

Validering van twee meetinstrumenten voor routine outcome monitoring in de psychiatrie: de HORVAN-studie

V.J.A. BUWALDA, S. DRAISMA, J.H. SMIT, J.A. SWINKELS, W. VAN TILBURG

ACHTERGROND Transparantie in de psychiatrie kan worden vergroot door de inzet van meetinstrumenten voor routine outcome monitoring (ROM). Instrumenten moeten snel en eenvoudig toepasbaar zijn, psychometrische kwaliteit bezitten en gevoelig zijn voor verandering. Ook moeten ze inzicht geven in klachten en in interpersoonlijk en sociaal-maatschappelijk functioneren.

DOEL Onderzoeken in hoeverre de combinatie van de Health of the Nation Outcome Scales (HoNOS) en de Outcome Questionnaire (OQ) in de Nederlandse situatie aan de genoemde kwaliteitseisen voldoet en hoe deze zich tot de Symptom Checklist (SCL-90) verhoudt.

METHODE Gegevens van 148 patiënten van drie meetmomenten waren bruikbaar voor analyse. De instrumenten werden onderzocht op psychometrische eigenschappen en sensitiviteit voor verandering.

RESULTATEN De drie schalen vertoonden hoge waarden voor interne consistentie (Cronbachs alfa). De HoNOS-totaalscore en de subschalen van de OQ correleerden respectievelijk redelijk tot goed met de SCL-90-totaalscore (convergente validiteit). Bij de eerste meting scoorden personen met comorbide diagnose het slechtst (discriminante validiteit). De klinisch significante verbetering tussen T1 en T2 en tussen T2 en T3 was voor de drie meetinstrumenten voldoende hoog.

CONCLUSIE De combinatie van het beoordelingsinstrument HoNOS en de zelfrapportagelijst OQ blijkt geschikt voor ROM in de psychiatrie.

[TIJDSCHRIFT VOOR PSYCHIATRIE 53(2011)10, 715-726]

TREFWOORDEN klinisch significante verbetering, routine outcome monitoring, valideringsonderzoek

In de afgelopen decennia ontstond in Nederland een toenemende vraag naar transparantie van het behandelproces in de gezondheidszorg. Er zijn meetinstrumenten ontwikkeld die het resultaat van behandeling op individueel niveau inzichtelijk maken of patiënttevredenheid meten. In de geestelijke gezondheidszorg wordt routine outcome monitoring (ROM) gebruikt (de Beurs & Zitman 2007). Met deze methode worden gegevens verza-

meld over het functioneren en welbevinden van de patiënt voor, tijdens en na de behandeling, middels zelfrapportage en/of beoordeling. De behandeling kan bijgestuurd worden wanneer blijkt dat er onvoldoende verbetering optreedt (Aartsen e.a. 2010).

Er zijn verschillende bruikbare meetinstrumenten om het beloop van therapie te meten, maar slechts enkele voldoen aan de eis om verbete-

ring ook valide te meten (Hermann 2005; Lambert e.a. 2001; Thornicroft e.a. 2005). Deze auteurs formuleren de volgende kwaliteitscriteria voor ROM-meetinstrumenten:

- ze zijn snel en eenvoudig toepasbaar;
- ze bevatten minimaal de volgende drie domeinen: klachten, interpersoonlijke problemen en sociaal-maatschappelijk functioneren (Lambert e.a. 2001);
- ze hebben voldoende psychometrische kwaliteit;
- ze zijn gevoelig voor verandering op korte termijn.

SCL-90

Het meest ingezette en onderzochte meetinstrument om behandelingseffecten in de klinische praktijk inzichtelijk te maken is de Symptom Checklist (SCL-90), een zelfrapportage-instrument dat het gehele ziektespectrum omvat (Arrindell & Ettema 1986; 2003; Derogatis 1977). De lijst heeft echter een aantal nadelen. Ten eerste wordt een grote groep patiënten uitgesloten omdat ze niet in staat zijn zelf de vragenlijst in te vullen, door ernstige ziekte, concentratieproblemen, cognitieve stoornissen of taalproblemen. Ten tweede ontbreken items over sociale interactie en het sociaal-maatschappelijk functioneren. Een derde nadeel is dat het invullen van de lijst relatief lang duurt.

HONOS EN OQ

Meetinstrumenten die deze nadelen niet vertonen, zijn de *Health of the Nation Outcome Scales* (HoNOS) en de *Outcome Questionnaire* (OQ). Dit zijn respectievelijk een beoordelingsmeetinstrument voor de ernstiger psychiatrische ziektebeelden, psychotische stoornissen en bipolaire stoornissen (Sharma e.a. 1999) en een zelfrapportage gericht op symptomen van angst en depressie (Lambert e.a. 2001).

In de HoNOS, OQROM Validiteit Nederland studie (HORVAN-studie) onderzoeken wij de HoNOS en de OQ in een ROM-constructie bij een Neder-

landse onderzoekspopulatie van ambulante patiënten. Daarbij toetsen we bruikbaarheid, psychometrische kwaliteit en gevoeligheid voor verandering. Deze ROM-configuratie vergelijken we met de SCL-90, omdat dat een veelgebruikt instrument is en het raadzaam is een zelfrapportage-instrument als 'gouden standaard' in te zetten (Oiesvold e.a. 2011).

De vraag die we met deze studie beogen te antwoorden luidt: in hoeverre voldoet de combinatie HoNOS en OQ in de Nederlandse situatie aan de vier geformuleerde kwaliteitseisen voor ROM-meetinstrumenten en hoe verhoudt de combinatie HoNOS-OQ zich tot de SCL-90?

Eerst bespreken we kenmerken van de drie meetinstrumenten en vervolgens beschrijven we methode en resultaten van de HORVAN-studie.

DE MEETINSTRUMENTEN EN KWALITEITSCRITERIA

Bruikbaarheid

De HoNOS is een beoordelingslijst met 12

TABEL 1	Subschalen van drie meetinstrumenten (aantal items) die gebruikt werden in HORVAN-studie
	<i>Health of the Nation Outcome Scales (HoNOS)</i> (12 items)
	Cognitieve en lichamelijke beperkingen (2)
	Gedragsproblemen (3)
	Symptomen (3)
	Sociale problemen/relaties/dagactiviteiten (4)
	<i>Outcome Questionnaire (OQ)</i> (45 items)
	Ernst interpersoonlijke relaties (11)
	Sociaal-maatschappelijke rol (9)
	Symptomen (25)
	<i>Symptom Checklist (SCL-90)</i> (90 items)
	Angst (10)
	Agorafobie (7)
	Depressie (16)
	Somativering (12)
	Insufficiënte denken en handelen (9)
	Interpersoonlijke sensitiviteit (18)
	Hostiliteit (6)
	Slaapproblemen (3)
	Psychoneuroticisme (9)

items, die door de professional over de patiënt wordt ingevuld aan de hand van een anamnese over de afgelopen 14 dagen. De HONOS is in Engeland ontwikkeld door Wing e.a. (1998) en wordt internationaal veelvuldig ingezet (Slade e.a. 1999; Trauer e.a. 1999). De afname vergt weinig tijd en het meetinstrument kan onafhankelijk van een specifiek ziektebeeld worden gebruikt. Wel moet de professional een training over de beoordelingsprocedure volgen.

De OQ is een zelfbeoordelingslijst met 45 items over de domeinen 'symptomen', 'interpersoonlijk functioneren' en 'sociaal functioneren',

gedurende de afgelopen week. De OQ is ontwikkeld in de VS (Lambert e.a. 1996).

De Symptom Checklist-90 (SCL-90) is een veelgebruikte multidimensionele zelfbeoordelingslijst ontwikkeld in de VS door Derogatis (1977). De lijst streeft met haar 90 items naar een volledige representatie van psychiatrische symptomen en de afnameduur is dan ook langer dan van de HONOS en OQ.

De verschillende subschalen van de drie meetinstrumenten zijn weergegeven in tabel 1.

Tabel 2 levert informatie over het gebruik en andere criteria. Duidelijk is dat de OQ het meest voldoet aan de eis om drie domeinen te omvatten

TABEL 2 Kenmerken van drie ROM-instrumenten, gebruikt in HORVAN-studie

Kenmerk	HONOS	OQ	SCL-90
Uitgangsbelden	psychose en bipolair	angst- en stemmingsstoornissen	alle psychiatrische ziektebeelden
Aantal items	12 (mogelijkheid 3 item addendum)	45	90
Likertschaal	0-4 (geen probleem-ernstig tot zeer ernstig probleem)	0-4 (nooit-bijna altijd)	1-5 (helemaal niet-heel erg)
Soort meetinstrument	beoordeling professional	zelfrating	zelfrating
Aantal subschalen	4	3	9
Tijdsinvestering; in min	5	10	15-20
Training nodig	ja	nee	nee
Beschikbaarheid	gratis downloadbaar	tegen betaling	tegen betaling
Betrouwbaarheid (Cronbachs α)	0,78 (Mulder e.a. 2004)	0,94 (Chapman 2003)	0,73-0,97 verschillende steekproeven (Arrindell e.a. 2004)
Interbeoordelaarsbetrouwbaarheid	0,29-0,82 op itemniveau (Orrel e.a. 1999) 0,71-0,86 verschillende steekproeven (Trauer e.a. 1999) 0,92 (Mulder e.a. 2004) 0,92 (Broersma & Sytema 2008)	0,79 (Chapman 2003)	-
Convergente validiteit (R andere meetinstrumenten)	Redelijk met GAF, QOL, MANSAS, CANSAS, CGI, BPRS (Mulder e.a. 2004) Goed met BPRS en RFS (Wing e.a. 1998)	Redelijk tot goed met BDI, STAI, SF36, SCL-90 (Umphress e.a. 1997) Redelijk met subschalen BASIS-32 (Doerfler e.a. 2002)	Hoog met GHQ-12 (Holi e.a. 2003) Hoog met GHQ (Koeter 1992)
Normgroepen	-	De Beurs e.a. 2005 De Jong e.a. 2007	Derogatis 1977 Arrindell & Ettema 1986
KSV-afkapwaarden	Alle itemwaarden < 2	Somscore < 55	Somscore < 124

HONOS = Health of the Nation Outcome Scales (HONOS); OQ = Outcome Questionnaire; SCL-90 = Symptom Checklist; GAF = Global Assessment of Functioning; MANSAS = Manchester Short Assessment of quality of life; CANSAS = Camberwell Assessment of Need short appraisal; CGI = Clinical Global Impression Scale; BPRS = Brief Psychiatric Rating Scale; RFS = Role Functioning Scales; BDI = Beck Depression Inventory; STAI = Spielberger State-Trait Anxiety Inventory; SF-36 = Short Form; BASIS-32 = Behavior and Symptom Identification Scale; GHQ = General Health Questionnaire; KSV = klinisch significante verbetering.

(klachten, interpersoonlijke problemen en sociaal-maatschappelijk functioneren). De HONOS bevat naast psychiatrische symptomen vier items over dagelijks en sociaal functioneren, terwijl de SCL-90 het accent legt op de ernst van klachten.

Psychometrische kwaliteit

In Nederland is de HONOS als goed beoordeeld op betrouwbaarheid en validiteit door Aartsen e.a. (2010), Broersma en Sytma (2008) en Mulder e.a. (2004). Voor de begripsvaliditeit, in termen van onderscheidend vermogen, vergeleken Mulder e.a. scores tussen klinische afdelingen. Opgenomen patiënten scoorden significant hoger dan patiënten die ambulante werden behandeld, patiënten in dagbehandeling scoorden er tussenin. Broersma en Sytma repliceerden deze resultaten voor de ouderenpsychiatrie. Daarnaast werden patiëntengroepen met psychotische stoornissen, depressieve stoornissen en angststoornissen met elkaar vergeleken: afzonderlijke items en subschalen vertoonden significante verschillen tussen deze groepen. Mulder e.a. vonden redelijke convergerende validiteit, in de vorm van de relatie van de HONOS-subschalen en totaalscores van andere meetinstrumenten (*Global Assessment of Functioning* (GAF), *Manchester Short Assessment of quality of life* (MANSA), *Camberwell Assessment of Need short appraisal* (CANSAS), *Clinical Global Impression Scale* (CGI) en *Brief Psychiatric Rating Scale* (BPRS)).

Er is voldoende onderzoek naar de psychometrische kwaliteit van de OQ beschikbaar (o.a. de Beurs e.a. 2005; Doerfler e.a. 2002; de Jong e.a. 2007; Umphress e.a. 1997). De betrouwbaarheid van de OQ is goed, zowel voor de Amerikaanse als Nederlandse situatie (de Beurs e.a. 2005; Chapman 2003; de Jong e.a. 2007).

Umphress e.a. (1997) lieten met verschillen in totaalscores van de OQ zien dat een psychiatrische patiëntengroep goed onderscheiden werd van respondenten uit de algemene populatie; voor Nederland werd dit onderzoek gerepliceerd door De Jong e.a. (2007). Ook vonden Umphress e.a. een goede convergentie met de *Beck Depression Inventory* (BDI),

de *Spielberger State-Trait Anxiety Inventory* (STAI), de *Short Form* (SF-36) en de SCL-90.

Doerfler e.a. (2002) vonden redelijke divergente en convergente validiteit van de OQ-subschalen met die van subschalen van de *Behavior and Symptom Identification Scale* (BASIS-32; psychiatrische symptomen en sociaal functioneren). Daarnaast differentieerden de subschalen verschillende diagnostische groepen goed.

Arrindell en Ettema (1986) gebruikten voor de interne consistentie van de SCL-90-schalen diverse steekproeven. Cronbachs alfa voor de verschillende subschalen varieerde van 0,73 tot 0,97. Holi e.a. (2003) vonden significant verschillende gemiddelde scores tussen steekproeven uit de algemene bevolking en ambulante patiënten. Koeter (1992) vergeleek de depressie- en angstschalen van de SCL-90 met die van de *General Health Questionnaire* (GHQ-28), en zette ze af tegen een DSM-III-diagnose in een sensitiviteits- en specificiteitsanalyse. De concurrente validiteit van deze subschalen van de SCL en de GHQ waren goed te noemen.

Gevoeligheid voor verandering

Voor het meten van verandering is de methode van Jacobson en Truax (1991) beschikbaar, de zogeheten klinisch significante verbetering (KSV). Deze houdt in dat (1) een statistisch betrouwbare verbetering (BV) optreedt en (2) een grenswaarde overschreden wordt die de overgang markeert van ziek naar herstel (KSV), de score van de (disfunctionele) patiënt dient de normscore van de functionele normpopulatie te passeren. Wordt alleen aan het eerste criterium voldaan, dan is er sprake van een betrouwbare verbetering, maar (nog) niet van herstel. Wordt alleen aan het tweede criterium voldaan, dan is er weliswaar een verschuiving van disfunctioneel of ziek naar functioneel ('gezond'), maar kunnen beide scores zich zo dicht bij de grensscore bevinden dat de verschuiving geen klinische betekenis heeft (Aarsse 2003; de Beurs e.a. 2005).

Parabiaghi e.a. (2005) onderzochten de KSV van de HONOS in een Italiaanse populatie. De

auteurs gebruikten echter een aangepast algoritme, zonder specifieke grenswaarde voor het onderscheid klinisch-functioneel. Als de patiënt op geen enkel item een waarde van 2 of hoger behaalde, kon deze als subklinisch (functioneel) beschouwd worden. In de steekproef herstelde 5,6%, en in een subgroep van patiënten met ernstige problematiek 14,4%.

Andere auteurs hebben veranderingen over de tijd gemeten met gebruikelijke statistische analysetechnieken. Wing e.a. (1998) toetsten op twee meetmomenten de verschillen tussen HONOS-scores met t-toetsen. Hunter e.a. (2009) gebruikten regressieanalyse bij patiënten met schizofrenie en vonden nauwelijks verandering.

Alle gevonden studies naar verandering met de OQ zijn gebaseerd op t-toetsen, niet uitgewerkt volgens de KSV-methode. Doerfler e.a. (2002) en Lambert e.a. (1996) vonden significante verbeteringen op alle subschalen tussen opname en uitschrijving van patiënten. Vermeersch e.a. (2000) onderzochten op itemniveau de gevoeligheid voor verandering.

Arrindell en Ettema (2003) geven een overzicht van studies met de SCL-90 in de periode 1957-1998 met de effecten van zowel korte als lange behandelingen. In alle besproken studies wordt verandering vastgesteld met gangbare statistische technieken, niet met KSV. Koeter (1992) beschikte over drie meetmomenten bij een steekproef ambulante patiënten, maar heeft geen expliciete verandering vastgesteld. Wilson e.a. (1997) gebruikten wel de KSV om herstel van patiënten met een posttraumatische stressstoornis bij behandeling met *eye movement desensitization and reprocessing* (EMDR) vast te stellen met de SCL-90: 56% was na 15 maanden klinisch significant verbeterd.

We kunnen concluderen dat er nog weinig onderzoek naar verbetering met de KSV-methode heeft plaatsgevonden voor de drie meetinstrumenten.

HORVAN-STUDIE: METHODE

Procedure en steekproef

De onderzoekspopulatie bestond uit ambulante patiënten die door huisartsen naar de polikliniek waren verwezen, in een middelgrote Nederlandse stad. Patiënten kregen een intakegesprek met de arts in opleiding tot psychiater (aios) of de psycholoog. Vervolgens werd een gestructureerd interview afgenomen van 30-40 minuten, met onder andere het Mini Internationaal Neuropsychiatrisch Interview (MINI; van Vliet & de Beurs 2007), door de gesuperviseerde aios of de psychiater. Met het MINI wordt een as I-diagnose bij de patiënt gesteld volgens de DSM-IV-classificatie. Verder werden de HONOS, OQ en SCL-90 afgenomen.

Een van de auteurs (V.J.A.B.) is als psychiater werkzaam op de polikliniek en beoordeelde – na een training – de patiënten door gebruik te maken van de HONOS in de periode januari 2002 tot november 2004. De psychiater trainde daarnaast de aios, die onder zijn supervisie werkte in het scoren van de HONOS. De HONOS-beoordelingen waren gebaseerd op de interviews met de patiënt na de afname van de MINI, aangevuld met de informatie die de patiënt gaf en de informatie van de hulpverlener van de patiënt. De patiënten vulden vervolgens de OQ en de SCL-90 in.

Tijdens de behandeling werd gemiddeld elke 10 weken aan de patiënt gevraagd een OQ en SCL-90 in te vullen en werden een HONOS, GAF en CGI afgenomen door de aios of psychiater. Op deze manier werd de verandering van ernst van de symptomen gevolgd tijdens de behandeling.

In totaal werd bij 370 patiënten een aanvangsmeting verricht, waarbij minimaal de drie meetinstrumenten werden afgenomen. Een deel van de patiënten nam deel aan vervolgmetingen ($n = 213$; 54%) gedurende een behandelperiode die maximaal 2,5 jaar duurde, met een maximum van 17 metingen bij een enkele patiënt. Het doel was patiënten elke 6 weken te beoordelen in het geval er medicatie werd ingezet als behandeling. Was

cognitieve gedragstherapie de gekozen behandel-methode, dan vond de beoordeling één keer per 3 maanden plaats. Een kleine groep patiënten ($n = 11$; 3%) werd tijdens de behandeling opgenomen in de kliniek. Uitvalredenen voor de 157 personen die niet voor een vervolgmeting op T2 kwamen, waren: het niet verschijnen op de gemaakte afspraken, het starten van een vervolgbehandeling buiten de polikliniek, een incomplete voormeting, geen diagnose of op een wachtlijst voor behandeling geplaatst worden.

Uiteindelijk waren de data van 148 patiënten bruikbaar voor statistische analyse van de eerste drie metingen. Van deze patiënten was er minimaal een vervolgmeting op tijdstip 2 (T2) van de drie meetinstrumenten voorhanden. Van 28% van deze personen ontbrak op het derde meetmoment minimaal een van de drie vragenlijsten ($n = 41$). Ontbrekende gegevens van deze personen bij de derde meting werden aangevuld middels imputatie volgens het principe *intention to treat, last value carried forward*.

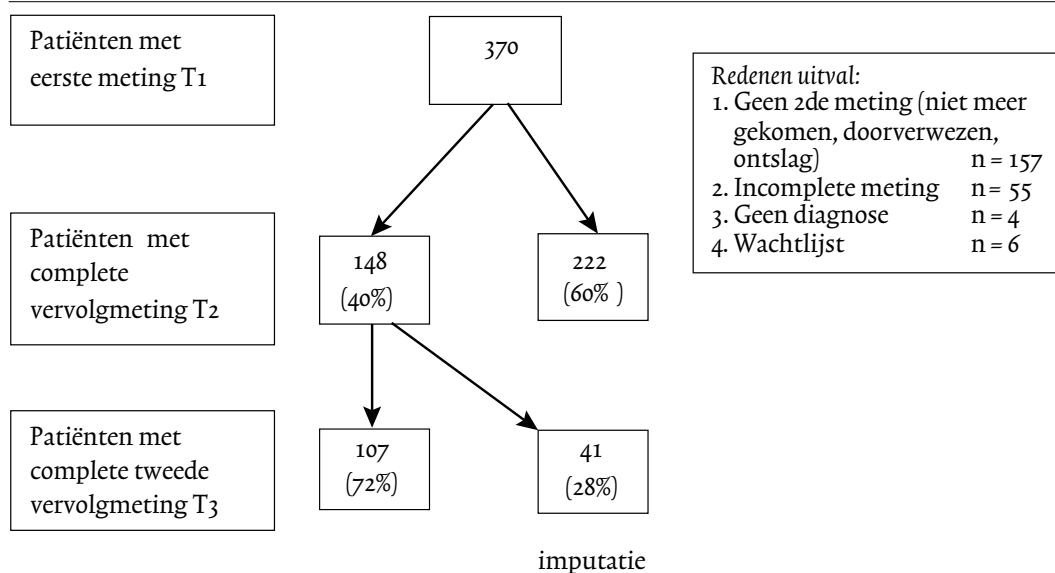
Analyse

Cronbachs alfa's werden berekend voor betrouwbaarheid. Correlaties tussen schalen voor

convergerende validiteit, t-toetsen, χ^2 -toetsen en variantieanalyse werden gebruikt om verschillen tussen groepen aan te geven. Betrouwbare verandering werd berekend met de formule $BV = (X_{t1} - X_{t0})/S_{diff}$, waarin $S_{diff} = \sqrt{2(SE)^2}$, dat wil zeggen: verandering uitgedrukt in verschillscore, gecorrigeerd voor de standaardmeetfout. Voor de standaardmeetfout werden de standaarddeviaties en Cronbachs alfa's van de eerste meting gebruikt. Vervolgens onderzochten wij hoeveel personen met een BV ook de normscore van disfunctioneel naar functioneel ('gezond') overschreden.

Voor de berekening van KSV waren normscores voor de OQ beschikbaar, lager dan 55 functioneel ('gezond') (de Jong e.a. 2007, p. 296), voor de HONOS en SCL-90 waren deze indirect beschikbaar. Voor de HONOS wil dit zeggen: geen item met een waarde hoger dan 1 (Parabiaghi e.a. 2005). Voor de SCL-90 betekent dit: somscore < 124 (Arrindell & Ettema 1986, p. 39, gemiddelden algemene populatie). Om de geschiktheid van de combinatie OQ-HONOS te bepalen werd een sensitiviteitsanalyse uitgevoerd met de veelgebruikte SCL-90 als gouden standaard.

FIGUUR 1 Stroomdiagram van inclusie in HORVAN-studie



TABEL 3 Demografische kenmerken onderzoekspopulatie van HORVAN-studie

Geslacht	
Man	53 (36)
Vrouw	92* (62)
Opleiding	
Laag	87 (74)
Hoog	30** (26)
Diagnose	
Angst	31 (21)
Depressie	34 (23)
Comorbide angst en depressie	59 (40)
Overig (psychose (14), verslaving (5), uitgesteld(5))	24 (16)
Gem. leeftijd; in j (SD; uitersten)	37,7 (13,1; 16-67)

*Van 3 personen was sekse onbekend.
**Bij 31 personen was opleiding onbekend; hoge opleiding was vanaf havo.

RESULTATEN

Onderzoekspopulatie, complete versus niet-complete follow-up

Van 148 personen waren de gegevens voor de eerste twee metingen compleet. Bij de derde meting ontbraken gegevens van 41 van deze 148 personen. Voor deze 41 werden de gegevens op T3 geïmputeerd (volgens het principe *last value carried forward*). Ook multipele imputatie werd toegepast (met demografische variabelen en scores van de eerste twee metingen als predictoren). Dat leverde echter geen noemenswaardige verschillen op. In tabel 3 worden de belangrijkste demografische kenmerken van de onderzoekspopulatie weergegeven. Na hergroepering van DSM-IV-diagnosen tot vier hoofdcategorieën ontstond de verdeling in de tabel.

De gemiddelde leeftijd was 37,7 jaar, een meerderheid van de patiënten was vrouw en hoogopgeleid. Degenen met complete gegevens ($n = 148$, minimaal 2 metingen) werden op beschikbare variabelen vergeleken met degenen met incomplete gegevens ($n = 222$, slechts aanvangsmeting). Er waren geen verschillen tussen degenen met complete gegevens en degenen met niet-complete

TABEL 4 Betrouwbaarheidscoëfficiënten (Cronbachs alfa) van de meetinstrumenten in HORVAN-studie

Subschaal (aantal items)	α^*	α	α
	T1**	T2	T3
OQ			
Symptomen (25)	0,90	0,93	0,96
Sociale rol (8)	0,60	0,69	0,78
Interpersoonlijke relaties (11)	0,76	0,82	0,89
Totaal (45)	0,92	0,95	0,97
HoNOS			
Gedragsproblemen (3)	0,31	0,38	0,37
Beperkingen (2)	0,55	0,53	0,32
Symptomen (3)	0,24	0,32	0,42
Sociale problemen (4)	0,49	0,60	0,53
Totaal (12)	0,64	0,74	0,73
SCL Totaal (90)	0,97	0,98	0,98

HoNOS = Health of the Nation Outcome Scales; OQ = Outcome Questionnaire; SCL-90 = Symptom CheckList.
*Split-halfbetrouwbaarheid gaf vergelijkbare waarden en volgordes.
**Gem. aantal dagen T1-T2 = 75 (SD 59); T2-T3 = 72 (SD 55).

follow-up in leeftijd ($T_{t1-t2} = 0,30$; $df = 256$; $p = 0,77$), sekseverdeling ($T_{t1-t2} = 1,17$; $df = 1$; $p = 0,32$) of verdeling over de diagnose hoofdgroepen ($F_{t1-t2} = 5,7$; $df = 1$; $p = 0,02$). Ook de gemiddelde (subschaal)scores op de drie meetinstrumenten verschilden niet significant tussen degenen met een complete en een niet-complete follow-up.

Betrouwbaarheid

In tabel 4 zijn de Cronbachs alfa's weergegeven voor OQ, de HoNOS en de SCL-90 op drie meetmomenten.

Drie van de vier OQ-schalen hadden hoge waarden voor interne consistentie (Cronbachs alfa $> 0,70$). Alleen voor 'Sociale rol' was de alfa lager dan 0,65. Deze subschaal bevatte slechts 9 items en leverde ook bij andere auteurs de laagste betrouwbaarheidswaarden. Voor de HoNOS-totaalschaal troffen we redelijk vergelijkbare alfa's aan met die van Mulder e.a. (2004), die 0,78 vonden. De subschalen van de HoNOS bereikten geen acceptabele alfa, maar het aantal items was dan ook zo klein

Subschalen	SCL-90		
	op T1	op T2	op T3
OQ45: Symptomen (OQSD)	0,84	0,91	0,82
OQ45: Sociale rol (OQSR)	0,51	0,63	0,61
OQ45: Interpersoonlijke relaties (OQIR)	0,63	0,71	0,68
OQ Totaal	0,84	0,90	0,90
HoNOS: totaal	0,64	0,75	0,80

HoNOS = Health of the Nation Outcome Scales; OQ = Outcome Questionnaire; SCL-90 = Symptom CheckList.
*Alle correlaties waren statistisch significant ($p < 0,01$).

dat dat moeilijk bereikbaar was. De HoNOS is ook niet zozeer gebaseerd op subschalen met correlerende items, elk afzonderlijk item representeert een complete dimensie. Voor de SCL-90 waren alle gevonden betrouwbaarheidscoëfficiënten hoog: op de subschalen waren deze 0,76-0,94 (meer dan voldoende). Dit kwam overeen met andere studies.

Convergente validiteit

In navolging van Umphress e.a. (1997) onderzochten wij de convergente validiteit door vergelijking van de subschalen van de OQ en de HoNOS met de SCL-90-totaalscore. De correlaties tussen de

Maten	OQ45	HoNOS	SCL-90
BV T1-T2	23,6	7,4	43,9
KSV T1-T2	18,9	3,4	12,8
BV T1-T3	37,2	11,5	54,1
KSV T1-T3	28,4	8,1	23,0

BV = betrouwbare verandering; KSV = klinisch significante verandering; HoNOS = Health of the Nation Outcome Scales; OQ = Outcome Questionnaire; SCL-90 = Symptom CheckList.

schalen van de verschillende instrumenten op de basismeting zijn gegeven in tabel 5.

Subschalen van de OQ correleerden hoog genoeg met de SCL-90-totaalscore, zodat gesproken kon worden van een goede convergente validiteit; de schalen indiceerden nagenoeg dezelfde constructen. Hoge correlaties van OQ en HoNOS met SCL-90 boden ondersteuning voor het idee dat beide instrumenten samen uitwisselbaar zijn met de SCL-90.

Discriminante validiteit

Voor verschillende diagnosegroepen waren verschillende gemiddelden en veranderingen in waarden verwacht. Zo zou de prognose voor verbetering van personen met stemmingsklachten gunstiger zijn dan voor diegenen met psychose. Voor zover de aantallen in de diagnosegroepen dat

	Meetmoment	Gemiddelde scores (uitersten)		
		OQ*	HoNOS*	SCL-90*
Angst (n = 31)	T1	68,8 (33-109)	5,6 (0-13)	177,7 (90-296)
	T2	58,6 (12-113)	4,4 (0-13)	165,2 (91-325)
	T3	55,3 (6-113)	4,1 (0-10)	156,2 (90-264)
Depressie (n = 34)	T1	69,4 (26-131)	6,4 (1-21)	185,3 (120-289)
	T2	68,1 (22-126)	5,6 (0-21)	176,2 (101-355)
	T3	56,7 (6-130)	4,8 (0-15)	156,8 (91-316)
Comorbide A-D (n = 59)	T1	83,0 (49-131)	8,8 (1-21)	227,8 (137-347)
	T2	67,3 (1-124)	6,3 (0-20)	183,7 (91-317)
	T3	63,2 (12-137)	5,8 (0-19)	176,8 (90-384)
Overig (n = 24)	T1	68,9 (24-140)	8,3 (0-23)	196,2 (102-344)
	T2	68,6 (25-119)	6,1 (2-13)	182,6 (106-330)
	T3	61,7 (24-124)	4,4 (1-16)	167,9 (114-347)

*Alleen gemiddelde scores op T1 verschilden significant over diagnoses ($F_{OQ} = 4,4$; $p = 0,004$; $F_{HoNOS} = 8,2$; $p = 0,01$; $F_{SCL} = 7,7$; $p = 0,00$).
HoNOS = Health of the Nation Outcome Scales; OQ = Outcome Questionnaire; SCL-90 = Symptom CheckList.

toelieten, onderzochten wij verschillen in gemiddelden voor de drie instrumenten.

De gemiddelde OQ-score was in de eerste meting beduidend hoger voor ‘comorbide angst en depressie’ dan voor andere diagnoses. Deze hogere score verdween in meting 2 en 3: het leek erop dat deze patiënten goed profiteerden van de behandeling. Dit patroon was ook aanwezig bij de HoNOS: patiënten met zowel angst als depressie begonnen met de ongunstigste score en deze score daalde het meest over de metingen. Ook de diagnose ‘overig’ startte met een relatief ongunstige HoNOS-score en daalde flink. Dat patroon zagen we ook in de resultaten van de SCL-90.

Klinisch significante verandering

Tabel 7 geeft percentages weer met betrouwbare verandering (BV, met de desbetreffende meetmomenten) en percentages die daarnaast ook klinisch verbeterden, d.w.z. de normscore overschreden (KSV).

Uitgedrukt in KSV op OQ-scores, verbeterde tussen de eerste en de tweede meting 1 op de 5 personen, en tussen de eerste en de derde meting zelfs 1 op de 3. Kortom, het beloop zoals gemeten met OQ werd gunstiger naarmate er meer tijd verstreek en de behandeling vorderde. Dit werd in mindere mate ook gevonden voor de HoNOS en SCL-90.

SCL-90 als gouden standaard

Een sensitiviteitsanalyse werd uitgevoerd voor OQ en HoNOS, met de SCL-90 als gouden standaard. Als afkapwaarden werden de normscores uit tabel 7 gebruikt. De gegevens van de derde meting werden gebruikt, omdat zich daarin relatief de meeste personen zonder klachten bevonden volgens SCL-90, namelijk 27,1%. Bij de eerdere metingen was de verdeling schever, wat een betrouwbare sensitiviteitsanalyse bemoeilijkt.

Voor de OQ was de sensitiviteit (terecht positieven) 0,93 en de specificiteit (terecht negatieven) 0,69. De sensitiviteit voor de HoNOS was 0,83 en de

specificiteit bedroeg 0,72. Voor de OQ vonden we een *area under the curve* (AUC) van 0,82 en voor de HoNOS was de AUC 0,77. De gevonden waarden voor zowel de OQ als de HoNOS waren goede voorspellers voor de uitkomsten op de SCL-90.

DISCUSSIE

De combinatie van het beoordelingsinstrument HoNOS met de zelfrapportagelijst OQ is voor ROM in de psychiatrie vruchtbaar, ter vervanging van de SCL-90. Wij geven hiervoor een aantal argumenten en doen daarbij aanbevelingen voor gebruik in de praktijk.

Ten eerste kost *afname* van de twee instrumenten samen ongeveer een kwartier, korter dan het invullen van een SCL-90.

Ten tweede blijkt uit de psychometrische analyse dat de twee instrumenten voldoende kwaliteit bezitten. Enerzijds zijn de betrouwbaarheden van beide instrumenten goed, en scores hangen voldoende samen met die op de SCL-90, wat bijdraagt aan de convergerende validiteit. Anderzijds verschillen de scores voor verschillende patiëntengroepen (discriminerend vermogen). Van de SCL-90 werd verondersteld dat deze de meer subjectieve gewaarwordingen van de patiënt bepaalt, terwijl de subschalen van de OQ en HoNOS samen konden worden gezien als meer objectieve maten voor psychische problemen. Vergelijking met de ‘gouden standaard’ SCL-90 geeft goede resultaten, zodat ook de criteriumvaliditeit van de twee instrumenten adequaat is.

Ten derde zijn beide instrumenten voldoende gevoelig voor verandering, blijktend uit de gevonden KSV-waarden. De HoNOS vertoont weliswaar een veel lagere KSV dan de twee andere meetinstrumenten, maar de 8% verbetering tussen metingen 1 en 3 is acceptabel. Parabiaghi e.a. (2005) vonden 5,6% en beschouwden dat als adequaat. Wel verbetert 7,4% van onze steekproef tussen T1 en T2, maar slechts 3,4% vertoont ook een klinisch significante verbetering.

Het is dan ook verstandig om deze cijfers (BV, KSV en aantal personen dat de grens passeert) in combinatie te gebruiken bij een beoordeling van verbetering en te bedenken dat KSV een strenge maat is en lage percentages oplevert.

Ook is de HoNOS veelal gebruikt als beoordelingsinstrument voor patiënten met ernstiger psychiatrische stoornissen (*severe mentally ill*). Een veel hogere veranderingswaarde is logischerwijs lastig: het instrument bevat slechts 12 items en het verbeteringscriterium is vrij streng (geen enkel item een waarde van 2 of hoger), vergeleken met de OQ en SCL-90. Wel moet men er in de praktijk rekening mee houden dat dit instrument slechts kleine veranderingen laat zien, die echter van betekenis kunnen zijn.

Daarnaast is het aan te bevelen alleen af te gaan op de HoNOS-totaalscore bij gebruik in onderzoek, omdat de subschalen niet voldoende betrouwbaar zijn. Van de OQ zijn ook de subschaalscores afzonderlijk te interpreteren in termen van verbetering of verslechtering.

Ten vierde vertegenwoordigen de twee instrumenten samen de belangrijkste domeinen, zowel functioneel als symptomatisch, waarvan we verwachtten dat behandeling daar een effect heeft.

Het combineren van 'klinisch oordeel' (HoNOS) met een zelfrapportage van subjectief welbevinden (OQ) geeft ook een praktisch voordeel, waarbij het wel of niet overeenstemmen van de afzonderlijke uitkomsten informatief kan zijn over de situatie van de patiënt. Informatie verkregen met beide instrumenten kan elkaar aanvullen. Zo melden Mulder e.a. (2010) dat bij ernstige psychiatrische aandoeningen de HoNOS en MANSa (een zelfrapportage-instrument) gecorreleerd, maar ook complementair zijn, omdat de HoNOS het perspectief van de behandelaar en de MANSa juist dat van de patiënt weergeeft. Afname van SCL-90 alleen voor ROM, zoals vaak gebeurt, is daarom dan ook minder informatief. Daarnaast is de SCL-90 expliciet ontwikkeld voor (epidemiologisch) onderzoek en daardoor ook minder gericht op de klinische praktijk.

Verder vonden wij dat patiënten met meerdere, of zwaardere diagnoses op de HoNOS het meest verbeteren; zij starten de behandeling met de slechtste scores en kunnen goed profiteren.

Een van de tekortkomingen van onze studie is dat de steekproef niet representatief is voor alle psychiatrische diagnoses. Voor toekomstig onderzoek lijkt het van belang de instrumenten te valideren voor steekproeven van patiënten met ernstiger psychiatrische aandoeningen, bijvoorbeeld een klinische setting. Daarnaast is het uitvalpercentage na de eerste meting hoog.

LITERATUUR

- Aarssen R. Betekenis van cliënttevredenheid als indicator voor kwaliteit van zorg. Amsterdam: Thela; 2003.
- Aartsen MJ, Spitsbaard AK, van Baarsen C, Dhondt ADF, Mascini M, Nefs A, e.a. Een multicenterstudie naar betrouwbaarheid, validiteit en gevoeligheid voor verandering van de HoNOS65+ binnen de ouderenpsychiatrie. *Tijdschr Psychiatr* 2010; 52: 543-53.
- Arrindell W, Boomsma A, Ettema H, Stewart R. Verdere steun voor het multi-dimensionale karakter van de SCL-90-R. *De Psycholoog* 2004; 39: 194-201.
- Arrindell WA, Ettema JHM. SCL-90: Handleiding bij een multidimensionele psychopathologie-indicator. Lisse: Swets & Zeitlinger; 1986/2003.
- Beurs E de, Den Hollander-Gijsman M, Buwalda V, Trijsburg W, Zitman F. De Outcome Questionnaire (OQ45): psychodiagnostisch gereedschap. *De Psycholoog* 2005; 40: 393-400.
- Broersma TW, Sytema S. Implementatie van het meetinstrument HoNOS65+. Onderzoek op een afdeling ouderenpsychiatrie. *Tijdschr Psychiatr* 2008; 50: 77-82.
- Chapman JE. Reliability and validity of the Progress Questionnaire: an adaptation of the Outcome Questionnaire. Philadelphia: Drexel University; 2003.
- Derogatis LR. The SCL-90 Manual; Scoring, administration and procedures for the SCL-90. Baltimore: Johns Hopkins University School of Medicine; 1977.
- Doerfler LA, Addis ME, Moran PW. Evaluating mental health outcomes in an inpatient setting: convergent and divergent validity of the OQ-45 and Basis-32. *J Behav Health Serv Res* 2002; 29: 394-403.
- Hermann RC. Quality Assessment and Improvement in a Changing

- Healthcare System. In: Hermann RC, red. Improving mental health care. A guide to measurement-based quality improvement. Arlington: American Psychiatric Publishing; 2005.
- Holi MM, Martunen 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: 233-8.
- Hunter R, Cameron R, Norrie J. Using patient reported outcomes in schizophrenia: the Scottish Schizophrenia Outcomes Study. *Psychiatr Serv* 2009; 60: 240-5.
- Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991; 59: 12-9.
- Jong K de, Nugter MA, Polak MG, Wagenborg, JEA, Spinhoven, P, Heiser, W. The Outcome Questionnaire (OQ-45) in a Dutch population: a cross-cultural validation. *Clin Psychol Psychother* 2007; 14: 288-301.
- Koeter M. Validity of the GHQ and SCL anxiety and depression scales: a comparative study. *J Affect Disord* 1992; 24: 271-80.
- Lambert MJ, Burlingame GM, Umphress V, Hansen NB, Vermeersch DA, Clouse GC, e.a. The reliability and validity of the Outcome Questionnaire. *Clin Psychol Psychother* 1996; 3: 249-58.
- Lambert M, Hansen NB, Finch AE. Client focused research: using client outcome data to enhance treatment effects. *J Consult Clin Psychol* 2001; 69: 159-72.
- Mulder CL, Staring ABP, Loos J, Buwalda VJA, Kuijpers D, Sytema S, e.a. De Health of the Nation Outcome Scales (HoNOS) als 'routine outcome assessment'. *Tijdschr Psychiatr* 2004; 46: 274-84.
- Mulder CL, van der Gaag M, Bruggeman R, Cahn W, Delespaul PAE, Dries P, e.a. Routine outcome monitoring voor patiënten met ernstige psychiatrische aandoeningen; een consensusdocument *Tijdschr Psychiatr* 2010; 52: 169-79.
- 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: 409-12.
- Oiesvold T, Bakkejord T, Sexton AJ. Concurrent validity of HoNOS compared with a patient derived measure (SCL-90-R) in out-patient clinics. *Psychiatry Res* 2011; 187: 297-300.
- Parabiaghi A, Barbato A, D'Avanzo P, Erlicher A, Lora A. Assessing reliable and clinically significant change on health of the nation outcome scales: methods for displaying longitudinal data. *Aust N Z J Psychiatry* 2005; 39: 719-25.
- Sharma VK, Wilkinson G, Fear S. Health of the nation outcome scales: a case study in general psychiatry. *Br J Psychiatry* 1999; 174: 395-8.
- Slade M, Beck A, Bindman T, Thornicroft G, Wright S. Routine clinical outcome measures for clients with severe mental illness: CANSAS and HoNOS. *Br J Psychiatry* 1999; 174: 404-8.
- Thornicroft G, Bebbington P, Leff J. Outcomes for long-term clients one year after discharge from a psychiatric hospital. *Psychiatry Serv* 2005; 56: 1416-22.
- Trauer T, Callaly T, Hantz P, Little J, Shields R, Smith J. Health of the Nation Outcome Scales. Results of the Victorian field trial. *Br J Psychiatry* 1999; 174: 380-8.
- Umphress VJ, Lambert MJ, Smart DJ. Concurrent and construct validity of the Outcome Questionnaire. *J Psychoeduc Assess* 1997; 15: 40-55.
- Vermeersch DA, Lambert MJ, Burlingame GM. Outcome Questionnaire: item sensitivity to change. *J Pers Assess* 2000; 74: 242-61.
- Vliet van IM, Beurs de E. Het Mini internationaal neuropsychiatrisch interview. Een kort gestructureerd diagnostisch psychiatrisch interview voor DSM-IV en ICD-10 stoornissen. *Tijdschr Psychiatr* 2007; 49: 393-7.
- Wilson SA, Becker LA, Tinker RH. Fifteen month follow-up of eye movement desensitization and reprocessing (EMDR) treatment for posttraumatic stress disorder and psychological trauma. *J Consult Clin Psychol* 1997; 65: 1047-56.
- Wing JK, Beevor AS, Curtis RH, Park BG, Hadden S, Burns H. Health of the Nation Outcome Scales (HoNOS): research and development. *Br J of Psychiatry* 1998; 172: 11-8.

AUTEURS

VICTOR BUWALDA is afdelingspsychiater, Altrecht, Utrecht, en was ten tijde van het onderzoek werkzaam bij de Robert Fleury Stichting (later onderdeel van Rivierduinen).

STASJA DRAISMA is methodoloog en datamanager, afdeling Onderzoek, GGZ inGeest Amsterdam.

JAN SMIT is hoogleraar Methodologie van longitudinaal psychiatrisch onderzoek, afdeling Psychiatrie en EMGO Instituut, VUmc en GGZ Ingeest, Amsterdam.

JAN SWINKELS is psychiater en hoogleraar Richtlijnontwikkeling in de geneeskunde aan de Universiteit van Amsterdam/Academisch Medisch Centrum en het Trimbos-instituut, Utrecht.

WILLEM VAN TILBURG is psychiater en emeritus hoogleraar Klinische Psychiatrie VUmc en Polikliniek Amsterdam, Leo Kanner Huis, Amsterdam.

Correspondentieadres: Victor Buwalda, Stadionweg 128-P, 1077
SW Amsterdam.
E-mail: imsep@xs4all.nl.

Geen strijdige belangen meegedeeld.

Het artikel werd voor publicatie geaccepteerd op 24-5-2011.

SUMMARY

Validation of two measuring instruments for routine outcome monitoring in psychiatry: the HORVAN study – V.J.A. Buwalda, S. Draisma, J.H. Smit, J.A. Swinkels, W. van Tilburg –

BACKGROUND Transparency in psychiatry can be increased by the use of routine outcome monitoring (ROM) instruments. Instruments should be easy to use and take very little time to complete; they also need to have psychometric qualities, be sensitive to change, and provide information about patients' symptoms, and about interpersonal and social functioning.

AIM To investigate to what extent the combination of Health of the Nation Outcome Scales (HoNOS) and the Outcome Questionnaire (OQ) in the Dutch situation meets the above-mentioned quality criteria and to examine how the combination relates to the Symptom CheckList (SCL-90).

METHOD Data for 148 patients collected at three measurement moments were available for analysis. The psychometric qualities of the instruments and their sensitivity to change were checked carefully.

RESULTS The three scales showed high values for internal consistency (Cronbach's alpha). The HoNOS total score and the subscales of the OQ correlated reasonably well with the SCL-90 total score (convergence validity). At the first measurements, patients with a comorbid diagnosis had the lowest scores (discrimination validity). The clinically significant change between T1 and T2 and between T2 and T3 was sufficiently high for all three measuring instruments.

CONCLUSION The combination of the HoNOS rating scale and the self-report list OQ seems to be suitable for ROM in psychiatry.

[TIJDSCHRIFT VOOR PSYCHIATRIE 53(2011)10, 715-726]

KEY WORDS clinically significant change, routine outcome monitoring (ROM), validation study