken en scripties is toegestaan, mits de bron duidelijk wordt vermeld.

Inhoud

Voorwoord	5
Samenvatting	7
1 Inleiding	9
2 Methode	13
2.1 Internationale literatuurdatabases	14
2.2 Inventarisatie van het Trimbos-instituut: inventarisatie meetinstrumenten in de GGZ	14
2.3 Inventarisatie voor signalering en behandeling van psychische problemen in de huisartsvoorziening	14
2.4 Websites van en gesprekken met aanbieders van geestelijke gezondheidszorg	14
2.5 Inventarisatie gebruik meetinstrumenten bij eerstelijnspsychologen	15
3 Resultaten	17
3.1 Generieke maten die symptomen of klachten meten	17
3.1.1 Vierdimensionele Klachtenlijst (4DKL)	17
3.1.2 Brief Symptom Inventory (BSI)	18
3.1.3 Clinical Global Impression (CGI)	20
3.1.4 General Health Questionnaire (GHQ)	21
3.1.5 Hospital Anxiety and Depression Scale (HADS)	22
3.1.6 Korte Klachten Lijst (KKL)	23
3.1.7 Mental Health Inventory (MHI)	24
3.1.8 Patient Health Questionnaire (PHQ)	25
3.1.9 De Symptom Checklist-90-R (SCL-90-R)	27
3.2 Generieke maten die ontwikkeld zijn voor Routine Outcome	28
3.2.1 Clinical Outcomes in Routine Evaluation – Outcome Measures (CORE-OM)	28
3.2.2 Eerstelijnsmonitor (ELM)	29
3.2.3 Health of Nation Outcome Scales (HONOS)	31
3.2.4 Outcome Questionnaire (OQ)	32
3.2.5 Outcome Rating Scale (ORS)	34
3.3 Generieke maten die het functioneren of kwaliteit van leven meten	36
3.3.1 EuroQol-5D (EQ-5D)	36
3.3.2 Global Assessment of Functioning (GAF)	37
3.3.3 Manchester Short Assessment of Quality of Life (MANSA)	38
3.3.4 (MOS) 36-item Short-Form Health Survey (SF-36)	39
3.3.5 World Health Organization – Disability Assessment Schedule-II (WHODAS-II)	40
3.3.6 WHO Quality of Life – BREF (WHOQOL-BREF)	42

4	Kleinschalig onderzoek naar de validiteit van de Outcome Rating Scale (ORS)	45
5	Discussie	49
	Bijlage 1: Enquête onder eerstelijnspsychologen	55
	Bijlage 2: Resultaten van enquête onder eerstelijnspsychologen	57
	Bijlage 3: De Outcome Rating Scale (ORS)	61
	Referentielijst	63

Voorwoord

Voor u ligt een rapport dat een uitgebreid overzicht geeft van het scala aan meetinstrumenten dat voorhanden is om psychische problematiek te meten, en mogelijk ingezet kan worden om behandeluitkomsten te meten in de eerstelijnspsychologische zorg.

Van deze instrumenten beschreven we wat hun psychometrische kwaliteit is, en of ze voldoen aan aanvullende eisen die gesteld worden vanuit de praktijk van de eerstelijnspsychologische zorg.

Dit rapport is tot stand gekomen met de hulp van vele mensen. In de eerste plaats willen wij de begeleidingscommissie bedanken voor het neerzetten van het juiste kader, en hun sturing en begeleiding tijdens dit proces. De begeleidingscommissie bestond uit dhr. H. Kamsma (voorzitter LVE), mw. D. Smit (beleidsmedewerker LVE en onderzoeker), dhr. D. Nieuwpoort (directeur LVE), dhr. H. Frencken (bestuurslid LVE) en mw. S. Bos (beleidsmedewerker MinVWS). Daarnaast zijn wij alle eerstelijnspsychologen dankbaar die via onze enquête hebben aangegeven welke meetinstrumenten zij in hun praktijk gebruiken, en voor de suggesties die zij gedaan hebben voor andere, potentieel geschikte uitkomstmaten. Tenslotte zijn wij zeer veel dank verschuldigd aan de eerstelijnspsychologen die bereid waren om deel te nemen aan ons kleinschalige onderzoek naar de samenhang naar de ORS en de SCL-90, en in dit kader extra tijd geïnvesteerd hebben in het verzamelen van gegevens bij hun cliënten.

NIVEL, Utrecht, februari 2010

Samenvatting

De eerstelijnspsycholoog is vrij toegankelijk, werkt nauw samen met de huisarts en wordt sinds 1999 gezien als één van de kerndisciplines in de eerstelijns geestelijke gezondheidszorg. Sinds 1 januari 2006 is de financiering van de eerstelijnspsychologische zorg onderdeel van de Zorgverzekeringswet, en sinds 1 januari 2008 wordt eerstelijnspsychologische zorg vergoed door de basisverzekering. Hierdoor is de eerstelijnspsycholoog toegankelijker, en tegelijkertijd neemt de vraag naar informatie over de kwaliteit van de geboden zorg toe. Op dit moment is er weinig informatie beschikbaar over de uitkomst van behandelingen die worden aangeboden binnen de eerstelijnspsychologische zorg. Dit rapport geeft een uitgebreid overzicht van potentieel geschikte uitkomstmaten voor de eerstelijnspsychologische zorg.

Om te komen tot dit overzicht van potentieel geschikte instrumenten, is op diverse manieren gezocht naar diagnostische vragenlijsten, en instrumenten die ontwikkeld zijn om de uitkomst van interventies te meten bij psychische problematiek. Er is gezocht naar relevante nationale en internationale literatuur, websites van aanbieders van geestelijke gezondheidszorg zijn geraadpleegd en er zijn gesprekken gevoerd met organisaties die software aanbieden voor het herhaaldelijk meten van behandeluitkomsten binnen de eerstelijnspsychologie. Tevens is een inventarisatie uitgevoerd onder eerstelijnspsychologen om in kaart te brengen welke meetinstrumenten op dit moment in de eerstelijnspsychologische praktijk gebruikt worden voor diagnostische doeleinden en/of om uitkomsten vast te leggen.

Van 21 generieke meetinstrumenten is de psychometrische kwaliteit en andere eigenschappen beschreven, waarbij een onderscheid is gemaakt tussen instrumenten die 1) klachten of symptomen in kaart brengen, 2) instrumenten die specifiek ontwikkeld zijn voor het routinematig vastleggen van uitkomsten, en 3) instrumenten die het functioneren of de kwaliteit van leven meten.

Hierbij wordt beschreven hoe werkbaar, beschikbaar, betrouwbaar, valide en responsief (=geschikt om bij herhaaldelijke afname veranderingen in functioneren in kaart te brengen) de diverse instrumenten zijn. De pagina's 50 tot en met 52 bevatten een zeer beknopt overzicht van de beschreven instrumenten en hun eigenschappen. Er is aanvullend onderzoek verricht naar de validiteit van de Outcome Rating Scale (ORS), die erg populair is bij eerstelijnspsychologen, maar waarbij tot op dit moment weinig onderzoek gedaan is naar de psychometrische eigenschappen.

De meeste beschreven instrumenten voldoen aan enkele van de reeds genoemde criteria, maar ze voldoen meestal niet aan alle wensen. Daarom kan er niet één instrument aangewezen worden dat bij uitstek geschikt is om behandeluitkomsten in de eerstelijnspsychologie te meten. Wel zijn enkele instrumenten af te raden vanwege

onvoldoende betrouwbaarheid en responsiviteit. Dit rapport stelt de Landelijke Vereniging van Eerstelijnspsychologen (LVE) in staat om zelf een keuze te maken voor één of meerdere uitkomstindicatoren die zij geschikt en zinvol vindt, en de toepassing hiervan in de dagelijkse praktijk kan adviseren aan haar leden. Het gebruik van uitkomstindicatoren biedt de eerstelijnspsycholoog inzicht in de resultaten van de verstrekte behandeling, geeft de mogelijkheid om de behandeling tussentijd bij te stellen indien nodig, en stelt eerstelijnspsychologen in staat om aan zorgverzekeraars te tonen dat zij door het gebruik van deze indicatoren extra kwaliteit bieden. In de toekomst kan de beroepsgroep zelf een (aantal) uitkomstindicatoren selecteren waarbij zij de kwaliteit van de eerstelijnspsychologische zorg kan presenteren.

1 Inleiding

De eerstelijnspsycholoog is vrij toegankelijk, werkt nauw samen met de huisarts en wordt sinds 1999 gezien als één van de kerndisciplines in de eerstelijns geestelijke gezondheidszorg. De eerstelijnspsycholoog biedt kortdurende behandeling bij psychische problematiek. Veel voorkomende klachten betreffen angsten, depressies, interpersoonlijke problemen en problemen in de aanpassing en verwerking. Driekwart van de cliënten is met maximaal acht sessies geholpen (LVE, 2009).

Op 1 januari 2008 werd de financiering van de geneeskundige geestelijke gezondheidszorg (GGZ), waar de eerstelijnspsychologische zorg toe behoort, overgeheveld van de AWBZ naar de Zorgverzekeringswet (MinVWS, 2007). Hoewel de geneeskundige GGZ al sinds 1 januari 2006 deel uitmaakte van de Zorgverzekeringswet, werd tot 2008 een bedrag vanuit het zorgverzekeringsfonds doorgeschoven naar de AWBZ ter dekking van de GGZ. Gelijktijdig met de overheveling naar de AWBZ, verschoof de bekostiging van de eerstelijnspsychologische zorg van de aanvullende verzekering naar de basisverzekering in de Zorgverzekeringswet.

Deze wijzigingen in financiering hebben een aantal consequenties met zich meegebracht. In de eerste plaats is de eerstelijnspsycholoog toegankelijker geworden voor veel mensen, omdat eerstelijnspsychologische zorg vanuit de basisverzekering betaald wordt, en cliënten de eerste acht consulten bij hun zorgverzekeraar kunnen declareren. Er wordt slechts een eigen bijdrage gevraagd van €10,- per sessie.

Ten tweede neemt hierdoor de aandacht toe voor de kwaliteit van de geboden zorg: in toenemende mate zijn diverse partijen geïnteresseerd in het zorgaanbod van eerstelijnspsychologen, en wat de prijs-kwaliteitverhouding is van wat zij te bieden hebben. Allereerst de cliënten: zij willen goed geïnformeerd een keuze kunnen maken voor een bepaalde eerstelijnspsycholoog; zij willen immers waar voor hun geld. Daarom is ook de eerstelijnspsycholoog te vinden op de website [kiesbeter.nl](http://www.kiesbeter.nl), die ruim 3 miljoen bezoekers per jaar trekt (www.kiesbeter.nl). Deze website biedt onder andere informatie over zorg, behandeling, patiëntenrechten en aanbieders van zorg, zoals eerstelijnspsychologen. Hierdoor wordt de consument de gelegenheid geboden om kenmerken van diverse zorgverleners met elkaar te vergelijken. [Kiesbeter.nl](http://www.kiesbeter.nl) wordt gemaakt door het RIVM (Rijksinstituut voor Volksgezondheid en Milieu), met financiering van het ministerie van Volksgezondheid, Welzijn en Sport (VWS).

Behalve cliënten, zijn ook zorgverzekeraars geïnteresseerd in wat een eerstelijnspsycholoog te bieden heeft: zij worden door de Zorgverzekeringswet gestimuleerd om, vaak door middel van contractering, de beste kwaliteit zorg zo goedkoop mogelijk in te kopen. Bovendien is kwaliteitsinformatie interessant voor de Inspectie voor de Gezondheidszorg om haar eigen werkwijzen te verbeteren, en voor de overheid om de kwaliteit van de aangeboden zorg te beoordelen in relatie tot het gevoerde beleid.

Deze roep om transparantie van aangeboden zorg speelt uiteraard niet alleen in de eerstelijnspsychologische zorg. Ook in andere sectoren van de (geestelijke) gezondheidszorg is deze trend zichtbaar. Het huidige kabinet heeft de transparantie van kwaliteit in alle zorgsectoren als speerpunt gesteld. In opdracht van het ministerie van VWS heeft de Inspectie van de Gezondheidszorg het programma “Zichtbare Zorg” opgericht. Het programma ondersteunt de sectoren in het meten en publiceren van de effectiviteit, veiligheid en klanttevredenheid van de zorg, en zet zich in om de gepubliceerde informatie betrouwbaar, valide en daadwerkelijk zichtbaar te maken (zie www.zichtbarezorg.nl voor meer informatie).

De kwaliteit van zorg kan zichtbaar gemaakt worden door middel van indicatoren. Een indicator is volgens Colsen en anderen (1995, p.297) ‘een meetbaar fenomeen dat een signalerende functie heeft met betrekking tot de kwaliteit van zorg en de kwaliteit van de organisatie’. Hoewel er diverse indicatoren zijn te onderscheiden, zoals structuurindicatoren of procesindicatoren, zullen we ons in dit rapport richten op uitkomstindicatoren (ook wel prestatie-indicatoren genoemd): zij vertellen iets over het resultaat van een behandeling. Dit kan zowel een uitspraak omvatten over de uitkomst vanuit het perspectief van de cliënt (klanttevredenheid), als bekeken worden vanuit professioneel oogpunt (zoals de score op een klinische maat).

In diverse sectoren wordt hard gewerkt aan de ontwikkeling van een set uitkomstindicatoren; bijvoorbeeld in de 2^e lijns-GGZ en verslavingszorg, gehandicaptenzorg en de sector verpleging, verzorging en thuiszorg. Voordat de resultaten gemeten met deze indicatorensets beschikbaar kunnen worden gesteld, is er echter nog een lange weg te gaan.

Er bestaan nog geen uitkomstindicatoren voor de eerstelijnspsychologische zorg, hoewel het in de toekomst van belang zal zijn dat ook de uitkomsten van interventies in de eerstelijnspsychologische zorg zichtbaar gemaakt kunnen worden. De enige informatie die beschikbaar is over de uitkomst van een behandeling door de eerstelijnspsycholoog betreft de score op de Global Assessment of Functioning (GAF), die bij aanvang en afsluiting van een behandeling door de psycholoog wordt geregistreerd in het LVE-codeboek. De GAF-score vormt as V van de DSM-IV (American Psychiatric Association, 1994), het diagnostisch handboek dat door veel psychologen en psychiaters wordt gebruikt om psychische stoornissen te diagnosticeren. De GAF-score is echter een zeer globale maat, die bovendien niet door de cliënt zelf, maar door de hulpverlener wordt ingevuld. Daarnaast zijn er twijfels over de kwaliteit van het instrument (Goldman et al., 1992; Vatnaland et al., 2007).

In de toekomst zal het echter steeds belangrijker worden om ook binnen de eerstelijnspsychologische zorg uitkomsten van diverse aanbieders te kunnen vergelijken, en om als beroepsgroep aan zorgverzekeraars te kunnen tonen welke kwaliteit eerstelijnspsychologen leveren. Om zich hier op voor te bereiden, heeft de LVE het NIVEL verzocht om een overzicht samen te stellen van bestaande meetinstrumenten die als potentiële uitkomstindicatoren kunnen dienen in de eerstelijnspsychologische zorg. Het onderscheid dat in de basisset prestatie-indicatoren GGZ en verslavingszorg 2009 gemaakt wordt, zal worden gehanteerd in dit rapport, namelijk: indicatoren die de ernst

van psychopathologie meten, indicatoren die betrekking hebben op het dagelijks functioneren, en indicatoren die kwaliteit van leven meten. In dit rapport zullen de volgende onderzoeksvragen beantwoord worden.

1. Welke meetinstrumenten zijn zoal voorhanden om de ernst van psychische problematiek, dagelijks functioneren en (aspecten van de) kwaliteit van leven van de cliënt te meten?
2. Hoe deugdelijk en bruikbaar zijn de verschillende instrumenten, waarbij gekeken wordt naar de volgende aspecten van het instrument:
 - a. Werkbaarheid: het aantal items en de afnameduur
 - b. Beschikbaarheid: waar is het instrument verkrijgbaar, tegen welke prijs, en is het vertaald in het Nederlands?
 - c. Betrouwbaarheid
 - d. Validiteit: meet het instrument wat het beoogt te meten?
 - e. Responsiviteit: is het instrument in staat om veranderingen in symptoomniveau, het functioneren of de kwaliteit van leven van de cliënt in kaart te brengen?

Opbouw van dit rapport

In hoofdstuk 2 zal de werkwijze besproken worden die gevolgd werd bij het beantwoorden van bovenstaande onderzoeksvragen: er komt aan de orde welke bronnen geraadpleegd werden om bestaande instrumenten op te sporen en welke criteria gehanteerd zijn om de kwaliteit en geschiktheid van deze instrumenten te bepalen. In hoofdstuk 3 zullen vervolgens de resultaten worden besproken van de literatuurstudie naar de kwaliteit en bruikbaarheid van bestaande instrumenten. Hoofdstuk 4 gaat in op het kleinschalige onderzoek dat is uitgevoerd naar de validiteit van de Outcome Rating Scale (ORS). Tenslotte zullen de belangrijkste bevindingen uit dit onderzoek worden samengevat en bediscussieerd in hoofdstuk 5.

2 Methode

Meetinstrumenten die (veranderingen in) ernst van psychische problemen, functioneren of kwaliteit van leven in kaart brengen kunnen voor verschillende doeleinden gebruikt worden. We onderscheiden 1) diagnostische instrumenten, 2) instrumenten die ontstaan zijn vanuit de Randomized Clinical Trials, en 3) instrumenten die gebruikt worden om behandeluitkomsten routinematig vast te leggen.

1) *Diagnostische vragenlijsten* worden binnen de psychodiagnostiek en psychotherapie gebruikt om tot nadere precisering van de diagnose te komen. Met name binnen de eerstelijnspsychologische praktijk zijn diverse vragenlijsten ingeburgerd die scores geven op een aantal (psychopathologische) dimensies. Deze instrumenten zijn relatief uitgebreid, en in principe geschikt om in de praktijk veranderingen in psychopathologie of functioneren te meten.

2) Met de opkomst van Evidence Based Medicine en de daaruit voortkomende *Randomized Controlled Trial (RCT)* is binnen de psychiatrische epidemiologie een grote vraag ontstaan naar gevalideerde uitkomstmaten: om het effect van een interventie te meten, ontstond een grote behoefte aan gevalideerde en gestandaardiseerde vragenlijsten. Hiervoor zijn zowel diagnosespecifieke meetinstrumenten ontwikkeld als meer generieke maten, die ook vaak als screeners voor algemene distress gebruikt worden.

3) Daarnaast is het *routinematig vastleggen van uitkomsten* sinds enige jaren in zwang geraakt binnen de ambulante geestelijke gezondheidszorg. Instrumenten die specifiek ontwikkeld zijn om herhaaldelijk veranderingen in het functioneren van cliënten te meten, zijn in principe relatief kort en a-specifiek: ze moeten immers geschikt zijn om met enige regelmaat, soms na elke zitting, te kunnen worden afgenomen.

In dit onderzoek zijn verschillende methoden gehanteerd om deze instrumenten op het spoor te komen:

1. Er is naar literatuur gezocht in internationale databases;
2. Een rapport van het Trimbos-instituut waarin binnen de GGZ geïnventariseerd werd welke uitkomstmaten daar gebruikt worden is bestudeerd (Wijngaarden et al., 2007);
3. Een inventarisatierapport van het NIVEL omtrent instrumenten om psychische problemen in de huisartsvoorziening te signaleren (Zwaanswijk et al., 2009);
4. Websites van aanbieders van geestelijke gezondheidszorg zijn bekeken, en met enkelen van deze aanbieders zijn gesprekken gevoerd;
5. Bij eerstelijnspsychologen is geïnventariseerd welke meetinstrumenten zij met enige regelmaat gebruiken ten behoeve van diagnostiek, en/of om uitkomsten van hun interventies vast te leggen.

2.1 Internationale literatuurdatabases

In eerste instantie is binnen de databases Pubmed, PsycInfo en Embase gezocht op de trefwoorden “psychotherapy” en “treatment outcome(s)” of “outcome assessment”, in de literatuur vanaf 1995. Aangezien deze zoekstrategie tot meer dan duizend treffers leidde met veel ruis, is besloten om binnen de huidige selectie te selecteren op studies die een Randomised Controlled Trial hebben uitgevoerd. In deze studies worden meetinstrumenten immers gebruikt om het effect van een interventie vast te stellen. Van de gevonden 531 (unieke) studies, werden de abstracten bekeken. Criterium om opgenomen te worden in het overzicht met mogelijk geschikte uitkomstindicatoren voor de eerstelijnspsychologie, is dat een instrument gericht is op de ernst van psychische problematiek, functioneren of kwaliteit van leven en bruikbaar is om verandering in dit opzicht te meten. Dit leidde tot 73 relevante studies. Daarnaast werden tevens 65 studies gevonden in een eigen verzameling van voornamelijk psychiatrisch-epidemiologische studies. Aan de hand van de abstracten van deze studies is een overzicht gemaakt van de gebruikte meetinstrumenten.

2.2 Onderzoek van het Trimbos-instituut: inventarisatie meetinstrumenten in de GGZ

Het Trimbos-instituut heeft in 2007 een vergelijkbare exercitie gedaan ten behoeve van de GGZ (van Wijngaarden en Kok, 2007). Het betrof een inventariserend onderzoek naar potentieel geschikte meetinstrumenten om de ernst van de problematiek, het dagelijks functioneren en kwaliteit van leven van cliënten in de GGZ in kaart te brengen. Deze thema's ontbraken op dat moment nog in de Basisset Prestatie-Indicatoren voor de GGZ en Verslavingszorg. Hiertoe is bij 36 GGZ-instellingen en RIAGG's geïnventariseerd welke instrumenten gebruikt worden om klinische uitkomsten en kwaliteit van leven te meten, en welke ervaringen men met de verschillende instrumenten had. Deze resultaten zijn bestudeerd.

2.3 Inventarisatie voor signalering en behandeling van psychische problemen in de huisartsvoorziening

In 2009 is er bij het NIVEL een rapport verschenen over geschikte signaleringsinstrumenten en behandelinterventies voor psychische problemen in de huisartspraktijk (Zwaanswijk et al., 2009). De besproken instrumenten zijn meegenomen in het huidige onderzoek.

2.4 Websites van en gesprekken met aanbieders van geestelijke gezondheidszorg

Websites zijn geraadpleegd van Scott Miller (Talking Cure) en CareMetrics. Beide organisaties hebben een instrument ontwikkeld om routinematig uitkomsten vast te leggen in de eerstelijnspsychologische zorg.

Tevens zijn gesprekken gevoerd met een aantal belangrijke personen en organisaties die ervaring hebben met het in kaart brengen van uitkomsten in de eerstelijnspsychologie. Zo hebben we gesproken met dhr. H. Kamsma, voorzitter van de LVE en eerstelijnspsycholoog, over de praktijk van de eerstelijnspsychologische zorg en welke rol uitkomstmetingen hierin (kunnen) spelen. Daarnaast zijn we ontvangen door Reflectum en door dhr. G.J. Kloens, oprichter van CareMetrics. Zowel Reflectum als CareMetrics hebben een softwaremodule op de markt gebracht waarmee bepaalde vragenlijsten (herhaaldelijk) digitaal bij cliënten kunnen worden afgenomen. De resultaten worden direct aan de psycholoog getoond en kunnen worden bewaard in het digitale dossier van de cliënt.

2.5 Inventarisatie gebruik meetinstrumenten bij eerstelijnspsychologen

Tevens is een korte enquête verspreid onder eerstelijnspsychologen die als zodanig geregistreerd staan op de website www.kiesbeter.nl, waar zorggebruikers en andere geïnteresseerden informatie kunnen vinden over zorgaanbieders, en waar diverse zorgaanbieders met elkaar vergeleken kunnen worden. Hen is gevraagd welke instrumenten zij gebruiken voor diagnostische doeleinden en/of om uitkomsten van hun interventies vast te leggen, en met welke frequentie deze instrumenten gebruikt worden. Tenslotte werd gevraagd of er bij de respondent meetinstrumenten bekend zijn (anderen dan die zij reeds gebruiken), die zij geschikt achten als uitkomstindicator in de eerstelijnspsychologie. De enquête is weergegeven in bijlage 1.

1042 eerstelijnspsychologen zijn per e-mail aangeschreven, waarvan 271 de vragenlijst hebben ingevuld. Dit is een respons van 26%.

De resultaten van deze enquête worden weergegeven in bijlage 2. De meest gebruikte instrumenten meten een specifiek domein van het functioneren, zoals de Utrechtse Coping Lijst (UCL), of een specifieke stoornis, zoals de Beck's Depression Inventory (BDI) (zie tabel 1 in bijlage 2). Wanneer alleen gekeken wordt naar generieke instrumenten, blijkt dat de Symptom Checklist-90 (SCL-90; 77%), de Outcome Rating Scale (ORS; 17%) en de 4-Dimensionele Klachtenlijst (4DKL; 8%) het frequentst gebruikt worden. Ook de Session Rating Scale (SRS; 14%) wordt veel genoemd, maar wordt hier verder buiten beschouwing gelaten omdat deze cliënttevredenheid meet, in plaats van symptomen, functioneren of kwaliteit van leven.

De SCL-90 wordt door evenveel psychologen alleen voor diagnostische doeleinden gebruikt, als voor zowel diagnostiek én het vastleggen van uitkomsten. De ORS wordt hoofdzakelijk gebruikt om uitkomsten te meten. De 4DKL wordt voornamelijk voor zowel diagnostische doeleinden ingezet als om de behandeluitkomst te meten (zie tabel 3 in bijlage 2). Tenslotte wordt de ORS door het frequentst genoemd wanneer gevraagd wordt naar een geschikt instrument om uitkomsten te meten, dat op dat moment door de psychologen zelf nog niet gebruikt werd (zie tabel 4 in bijlage 2).

Van alle meetinstrumenten die volgens bovenstaande strategieën verzameld zijn, werd een overzicht gemaakt. In overleg met de begeleidingscommissie werd vervolgens besloten om alleen de generieke instrumenten nader te onderzoeken, omdat zij niet diagnose-specifiek zijn, en bij elke cliënt in de eerstelijnspsychologische zorg kunnen

worden afgenomen. De instrumenten die zich richten op één specifiek domein van het functioneren, zoals coping, of op een specifieke stoornis, zoals depressie, zijn vervolgens buiten beschouwing gelaten.

De generieke instrumenten zijn vervolgens ingedeeld in instrumenten die 1) klachten of symptomen meten; 2) instrumenten die specifiek ontwikkeld zijn om op herhaaldelijke basis de uitkomst van een behandeling te meten (routine outcome monitoring); en 3) instrumenten die het dagelijks functioneren of de kwaliteit van leven van de cliënt in kaart brengen.

Samen met de begeleidingscommissie is besloten om de generieke instrumenten op de volgende criteria te beoordelen:

1. De *werkbaarheid* van een instrument: hoeveel items heeft het instrument, en wat is de afnameduur?
2. De *beschikbaarheid* van het instrument: is het instrument verkrijgbaar, zijn er kosten aan verbonden, en is het in het Nederlands vertaald?
3. De *betrouwbaarheid* van het instrument: hiervoor wordt de Cronbach's alpha (α) van het instrument beschreven. Cronbach's alpha is een maat voor de interne consistentie van de items van een vragenlijst, en geeft aan hoe betrouwbaar het instrument meet. Een alpha van minimaal .70 wordt als voldoende beschouwd.
4. *Face validiteit*: de mate waarin het instrument op het eerste gezicht lijkt te meten wat het behoort te meten. Er wordt een indruk gegeven van de face validiteit door te kijken naar de frequentie van het gebruik van dit instrument in wetenschappelijk onderzoek.
5. *Criterium- en constructvaliditeit*: criterium validiteit betreft de mate waarin een test een voorspellende waarde heeft; construct validiteit betreft de mate waarin de test de juiste weergave geeft van een te meten construct. Om dit laatste te onderzoeken wordt de relatie bekeken met reeds gevalideerde instrumenten die hetzelfde construct meten (convergente validiteit) of een geheel ander construct (divergente validiteit). Deze samenhang wordt vastgesteld door middel van Pearsons correlatiecoëfficiënten. De waardering van een correlatiecoëfficiënt die in dit rapport gehanteerd wordt, is gebaseerd op Guilford (1965): een verband is zeer laag ($r < .20$), laag ($r = .20 - .40$), matig ($r = .40 - .70$), hoog ($r = .70 - .90$) of zeer hoog ($r > .90$).
6. De *responsiviteit* van het instrument: bij het bepalen van de geschiktheid van een instrument om op routinematige basis gebruikt te kunnen worden als uitkomstmaat, is het van belang om naast bovenstaande aspecten tevens te kijken naar de responsiviteit van het instrument. Responsiviteit betreft de mate waarin een instrument in staat is om veranderingen in het construct (bijvoorbeeld functioneren of kwaliteit van leven) goed vast te stellen.

Om tot een beschrijving te komen van bovenstaande eigenschappen van de generieke uitkomstmaten, is met behulp van een speciaal Pubmed-filter gezocht naar literatuur over de psychometrische eigenschappen van de geïdentificeerde meetinstrumenten. Dit Pubmed-filter is ontwikkeld om op eenvoudige wijze artikelen te vinden die de psychometrische eigenschappen van meetinstrumenten hebben onderzocht (Terwee et al., 2009).

3 Resultaten

De bevindingen worden per meetinstrument beschreven, welke zijn ingedeeld in maten die 1) symptomen of klachten meten, 2) maten die ontwikkeld zijn t.b.v. het routinematig vastleggen van behandeluitkomsten, en 3) maten die het functioneren of de kwaliteit van leven vaststellen. Binnen deze categorieën worden de instrumenten in alfabetische volgorde behandeld.

3.1 Generieke maten die symptomen of klachten meten

3.1.1 Vierdimensionele Klachtenlijst (4DKL)

De Vierdimensionele KlachtenLijst (4DKL; Terluin, 1996) is een vragenlijst die is ontwikkeld in de huisartspraktijk, in een onderzoek naar surmenage. De 4DKL bestaat uit 50 items, die vragen naar het vóórkomen van diverse klachten in de afgelopen week, zoals hoofdpijn en angst om alleen het huis uit te gaan. De items zijn verdeeld over de schalen distress, depressie, angst en somatisatie, en kunnen worden beantwoord op een vijfpuntsschaal (variërend van 'nee' tot 'heel vaak of voortdurend'). De 4DKL wordt door de cliënt zelf ingevuld.

Werkbaarheid

De 4DKL bevat 50 items. De afnameduur bedraagt circa 10 minuten.

Beschikbaarheid

De 4DKL is Nederlandstalig, en gratis te downloaden via <http://www.emgo.nl/researchtools/4dsq.asp>.

Betrouwbaarheid

De 4DKL-schalen hebben een hoge interne consistentie: Cronbach's alpha varieert van .84 tot .94 (Terluin, 1996).

Face validiteit

Er zijn een zestal artikelen verschenen over de psychometrische eigenschappen van de 4DKL, allen geschreven door B. Terluin als (eerste) auteur, die de 4DKL ontwikkeld heeft. De 4DKL is in enkele onderzoeksstudies gebruikt, waaronder in de Nederlandse Studie naar Depressie en Angst (NESDA) (Penninx et al., 2008).

Construct- en criteriumvaliditeit

Wat betreft de criteriumvaliditeit blijken de schalen 'distress' en 'depressie' goed in staat om, respectievelijk, psychosociale diagnoses en de diagnose depressie te voorspellen.

Diagnosen angst en somatisatie werden matig voorspeld door de schalen ‘angst’ en ‘somatisatie’.

De constructvaliditeit van de 4DKL is goed: er worden hoge correlaties gevonden tussen de 4DKL-schalen en enkele relevante meetinstrumenten: tussen de schaal ‘distress’ en de Maastrichtse Vragenlijst (MV; Appels et al., 1987), welke vitale uitputting meet ($r = .76$); tussen de schaal ‘depressie’ en de SCL-90-depressie en de BDI ($r = .71 - .72$); tussen de schaal ‘angst’ en de SCL-90-angst en –agorafobie en de Beck’s Anxiety Inventory (BAI) ($r = .80 - .81$); en tussen de schaal ‘somatisatie’ en de SCL-90-somatisatie ($r = .82$) (Terluin et al., 2006). Ook de validering in een steekproef met uitsluitend werkende personen leverde bevredigende resultaten op (Terluin et al., 2004).

Responsiviteit

De schaal ‘distress’ blijkt goed responsief te zijn om globale verbeteringen vast te stellen. De responsiviteit bij de andere schalen is niet vastgesteld (Terluin et al., 2006).

Werkbaarheid	Bevat 50 items; afnameduur circa 10 minuten
Zelfrapportage	Ja
Beschikbaarheid	Nederlandstalig; gratis beschikbaar
Betrouwbaarheid	Zeer goed: $\alpha = .84 - .94$
Face validiteit	Enkele studies verschenen over de psychometrische eigenschappen van de 4DKL, allen geschreven door B. Terluin (ontwikkelaar 4DKL). Het instrument wordt in enkele onderzoeksstudies gebruikt.
Construct- en criteriumvaliditeit	Hoge correlaties gevonden met relevante instrumenten ($r = .71 - .82$)
Responsiviteit	Voldoende bij de schaal ‘distress’; bij overige schalen niet vastgesteld

3.1.2 Brief Symptom Inventory (BSI)

De Brief Symptom Inventory (BSI; Derogatis et al., 1983) is een verkorte versie van de SCL-90 (Derogatis et al., 1973). Hij bestaat uit 53 items: van alle SCL-90 dimensies, zijn de items opgenomen die volgens een factoranalyse het sterkst laadden op deze dimensies. De items zijn verdeeld over negen schalen: somatische klachten, cognitieve klachten, interpersoonlijke gevoeligheid, depressie, angst, hostiliteit, fobische klachten, paranoïde gedachten en psychoticisme. Daarnaast zijn er vier items opgenomen die geen schaal vormen. De items worden gescoord op een vijfpuntsschaal, variërend van ‘helemaal niet’ tot ‘heel erg’. De BSI wordt, netzoals de SCL-90, gebruikt als screeningsinstrument, maar ook om behandel effecten vast te stellen. Het instrument kan worden afgenomen bij personen vanaf 13 jaar. De BSI wordt door de cliënt zelf ingevuld.

Werkbaarheid

De BSI bevat 53 items, en kan worden afgenomen in circa 10 minuten (Derogatis et al., 1983)

Beschikbaarheid

De BSI is (opnieuw) vertaald in het Nederlands door Beurs en Zitman (de Beurs et al., 2006). Scoreformulieren zijn verkrijgbaar via uitgeverij PITS, en kosten €1,64 per stuk.

Betrouwbaarheid

Derogatis e.a. (1983) rapporteren voor de diverse schalen een Cronbach's alpha die varieert tussen de .71 en .85. Voor de Nederlandse vertaling van de BSI werden tevens acceptabele waarden gevonden: Cronbach's alpha loopt uiteen van .71 tot .96 (de Beurs et al., 2006).

Face validiteit

De BSI wordt veel in wetenschappelijke studies gebruikt.

Constructvaliditeit

Uit verschillende onderzoeken blijkt een zeer hoge correlatie tussen scores op de SCL-90 en BSI, wat logischerwijs ook verwacht mag worden. Er worden correlaties gerapporteerd variërend van .92 tot .99 (Derogatis, 1983), hoewel het verband tussen de Nederlandse vertalingen van de SCL-90-R en de BSI wat lager uitvallen: $r = .70 - .79$ (De Beurs & Zitman, 2006). Dit wordt vermoedelijk veroorzaakt doordat beide instrumenten twee weken na elkaar zijn afgenomen, en doordat de items van de BSI opnieuw vertaald zijn vanuit het Engels, in plaats van direct overgenomen uit de Nederlandstalige SCL-90-R. Hierdoor zijn sommige items anders verwoord. De Beurs en Zitman (2006) concluderen dat er sprake is van twee parallelle meetinstrumenten.

Uiteraard is ook gekeken naar de samenhang tussen subschalen van de BSI en instrumenten die een soortgelijk construct meten. Er worden hoge correlaties gerapporteerd tussen de relevante BSI-subschalen en de Korte Klachten Lijst ($r = .73$), de Beck Depression Inventory ($r = .78$), de Montgomery Åsberg Depression Rating Scale ($r = .72$), de Agorafobische Angstschaal ($r = .77$) (Beurs & Zitman, 2006) en diverse subschalen van de MMPI ($r = .32 - .72$, Derogatis, 1983). De BSI-totaalscore correleert matig met de totaalscore op de Brief Psychiatric Rating Scale ($r = .55$, Morlan 1998).

Responsiviteit

De Nederlandse vertaling van de BSI is in staat om veranderingen in het functioneren op te pikken, zo bleek na herhaaldelijke metingen bij poliklinische patiënten. De Beurs en Zitman (2006) concluderen dat de BSI zeer geschikt is als uitkomstmaat voor therapie-effect.

Werkbaarheid	Bevat 53 items, afnameduur bedraagt circa 10 minuten
Zelfrapportage	Ja
Beschikbaarheid	Vertaald in het Nederlands, kosten zijn €1,64 per scoreformulier
Betrouwbaarheid	$\alpha = .71 - .96$
Face validiteit	Veel gebruikt in wetenschappelijk onderzoek
Constructvaliditeit	Hoge correlaties met de SCL-90-R, en matige tot hoge correlaties tussen BSI-subschalen en meetinstrumenten die een soortgelijk construct meten.
Responsiviteit	In orde

3.1.3 *Clinical Global Impression (CGI)*

De Clinical Global Impression (CGI; Guy, 1976) is een schaal die bestaat uit twee items: één item meet de globale ernst van de aandoening gedurende de laatste week (CGI – severity) op een zevenpuntsschaal (van ‘normaal, niet ziek’ tot ‘zeer ernstig ziek’); één item meet de globale verandering ten opzichte van het begin van de behandeling (CGI – improvement) op een zevenpuntsschaal (van ‘zeer veel verbeterd’ tot ‘zeer veel slechter’). De CGI wordt door de hulpverlener ingevuld na afloop van het gesprek met de cliënt.

Werbaarheid

De CGI bestaat uit twee items, en vergt daardoor een minimale afnameduur.

Beschikbaarheid

De CGI is vertaald in het Nederlands, en gratis te downloaden via <http://www.tijdschriftvoorpsychiatrie.nl/meetinstrumenten/download.php?id=29>

Betrouwbaarheid

De interbeoordelaarsbetrouwbaarheid van de CGI ligt tussen .64 en .94, wat overeenkomt met goede tot zeer goede overeenstemming tussen de diverse beoordelaars (Kadouri et al., 2007).

Face validiteit

De CGI wordt veel gebruikt in wetenschappelijke studies, hoofdzakelijk naar schizofrenie.

Constructvaliditeit

Uit het onderzoek van Kadouri e.a. (2007) naar de validiteit van de CGI bij een populatie van depressieve patiënten laat een goede constructvaliditeit zien: scores op de CGI correleren matig met de SCL-90 ($r = .62$), en hoog met de HDRS ($r = .81$). Uit de studie van Berk e.a. bij intramurale GGZ-patiënten (2008) komen hele andere resultaten naar voren: de CGI correleert laag tot zeer laag met de HoNOS ($r = .12$ en $.31$) en met de MHQ-14 ($r = -.13$).

Responsiviteit

Er zijn nog geen resultaten gepubliceerd over de responsiviteit van de CGI.

Overig

Een belangrijke beperking van dit instrument is dat de hulpverleners de cliënt goed moet kennen om de CGI op een valide manier in te kunnen vullen (Berk et al., 2008).

Werkbaarheid	De CGI bevat slechts 2 items en kent daardoor een minimale afnameduur
Zelfrapportage	Nee, wordt door de hulpverlener ingevuld
Beschikbaarheid	Een Nederlandstalige versie is beschikbaar en gratis te downloaden
Betrouwbaarheid	Interbeoordelaarsbetrouwbaarheid is voldoende
Face validiteit	De CGI wordt zeer veel gebruikt in wetenschappelijke studies
Validiteit	De constructvaliditeit is goed in onderzoek van Kadouri e.a., maar slecht in een studie van Berk e.a.
Responsiviteit	Er is niets gepubliceerd over de responsiviteit van de CGI. Een belangrijke beperking is dat de hulpverlener de cliënt goed moet kennen om de vragenlijst in te vullen.
Overig	De hulpverlener moet de cliënt wel goed kennen om het instrument op valide wijze in te vullen

3.1.4 General Health Questionnaire (GHQ)

De General Health Questionnaire (GHQ; Goldberg, 1972) is een veelvuldig gebruikt instrument om de algehele mentale gezondheid van patiënten in de eerstelijnsgezondheidszorg vast te stellen. De GHQ is ontwikkeld voor volwassenen. De oorspronkelijke versie bevat 60 items, maar er zijn later een aantal kortere versies ontwikkeld, namelijk de GHQ-30, GHQ-28 (geschaalde versie) en GHQ-12. Het instrument is in het Nederlands vertaald door Koeter en Ormel (1991), en wordt door de cliënt zelf ingevuld. De GHQ vraagt naar ervaren veranderingen in het psychische belevingen en functioneren, bijvoorbeeld: bent u de laatste tijd door zorgen veel slaap te kort gekomen? Antwoorden kunnen gescoord worden in vier antwoordcategorieën, variërend van ‘helemaal niet’ tot ‘veel meer dan gewoonlijk’.

Werkbaarheid

Afname duurt 10 minuten voor de GHQ-28 en GHQ-30, en 5 minuten voor de GHQ-12.

Beschikbaarheid

De GHQ is vertaald in het Nederlands. Het instrument werd in het verleden uitgegeven door Pearson, maar is niet meer leverbaar.

Betrouwbaarheid

Goldberg e.a. (1997) beschrijven in een overzicht van diverse studies naar de psychometrische eigenschappen van de GHQ dat de gevonden Cronbach's alpha's variëren tussen de .82 en .94. Studies die zich specifiek richten op de GHQ-12 rapporteren ook een hoge interne consistentie van .88 en .91, onderzocht bij een populatie van Spaanse ouderen en Duitse patiënten in de eerstelijns, respectievelijk (Salama-Younes et al., 2009) (Schmitz et al., 1999b). Enige uitzondering hier op vormt een studie die is uitgevoerd bij een Spaanse volwassen populatie; hier werd een alpha gevonden van .76 (Sanchez-Lopez et al., 2008).

Face validiteit

De GHQ wordt veelvuldig gebruikt in wetenschappelijke studies.

Construct- en criteriumvaliditeit

Uit een grootschalige WHO-study waarin de GHQ-28 en de GHQ-12 in 15 verschillende landen werd gevalideerd, blijkt dat de GHQ-12 net zo robuust en valide is als de langere GHQ-28 (Goldberg et al., 1997). Om de criteriumvaliditeit te bepalen werden diagnoses gebaseerd op DSM-criteria vergeleken met GHQ-scores. De GHQ bleek goed in staat te zijn om patiënten met een DSM-diagnose te onderscheiden van degenen zonder diagnose. Deze studie concludeerde bovendien dat er geen verschillen zijn in validiteit tussen de (oorspronkelijke) Engelse versie van de GHQ, en de vertaalde versies.

De constructvaliditeit werd bepaald door de relatie tussen de GHQ en andere bestaande maten van mentale gezondheid te onderzoeken. Er werden matige verbanden gevonden tussen de GHQ en de ISRA (Inventory of Situations and Responses of Anxiety; $r=.57$) (Sanchez-Lopez et al., 2008) en de schalen van de SCL-90 (Symptom Checklist-90; $r = .37 - .73$) (Schmitz et al., 1999a).

Hoewel de GHQ bedoeld is voor afname bij volwassenen, wordt het instrument ook wel gebruikt bij adolescenten. Een Australische studie onder 11 tot 15-jarigen toonde aan dat de GHQ binnen deze groep een valide instrument is. Er werden matig tot hoge correlaties gevonden met maten van angst, stress, depressie en negatief affect ($r=.54 - .75$). Een beperkte, negatieve relatie werd gevonden met self-efficacy ($r= -.37$) (Tait et al., 2003).

Responsiviteit

De GHQ is voldoende responsief bij urologiepatiënten (Quek et al., 2001).

Werkbaarheid	Bevat 30, 28 of 12 items; afnameduur is respectievelijk 10 en 5 minuten
Zelfrapportage	Ja
Beschikbaarheid	In het Nederlands vertaald; niet meer leverbaar
Betrouwbaarheid	Goed: $\alpha = .76 - .94$
Face validiteit	Veelvuldig gebruikt in wetenschappelijke studies
Construct- en criteriumvaliditeit	Criteriumvaliditeit: goed Construct validiteit: goed: matige tot hoge relaties zijn aangetoond met andere maten van mentale gezondheid. Ook bij adolescenten.
Responsiviteit	Voldoende, gemeten bij urologiepatiënten

3.1.5 Hospital Anxiety and Depression Scale (HADS)

De Hospital Anxiety and Depression Scale (HADS; Zigmond et al., 1983) is ontwikkeld als een instrument om te screenen op de mogelijke aanwezigheid van angst- of depressiestoornissen. De HADS bevat twee subschalen: angst (HADS-A) en depressie (HADS-B), die ieder uit zeven items bestaan. De items worden gescoord op een vierpuntsschaal, variërend van 'bijna nooit' tot 'bijna altijd'. De HADS wordt door de cliënt zelf ingevuld.

Werkbaarheid

De HADS bestaat uit 14 items. Afname kost vijf minuten.

Beschikbaarheid

De HADS is vertaald in het Nederlands door Bonke en Serlie, en gratis beschikbaar via <http://www.tijdschriftvoorpsychiatrie.nl/meetinstrumenten/download.php?id=136>

Betrouwbaarheid

Bjelland e.a. (2002) concluderen in een literatuurstudie naar de psychometrische eigenschappen van de HADS, dat dit instrument een goede interne consistentie heeft. Cronbach's alpha voor de angstschaal varieert van .76 tot .93; de alpha voor de depressieschaal varieert van .72 tot .90 in diverse studies. Alleen een Portugese studie rapporteerde alpha's van respectievelijk .68 en .67.

Face validiteit

De HADS wordt zeer veel gebruikt in wetenschappelijke studies.

Constructvaliditeit

Er worden door diverse studies matige tot hoge correlaties gevonden met instrumenten zoals de BDI, GHQ, STAI, MADRS, CAS en SCL-90 (Bjelland et al., 2002). In Nederland is de HADS gevalideerd door een vergelijking te maken met de SCL-90 subschalen angst en depressie. De gevonden relatie is matig ($r = .49$ voor de angstschaal en $.69$ voor de depressieschaal).

Responsiviteit

De HADS is voldoende responsief (Cameron et al., 2008)

Werkbaarheid	Bevat 14 items; afnameduur is 5 minuten
Zelfrapportage	Ja
Beschikbaarheid	Nederlandstalige versie; gratis beschikbaar
Betrouwbaarheid	Goed: $\alpha = .76 - .93$. Slechts één studie vond onvoldoende alpha's.
Face validiteit	De HADS wordt zeer veel gebruikt in wetenschappelijke studies
Validiteit	Er worden matige tot hoge correlaties gevonden met relevante meetinstrumenten.
Responsiviteit	Voldoende

3.1.6 Korte Klachten Lijst (KKL)

De Korte Klachten Lijst (KKL; Lange et al., 2007) maakt onderdeel uit van de routine-outcome-monitoring van PsyQ, en omvat 14 items om de mate van psychisch lijden vast te stellen op diverse dimensies, zoals angst en depressie. De KKL wordt door de cliënt zelf ingevuld.

Werkbaarheid

De KKL is in enkele minuten in te vullen.

Beschikbaarheid

De KKL is Nederlandstalig, en zowel in een paper-pencil versie als digitaal verkrijgbaar via Bohn Stafleu van Loghum, voor €1,10 per stuk.

Betrouwbaarheid

De interne consistentie van de KKL is .85 bij psychiatrische patiënten, en .79 bij de algemene bevolking (Lange et al., 2007).

Face validiteit

De KKL wordt (nog) niet gebruikt in (internationale) wetenschappelijke studies.

Constructvaliditeit

Om de constructvaliditeit te bepalen zijn de scores op de KKL vergeleken met scores op de SCL-90. De gevonden validiteit is hoog: $r=.86$.

Responsiviteit

De KKL is onder andere ontwikkeld om het effect van behandelingen te meten. Uit onderzoek blijkt dat scores op de KKL, gemeten vóór en na behandeling, significant van elkaar verschillen (Lange et al., 2007).

Werkbaarheid	Afnameduur is enkele minuten
Zelfrapportage	Ja
Beschikbaarheid	Nederlandstalig; €1,10 per stuk
Betrouwbaarheid	Voldoende betrouwbaarheid bij psychiatrische patiënten en bij mensen in de gehele bevolking: $\alpha=.85$ en $.79$
Face validiteit	De KKL wordt niet gebruikt in (internationale) wetenschappelijke studies
Validiteit	Hoge correlatie met de SCL-90, $r=.86$
Responsiviteit	Onder meer ontwikkeld om behandelings-effect te meten; significante verschillen tussen voor- en na-meting bij behandeling.

3.1.7 Mental Health Inventory (MHI)

De Mental Health Inventory (MHI) is de 'subschaal 'psychische gezondheid' van de SF-36 en de SF-20 (Ware et al., 1992), en is onafhankelijk van de andere schalen af te nemen. De MHI bestaat uit 5 items, zoals 'hoe vaak gedurende de afgelopen vier weken voelde u zich erg zenuwachtig?', die gescoord kunnen worden op een zespuntsschaal, variërend van 'voortdurend' tot 'nooit'. Een lagere score gaat gepaard met een slechtere gezondheid. De MHI wordt door de cliënt zelf ingevuld.

Werkbaarheid

De afnameduur van de MHI bedraagt tussen de 2 en 3 minuten.

Beschikbaarheid

De SF-36 is vertaald in het Nederlands door Aaronson e.a. (Aaronson et al., 1998). De SF-36 is gratis te downloaden via <http://www.memic.unimaas.nl:8888/daisy/SMILE-CMS/228-SMILE.html?branch=1&language=1>

Betrouwbaarheid

De interne consistentie van de MHI is hoog: Cronbach's alpha is .90 in onderzoek met de oorspronkelijke versie (McHorney et al., 1994), en varieert tussen de .82 en .87 voor diverse Nederlandse populaties (Aaronson et al., 1998).

Face validiteit

De MHI wordt veel gebruikt in wetenschappelijke studies.

Construct- en criteriumvaliditeit

Wat betreft de validiteit van de MHI is hoofdzakelijk gekeken of deze geschikt is als screener voor veelkomende psychische stoornissen (criteriumvaliditeit). De MHI is geschikt als screener voor zowel depressie als sommige angststoornissen (Rumpf et al., 2001; Cuijpers et al., 2009), maar niet voor het screenen op andere DSM-stoornissen zoals somatoforme en middelenafhankelijkheid (Rumpf et al., 2001). Daarnaast blijkt deze subschaal van de SF-36 het beste in staat te zijn om een onderscheid te maken tussen respondenten mét en zonder psychische klachten, en tussen respondenten met milde en ernstige psychische klachten (McHorney et al., 1993). In een andere studie werd een matige correlatie gevonden tussen de MHI en de GHQ ($r = .64$) (Hoeymans et al., 2004).

Responsiviteit

De MHI is voldoende responsief (Garraat et al., 1994).

Werkbaarheid	Bevat 5 items, afnameduur bedraagt 2-3 minuten
Zelfrapportage	Ja
Beschikbaarheid	Nederlandstalige versie; gratis beschikbaar
Betrouwbaarheid	Goed: $\alpha = .82-.90$
Face validiteit	De MHI wordt veel gebruikt in wetenschappelijke studies
Construct- en criteriumvaliditeit	Voldoende geschikt als screener voor depressie en sommige angststoornissen, matig verband met GHQ ($r = .64$)
Responsiviteit	Voldoende

3.1.8 Patient Health Questionnaire (PHQ)

De Patient Health Questionnaire (PHQ; Spitzer et al., 1999) is ontwikkeld voor het screenen op vijf van de meest voorkomende psychische stoornissen in de eerstelijnsgezondheidszorg aan de hand van DSM-criteria: depressie, angst, alcoholafhankelijkheid, somatoforme stoornissen en eetstoornissen. De PHQ is de zelfrapportageversie van de originele PRIME-MD (Primary Care Evaluation of Mental Disorders), die bestaat uit een deel dat door de patiënt zelf ingevuld wordt, en een deel dat door de hulpverlener wordt gecoörd. De PHQ bestaat uit 59 items, en kan worden uitgebreid met een aantal aanvullende vragen over menstruatie, zwangerschap en psychosociale stress. Tevens is de Brief Patient Health Questionnaire verkrijgbaar, die alleen de vragen over stemming en angst bevat (28 items). Daarnaast is de PHQ-9 beschikbaar, die uitsluitend bestaat uit vragen over depressie. Deze depressieschaal is verder ingekort tot 2 items in de PHQ-2. De PHQ-15 is ontwikkeld voor de herkenning van somatisatie. Speciaal voor adolescenten is er de PHQ-A.

De responscategorieën voor depressie-, somatoforme en enkele vragen over angststoornissen bestaan uit een 3- of 4-punts Likert Schaal; de overige vragen worden beantwoord op een 2-puntsschaal.

Aangezien deze studie zich richt op generieke maten die gebruikt kunnen worden in de eerstelijnspsychologie, zal nadere bespreking van de PHQ zich richten op de volledige PHQ en de Brief PHQ.

Werkbaarheid

De PHQ die als generieke maat beschouwd kan worden bevat 59 of 28 items. De afnameduur van de PHQ is circa 15 en 10 minuten, respectievelijk.

Beschikbaarheid

De volledige PHQ is vertaald in het Nederlands, en gratis te downloaden via www.phqscreeners.com. De verkorte PHQ is niet als zodanig in het Nederlands beschikbaar, maar kan worden afgeleid van de volledige versie.

Betrouwbaarheid

Er is geen literatuur gevonden over de betrouwbaarheid van de volledige of verkorte PHQ in de algemene bevolking of een eerstelijnsample. Wel zijn er studies gedaan naar de betrouwbaarheid van de depressieschaal (PHQ-9), waarbij een goede interne consistentie wordt gerapporteerd: Cronbach's alpha varieert van .80 tot .92 (Cameron et al., 2008; Lee et al., 2007).

Face validiteit

De PHQ wordt veelvuldig in wetenschappelijke studies gebruikt.

Constructvaliditeit

Er werden significante verbanden gevonden tussen scores op de PHQ en schalen van de SF-20 (Spitzer et al., 1999). Dit verband is ook aangetoond voor de PHQ-A (Johnson et al., 2002).

Responsiviteit

De responsiviteit van de PHQ is voldoende (Cameron et al., 2008). Ook uit ander onderzoek blijkt dat de vragenlijst geschikt voor het monitoren van uitkomsten over een bepaalde tijdsperiode (Spitzer et al., 1999).

Werkbaarheid	Bevat 50 of 28 items; afnameduur circa 15 en 10 minuten, respectievelijk
Zelfrapportage	Ja
Beschikbaarheid	Beschikbaar in het Nederlands; gratis te downloaden
Betrouwbaarheid	PHQ-9: goed: $\alpha = .80 - .92$
Face validiteit	Veelvuldig in onderzoek gebruikt
Constructvaliditeit	Significante verbanden met de SF-20
Responsiviteit	Voldoende

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

De Symptom Checklist-90-R (SCL-90-R; Derogatis et al., 1973) is een multidimensionele klachtenlijst. De SCL-90-R wordt gebruikt als screeningsinstrument, maar ook om behandel-effecten vast te stellen. De SCL-90-R bevat 90 items, die verdeeld zijn over acht schalen: angst, depressie, somatische klachten, insufficiëntie van denken en handelen, wantrouwen en interpersoonlijke sensitiviteit, woede-hostiliteit en slaapproblemen. Daarnaast zijn er negen items die geen schaal vormen, maar wel bijdragen aan de totaalscore (Global Severity Index, GSI). Hiermee wijkt de Nederlandse versie af van de internationale versie, die tevens de schalen 'psychoticisme' en 'paranoïde gedachten' bevat, en waarbij de dimensie 'slaapproblemen' ontbreekt. De items worden gescoord op een vijf-puntsschaal, variërend van 'helemaal niet' tot 'heel erg'. De SCL-90-R kan worden afgenomen bij personen vanaf 12 jaar. Behalve de 90-item versie, is er tevens een SCL-45 beschikbaar. De SCL-90-R wordt door de cliënt zelf ingevuld.

Werkbaarheid

De afnameduur van de SCL-90-R bedraagt circa 20 minuten.

Beschikbaarheid

De SCL-90-R is in het Nederlands vertaald door Arrindell en Ettema (Arrindell et al., 1986). Het instrument is verkrijgbaar bij Pearson; een scoreformulier kost €1,55 per stuk.

Betrouwbaarheid

Uit Nederlands onderzoek naar de psychometrische eigenschappen van de SCL-90-R in een steekproef van cliënten in de eerstelijns, blijkt dat alle schalen een hoge interne consistentie vertonen: Cronbach's alpha is voor elke schaal groter dan .81 (Kloens, 2005). De schalen 'depressie', interpersoonlijke sensitiviteit en de negen ongeschaalde items laten een zeer hoge interne consistentie zien (alpha = .91-.97). Ook Fins onderzoek laat een acceptabele interne consistentie zien (alpha = .77-.90) (Holi et al., 1998).

Face validiteit

De SCL-90-R wordt zeer veel in wetenschappelijke studies gebruikt.

Construct- en criteriumvaliditeit

De SCL-90-R blijkt goed in staat om onderscheid te maken tussen psychiatrische patiënten en personen uit de alghele bevolking (Holi et al., 1998). Daarnaast bleek de SCL-90 net zo goed in het herkennen van patiënten met een psychiatrische diagnose als de GHQ (Holi et al., 2003). In het onderzoek van Kloens naar de Nederlandstalige SCL-90 worden matige correlaties gevonden tussen scores op de SCL-90 subschalen en enkele schalen van de Nederlandse PersoonlijkheidsVragenlijst (NPV), Nederlands Verkorte MMPI (NVM) en de Utrechtse Coping Lijst (UCL) (Kloens, 2005).

Responsiviteit

Nederlands onderzoek laat zien dat de SCL-90 in staat is om veranderingen in het functioneren van eerstelijnscliënten in kaart te brengen (Kloens, 2005): er is een significante afname van SCL-90 scores waarneembaar tussen de aanvangsmeting, de meting bij afsluiting van de behandeling, en de meting die twee maanden na afronding van de behandeling plaatsvond.

Werkbaarheid	Afnameduur bedraagt 20 minuten
Zelfrapportage	Ja
Beschikbaarheid	Er is een Nederlandstalige versie beschikbaar. Kosten: €1,55 per formulier
Betrouwbaarheid	Goed: $\alpha = .77 - .97$
Face validiteit	De SCL-90 wordt zeer veel gebruikt in wetenschappelijke studies
Construct- en criteriumvaliditeit	Goed in staat onderscheid te maken tussen psychiatrische patiënten en niet-patiënten, matige relaties met NPV, NVM, UCL
Responsiviteit	De SCL-90 is responsief

3.2 Generieke maten die ontwikkeld zijn voor Routine Outcome Monitoring (ROM)

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

De Clinical Outcomes in Routine Evaluation – Outcome Measures (CORE-OM; Evans et al., 2000) is ontwikkeld als korte vragenlijst om effectiviteit en efficacy van psychotherapeutische behandelingen te meten. De CORE-OM bevat 34 items, verdeeld over vier schalen: subjectief welzijn, problemen/symptomen, dagelijks functioneren en risico's t.o.v. zichzelf en anderen. De items worden gescoord op een vijf puntsschaal, variërend van 'helemaal niet' tot 'meestal of altijd'. Er zijn ook versies beschikbaar met respectievelijk 18, 10 en 5 items. De auteurs geven aan dat de CORE-OM bij uitstek geschikt is om voorafgaand en na afloop van de behandeling af te nemen. Het instrument is daarnaast ook geschikt om tussentijds afgenomen te worden. De CORE-OM wordt door de cliënt zelf ingevuld.

Werkbaarheid

De afnameduur van de CORE-OM varieert van 10 minuten (34 items) tot enkele minuten (5 items).

Beschikbaarheid

De Engelstalige versie is gratis te downloaden via www.coreims.co.uk. In Nederland is de CORE-OM uitsluitend digitaal beschikbaar, en wordt aangeboden door Mentaal Beter Services/Reflectum. Deze software-ontwikkelaar richt zich op praktijkautomatisering in de 1^e en 2^e lijns GGZ, en biedt een applicatie aan ten behoeve van Routine Outcome Monitoring (ROM) waarin de CORE-OM is opgenomen. Deze applicatie kan aan elk digitaal LVE-codeboek gehangen worden.

Betrouwbaarheid

De interne consistentie van de 34-item én de 18-item CORE-OM totaalscore is hoog: alpha is .94. Cronbach's alpha voor de subschalen varieert van .75 tot .90 (Evans et al., 2002; Barkham et al., 2001).

Face validiteit

De CORE-OM is in een tiental wetenschappelijke studies gebruikt.

Constructvaliditeit

Wat betreft de constructvaliditeit werden hoge correlaties worden gevonden tussen de scores op de CORE-OM en de meetinstrumenten die het functioneren of ervaren klachten in beeld brengen (de BDI, BSI, GHQ en SCL-90): $r = .81 - .88$). Een matig verband werd gevonden met de Beck's Anxiety Inventory (BAI) en de Inventory of Interpersonal Problems (IIP) ($r = .55-.65$) (Evans et al., 2002; Barkham et al., 2001).

Responsiviteit

Uit onderzoek blijkt dat de CORE-OM voldoende responsief is: diverse populaties die psychotherapie ontvingen lieten substantiële, significante verbeteringen zien op de CORE-OM (Evans et al., 2002)

Werkbaarheid	Bevat 34, 18, 10 of 5 items; afnameduur varieert van tien tot enkele minuten
Zelfrapportage	Ja
Beschikbaarheid	De Engelstalige versie is gratis beschikbaar; in Nederland wordt de CORE-OM alleen digitaal (en tegen kosten) aangeboden in een ROM-applicatie.
Betrouwbaarheid	Hoge interne consistentie voor de totaalscore; subschalen zijn voldoende betrouwbaarheid
Face validiteit	De CORE-OM is in een tiental wetenschappelijke studies gebruikt.
Constructvaliditeit	Hoge correlaties gevonden met meetinstrumenten die klachten/functioneren meten.
Responsiviteit	De CORE-OM is voldoende responsief

3.2.2 Eerstelijnsmonitor (ELM)

De EersteLijnsMonitor (ELM) is ontwikkeld door CareMetrics, een organisatie opgezet door dr. Gert Jan Kloens, en richt zich o.a. op de diagnostiek en screening binnen de gezondheidszorg. De ELM maakt deel uit van een uitgebreide set vragenlijsten voor o.a. screening, diagnostiek en routine-outcome-monitoring (ROM). De ELM kan binnen de beveiligde internetomgeving van het CareMetrics Healthcare Platform worden afgenomen. Het instrument bevat maximaal 124 items: de Nederlandse Relatie Vragenlijst (NRV) wordt alleen afgenomen als de respondent een intieme partner heeft. De items zijn verdeeld in dertien schalen, die gegroepeerd zijn in vier hoofdcategorieën: psychisch welzijn, persoonlijkheid, coping en relatiekwaliteit. Items worden beantwoord op een vier- of vijf puntsschaal. Enkele voorbeelden zijn:

- Angst op straat of in open ruimten
- Staat u 's ochtends moe en niet uitgerust op?

De ELM wordt door de cliënt zelf ingevuld.

Beschikbaarheid

De ELM is Nederlandstalig, en wordt gratis aangeboden aan vrijgevestigde zorgverleners die een overeenkomst hebben met zorgverzekeraar UVIT, ook voor niet-UVIT verzekerden. Zorgverleners zonder UVIT-contract betalen een abonnements-tarief van €200 per jaar, waarbij het gebruik bij de eerste twintig cliënten gratis is. Daarna wordt een tarief gehanteerd van €20,- per nieuwe cliënt. Het abonnementsgeld voor LVE-leden is €50 per jaar, met dezelfde voorwaarden.

Werbaarheid

Hoewel de ELM (maximaal) 124 items bevat, geeft de auteur zelf aan dat de afnameduur tien minuten bedraagt.

Betrouwbaarheid

De interne consistentie is voor de meeste schalen hoog, zowel bij een niet-klinische populatie als bij cliënten die zorg krijgen in de eerstelijns: Cronbach's alpha varieert van .82 tot .91. Alleen de subschaal 'passieve coping' laat een onvoldoende interne consistentie zien ($\alpha = .56 - .66$) (Kloens et al., 2008).

Face validiteit

Er is één artikel gepubliceerd over de ELM, door Kloens zelf (de auteur van de ELM). Het instrument is niet in andere wetenschappelijke studies gebruikt.

Constructvaliditeit

De constructvaliditeit van de ELM is bepaald door de scores op de dertien ELM-schalen te vergelijken met scores op veel gebruikte vragenlijst zoals de SCL-90, de Negativisme, Psychopathologie en Somatiserings Test (NPST), de Nederlandse Persoonlijkheidsvragenlijst (NPV), de Utrechtse Coping Lijst (UCL) en de Nederlandse Relatievragenlijst (NRV). Er worden hoge correlaties gevonden, tussen de ELM-hoofdcategorie 'psychisch welzijn' en de SCL-90, tussen de subschalen waarbij dit ook te verwachten valt ($r = .77 - .86$). Bijvoorbeeld, de subschaal 'depressieve klachten' van de ELM vertoont een hoge correlatie met de subschalen 'depressie' en 'psychoneuroticisme' van de SCL-90. Matige tot hoge correlaties worden gevonden tussen de hoofdcategorie 'psychisch welzijn' en de Utrechtse Burn-Out Schaal (UBOS), tussen de hoofdcategorie 'persoonlijkheid' en de NPST en NPV, tussen de hoofdcategorie 'coping' en de UCL, en ten slotte tussen de hoofdcategorie 'relatiekwaliteit' en de NRV (Kloens et al., 2008). Deze beschreven validiteitsmeting is gebaseerd op 44 respondenten, een erg klein aantal. Overigens zijn er geen significante verschillen gevonden in achtergrondkenmerken tussen deze respondenten die zowel de ELM als de te vergelijken vragenlijsten hebben ingevuld en de respondenten die alleen de ELM invulden. De auteurs concluderen daarom dat het hoogstwaarschijnlijk is dat de gevonden validiteit te generaliseren is naar de gehele populatie eerstelijnscliënten.

Responsiviteit

De ELM is ontworpen om therapieresultaten te meten op drie meetmomenten. Echter, in bovengenoemde publicatie is de ELM eenmalig afgenomen bij een klinische en niet-klinische groep. Er is niets beschreven over de responsiviteit van de ELM.

Werkbaarheid	De ELM bevat max. 124 items; afnameduur bedraagt 10 minuten volgens de auteur
Zelfrapportage	Ja
Beschikbaarheid	Nederlandstalig; wordt gratis aangeboden aan vrijgevestigde zorgverleners met een UVIT-contract; in andere gevallen zijn er kosten aan verbonden
Betrouwbaarheid	M.u.v. de subschaal 'passieve coping' werd een hoge interne consistentie gevonden ($\alpha = .82 - .91$)
Face validiteit	Alleen Kloens (auteur ELM) zelf heeft onderzoek gedaan naar de psychometrische eigenschappen van de ELM. Het instrument is niet in wetenschappelijke studies gebruikt.
Constructvaliditeit	Voor alle hoofdcategorieën werden hoge tot matige correlaties gevonden met relevante meetinstrumenten
Responsiviteit	De ELM is ontwikkeld om op drie meetmomenten af te nemen; er is geen wetenschappelijk onderzoek gedaan naar de responsiviteit van het instrument.

3.2.3 Health of Nation Outcome Scales (HONOS)

De Health of Nation Outcome Scales (HoNOS; Wing et al., 1998) is ontwikkeld in opdracht van het Engelse ministerie van Volksgezondheid, met als doel om routinematig de mentale gezondheid en het sociaal functioneren van mensen met psychische klachten in kaart te brengen. De HoNOS bestaat uit 12 items, die verdeeld zijn in 4 subschalen: gedragsproblemen, beperkingen, symptomatologie en sociale problemen. De items worden gescoord op een vijfpuntsschaal, variërend van 'geen probleem' tot 'ernstig tot zeer ernstig probleem'. De HoNOS wordt door de hulpverlener ingevuld. Behalve voor volwassenen, is er ook een HoNOS-versie ontwikkeld voor kinderen.

Werkbaarheid

De HoNOS bevat 12 items. De afnameduur bedraagt circa 5 tot 15 minuten, afhankelijk van hoe goed de hulpverlener de cliënt kent.

Beschikbaarheid

De HoNOS is vertaald in het Nederlands (Mulder et al., 2004b), en gratis te downloaden via <http://www.o3-onderzoekcentrumggz.nl/honosnl/index.htm>.

Betrouwbaarheid

Diverse onderzoeken zijn het niet eens over de interbeoordelaarsbetrouwbaarheid van de HoNOS. Deze wordt als voldoende tot goed ($ICC = .77 - .86$) beschouwd in een studie van Wing (auteur van de HoNOS; 1998). Orrell e.a. beschrijven dat dit voor acht van de 12 items geldt (1999). Daarentegen concluderen Bebbington e.a. dat de beoordelingen van hulpverleners in het algemeen slecht overeenkomen met die van onderzoekers (1999). Mulder e.a. vinden dat de interbeoordelaarsbetrouwbaarheid goed is voor de totaalscore en de subschalen, met uitzondering van de subschaal 'beperkingen' (2004a). De interne consistentie van de HoNOS is goed in een Nederlandse studie ($\alpha = .78$;

(Mulder et al., 2004a), maar onvoldoende in een Brits onderzoek ($\alpha = .65$; Orrell et al., 1999).

Face validiteit

De HoNOS wordt niet veelvuldig gebruikt in wetenschappelijke studies. Daarentegen wordt de HoNOS wel in Engeland en Australië veel in de praktijk gebruikt als uitkomstmaat (beschreven in Mulder et al., 2004a).

Constructvaliditeit

Er worden matige correlaties gevonden tussen de HoNOS-totaalscore en de GAF, CGI en QOL ($r = .51 - -.69$). Hoge correlaties worden gerapporteerd met de CANSAS (Camberwell Assessment of Need Short Appraisal Schedule) en de BPRS (Brief Psychiatric Rating Scale) ($r = .71 - .75$). De HoNOS subschalen 'symptomatologie' en 'sociale problemen' laten een matig verband zien met de hiervoor genoemde instrumenten; dit verband is laag voor de subschalen 'gedragsproblemen' en 'beperkingen' (Mulder et al., 2004a). Ook Wing e.a. (1998) vinden matige tot hoge correlaties tussen de HoNOS en relevante items van de BPRS en RFS (Role Functioning Scale) ($r = .65 - .85$).

Responsiviteit

Audin e.a. (2001) hebben een studie uitgevoerd naar de bruikbaarheid van de HoNOS als uitkomstmaat in psychotherapeutische behandelingen. Zij concluderen dat de HoNOS onvoldoende responsief is: slechts drie van de twaalf items zijn voldoende in staat om veranderingen in het functioneren te registreren. Ook Bebbington e.a. geven aan dat de HoNOS slecht presteert wanneer veranderingen in het functioneren als uitkomstmaat worden gebruikt (Bebbington et al., 1999).

Werkbaarheid	Bevat 12 items; afnameduur 5-15 minuten
Zelfrapportage	Nee, wordt door hulpverlener ingevuld
Beschikbaarheid	Nederlandstalig; gratis beschikbaar
Betrouwbaarheid	Wisselende bevindingen over zowel de interbeoordelaarsbetrouwbaarheid als de interne consistentie
Face validiteit	Veel onderzoek gedaan naar psychometrische eigenschappen, maar het instrument wordt weinig gebruikt in wetenschappelijke studies
Constructvaliditeit	Matig tot goed ($r = .51 - .85$)
Responsiviteit	Uit twee studies blijkt dat de HoNOS onvoldoende responsief is

3.2.4 Outcome Questionnaire (OQ)

De Outcome Questionnaire (OQ; Lambert et al., 1996) is een veelvuldig gebruikt instrument om behandelresultaten te meten. De OQ-45 bestaat uit 45 items, die zijn verdeeld in drie subschalen: 1) symptomatische distress (bestaande uit items die betrekking hebben op de meest voorkomende psychiatrische stoornissen zoals angst, depressie en afhankelijkheid van middelen), 2) interpersoonlijke relaties en 3) sociale rollen. De items worden gescoord op een vijfpuntsschaal, van 'nooit' (0) tot 'bijna altijd' (4). Er is ook een OQ-30 met 30 items en een OQ-10 met 10 items ontwikkeld. Daarnaast

is er ook een OQ ontwikkeld voor kinderen: de Y-OQ wordt gescoord door de ouders van kinderen in de leeftijd van 4 tot en met 17 jaar; de Y-OQ-SR kan door jongeren van 12 tot en met 18 zelf worden ingevuld. Beide bestaan uit 64 items. De Y-OQ-30 bevat 30 items.

Beschikbaarheid

De OQ-45, Y-OQ en Y-OQ-SR zijn vertaald in het Nederlands. Van de OQ-30, OQ-10 en de Y-OQ-30 is op dit moment geen Nederlandse versie beschikbaar. De OQ is verkrijgbaar via OQ-measures: www.oqmeasures.com, waar zowel de paper-pencil versie als een elektronische versie besteld kan worden. De kosten worden gerekend per cliënt (in plaats van per afname), en bedragen circa US\$ 1,- (= € 0,73; febr 2010).

Werkbaarheid

Ondanks dat de OQ-45 45 items bedraagt, wordt aangegeven dat het instrument in vijf minuten is af te nemen.

Betrouwbaarheid

De interne consistentie van de Nederlandse versie van de OQ-45, gemeten bij een populatie van studenten, gezonde personen en patiënten is goed (alpha varieert van .82 tot .96). Wanneer naar de subschalen wordt gekeken, is de betrouwbaarheid hoog voor de symptomatische distress schaal ($\alpha = .89 - .91$), wat lager maar voldoende voor de interpersoonlijke relatie schaal ($\alpha = .74 - .80$) en onvoldoende voor de sociale rol schaal ($\alpha = .53 - .69$) (Jong et al., 2008). Vergelijkbare resultaten zijn gevonden in Amerikaans onderzoek. Daar werd echter een hogere betrouwbaarheid gevonden voor de sociale rol schaal ($\alpha = .71-.82$) (Lambert et al., 1996).

Face validiteit

Een tiental wetenschappelijke studies maakt gebruik van de OQ-45.

Constructvaliditeit

Om de constructvaliditeit van de Nederlandstalige versie van de OQ-45 te bepalen, is een vergelijking gemaakt met de scores op de GSI (global severity index van de SCL-90-R), de Depression, Anxiety and Stress Scales (DASS) (Beurs et al., 2001) bij een patiëntgroep en de Groningse Vragenlijst Sociaal Gedrag (GVSG-45) (de Jong et al., 2001) bij een studentengroep. De gevonden relaties met de symptomatische distress schaal zijn hoog in de patiëntengroep ($r=.72-.80$) en matig tot hoog in de studentengroep ($r=.42-.78$). Zoals verwacht is de relatie tussen voorgenoemde meetinstrumenten en de interpersoonlijke relatie schaal en de sociale rol schaal van de OQ-45 wat minder uitgesproken: respectievelijk $r=.38-.62$ en $r=.38-.55$ (de Jong et al., 2008). Ook in Amerikaans onderzoek werd een matig tot hoge correlatie gevonden tussen de OQ-totaalscores en subschaalscores met relevante meetinstrumenten (Lambert et al., 1996).

Responsiviteit

Vermeersch e.a. (2000) hebben uitgebreid onderzoek gedaan naar de gevoeligheid van de OQ-45 om veranderingen in het functioneren van patiënten te meten. De auteurs

concluderen dat zowel de totaalscores, de subschaalscores en de meerheid van de items (37 van de 45) voldoende responsief zijn: de scores veranderen in de verwachte richting over tijd (ze nemen toe dan wel af waar dat verwacht wordt), en de scores laten een sterkere verbetering zien over tijd bij patiënten dan bij niet-patiënten. Ook onderzoek naar de responsiviteit van de Nederlandstalige OQ-45 laat bevredigende resultaten zien (de Jong et al., 2008).

Werkbaarheid	Bevat 45, 30 of 10 items; afnameduur van de OQ-45 is 5 minuten
Zelfrapportage	Ja
Beschikbaarheid	Er is een Nederlandse versie beschikbaar van de OQ-45 en de Y-OQ; kosten circa US\$ 1,- per cliënt (= € 0,73; febr 2010)
Betrouwbaarheid	Goede betrouwbaarheid van de OQ-totaalscore en subschalen 'symptomatische distress' en 'interpersoonlijke relaties'. De subschaal 'sociale rol' laat een lagere betrouwbaarheid zien in Nederlands onderzoek, maar komt beter naar voren in Amerikaans onderzoek
Face validiteit	Een tiental wetenschappelijke studies heeft de OQ gebruikt.
Constructvaliditeit	Er worden matige tot hoge correlaties gevonden tussen OQ-totaalscores en subschaalscores met relevante meetinstrumenten
Responsiviteit	De OQ-45 is speciaal ontworpen om behandel-effecten te meten. Onderzoek toont aan het instrument goed in staat is om veranderingen in het functioneren van de patiënt op te pikken.

3.2.5 Outcome Rating Scale (ORS)

De Outcome Rating Scale (ORS) is ontwikkeld door Scott Miller, en gebaseerd op de Outcome Questionnaire-45 (OQ-45). De ORS bestaat uit vier items, die vragen naar het individueel, relationeel, sociaal en algeheel welbevinden. Er wordt gebruik gemaakt van een visueel-analoge schaal: elk item wordt gescoord door een markeerpunt te plaatsen op een lijn van 10 centimeter, waarbij lage scores zich aan de linkerkant bevinden, en hoge scores aan de rechterkant. De score wordt bepaald door het markeerpunt vanaf het begin van de schaal tot op de millimeter nauwkeurig op te meten met een lineaal. Een hogere score is geassocieerd met een beter welbevinden. Vervolgens worden de scores op de vier items opgeteld tot een totaalscore. Wanneer de ORS na elke sessie wordt afgenomen, kan er een grafiek gemaakt worden waarin de totaalscore wordt afgezet tegen het aantal sessies, die snel inzichtelijk maakt of een cliënt voldoende vooruitgang boekt. Wanneer er na de derde sessie onvoldoende verandering zichtbaar is (minder dan 5 punten verbetering), is er sprake van een verhoogd risico op afbreken van de behandeling, en wordt een beroep gedaan op de behandelaar om de therapie bij te sturen en de therapeutische relatie bespreekbaar te maken. Er is tevens een versie voor kinderen beschikbaar (Child Outcome Rating Scale – CORS). De (C)ORS wordt door de cliënt zelf ingevuld.

Werkbaarheid

De ORS en CORS bestaan uit slechts 4 items, en deze zijn in enkele minuten in te vullen. Onderzoek naar de bruikbaarheid van dit instrument in een klinische setting liet zien dat 86% van de hulpverleners gebruik maakten van de ORS, versus 25% die de OQ-45 gebruikten (Miller et al., 2003).

Beschikbaarheid

De ORS en CORS zijn in het Nederlands vertaald door Asmus, Crouzen en van Oenen (2009), en gratis te downloaden via www.talkingcure.com.

Betrouwbaarheid

Er zijn slechts twee studies uitgevoerd naar de psychometrische kenmerken. Beide onderzoeken vinden een hoge interne betrouwbaarheid van $\alpha = .93$ (Miller et al., 2003) en $.97$ (Bringhurst et al., 2006) in een niet-klinische populatie.

Face validiteit

Tot op heden zijn er slechts twee studies verschenen over de psychometrische eigenschappen van de ORS, beide gepubliceerd door Scott Miller (auteur ORS) e.a.. Het instrument is niet gebruikt in wetenschappelijke studies (m.u.v. een studie die het instrument in gebarentaal heeft vertaald).

Constructvaliditeit

De constructvaliditeit is bepaald door de totaalscores te vergelijken op de ORS en OQ-45. Er werden matige correlatiecoëfficiënten gevonden, variërend van $.57$ tot $.69$ (Bringhurst et al., 2006; Miller et al., 2003). De auteurs zelf geven aan dat dit waarschijnlijk verklaard wordt door het feit dat de ORS slechts 4 items bevat. De validiteit is alleen vastgesteld in een niet-klinische populatie.

Gezien de weinig gegevens over de validiteit van de ORS, maar de grote populariteit die het instrument onder eerstelijnspsychologen geniet, hebben wij zelf een kleinschalig onderzoek opgezet om de samenhang tussen scores op de ORS en de SCL-90 bij cliënten van eerstelijnspsychologen in kaart te brengen. Meer informatie over dit onderzoek en de resultaten daarvan worden beschreven in hoofdstuk 4.

Responsiviteit

De ORS is in staat om veranderingen in het functioneren van de respondent op te pikken (responsiviteit), waardoor het een geschikt instrument is om het effect van een behandeling te meten. Hiervan is sprake wanneer de scores op het instrument relatief stabiel blijven in een niet-klinische populatie, en veranderen in een klinische populatie die behandeling krijgen voor hun problemen. Uit onderzoek blijkt dat er geen significante verschillen bestaan tussen de voor- en nameting van de ORS bij een niet-klinische groep; daarentegen zijn er wel significante verschillen gevonden tussen beide metingen in een klinische groep die counseling ontvingen (Miller et al., 2003).

Werkbaarheid	Bevat 4 items; afnameduur is enkele minuten
Zelfrapportage	Ja
Beschikbaarheid	Nederlandstalige versie beschikbaar, gratis
Face validiteit	Er zijn slechts 2 studies verschenen over de psychometrische eigenschappen, beide geschreven door de auteur van de ORS. Het instrument is niet gebruikt in wetenschappelijke studies.
Betrouwbaarheid	Hoge interne consistentie ($\alpha .93 - .97$), maar alleen in niet-klinische populatie vastgesteld
Constructvaliditeit	Matig; alleen in niet-klinische populatie vastgesteld
Responsiviteit	De ORS is responsief

3.3 Generieke maten die het functioneren of kwaliteit van leven meten

3.3.1 EuroQol-5D (EQ-5D)

De EuroQol-5D (EQ-5D; Brooks, 1996) is een korte vragenlijst, bedoeld voor het in kaart brengen van gezondheidsgerelateerde kwaliteit van leven. Het instrument wordt veel gebruikt voor kosten-effectiviteit analyses. De EQ-5D-index bestaat uit vijf items over mobiliteit, zelfzorg, dagelijkse activiteiten, pijn/klachten en angst/depressie. Elk item wordt gescoord op een driepuntsschaal, van 'geen problemen' tot 'ernstige problemen'. Daarnaast is een visueel-analoge schaal (EQ-5D VAS) waarop de respondent aan kan geven hoe goed of slecht zijn/haar gezondheidstoestand is, variërend van 0 (slechts voorstelbare gezondheidstoestand) tot 100 (best voorstelbare gezondheidstoestand).

Werkbaarheid

De EQ-5D bevat vijf items, en is binnen enkele minuten in te vullen.

Beschikbaarheid

Van de EQ-5D is een officiële Nederlandse versie beschikbaar. Deze is op te vragen via <http://www.euroqol.org/eq-5d/how-to-obtain-eq-5d.html/>

Face validiteit

De EQ-5D wordt veel gebruikt in wetenschappelijke studies.

Betrouwbaarheid

De interne consistentie van de EQ-5D-index is voldoende: $\alpha = .73 - .74$ (Savoia et al., 2006; Kontodimopoulos et al., 2008).

Constructvaliditeit

Er worden matige correlaties gevonden tussen scores op de EQ-5D-index en de physical component summary ($r = .66$) en de mental component summary ($r = .41$) van de SF-12. Een vrijwel gelijk verband wordt gevonden tussen de EQ-5D-VAS en de twee subschalen van de SF-12 (Johnson et al., 2000; Johnson et al., 1998).

Responsiviteit

De EQ-5D blijkt in staat om veranderingen in gezondheidsgerelateerde kwaliteit van leven te registreren (Lamers et al., 2006).

Werkbaarheid	Afnameduur bedraagt enkele minuten
Zelfrapportage	Ja
Beschikbaarheid	Nederlandse versie is beschikbaar
Betrouwbaarheid	Goed: $\alpha = .73-.74$
Face validiteit	De EQ-5D wordt veel gebruikt in wetenschappelijke studies
Validiteit	Matig: $r = .41 - .66$
Responsiviteit	EQ-5D is voldoende responsief

3.3.2 *Global Assessment of Functioning (GAF)*

As V van DSM-IV wordt gevormd door de Global Assessment of Functioning (GAF; American Psychiatric Association, 1994), en is één van de meest gebruikte instrumenten om het psychisch, sociaal en beroepsmatig functioneren van cliënten met psychische problematiek in kaart te brengen. De GAF bestaat uit 10 beschrijvingen, waarbij de hulpverlener de beschrijving moet kiezen die het best het globale functioneren van de cliënt weergeeft. Elk item bestaat uit twee onderdelen: het eerste deel gaat over de ernst van de problematiek, het tweede deel betreft het niveau van functioneren. Wanneer de gekozen ernst van de problematiek en het niveau van functioneren twee verschillende beschrijvingen betreft, wordt de GAF-score bepaald door de ernstigste van deze twee. De GAF-score ligt tussen 0 en 100, waarbij een hogere score een beter functioneren weerspiegelt.

Werkbaarheid

De GAF-score is binnen enkele minuten te bepalen.

Beschikbaarheid

De GAF maakt deel uit van de DSM-IV. De GAF-score is tevens opgenomen in het LVE-codeboek.

Betrouwbaarheid

Er is veel onderzoek gedaan naar de interbeoordelaarsbetrouwbaarheid van de GAF, aangezien dit instrument door de hulpverlener wordt ingevuld. In een review over de psychometrische kenmerken van de DSM-III concluderen Goldman e.a. dat de interbeoordelaarsbetrouwbaarheid van de GAF matig is (1992). Hilsenroth e.a. vinden daarentegen in hun studie naar as V van DSM-IV wel een uitstekende interbeoordelaarsbetrouwbaarheid ($ICC > .74$) (2000). Echter, dit betreffen meestal studies waarbij hulpverleners de GAF invullen op basis van een vignette, of de scoring wordt door onderzoekers gedaan. Vatnaland e.a. tonen aan dat de interbeoordelaarsbetrouwbaarheid van GAF-scores door onderzoekers ingevuld inderdaad prima zijn ($ICC = .81$ en $.85$), maar dat de betrouwbaarheid van het instrument dat is ingevuld door hulpverleners in de dagelijkse praktijk onvoldoende is ($ICC = .39$ en $.59$) (Vatnaland et al., 2007).

Face validiteit

De GAF wordt zeer veel gebruikt in wetenschappelijke studies.

Constructvaliditeit

Goldberg en anderen rapporteren in hun literatuur review dat de GAF een redelijke validiteit heeft, waarbij de scores tussen diverse patiëntengroepen met elkaar vergeleken zijn. Hilsenroth e.a. vinden een matige correlatie met de SCL-90 ($r = -.46$) (Hilsenroth et al., 2000). In een andere studie wordt een matig verband aangetoond tussen de GAF en een instrument dat klinische symptomen meet (Psychiatric Symptom Assessment Scale, PSAS, $r = -.44$), maar slechts een zeer beperkt verband met andere maten van functioneren zoals de Lehman Quality of Life Scale ($r = .22$).

Responsiviteit

Uit een studie van Piersma en Boes (1997) komt naar voren dat er geen enkel significant verschil gevonden werd tussen GAF-scores gemeten bij opname en bij ontslag, bij volwassenen en adolescenten die in een (psychiatrisch) ziekenhuis verbleven. Het Trimbos Instituut heeft in 2008 een inventarisatie uitgevoerd van instrumenten die potentieel geschikt zijn om op te worden genomen in de basisset prestatie-indicatoren in de GGZ en verslavingszorg. Hiervoor is onder andere bij hulpverleners in GGZ-instellingen nagegaan welke instrumenten zij gebruiken, en wat hun ervaringen hiermee zijn. 22 van de 23 hulpverleners die gebruik maakten van de GAF en tijdens de inventarisatie hier een opmerking over maakten, waren ontevreden over dit instrument. Er werd onder andere aangegeven dat de GAF weinig specifieke informatie oplevert en dat het instrument ongevoelig is voor veranderingen (Wijngaarden et al., 2008).

Werkbaarheid	Afnameduur vergt enkele minuten
Zelfrapportage	Nee, wordt door de hulpverlener ingevuld
Beschikbaarheid	Onderdeel DSM-IV, en maakt reeds deel uit van het LVE-codeboek
Betrouwbaarheid	De interbeoordelaarsbetrouwbaarheid wordt wisselend als goed of onvoldoende gerapporteerd
Face validiteit	De GAF wordt zeer veel gebruikt in wetenschappelijke studies
Constructvaliditeit	Matig
Responsiviteit	De GAF lijkt niet goed in staat om veranderingen in het functioneren te registreren. Daarnaast wordt het instrument niet goed beoordeeld door medewerkers in de GGZ.

3.3.3 *Manchester Short Assessment of Quality of Life (MANSA)*

De Manchester Short Assessment of Quality of Life (MANSA; Priebe et al., 1999) is een verkorte versie van de Lancashire Quality of Life Profile (LQLP; Oliver, 1991), en is speciaal ontwikkeld voor het meten van kwaliteit van leven bij patiënten met psychische problematiek. Het instrument bestaat uit 16 items, en meet onder andere de tevredenheid met werk/werkloosheid, met de kwaliteit van vriendschappen en met psychische gezondheid op een zevenpuntsschaal (van 'kan niet slechter' tot 'kan niet beter'). Daarnaast zijn er een viertal vragen die met 'ja' of 'nee' beantwoord worden, zoals de aanwezigheid van een goede vriend of vriendin. De MANSA wordt door de cliënt zelf ingevuld.

Werkbaarheid

De MANSA bevat 16 items. Afnameduur is circa 10 minuten.

Beschikbaarheid

De MANSA is in het Nederlands vertaald door Van Nieuwenhuizen, Schene en Koeter (2000), en gratis te downloaden via [http://www.trimbos.nl/Downloads/Depressie-initiatief/B4%2031%20-%20MANSA%20\(Vragenlijst\).pdf](http://www.trimbos.nl/Downloads/Depressie-initiatief/B4%2031%20-%20MANSA%20(Vragenlijst).pdf)

Betrouwbaarheid

De interne consistentie van de MANSA is voldoende ($\alpha = .74$; Priebe et al., 1999) tot goed ($\alpha = .81$; Bjorkman et al., 2005).

Face validiteit

De MANSA is in enkele wetenschappelijke studies gebruikt.

Constructvaliditeit

Wat betreft de constructvaliditeit werden hoge correlatiecoëfficiënten gevonden tussen de MANSA en de LQLP ($r = >.82$). De relatie met totaalscores op de Brief Psychiatric Rating Scale (BPRS; Ventura et al., 1993) is met $-.49$ matig. Bjorkman e.a. (2005) vonden een positief verband tussen de MANSA en sociaal netwerk, empowerment en psychosocial functioneren. Een negatieve relatie werd gevonden met psychopathologie, zorgbehoefte, ervaren discriminatie en ervaringen van afwijzing.

Responsiviteit

Er is nog geen studie gepubliceerd over de responsiviteit van de MANSA.

Werkbaarheid	16 items, afnameduur circa 10 minuten
Zelfrapportage	Ja
Beschikbaarheid	Nederlandstalige versie beschikbaar, gratis te downloaden
Betrouwbaarheid	Goed: $\alpha = .74 - .81$
Face validiteit	De MANSA is in enkele wetenschappelijke studies gebruikt
Constructvaliditeit	Er is een sterke relatie tussen de MANSA en de LQLP; daarnaast zijn er ook verbanden gevonden met andere relevante meetinstrumenten.
Responsiviteit	Er is nog niets gepubliceerd over de responsiviteit van de MANSA.

3.3.4 (MOS) 36-item Short-Form Health Survey (SF-36)

De Medical Outcomes Study (MOS) 36-item Short-Form Health Survey (SF-36; Ware et al., 1992) is de uitgebreide variant van de MOS SF-20 (Tarlov et al., 1989). De SF-36 heeft 36 items, die samen acht subschalen vormen: fysiek functioneren, rolbeperkingen door een fysiek probleem, lichamelijke pijn, algemene gezondheidsbeleving, vitaliteit, sociaal functioneren, rolbeperkingen door een emotioneel probleem en geestelijke gezondheid. Daarnaast is één item opgenomen over de veranderingen in gezondheid. Afhankelijk van het item, worden vragen beantwoord op een twee-, drie-, vijf- of zespuntsschaal. Een lagere score gaat gepaard met een slechtere gezondheid. De SF-36 wordt door de cliënt zelf ingevuld.

Werkbaarheid

De afnameduur van de SF-36 is tien minuten.

Beschikbaarheid

De SF-36 is vertaald in het Nederlands door Aaronson e.a. (1998). De SF-36 is gratis te downloaden via <http://www.memic.unimaas.nl:8888/daisy/SMILE-CMS/228-SMILE.html?branch=1&language=1>

Betrouwbaarheid

Alle subschalen van de SF-36 in de oorspronkelijke versie laten een hoge interne

consistentie zien, variërend van .78 tot .93 (McHorney et al., 1994). De betrouwbaarheid van de Nederlandse SF-36 is vastgesteld door Aaronson e.a. (1998), waarbij de SF-36 werd afgenomen bij twee groepen uit de algemene bevolking, een groep migrainepatiënten en een groep kankerpatiënten. Alle subschalen lieten een voldoende interne consistentie zien, variërend van .70 tot .93. Alleen de subschaal 'sociaal functioneren' liet een onvoldoende betrouwbaarheid zien bij de groep kankerpatiënten ($\alpha = .66$).

Face validiteit

De SF-36 wordt zeer veel in wetenschappelijk onderzoek gebruikt.

Criteriumvaliditeit

De SF-36 blijkt goed in staat om patiënten met diverse stoornissen en een variërende ernst in stoornissen te onderscheiden (McHorney et al., 1993). Ook in onderzoek naar de validiteit van de Nederlandse SF-36 is de validiteit bepaald door te onderzoeken of bepaalde groepen anders scoren op de SF-36 dan andere groepen. In dit geval blijken ouderen en vrouwen lager te scoren op de SF-36 (d.w.z.: een slechtere gezondheid) dan jongeren en mannen, wat overeenkomt met vooraf opgestelde en op eerder onderzoek gebaseerde hypothesen. Bovendien bleek de SF-36 in staat om respondenten die geen, één, twee, en twee of meer chronische gezondheidsklachten presenteerden van elkaar te onderscheiden (Aaronson et al., 1998).

Responsiviteit

De SF-36 is voldoende responsief (Garraat et al., 1994).

Werkbaarheid	Bevat 36 items; afnameduur bedraagt 10 minuten
Zelfrapportage	Ja
Beschikbaarheid	Nederlandstalige versie
Betrouwbaarheid	Goed: $\alpha = .78 - .93$, m.u.v. de de subschaal 'sociaal functioneren' bij kankerpatiënten
Face validiteit	De SF-36 wordt zeer veel gebruikt in wetenschappelijke studies
Criteriumvaliditeit	De SF-36 is in staat om verschillende respondentengroepen van elkaar te onderscheiden
Responsiviteit	Voldoende

3.3.5 World Health Organization – Disability Assessment Schedule-II (WHODAS-II)

De World Health Organization Disability Assessment Schedule-II (WHODAS-II; Trompenaars et al., 2005) is ontwikkeld om het functioneren in de afgelopen dagen te meten bij respondenten met fysieke en psychische problematiek. De WHODAS-II bevat 36 items, die verdeeld zijn in zes subschalen: communicatie, fysieke mobiliteit, zelfzorg, interpersoonlijke interactie, huishoudelijke verantwoordelijkheden en werk, en participatie in de maatschappij. Elk item wordt gescoord m.b.v. een vijfpuntsschaal, variërend van 'geen problemen' tot 'extreem/ kan niet'. Er is ook een 12-item versie ontwikkeld. De WHODAS kan worden afgenomen tijdens een interview, of zelfstandig door de respondent worden ingevuld.

Werkbaarheid

De afnameduur van de WHODAS-II is 8 tot 10 minuten.

Beschikbaarheid

De WHODAS-II is vertaald in het Nederlands. Het instrument is gratis beschikbaar, en kan worden aangevraagd via de WHO: <http://www.who.int/icidh/whodas/index.html>.

Betrouwbaarheid

De WHODAS-II laat een hoge interne consistentie zien van $\alpha = .95$ voor de gehele schaal, en variërend van $.75 - .91$ voor de subschalen. Alleen de interne consistentie van de subschaal zelfzorg is benedenmaats: α is $.65 - .68$ (Chwastiak et al., 2003). In Nederlands onderzoek, waarbij overigens een zelfgeconstrueerde versie van 22 items van de WHODAS-II gebruikt werd, blijkt de interne consistentie van de subschalen hoog: $\alpha = .89 - .94$, ook voor de subschaal zelfzorg (Buist-Bouwman et al., 2008).

Face validiteit

De WHODAS-II wordt in enkele wetenschappelijke studies gebruikt, waaronder de Nederlandse Studie naar Depressie en Angst (NESDA) (Penninx et al., 2008), en in de Australische National Survey of Mental Health and Wellbeing (2007).

Construct- en criteriumvaliditeit

Er worden hoge correlaties gevonden tussen scores op de WHODAS-II en de SF-36: $-.71$ en $-.72$ (Chwastiak et al., 2003). Een matig verband is gerapporteerd met de QWB: $r = .38 - .63$ (Quality of Well-Being; Kaplan et al., 1989). Wat betreft discriminerende validiteit wordt gevonden dat de WHODAS-II goed in staat is om onderscheid te maken tussen patiëntengroepen zonder DSM-diagnose, patiëntengroepen met een life-time DSM-diagnose en degenen met een zesmaands DSM-diagnose (Buist-Bouwman et al., 2008).

Responsiviteit

Respondenten laten tijdens de driemaands-follow-up een significante afname zien van hun score op de WHODAS-II. Dit instrument blijkt net zo goed in staat om veranderingen in het functioneren te registreren als de SF-36 (Chwastiak et al., 2003). Ook Perini e.a. bevestigen dat de responsiviteit van de WHODAS-II tenminste net zo goed is als die van de SF-12 en Kessler-10 in het meten van veranderingen bij angstpatiënten (Perini et al., 2006).

Werkbaarheid	Bevat 36 of 12 items; de afnameduur bedraagt 8-10 minuten voor de 36-item-versie
Zelfrapportage	Ja
Beschikbaarheid	Vertaald in het Nederlands; gratis beschikbaar
Betrouwbaarheid	Goed: $\alpha = .75 - .94$. De interne consistentie van de schaal 'zelfzorg' is onvoldoende in Amerikaans onderzoek, maar voldoende in een Nederlandse studie.
Face validiteit	Enkele wetenschappelijke studies maken gebruik van de WHODAS-II
Construct- en criteriumvaliditeit	Discriminerende en constructvaliditeit zijn goed.
Responsiviteit	De WHODAS-II is voldoende responsief om veranderingen te registreren

3.3.6 WHO Quality of Life – BREF (WHOQOL-BREF)

WHO Quality of Life – BREF (WHOQOL-BREF; The WHOQOL Group, 1998) is een generiek instrument ontwikkeld door de World Health Organization, en meet de kwaliteit van leven. Resultaten zijn cross-cultureel vergelijkbaar. De oorspronkelijke WHOQOL bestond uit 100 items, maar de kortere WHOQOL-BREF bevat 26 items. De items zijn verdeeld in vier domeinen: fysieke gezondheid, mentale gezondheid, sociale relaties en omgeving, en worden door de cliënt zelf beantwoord op een vijf-puntsschaal.

Werkbaarheid

De WHOQOL-BREF bevat 26 items. De afnameduur bedraagt tussen de 10 en 15 minuten.

Beschikbaarheid

Het instrument is vertaald in het Nederlands (Vries et al., 1996), en gratis beschikbaar via <http://www.tijdschriftvoorpsychiatrie.nl/meetinstrumenten/info.php?id=126>.

Betrouwbaarheid

In een Nederlands onderzoek naar de validiteit van de WHOQOL-BREF bij psychiatrische patiënten die ambulante zorg ontvingen blijkt dat de interne consistentie van de domeinen fysiek en mentaal functioneren en omgeving voldoende is: Cronbach's alpha is respectievelijk .80, .74 en .73. Alleen het domein sociale relaties scoort onder de maat: $\alpha = .66$ (Trompenaars et al., 2005).

Face validiteit

De WHOQOL-BREF wordt veel gebruikt in wetenschappelijke studies.

Constructvaliditeit

In een Nederlands onderzoek naar de validiteit van de WHOQOL-BREF bij psychiatrische patiënten werden scores op de WHOQOL vergeleken met de SCL-90 (Trompenaars et al., 2005). Er wordt een matig verband gevonden tussen de GSI (totaalscore SCL-90) en het WHOQOL-domein fysiek functioneren ($r = -.49$) en het domein mentaal functioneren ($r = -.44$). Een sterker verband wordt gevonden wanneer specifiek gekeken wordt naar de relatie tussen het domein fysiek functioneren en de schaal 'somatiseren' van de SCL-90 ($r = -.59$) en tussen het domein mentaal functioneren en de schaal 'depressie' ($r = -.62$). Het verband met de domeinen sociale relaties en omgeving en de GSI is laag ($r = -.21$ respectievelijk $-.35$). Een sterker verband worden gevonden tussen het domein sociale relaties en de PSSS (Perceived Social Support Scale; Procidano et al., 1983) ($r = .56$).

Responsiviteit

De sensitiviteit van de WHOQOL-BREF is niet onderzocht in een sample van patiënten met psychische klachten. Wel duidt onderzoek bij Taiwanese ouderen en reumapatiënten in Nieuw-Zeeland er op dat de responsiviteit goed is, met name van het domein fysiek functioneren (Hwang et al., 2003; Taylor et al., 2004).

Werkbaarheid	Afnameduur bedraagt 10-15 minuten
Zelfrapportage	Ja
Beschikbaarheid	Nederlandstalige versie beschikbaar, gratis
Betrouwbaarheid	Goed, behalve voor het domein 'sociale relaties'
Face validiteit	De WHOQOL-BREF wordt veel gebruikt in wetenschappelijke studies.
Constructvaliditeit	Er worden lage tot matige correlaties gevonden met de SCL-90-subschalen en sociale steun (PSSS)
Responsiviteit	Responsiviteit is goed, maar nog niet onderzocht bij personen met psychische klachten.

4 Kleinschalig onderzoek naar de validiteit van de Outcome Rating Scale (ORS)

Van de meeste instrumenten die beschreven werden in hoofdstuk 3, zijn meerdere studies gepubliceerd die een uitspraak doen over de psychometrische eigenschappen. Een belangrijke uitzondering hier op wordt gevormd door de Outcome Rating Scale (ORS) van Scott Miller. De enige publicaties die een uitspraak doen over de betrouwbaarheid en validiteit van de ORS zijn van Scott Miller zelf. Dit staat in schril contrast met de grote populariteit die het instrument geniet onder eerstelijnspsychologen: uit de resultaten van onze enquête onder eerstelijnspsychologen blijkt dat de ORS op de zesde plaats staat van de meest gebruikte instrumenten. Bovendien wordt de ORS het frequentst geadviseerd als geschikte uitkomstmaat door degenen die het instrument (nog) niet gebruiken.

De ORS lijkt een betrouwbaar instrument te zijn: er worden alpha's gepresenteerd van .93 en .97 (Miller et al., 2003; Bringham et al., 2006). De validiteit van de ORS is echter uitsluitend onderzocht in relatie tot de Outcome Questionnaire, waarvan de ORS is afgeleid. Juist de validiteit van een instrument is zeer belangrijk: het is van het grootste belang dat men weet welk construct een instrument precies meet. Op dit moment is er dan ook te weinig informatie om een goede uitspraak te kunnen doen over de kwaliteit van de ORS, hoewel er in de praktijk veel vraag naar is.

Daarom is besloten om een kleinschalig onderzoek uit te voeren naar de samenhang van de ORS met de SCL-90: het instrument dat het meest gebruikt wordt door eerstelijnspsychologen, en waarvan de goede psychometrische eigenschappen ruim bewezen zijn. De Outcome Rating Scale is toegevoegd in bijlage 3.

Om een uitspraak te kunnen doen over de samenhang tussen de ORS en de SCL-90, zijn alle eerstelijnspsychologen benaderd die in de eerder genoemde enquête hadden aangegeven de SCL-90 te gebruiken. Hen werd gevraagd om bij de eerste tien nieuwe cliënten die zich bij hen aanmelden, zowel een ORS als een SCL-90 af te nemen tijdens het eerste gesprek / intake. Ook werden een viertal achtergrondkenmerken van de cliënt in kaart gebracht: leeftijd, geslacht, opleidingsniveau en de voornaamste klacht waarmee de cliënt zich aanmeldt. 191 psychologen zijn met dit verzoek aangeschreven, waarvan er precies 50 hun deelname hebben toegezegd. Uiteindelijk hebben 40 psychologen gegevens aangeleverd, van in totaal 267 cliënten.

Van de cliënten van wie gegevens zijn aangeleverd voor dit onderzoek was 66% vrouw (N=177). De gemiddelde leeftijd van de cliënten bedroeg 39 jaar (sd=.83). Circa 50% heeft een diploma behaald in het hoger onderwijs. Het overgrote deel van de cliënten betrof volwassenen; zij melden zich hoofdzakelijk bij de eerstelijnspsycholoog met problemen met aanpassing en verwerking (19.5%), interpersoonlijke problemen (17.2%),

depressie (14.2%) of angst (14.2%). Deze cliëntengroep vormt een goede afspiegeling van de totale cliëntengroep die de eerstelijnspsycholoog bezoekt (Jaarbericht LVE, 2008).

In de eerste plaats is de interne betrouwbaarheid van de ORS berekend. Deze is met $\alpha = .82$ lager dan de resultaten die door Scott Miller e.a. zelf gepubliceerd zijn (Miller et al., 2003; Bringhurst et al., 2006), maar is zeer acceptabel.

Tabel 4.1 toont de samenhang tussen de ORS en SCL-90. Uit de resultaten blijkt dat er hoofdzakelijk lage ($r = .20 - .40$) tot matige correlaties ($r = .40 - .70$) bestaan tussen de SCL-90 en de ORS. Er is sprake van een negatieve correlatie, omdat een hogere score op de ORS (een hogere score is gerelateerd aan een beter functioneren) samenhangt met een lagere score op de SCL-90 (een lagere score is gerelateerd aan beter functioneren). Ter verduidelijking zijn verbanden vanaf $r = -.40$ dikgedrukt.

Tabel 4.1 Pearsons correlatiecoëfficiënten die de samenhang tussen de Outcome Rating Scale (ORS) en de Symptom Checklist-90 (SCL-90) uitdrukken. Correlaties vanaf $r = -.40$ zijn dikgedrukt.

	Outcome Rating Scale (ORS)				
	Individueel functioneren	Relationeel functioneren	Sociaal functioneren	Algeheel functioneren	Totaalscore
<i>Symptom Checklist-90</i>					
Angst (AN)	-.46*	-.23*	-.40*	-.49*	-.48*
Agorafobie (AGO)	-.34*	-.14*	-.33*	-.33*	-.35*
Depressie (DEP)	-.54*	-.40*	-.51*	-.61*	-.63*
Somatische klachten (SOM)	-.37*	-.16*	-.42*	-.41*	-.41*
Insufficiëntie van denken en handelen (IN)	-.43*	-.34*	-.53*	-.51*	-.56*
Sensitiviteit en wantrouwen (SEN)	-.39*	-.38*	-.49*	-.43*	-.52*
Hostiliteit (HOS)	-.28*	-.22*	-.39*	-.33*	-.38
Slaapproblemen (SLA)	-.37*	-.16*	-.28*	-.35*	-.35*
Overig (OVER)	-.31*	-.32*	-.42*	-.43*	-.46*
Totaalscore	-.53*	-.39*	-.55*	-.58*	-.63*

* $p < .05$

Uit tabel 4.1 blijkt dat er een duidelijk verband bestaat tussen de totaalscore op de SCL-90, en de items individueel functioneren ($r = -.53$), relationeel functioneren ($r = -.39$), sociaal functioneren ($r = -.55$) en algeheel functioneren ($r = -.58$). Het sterkste verband bestaat tussen de totaalscores van de ORS en de SCL-90: $r = -.63$. Daarnaast is zichtbaar dat er duidelijke verbanden bestaan tussen de items van de ORS en de schalen angst, depressie, insufficiëntie van denken en handelen en sensitiviteit en wantrouwen van de SCL-90. Ook wordt duidelijk dat de items van de ORS een sterker verband vertonen met soortgelijke schalen van de SCL-90. Relationeel functioneren (ORS) blijkt bijvoorbeeld sterker gerelateerd te zijn aan de SCL-90-schaal sensitiviteit en wantrouwen ($r = -.38$), dan aan de SCL-90-schaal somatische klachten ($r = -.16$). Tevens valt op dat de items 'sociaal functioneren' en 'algeheel functioneren' van de ORS een relatief sterke samenhang vertonen met veel schalen van de SCL-90 ($r = -.28$ tot en met $-.55$ en $r = -.33$ tot en met $-.61$, respectievelijk).

Gezien de gevonden correlaties wordt gesproken over een matig verband. Bij het interpreteren van de gegevens is het van belang om te realiseren dat de ORS specifiek ontwikkeld is voor het routinematig vastleggen van veranderingen in het functioneren van de cliënt: het instrument is gemaakt om na afloop van elk consult ingevuld te worden. Vanwege de kleinschalige opzet van dit onderdeel van het onderzoek, is de ORS slechts éénmaal afgenomen, en zijn de scores vergeleken met scores op de SCL-90: een instrument dat zeer uitgebreid het functioneren in kaart brengt, en duidelijk niet bedoeld is om na afloop van elke zitting afgenomen te worden. Met deze kanttekening in gedachten, kan gesteld worden dat de ORS redelijk goed in staat is om een globale indruk te geven van het functioneren van de cliënt. De ORS-totaalscore vertoont de hoogste samenhang met de totaalscore van de SCL-90.

5 Discussie

In dit rapport werden de psychometrische eigenschappen en andere relevante kenmerken van 21 generieke meetinstrumenten beschreven, die potentieel geschikt zijn om uitkomsten te meten in de eerstelijnspsychologische zorg. Het betreffen instrumenten die symptomen of klachten meten, instrumenten die speciaal ontwikkeld werden om op routinematige basis uitkomsten van interventies in kaart te brengen, en het gaat om instrumenten die de kwaliteit van leven van cliënten meten. Sommige instrumenten worden door de behandelaar ingevuld; de meeste instrumenten vragen de cliënt om het niveau van klachten, functioneren of kwaliteit van leven zelf in kaart te brengen.

Uit voorgaande hoofdstukken blijkt dat de verscheidenheid aan en kwaliteit van instrumenten groot is. De instrumenten zijn op diverse criteria beoordeeld: de werkbaarheid (het aantal items en de afnameduur); de beschikbaarheid (waar is het instrument verkrijgbaar, tegen welke prijs, en is het vertaald in het Nederlands?); de betrouwbaarheid; de validiteit (meet het instrument wat het beoogt te meten?); en de responsiviteit (is het instrument in staat om veranderingen in symptoomniveau, het functioneren of de kwaliteit van leven van de cliënt in kaart te brengen?).

Het instrument dat bij uitstek geschikt is om uitkomsten te meten in de eerstelijnspsychologische praktijk, voldoet aan al deze criteria: het is een gedegen instrument met een goede betrouwbaarheid en validiteit en het is goed beschikbaar en in korte tijd af te nemen. Bovendien is het een instrument dat niet alleen tijdens de intake en de laatste zitting wordt afgenomen om de uitkomst van de behandeling in kaart te brengen, maar het instrument is tevens tussentijds inzetbaar: het is voldoende responsief. Hierdoor wordt de eerstelijnspsycholoog de mogelijkheid geboden om de behandeling tijdig bij te sturen, mocht dit nodig zijn.

Tabel 5.1 toont een beknopt overzicht van de score van alle besproken instrumenten op de geformuleerde criteria. Negen van de 21 vragenlijsten hebben meer dan 15 items, en zijn daarom minder geschikt om herhaaldelijk af te kunnen nemen. Het betreft hier onder andere de SCL-90 en de 4DKL, die beiden goede psychometrische eigenschappen laten zien. Voor 7 van de 21 instrumenten moet betaald worden voor het gebruik, waardoor zij vermoedelijk niet de voorkeur verdienen om op grote schaal, en op routinematige basis door eerstelijnspsychologen afgenomen te worden. Het gaat bijvoorbeeld om de CORE-OM, die specifiek ontwikkeld is voor routine outcome monitoring, en de Korte Klachten Lijst. De GHQ wordt niet meer uitgegeven. Daarnaast laten verschillende instrumenten een onvoldoende betrouwbaarheid, validiteit of responsiviteit zien. De meest in het oog springende instrumenten met een onvoldoende betrouwbaarheid én responsiviteit, zijn de GAF en de HoNOS. Opvallend, omdat zij beiden zeer veel gebruikt worden in de

Tabel 5.1: Overzicht van de mate waarin de generieke uitkomstmaten voldoen aan de gestelde criteria

Instrument	Aantal items	Afnameduur in minuten	Zelfrapportage	Nederlandstalige versie	Beschikbaarheid	Betrouwbaarheid ^a	Face validiteit ^d	Construct-validiteit ^e	Responsiviteit
<i>Symptomen/klachten</i>									
4DKL - 4 dimensionele klachten lijst	50	10	Ja	Ja	Gratis	+ / ++	?	++	+
BSI – Brief Symptom Inventory	53	10	Ja	Ja	€1,64 p/s	+ / ++	++	+/- / + / ++	+
CGI - Clinical Global Impression	2	1	Nee	Ja	Gratis	+ / ++ ^b	+	-- / - / + / ++	?
GHQ - General Health Questionnaire	30/28/12	10/5	Ja	Ja	Niet meer leverbaar	+ / ++	+	- / +/- / +	+
HADS - Hospital Anxiety Depression Scale	14	5	Ja	Ja	Gratis	- / + / ++	++	+	+
KKL – Korte Klachten Lijst	14	Enkele minuten	Ja	Ja	€1,10 p/s	+	?	+	+
MHI – Mental Health Inventory	5	2-3	Ja	Ja	Gratis	+	+	+/-	+
PHQ – Patient Health Questionnaire	59/28	15/10	Ja	Ja	Gratis	+ / ++	+	+	+
SCL-90 – Symptom Checklist-90	90/45	20/10	Ja	Ja	€1,55 p/s	+ / ++	++	+/- / + / ++	+

Instrument	Aantal items	Afnameduur in minuten	Zelfrapportage	Nederlandstalige versie	Beschikbaarheid	Betrouwbaarheid ^a	Face validiteit ^d	Construct-validiteit ^e	Responsiviteit
<i>Routine Outcome Monitoring</i>									
CORE-OM Clinical Outcomes in Routine Evaluation – Outcome Measures	34/18/10/5	10 / enkele minuten	Ja	Ja	Uitsluitend digitaal verkrijgbaar via Mentaal Beter Services Via CareMetrics	+ / ++	?	+/- / +	+
ELM – Eerstelijnsmonitor*	Max. 124	10	Ja	Ja		- ^c / + / ++	?	+	?
HoNOS – Health of Nation Outcome Scales	12	5-15	Nee	Ja		- / + ^b	?	+/- / + / ++	-
OQ – Outcome Questionnaire	45/30	5	Ja	Ja	\$1,- p/cliënt	- / + / ++	+	- / +/- / +	+
ORS – Outcome Rating Scale*	4	Enkele minuten	Ja	Ja	Gratis	+ / ++	?	+/- / +	+
<i>Functioneren/kwaliteit van leven</i>									
EQ-5D – EuroQol-5D	5	Enkele minuten	Ja	Ja	Aanvragen, kosten afhankelijk van gebruik	+	+	+/-	+
GAF – Global Assessment of Functioning	10	Enkele minuten	Nee	Ja		Onderdeel DSM-IV-TR	++	- / + ^f	-

Instrument	Aantal items	Afnameduur in minuten	Zelfrapportage	Nederlandstalige versie	Beschikbaarheid	Betrouwbaarheid^a	Face validiteit^d	Construct-validiteit^e	Responsiviteit
MANSA – Manchester Short Assessment of Quality of Life	16	10	Ja	Ja	Gratis	+	?	+/- / +	?
SF-36 36-item Short Form Health Survey	36	10	Ja	Ja	Gratis	+ / ++	++	+ ^f	+
Whodas-II WHO Disability Assessment Schedule	36/12	8-10	Ja	Ja	Gratis	- / + / ++	?	- / + / ++	+
WHOQOL-BREF WHO Quality of Life – BREF	26	10-15	Ja	Ja	Gratis	- / +	+	- / +/-	+

a Tenzij anders aangegeven, betreft het hier de interne consistentie (α): - ($\alpha < .70$), + ($\alpha = .70 - .90$), ++ ($\alpha > .90$)

b Betreft alleen de subschaal 'passieve coping'

c Het betreft de interbeoordelaarsbetrouwbaarheid: - (ICC < .50), + (ICC = .50-.75), ++ (ICC > .75)

d Face validiteit: beoordeeld middels de frequentie waarmee een instrument gebruikt wordt in wetenschappelijke studies? + veel; ++ zeer veel; ? weinig studies, face validity daardoor onbekend

e Constructvaliditeit: -- ($r < .20$), - ($r = .20 - .40$), +/- ($r = .40 - .70$), + ($r = .70 - .90$), ++ ($r > .90$), volgens Guilford (1965).

f Criteriumvaliditeit: onvoldoende (-) of voldoende (+) in staat om diverse subgroepen van elkaar te onderscheiden

* psychometrische eigenschappen zijn alleen door de auteur van het instrument (en eventuele collega's) onderzocht

klinische praktijk: de GAF vormt een onderdeel van de DSM-IV en is daardoor breed ingevoerd en goed voorhanden. De HoNOS is in Engeland ontwikkeld in samenwerking met de Engelse overheid, en wordt daar zeer veel toegepast in de praktijk. Beide zijn overigens instrumenten die door de hulpverlener worden ingevuld. Tevens zijn er instrumenten waarvan de psychometrische eigenschappen nog weinig onderzocht zijn, waaronder de ELM en de ORS. Ons aanvullend onderzoek liet overigens zien dat er een matige samenhang bestaat tussen scores op de ORS en de SCL-90, waardoor verondersteld kan worden dat de ORS redelijk in staat is om op globale wijze het functioneren van de cliënt in kaart te brengen.

Incidenteel negatieve ervaringen met betrouwbaarheid of validiteit werden ook gemeld over GHQ, CGI, WHODAS, HADS en voor een subschaal van de WHOQOL-BREF, ELM en OQ-45.

Kortom, het is lastig om één instrument aan te wijzen dat bij uitstek geschikt is als uitkomstindicator in de eerstelijnspsychologie. Wel wordt uit bovenstaande duidelijk dat er enkele instrumenten kunnen worden aangeraden, zoals de GAF en de CGI. De Landelijke Vereniging van Eerstelijnspsychologen kan dit rapport gebruiken om zelf een keuze te maken voor uitkomstindicatoren die zij geschikt en zinvol vindt voor gebruik in de eerstelijnspsychologie, en de toepassing hiervan in de dagelijkse praktijk kan adviseren aan haar leden. Het gebruik van uitkomstindicatoren kan de eerstelijnspsycholoog inzicht bieden in de resultaten van de verstrekte behandeling, zowel op individueel als landelijk niveau. Bovendien stelt het de eerstelijnspsycholoog in staat om, zonder direct de resultaten van de uitkomstmetingen beschikbaar te stellen, zorgverzekeraars te tonen dat er door het gebruik van uitkomstindicatoren extra kwaliteit geboden wordt in de behandeling van cliënten. Uit onderzoek blijkt dat tussentijds gebruik van uitkomstmaten, waarbij de resultaten door behandelaar en cliënt samen besproken worden, kan leiden tot een verhoogde effectiviteit van de behandeling. Het stelt de behandelaar namelijk in staat om tijdig te ontdekken wanneer een cliënt geen of onvoldoende vooruitgang boekt, en geeft mogelijkheden om de behandeling bij te sturen. Onderzoek met herhaaldelijke afname van de OQ-45 en terugkoppeling van de resultaten naar de therapeut leidde tot een verbeterde therapie-uitkomst, vergeleken met cliënten wiens therapeut geen feedback ontving (Lambert et al., 2001). Reese e.a. (2009) toonden aan dat cliënten die tijdens elke behandelsessie de Outcome Rating Scale (ORS) en de Session Rating Scale (SRS; bedoeld om de tevredenheid van de cliënt over de behandelaar en de sessie in kaart te brengen) invulden, een significant betere uitkomst lieten zien na afloop van de therapie. In 2010 zal de LVE een advies uitbrengen naar haar leden over geschikte uitkomstindicatoren. Hierdoor stelt zij zich in staat om als beroepsgroep extra kwaliteit in haar zorgaanbod te bieden, en in de toekomst de kwaliteit van de eerstelijnspsychologische zorg kan presenteren aan geïnteresseerden, waaronder zorgverzekeraars.

Bijlage 1: Enquête onder eerstelijnspsychologen

1. Welke gestandaardiseerde meetinstrumenten worden regelmatig door u gebruikt om de (voorlopige) uitkomst of het resultaat van uw therapeutisch handelen vast te leggen? U kunt hierbij denken aan bijvoorbeeld de Symptom Checklist-90 (SCL-90), Routine Outcome Measurement (ROM), de Beck Depression Inventory (BDI) etc.
2. Wilt u van elk genoemd meetinstrument aangeven of u dit instrument routinematig of in voorkomende gevallen gebruikt?

	Frequentie van gebruik	
Meetinstrument	Routinematig / altijd	In voorkomende gevallen
.....		
.....		
....		

3. Wilt u van elk genoemd meetinstrument aangeven of u dit instrument voor diagnostische doeleinden gebruikt, of voor het meten van de uitkomst van uw therapeutisch handelen?

	Doeleinde gebruik	
Meetinstrument	Diagnostisch	Uitkomst vastleggen
.....		
.....		
....		

4. Zijn er volgens u nog andere meetinstrumenten geschikt als outcome- en prestatie-indicatoren, die u nog niet genoemd heeft?

Bijlage 2: Resultaten van enquête onder eerstelijnspsychologen

Tabel 1 Top-20 van veelgebruikte meetinstrumenten door eerstelijnspsychologen (N=271)

Naam instrument	Naam instrument voluit	N	%
SCL-90	Symptom Checklist-90	209	77
BDI	Beck's depression inventory	112	41
UCL	Utrechtse Coping Lijst	102	38
NPV	Nederlandse Persoonlijkheidsvragenlijst	84	31
NVM	Nederlandse Verkorte MMPI	51	19
ORS	Outcome Rating Scale	46	17
SRS	Session Rating Scale	38	14
UBOS	Utrechtse BurnOut Schaal	36	13
MMPI	Minisota Multiphasic Personality Inventory	23	8
4DKL	4-dimensionele klachtenlijst	23	8
NRV	Nederlandse Relatie Vragenlijst	20	7
SIG	Schaal voor Interpersoonlijk Gedrag	18	7
NPST	Negativisme, Ernstige Psychopathologie en Somatisering Test	18	7
CBCL	Child Behavior Checklist	17	6
Zung Depressie	Zung Self-Rating Depression Scale	15	6
YSR / YRF	Youth Self Report Form	12	4
Eigen vragenlijst / methode		11	4

Tabel 2 Frequentie waarmee veelgenoemde meetinstrumenten gebruikt worden door eerstelijnspsychologen

Instrument (afkorting)	Instrument (voluit)	routinematig / altijd	in voorkomende gevallen
		N	N
SCL-90	Symptom Checklist-90	99	108
UCL	Utrechtse Coping Lijst	41	60
ORS	Outcome Rating Scale	34	11
SRS	Session Rating Scale	26	12
NPV	Nederlandse Persoonlijkheidsvragenlijst	24	58
NVM	Nederlandse Verkorte MMPI	24	26
4DKL	4-dimensionele klachtenlijst	11	12
BDI	Beck's depression inventory	8	102
MMPI	Minisota Multiphasic Personality Inventory	6	16
CBCL	Child Behavior Checklist	6	11
Eigen vragenlijst / methode		6	4
ROM	Routine Outcome Measurement (not specified)	5	2
Zung Depressie	Zung Self-Rating Depression Scale	4	11
NPST	Negativisme, Ernstige Psychopathologie en Somatisering Test	3	15
Young Schema Vragenlijst		3	6
SEV	Sociaal Emotionele Vragenlijst	3	5
OQ-45	Outcome-Questionnaire-45	3	3
YSR / YRF	Youth Self Report Form	2	10
TRF	Teacher Report Form	2	7
CDI	Child depression inventory	2	6
CBSK-CBSA	Competentie BelevingsSchaal voor Kinderen / Adolescenten	2	6

Tabel 3: Doeleinde van veelgebruikte meetinstrumenten door eerstelijnspsychologen

		Doeleinde gebruik								
		Alleen diagnostisch		Alleen uitkomst vastleggen		Beide		Geen van beide		Totaal
Naam instrument	Naam instrument (voluit)	N	%	N	%	N	%	N	%	N
SCL-90	Symptom Checklist-90	90	43	13	6	100	48	6	3	209
BDI	Beck's depression inventory	60	54	5	5	42	38	5	5	112
UCL	Utrechtse Coping Lijst	81	79	1	1	14	14	6	6	102
NPV	Nederlandse Persoonlijkheidsvragenlijst	75	89	0	-	4	5	5	6	84
NVM	Nederlandse Verkorte MMPI	46	90	0	-	2	4	3	6	51
ORS	Outcome Rating Scale	0	-	34	74	10	22	2	4	46
SRS	Session Rating Scale	3	8	25	66	8	21	2	5	38
UBOS	Utrechtse BurnOut Schaal	22	61	4	11	8	22	2	6	36
MMPI	Minisota Multiphasic Personality Inventory	21	91	0	-	1	4	1	4	23
4DKL	4-dimensionele klachtenlijst	5	22	4	17	14	61	0	-	23
NRV	Nederlandse Relatie Vragenlijst	15	75	2	10	2	10	1	5	20
NPST	Negativisme, Ernstige Psychopathologie en Somatisering Test	18	100	0	-	0	-	0	-	18
SIG	Schaal voor Interpersoonlijk Gedrag	16	89	1	6	1	6	0	-	18
CBCL	Child Behavior Checklist	7	41	0	-	8	47	2	12	17
Zung Depressie	Zung Self-Rating Depression Scale	6	40	1	7	8	53	0	-	15
YSR / YRF	Youth Self Report Form	8	67	0	-	3	25	1	8	12
eigen vragenlijst / methode		1	9	6	55	2	18	2	18	11
Young Schema Vragenlijst		9	100	0	-	0	-	0	-	9
PMT	Prestatie Motivatie Test	8	89	0	-	1	11	0	-	9
TRF	Teacher Report Form	7	78	0	-	2	22	0	-	9
CDI	Child depression inventory	4	50	1	13	2	25	1	13	8

Tabel 4: Instrumenten die eerstelijnspsychologen geschikt achten voor het meten van uitkomstmaten, die zij zelf niet (regelmatig) gebruiken

Naam instrument	Naam instrument voluit	N
ORS	Outcome Rating Scale	22
SRS	Session Rating Scale	13
eigen vragenlijst/methode		6
ELM	Eerstelijnsmonitor	6
SCL-90	Symptom Checklist-90	6
UBOS	Utrechtse Burn-Out Schaal	4
4DKL	4-Dimensionele Klachtenlijst	3
BDI	Beck's Depression Inventory	3
IOA	Inventarisatie Omgaan met Anderen	3
OQ-45	Outcome Questionnaire-45	3
ROM	(Routine Outcome Monitoring; niet gespecificeerd)	3

Bijlage 3: De Outcome Rating Scale (ORS)

Outcome Rating Scale: Hoe gaat het met u?

Datum: _____

Hoe is het met u gegaan de afgelopen week, of sinds het laatste behandelcontact -
inclusief vandaag -?

Zet op elke lijn een kruisje. Links is 'laag' en rechts is 'hoog'.

Individueel

(persoonlijk welbevinden)

I-----I

Relationeel

(familie, intieme vrienden)

I-----I

Sociaal

(werk, opleiding, sociale contacten)

I-----I

Algeheel

(algemeen welbevinden)

I-----I

Institute for the Study of Therapeutic Change www.talkingcure.com

© 2000, Scott D. Miller, Barry L. Duncan, & Lynn Johnson

vertaling V.4.0: Frank Asmus, Mark Crouzen & Flip Jan van Oenen

Email: vertaling@yahoo.com

Referentielijst

- Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R et al. *Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations.* J Clin Epidemiol, 1998; 51(11): p.1055-1068
- American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (4th ed.)*. American Psychiatric Association, 1994;
- Appels A, Höppener P, Mulder P. *A questionnaire to assess premonitory symptoms of myocardial infarction.* International Journal of Cardiology, 1987; 17: p.15-24
- Arrindell WA, Ettema JHM. *Handleiding bij een multidimensionele psychopathologie-indicator*. Lisse: Swets, Test Publishers, 1986
- Audin K, Margison FR, Clark JM, Barkham M. *Value of HoNOS in assessing patient change in NHS psychotherapy and psychological treatment services.* Br J Psychiatry, 2001; 178: p.561-566
- Barkham M, Margison F, Leach C, Lucock M, Mellor-Clark J, Evans C et al. *Service profiling and outcomes benchmarking using the CORE-OM: toward practice-based evidence in the psychological therapies. Clinical Outcomes in Routine Evaluation-Outcome Measures.* J Consult Clin Psychol, 2001; 69(2): p.184-196
- Bebbington P, Brugha T, Hill T, Marsden L, Window S. *Validation of the Health of the Nation Outcome Scales.* Br J Psychiatry, 1999; 174: p.389-394
- Berk M, Ng F, Dodd S, Callaly T, Campbell S, Bernardo M et al. *The validity of the CGI severity and improvement scales as measures of clinical effectiveness suitable for routine clinical use.* J Eval Clin Pract, 2008; 14(6): p.979-983
- Beurs Ed, Dyck Rv, Marquenie LA, Lange A, Blonk RWB. *De DASS: een vragenlijst voor het meten van depressie, angst en stress.* Gedragstherapie, 2001; 34: p.35-53
- Bjelland I, Dahl AA, Haug TT, Neckelmann D. *The validity of the Hospital Anxiety and Depression Scale. An updated literature review.* J Psychosom Res, 2002; 52(2): p.69-77
- Bjorkman T, Svensson B. *Quality of life in people with severe mental illness. Reliability and validity of the Manchester Short Assessment of Quality of Life (MANSA).* Nord J Psychiatry, 2005; 59(4): p.302-306
- Bringinghurst DL, Watson CW, Miller SD, Duncan BL. *The reliability and validity of the Outcome Rating Scale: a replication study of a brief clinical measure.* Journal of Brief Therapy, 2006; 5(1): p.23-30
- Brooks R. *EuroQol: The current state of play.* Health Policy, 1996; 37: p.53-72
- Buist-Bouwman MA, Ormel J, de Graaf R, Vilagut G, Alonso J, Van Sonderen E et al. *Psychometric properties of the World Health Organization Disability Assessment Schedule used in the European Study of the Epidemiology of Mental Disorders.* Int J Methods Psychiatr Res, 2008; 17(4): p.185-197

- Cameron IM, Crawford JR, Lawton K, Reid IC. *Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care*. Br J Gen Pract, 2008; 58(546): p.32-36
- Chwastiak LA, Von Korff M. *Disability in depression and back pain: evaluation of the World Health Organization Disability Assessment Schedule (WHO DAS II) in a primary care setting*. J Clin Epidemiol, 2003; 56(6): p.507-514
- Colsen PJA, Casparie AF. *Indicatorregistratie: een model ten behoeve van integrale kwaliteitszorg in een ziekenhuis*. Medisch Contact, 1995; 50: p.297-299
- Cuijpers P, Smits N, Donker T, Ten Have M, de Graaf R. *Screening for mood and anxiety disorders with the five-item, the three-item, and the two-item Mental Health Inventory*. Psychiatry Res, 2009; 168(3): p.250-255
- de Beurs E, Zitman F. *De Brief Symptom Inventory (BSI)*. Maandblad Geestelijke Volksgezondheid, 2006; 61(02): p.120-141
- Derogatis LR, Lipman RS, Covi L. *SCL-90: An outpatient psychiatric rating scale: preliminary report*. Psychopharmacology Bulletin, 1973; 9: p.13-27
- Derogatis LR, Melisaratos N. *The Brief Symptom Inventory: an introductory report*. Psychol Med, 1983; 13(3): p.595-605
- Evans C, Connell J, Barkham M, Margison F, McGrath G, Mellor-Clark J et al. *Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM*. Br J Psychiatry, 2002; 180: p.51-60
- Evans C, Mellor-Clark J, Margison F, Barkham M, Audin K, Connell J et al. *CORE: Clinical outcomes in routine evaluation*. Journal of Mental Health, 2000; 9: p.247-255
- Goldberg DP. *The detection of psychiatric illness by questionnaire*. London: Oxford University Press, 1972
- Goldberg DP, Gater R, Sartorius N, Ustun TB, Piccinelli M, Gureje O et al. *The validity of two versions of the GHQ in the WHO study of mental illness in general health care*. Psychol Med, 1997; 27(1): p.191-197
- Goldman HH, Skodol AE, Lave TR. *Revising axis V for DSM-IV: a review of measures of social functioning*. Am J Psychiatry, 1992; 149(9): p.1148-1156
- Guy W. *ECDEU Assessment Manual for Psychopharmacology - Revised*. Rockville, MD: Nation Institute for Mental Health, 1976
- Guilford JP. *Fundamental Statistics in Psychology and Education*. New york: McGraw-Hill, 1965
- Hilsenroth MJ, Ackerman SJ, Blagys MD, Baumann BD, Baity MR, Smith SR et al. *Reliability and validity of DSM-IV axis V*. Am J Psychiatry, 2000; 157(11): p.1858-1863
- Hoeymans N, Garssen AA, Westert GP, Verhaak PF. *Measuring mental health of the Dutch population: a comparison of the GHQ-12 and the MHI-5*. Health Qual Life Outcomes, 2004; 2: p.p.23
- Holi MM, Marttunen M, Aalberg V. *Comparison of the GHQ-36, the GHQ-12 and the SCL-90 as psychiatric screening instruments in the Finnish population*. Nord J Psychiatry, 2003; 57(3): p.233-238
- Holi MM, Sarnallahti PR, Aalberg VA. *A Finnish validation study of the SCL-90*. Acta Psychiatr Scand, 1998; 97(1): p.42-46
- Hwang HF, Liang WM, Chiu YN, Lin MR. *Suitability of the WHOQOL-BREF for community-dwelling older people in Taiwan*. Age Ageing, 2003; 32(6): p.593-600

- Johnson JA, Coons SJ. *Comparison of the EQ-5D and SF-12 in an adult US sample*. Qual Life Res, 1998; 7: p.155-166
- Johnson JA, Pickard AS. *Comparison of the EQ-5D and SF-12 health surveys in a general population survey in Alberta, Canada*. Med Care, 2000; 38: p.115-121
- Johnson JG, Harris ES, Spitzer RL, Williams JB. *The patient health questionnaire for adolescents: validation of an instrument for the assessment of mental disorders among adolescent primary care patients*. J Adolesc Health, 2002; 30(3): p.196-204
- Jong A de, Lubbe PM vd. *Groningse Vragenlijst voor Sociaal Gedrag: Zelfbeoordelingsvragenlijsten voor het vaststellen van problemen in het interpersoonlijk functioneren (handleiding)*. Groningen: Rob Giel Onderzoekscentrum / Rijksuniversiteit Groningen, 2001
- Jong A de, Nugter A, Polak M, Spinhoven P, Heiser W, Wagenborg H. *The Dutch version of the Outcome Questionnaire (OQ-45): A cross-cultural validation*. Psychologie & Gezondheid, 2008; 36(1): p.35-45
- Kadouri A, Corruble E, Falissard B. *The improved Clinical Global Impression Scale (iCGI): development and validation in depression*. BMC Psychiatry, 2007; 7: p.p.7
- Kaplan RM, Anderson JP, Wu AW, Mathews WC, Kozin F, Orenstein D. *The Quality of Well-Being Scale: Applications in AIDS, cystic fibrosis and arthritis*. Med Care, 1989; 27: p.p.S27-S43
- Kloens GJ. *Diagnostiek met vragenlijsten in de eerstelijns*. Groningen: Rijksuniversiteit Groningen, 2005
- Kloens,GJ et.al. *Diagnostiek en indicatiestelling met behulp van vragenlijsten*
CareMetrics<http://www.caremetrics.nl/cms3/theoretisch-kader/artikelen-publicaties/artikelen-dr-ger-jan-kloens-collega-s/>
- Koeter MWJ, Ormel J. *General Health Questionnaire, Nederlandse bewerking: handleiding*. Lisse: Swets, Test Services, 1991
- Kontodimopoulos N, Pappa E, Niakas D, Yfantopoulos J, Dimitrakaki C, Tountas Y. *Validity of the EuroQoL (EQ-5D) instrument in a Greek general population*. Value Health, 2008; 11(7): p.1162-1169
- Lambert MJ, Burlingame GM, Umphress V, Hanssen NB, Vermeersch DA. *The reliability and validity of the Outcome Questionnaire*. Clinical Psychology & Psychotherapy, 1996; 3: p.249-258
- Lambert MJ, Hansen, NB, Finch, AE. *Patient-focused research: Using patient outcome data to enhance treatment effects*. Journal of Consulting and Clinical Psychology, 2001; 69(2): p.159-172
- Lamers LM, Bouwmans CAM, van Straten A, Donker MCH, Hakkaart L. *Comparison of EQ-5D and SF-6D utilities in mental health patients*. Health Economics, 2006; 15: p.p. 1229-1236
- Landelijke Vereniging van Eerstelijnspsychologen (LVE). *Jaarbericht 2008*. Amsterdam: LVE, 2009
- Lange A, Appelo M. *De Korte Klachten Lijst (KKL): Handleiding*. Houten: Bohn Stafleu van Loghum, 2007
- Lee PW, Schulberg HC, Raue PJ, Kroenke K. *Concordance between the PHQ-9 and the HSCL-20 in depressed primary care patients*. J Affect Disord, 2007; 99(1-3): p.139-145
- McHorney CA, Ware JE, Jr., Lu JF, Sherbourne CD. *The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups*. Med Care, 1994; 32(1): p.40-66

- McHorney CA, Ware JE, Jr., Raczek AE. *The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs*. Med Care, 1993; 31(3): p.247-263
- Miller SD, Duncan BL, Brown J, Sparks JA, Claud DA. *The Outcome Rating Scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure*. Journal of Brief Therapy, 2003; 2(2): p.91-100
- MinVWS. *Overheveling GGZ per 1 januari 2008*. ,7.
- Mulder CL, Staring ABP, Loos J, Buwalda V, Kuijpers D, Sytema S et al. *De Health of the Nation Outcome Scale (HoNOS) als instrument voor 'routine outcome assessment'*. Tijdschrift voor Psychiatrie, 2004a; 46(5): p.273-284
- Mulder CL, Staring ABP, Loos J, Buwalda VJA, Kuijpers D, Sytema S et al. *De Health of the Nation Outcome Scales (HoNOS) in Nederlandse bewerking*. Rotterdam: 2004b
- Oliver JPJ. *The social care directive: Development of a quality of life profile for use in the community services for the mentally ill*. Social Work & Social Sciences Review, 1991; 3: p.5-45
- Orrell M, Yard P, Handysides J, Schapira R. *Validity and reliability of the Health of the Nation Outcome Scales in psychiatric patients in the community*. Br J Psychiatry, 1999; 174: p.409-412
- Penninx BW, Beekman AT, Smit JH, Zitman FG, Nolen WA, Spinhoven P et al. *The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods*. Int J Methods Psychiatr Res, 2008; 17(3): p.121-140
- Perini SJ, Slade T, Andrews G. *Generic effectiveness measures: sensitivity to symptom change in anxiety disorders*. J Affect Disord, 2006; 90(2-3): p.123-130
- Piersma HL, Boes JL. *The GAF and psychiatric outcome: a descriptive report*. Community Ment Health J, 1997; 33(1): p.35-41
- Priebe S, Huxley P, Knight S, Evans S. *Application and results of the Manchester Short Assessment of Quality of Life (MANSA)*. Int J Soc Psychiatry, 1999; 45(1): p.7-12
- Procidano ME, Heller K. *Measures of perceived social support from friends and from family: three validation studies*. Am J Community Psychol, 1983; 11(1): p.1-24
- Quek KF, Low WY, Razack AH, Loh CS. *Reliability and validity of the General Health Questionnaire (GHQ-12) among urological patients: a Malaysian study*. Psychiatry Clin Neurosci, 2001; 55(5): p.509-513
- Reese RJ, Norsworthy LA, Rowlands, SR. *Does a continuous feedback system improve psychotherapy outcome?* Psychotherapy Theory, Research, Practice, Training, 2009, 46(4): p. 418-431
- Rumpf HJ, Meyer C, Hapke U, John U. *Screening for mental health: validity of the MHI-5 using DSM-IV Axis I psychiatric disorders as gold standard*. Psychiatry Res, 2001; 105(3): p.243-253
- Salama-Younes M, Montazeri A, Ismail A, Roncin C. *Factor structure and internal consistency of the 12-item General Health Questionnaire (GHQ-12) and the Subjective Vitality Scale (VS), and the relationship between them: a study from France*. Health Qual Life Outcomes, 2009; 7: p.p.22
- Sanchez-Lopez MP, Dresch V. *The 12-Item General Health Questionnaire (GHQ-12): reliability, external validity and factor structure in the Spanish population*. Psicothema, 2008; 20(4): p.839-843

- Savoia E, Fantini MP, Pandolfi PP, Dallolio L, Collina N. *Assessing the construct validity of the Italian version of the EQ-5D: preliminary results from a cross-sectional study in North Italy*. Health Qual Life Outcomes, 2006; 4: p.p.47
- Schmitz N, Kruse J, Heckrath C, Alberti L, Tress W. *Diagnosing mental disorders in primary care: the General Health Questionnaire (GHQ) and the Symptom Check List (SCL-90-R) as screening instruments*. Soc Psychiatry Psychiatr Epidemiol, 1999a; 34(7): p.360-366
- Schmitz N, Kruse J, Tress W. *Psychometric properties of the General Health Questionnaire (GHQ-12) in a German primary care sample*. Acta Psychiatr Scand, 1999b; 100(6): p.462-468
- Spitzer RL, Kroenke K, Williams JB. *Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire*. JAMA, 1999; 282(18): p.1737-1744
- Tait RJ, French DJ, Hulse GK. *Validity and psychometric properties of the General Health Questionnaire-12 in young Australian adolescents*. Aust N Z J Psychiatry, 2003; 37(3): p.374-381
- Tarlov AR, Ware JE, Greenfield S, et al. *The Medical Outcomes Study: An application of methods for monitoring the results of medical care*. JAMA, 1989; 262(7): p.925-930
- Taylor WJ, Myers J, Simpson RT, McPherson KM, Weatherall M. *Quality of life of people with rheumatoid arthritis as measured by the World Health Organization Quality of Life Instrument, short form (WHOQOL-BREF): score distributions and psychometric properties*. Arthritis Rheum, 2004; 51(3): p.350-357
- Terluin B. *De Vierdimensionele Klachtenlijst (4DKL)*. Huisarts en Wetenschap, 1996; 39(12): p.538-547
- Terluin B, van Marwijk HW, Ader HJ, de Vet HC, Penninx BW, Hermens ML et al. *The Four-Dimensional Symptom Questionnaire (4DSQ): a validation study of a multidimensional self-report questionnaire to assess distress, depression, anxiety and somatization*. BMC Psychiatry, 2006; 6: p.p.34
- Terluin B, Van Rhenen W, Schaufeli WB, De Haan M. *The Four-Dimensional Symptom Questionnaire (4DSQ): Measuring distress and other mental health problems in a working population*. [References]. Work & Stress, 2004; 18(3): p.187-207
- Terwee CB, Jansma EP, Riphagen II, de Vet HC. *Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments*. Qual Life Res, 2009; 18(8): p.1115-1123
- The WHOQOL Group. *Development of the World Health Organization WHOQOL-BREF quality of life assessment*. Psychol Med, 1998; 28(3): p.551-558
- Trompenaars FJ, Masthoff ED, Van Heck GL, Hodiament PP, De Vries J. *Content validity, construct validity, and reliability of the WHOQOL-Bref in a population of Dutch adult psychiatric outpatients*. Qual Life Res, 2005; 14(1): p.151-160
- Vatnaland T, Vatnaland J, Friis S, Opjordsmoen S. *Are GAF scores reliable in routine clinical use?* Acta Psychiatr Scand, 2007; 115(4): p.326-330
- Ventura J, Lukoff D, Nuechterlein KH, Liberman RP, Green MF, Shaner A. *Training and Quality Assurance with the BPRS Brief Psychiatric Rating Scales (BPRS): Expanded Version (4.0)*. International Journal of Methods in Psychiatric Research, 1993; 3: p.227-243

- Vermeersch DA, Lambert MJ, Burlingame GM. *Outcome Questionnaire: item sensitivity to change*. J Pers Assess, 2000; 74(2): p.242-261
- Vries Jd, Heck GLv. *Nederlandse WHOQOL-Bref*. Tilburg: Universiteit van Tilburg, 1996
- Ware JE, Sherbourne CD. *The MOS Short-Form Health Survey (SF-36): I. Conceptual framework and item selection*. Medical Care, 1992; 30: p.473-483
- Wijngaarden Bv, Kok I. *Een inventarisatie van potentieel geschikte instrumenten voor de Basisset Prestatie-indicatoren geestelijke gezondheidszorg en verslavingszorg*. Trimbos Instituut.
- Wing JK, Beevor AS, Curtis RH, Park SB, Hadden S, Burns A. *Health of the Nation Outcome Scales (HoNOS). Research and development*. Br J Psychiatry, 1998; 172: p.11-18
- Zigmond AS, Snaith RP. *The Hospital Anxiety and Depression Scale*. Acta Psychiatr Scand, 1983; 67: p.361-370
- Zwaanswijk M, Verhaak PFM. *Effectieve kortdurende interventies voor psychische problemen*. Utrecht: Nivel, 2009