

ROM in de volwassenenpsychiatrie: een evaluatie van meetinstrumenten

V.J.A. BUWALDA, M.A. NUGTER, S.Y. VAN DE BRUG, S. DRAISMA, J.H. SMIT, J.A. SWINKELS, W. VAN TILBURG

ACHTERGROND Er is behoefte aan een evaluatie van de momenteel meest gebruikte ROM-instrumenten in de Nederlandse psychiatrie voor ernstige en niet ernstige psychiatrische aandoeningen.

DOEL Een overzicht geven van kenmerken en kwaliteit van instrumenten.

METHODE Middels een literatuurstudie werden 6 instrumenten geselecteerd die we beschrijven aan de hand van psychometrie, praktische en klinische kenmerken.

RESULTATEN De meeste instrumenten zijn geschikt en hebben voldoende kwaliteit.

DISCUSSIE De vraag is welke combinatie de meeste toekomst heeft voor zowel ernstige als niet ernstige psychiatrische aandoeningen.

[TIJDSCHRIFT VOOR PSYCHIATRIE 54(2012)2, 147-152]

TREFWOORDEN psychometrische kwaliteit, ROM-meetinstrumenten

Er bestaan geen vaste richtlijnen voor de keuze van meetinstrumenten voor routine outcome monitoring (ROM). In het inleidende artikel van Nugter en Buwalda (elders in dit nummer) worden klinische, praktische en wetenschappelijke criteria genoemd.

In dit artikel bespreken we de generieke ROM-instrumenten die in Nederland veel worden gebruikt om zowel het beloop van klachten als het sociaal en maatschappelijk functioneren te meten in de volwassenenpsychiatrie. De vraag die we trachten te beantwoorden is, in hoeverre deze veelgebruikte instrumenten voldoen aan wetenschappelijke en praktische criteria en dus geschikt zijn om toestandsveranderingen bij patiënten te meten.

WETENSCHAPPELIJKE CRITERIA

Bruikbare ROM-instrumenten zullen aan de minimale eisen van betrouwbaarheid en validiteit moeten voldoen zoals die bijvoorbeeld zijn omschreven in de COTAN-handleiding (Evers e.a. 2010) en door Van Beljouw en Verhaak (2010).

Recentelijk zijn er wetenschappelijke criteria ontwikkeld om klinisch betekenisvolle veranderingen in de ernst van psychische aandoeningen vast te stellen. Op geaggregeerd niveau wordt meestal de effectgrootte (*effect size*; ES) van het scoreverschil op verschillende tijdstippen in het behandelproces gebruikt. De ES wordt berekend door het scoreverschil te delen door de standaarddeviatie van de voormeting. Als indicatie noemt men een ES van 0,2 klein, een ES van 0,5 is gemiddeld en een ES van 0,8 is groot.

Voor het vaststellen van klinisch relevante veranderingen op individueel niveau ontwikkel-

den Jacobson en Truax (1991) de begrippen 'reliable change index' (betrouwbare verandering; BV) en 'clinical significant change' (klinisch significante verbetering; KSV). Een verandering is betrouwbaar als het verschil tussen de metingen significant is en gecorrigeerd voor meetfouten van het instrument.

Een dergelijke verandering hoeft echter niet klinisch relevant te zijn; voor een klinisch significante verbetering moet de patiënt bij de nameting in de scorerange van normaal functionerende personen vallen. Om behandelbeslissingen te funderen, zijn dus normscores nodig van referentiegroepen van gezonde personen, maar ook van verschillende – meer en minder ernstige – patiëntengroepen.

BESPREKING MEETINSTRUMENTEN

Een zestal meetinstrumenten komt voor de Nederlandse ROM in aanmerking: de beoordelingsinstrumenten 'Clinical Global Impression Scale' (CGI) en de 'Health of the Nation Outcome Scales' (HONOS), en de zelfrapportage-instrumenten: 'Outcome Rating Scale' (ORS), 'Symptoms Checklist-90' (SCL-90), 'Brief Symptom Index' (BSI) en 'Outcome Questionnaire' (OQ-45) (Mulder e.a. 2010).

In de volgende tekst evalueren wij de psychometrische kwaliteit van deze instrumenten en gaan we kort in op klinische en praktische kenmerken. Bij de psychometrische kwaliteit beschrijven we zowel de klassieke eisen van validiteit en betrouwbaarheid, als de mate waarin gevoeligheid voor verandering is onderzocht. Ook geven we aan of men het instrument kan gebruiken voor patiënten met ernstige psychiatrische aandoeningen (EPA) of patiënten die geen EPA hebben (N-EPA).

CGI

Algemeen De algemeenste beoordelingschaal is de CGI, die uit twee vragen met een zevenpuntsantwoordschaal bestaat. De CGI-S (Severity of Illness Scale) meet de ernst van de aandoening gedurende de laatste week en de CGI-I

(Global Improvement Scale) de veranderingen ten opzichte van het begin van de behandeling.

Psychometrie Psychometrisch onderzoek van de CGI in Nederland is niet bekend. Kadouri e.a. (2007) vonden interbeoordelaarsbetrouwbaarheidscoëfficiënten (ICC's) van 0,64 tot 0,94. Berk e.a. (2008) berekenden verschilcores voor klachtenverbetering, door de CGI-I direct te scoren en het verschil te berekenen tussen opname- en ontslagscores op de CGI-S. De correlatie tussen het CGI-S-verschil en de CGI-I-score was 0,71 en significant. Qua verandering werden bij patiënten met bipolaire, depressieve en psychotische stoornissen aanmerkelijke verbeteringen tussen opname en ontslag aangetroffen. De convergente validiteit was goed, waaronder de correlatie met de SCL-90 (Kadouri e.a. 2007) en de HONOS (Berk e.a. 2008). Er zijn geen normscores of effect sizes gerapporteerd.

Praktisch Deze beoordelingslijst is kort (2 items) en gratis beschikbaar. Een korte training is voldoende voor gebruik.

Klinisch De CGI levert uitsluitend twee algemene functioneringsscores op. Het instrument wordt zowel voor patiënten met EPA als met N-EPA ingezet.

HONOS

Algemeen De Health of the Nation Outcome Scales (HONOS) is een beoordelingsmeetinstrument voor EPA en wordt afgenomen door getrainde professionals. Twaalf items, gescoord op een vijfpuntsschaal, meten vier functioneringsgebieden: gedragsproblemen, beperkingen, symptomen en sociale problemen.

Psychometrie Mulder e.a. (2004) vonden een acceptabele Cronbachs alfa van 0,78 voor de totaalscore en een ICC van 0,92. De convergente validiteit, samenhang met andere instrumenten, bleek goed. Bij het vaststellen van de discrimi-

nante validiteit bleek dat totaalscores significant verschilden bij patiënten met verschillende behandelingenintensiteiten en diagnoses. Broersma en Sytema (2008) en Aartsen e.a. (2010) bevestigden dit beeld voor de ouderenpsychiatrie, waarbij Aartsen e.a. ook verandering bepalen met verschillen, maar deze niet toetsen. Broersma en Sytema (2010) rapporteren voor de HoNOS 65+ een hoge effect size van 0,84 op de totaalscore en de subschaal symptomen en een matige ES op de subschaal sociale problemen.

In internationaal onderzoek naar verandering vinden Sharma e.a. (1999) voor verschillende diagnoses verbetering op de subschalen sociale problemen en symptomen, met t-toetsen. Audin e.a. (2001) vonden verbetering, een goede ES (0,69) en KSV (24%) bij ambulante patiënten, die vooral gedragen werd door drie items (depressie, andere problemen, sociale contacten). Zij concluderen dat de HoNOS minder bruikbaar is voor N-EPA. Buwalda e.a. (2011) vinden de HoNOS wel bruikbaar voor ambulante patiënten met N-EPA. Wel blijkt de KSV-methode streng, resulterend in lage percentages patiënten met KSV.

Praktisch De HoNOS is kort. Er zijn geen kosten aan verbonden, maar wel is een training nodig.

Klinisch Het instrument omvat meerdere dimensies die relevant zijn voor behandeling en kan zowel voor EPA als N-EPA gebruikt worden.

ORS

Algemeen De Outcome Rating Scale is een zelfrapportagelijst van vier items met een visuele-analoge antwoordschaal, gescoord van 0-10 (Miller e.a. 2003). De items vragen naar individueel, relationeel, sociaal en algeheel functioneren. De domeinen zijn afgeleid van de OQ-45. Het is de bedoeling om de ORS na elk consult af te nemen.

Psychometrie De ORS is in Nederland door Van Beljouw en Verhaak (2010) onderzocht onder

eerstelijnspsychologen. Zij vonden een alfa van 0,82 en matige correlaties met de SCL-90. Miller e.a. (2003) en Bringham e.a. (2006) vonden hogere alfawaarden. Hafkenscheid e.a. (2010) onderzochten verandering via de KSV-methode en stelde vast dat men met de ORS verandering zichtbaar gemaakt kan worden.

Praktisch De ORS is kort en er zijn twee versies beschikbaar. Aan de vertaalde versie van Scot Miller zijn kosten verbonden.

Klinisch Deze schaal wordt vooral gebruikt in psychotherapie, maar gaat niet expliciet in op psychiatrische stoornissen.

SCL-90

Algemeen De Symptom Checklist-90 (SCL-90; Arindell & Ettema 1986; 2003; Derogatis e.a. 1973) is een multidimensionele zelfbeoordelingsschaal van 90 items met 5-puntsantwoordschalen voor de belangrijkste psychopathologische klachtengebieden. De SCL-90 wordt in Nederland veel gebruikt. De Nederlandse versie meet acht symptomdimensies: wantrouwen en interpersoonlijke sensitiviteit, somatische klachten, depressie, angst, agorafobie, woede-hostiliteit, insufficiëntie van denken en handelen en slaapproblemen.

Psychometrie De Cronbachs alfa's voor de subschalen variëren van 0,73 tot 0,97 en test-hertestcoëfficiënten van 0,62 tot 0,91 (Arindell & Ettema 2003). De auteurs beschrijven ook normgroepen. Zij vinden dat de effecten van zowel korte als lange behandelingen met de SCL-90 gemeten kunnen worden en beschrijven studies waarin verandering wordt gemeten met t-toetsen. Buwalda e.a. (2011) vonden bij één op de vier ambulante patiënten klinisch significante verbetering.

Praktisch De afnameduur is 20 minuten, een scoreformulier kost € 1,66.

Klinisch De SCL-90 bestrijkt meerdere gebieden uit de psychiatrische stoornissen en wordt vooral gebruikt voor N-EPA.

BSI

Algemeen en psychometrie De Brief Symptom Index (BSI; De Beurs & Zitman 2006), een zelfbeoordelingsschaal, is een afgeleide van de SCL-90 met 53 items en heeft vergelijkbare cijfers voor de psychometrische eigenschappen (alfa hoger dan 0,71, goede correlaties met andere instrumenten).

Praktisch De afnameduur van de BSI is korter dan de SCL-90 en een scoreformulier kost € 1,75. Een Nederlandse, gevalideerde versie van de BSI is beschikbaar.

Klinisch De BSI bestrijkt meerdere gebieden en wordt vooral gebruikt voor N-EPA.

OQ-45

Algemeen De Outcome Questionnaire-45 (OQ-45; Lambert e.a. 2004), een zelfbeoordelingsinstrument, bestaat uit 45 vragen gescoord op een vijfpuntsschaal, en meet drie domeinen: symptomen, interpersoonlijke relaties en sociaal-maatschappelijke rol.

Psychometrie In Nederland is de OQ onderzocht door De Beurs e.a. (2005), De Jong en Nugter (2004) en De Jong e.a. (2007). Cronbachs alfa's voor subschalen en totaalscore variëren van 0,68-0,95. Test-hertestbetrouwbaarheid varieert van 0,68-0,83. Ook werden correlaties hoger dan 0,75 gevonden voor samenhang met andere instrumenten. De Jong e.a. (2007) en De Beurs e.a. (2005) constateerden een goede discriminante validiteit en leverden normscores voor personen zonder klachten. Ook vonden zij verandering na een korte behandeling op de totaalscore, met een t-toets en een effectgrootte van 1,33. Buwalda e.a. (2011) vonden bij één op de drie ambulante patiënten klinisch significante verbetering.

Praktisch De OQ is in te vullen in 10 minuten; kosten zijn per patiënt € 0,73 (meerdere afnames mogelijk).

Klinisch Meerdere gebieden worden gemeten, in Nederland is zij vooralsnog niet voor EPA onderzocht.

DISCUSSIE

De zes besproken meetinstrumenten bestrijken samen het klachtenspectrum op as I van de DSM-IV-TR-classificatie en komen vanwege voldoende betrouwbaarheid en validiteit in nationale en internationale literatuur in aanmerking voor gebruik voor ROM in Nederland. De zelfrapportageinstrumenten SCL-90 en BSI pretenderen een breed meetbereik te hebben, maar in de praktijk worden deze meetinstrumenten zelden voor patiënten met EPA (zoals psychotische stoornissen) gebruikt. Daarnaast zijn ze duur in gebruik. Ook de OQ-45 wordt ingezet voor de lichtere categorie patiënten en is niet gratis. De ORS en CGI zijn (nog) niet voldoende onderzocht in Nederland. De HONOS is gratis, maar er is wel een training nodig voor afname. De beoordelingsinstrumenten voor patiënten met EPA zijn gratis en de zelfrapportage-instrumenten voor de groep met N-EPA niet.

Een combinatie van de HONOS en de OQ (resp. beoordeling en zelfrapportage) zou niet alleen het breedste stoornisspectrum bestrijken, maar zou behalve in de klachten ook inzicht geven in het sociaal en maatschappelijk functioneren van de patiënt. Mogelijkerwijs is het tevens het goedkoopste duo. Validerend onderzoek naar de losse instrumenten is klaar. Onderzoek waarbij de combinatie van de HONOS en OQ-45 wordt onderzocht, is in een afrondende fase (Buwalda e.a. 2011) en de resultaten lijken deze combinatie te ondersteunen.

LITERATUUR

- Aartsen MT, 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; 53: 543-53.
- Arindell WA, Ettema JHM. SCL-90: Handleiding bij een multidimensionele psychopathologie-indicator. Lisse: Swets & Zeitlinger; 2003.
- Audin K, Margison FR, Clark JM, Barkham M. Value of HoNOS in assessing patient change in NHS psychotherapy and psychological treatment service. *Br J Psychiatry* 2001; 178: 561-6.
- Beljouw IMJ van, Verhaak PFM. Geschiede uitkomstmaten voor routinematige registratie door eerstelijnspsychologen. Utrecht: NIVEL; 2010.
- Berk M, Ng F, Dodd S, Callaly T, Campbell S, Bernardo M, e.a. 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: 979-83.
- Beurs E de, Hollander-Gijsman M den, Buwalda VJA, Trijsburg W, Zitman, F. De Outcome Questionnaire (OQ45): psychodiagnostisch gereedschap. *De Psycholoog* 2005; 40: 393-400.
- Beurs E de, Zitman F. De Brief Symptom Inventory (BSI). De betrouwbaarheid en validiteit van een handzaam alternatief voor de SCL-90. *Maandblad Geestelijke volksgezondheid* 2006; 61: 120-41.
- Broersma TW, Sytema S. Implementatie van het meetinstrument HoNOS65+. Onderzoek op een afdeling voor ouderenpsychiatrie. *Tijdschr Psychiatr* 2008; 50: 77-82.
- Broersma TW, Sytema S. De Health of the Nation Outcome Scales als effectmaat in de ouderenpsychiatrie. *Tijdschr Gerontol Geriatri* 2010; 41: 13-8.
- Bringhurst MDL, Watson CW, Miller SD, Duncan BL. The reliability and validity of the Outcome Rating Scale: a replication study of a brief clinical measure. *J Brief Therapy* 2006; 5: 23-8.
- Buwalda VJA, Draisma S, Smit JH, Swinkels JA, van Tilburg W. Validering van twee meetinstrumenten voor routine outcome monitoring in de psychiatrie: de HORVAN-studie. *Tijdschr Psychiatr* 2011; 53: 715-26.
- Derogatis LR, Lipman RS, Covi L. SSCLx90: an outpatient psychiatric rating scale- preliminary report. *Psychopharmacol Bull* 1973; 9: 13-27.
- Evers A, Lucassen W, Meijer R, Sijtsma K. COTAN Beoordelingssysteem voor de kwaliteit van tests. Geheel herziene versie. Amsterdam: NIP; 2010.
- Hafkenscheid A, Duncan BL, Miller SD. The Outcome and Session Rating Scales: A cross-cultural examination of the psychometric properties of the Dutch translation. *J Brief Psychother* 2010; 7.
- 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 WJ. The Outcome Questionnaire (OQ-45) in a Dutch population: a cross-cultural validation. *Clin Psychol Psychother* 2007; 14: 288-301.
- Jong K de, Nugter AM. De Outcome Questionnaire: psychometrische kenmerken van de Nederlandse vertaling. *Nederlands Tijdschrift voor Psychologie* 2004; 59: 76-9.
- Kadouri A, Corruble E, Falissard B. The improved Clinical Global Impression Scale (iCGI): development and validation in depression. *BMC Psychiatry* 2007; 7: 7.
- Lambert MJ, Morton JJ, Hatfield DR, Harmon C, Hamilton S, Shimokawa K, e.a. Administration and scoring manual for the OQ-45.2 (Outcome Questionnaire) (3de druk). Wilmington: American Professional Credentialing Services LLC; 2004.
- Miller SD, Duncan BL, Brown J, Sparks JS, Claud DA. The Outcome Rating Scale: A preliminary study on the reliability, validity and feasibility of a brief visual analog measure. *J Brief Therapy* 2003; 2: 91-100.
- Mulder CL, Staring ABP, Loos J, Buwalda VJA, Kuijpers D, Sytema S, e.a. De Health of the Nation Outcome Scale (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.
- 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.

AUTEURS

VICTOR BUWALDA is psychiater, Altrecht GGZ, en onderzoeker, VUmc/GGZ InGeest, Amsterdam.

ANNET NUGTER is hoofd afdeling Onderzoek en Monitoring, GGZ Noord-Holland-Noord, Heerhugowaard.

SUSAN VAN DE BRUG is psycholoog en onderzoeker, Altrecht GGZ.

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.

JAN SWINKELS is psychiater en hoogleraar Richtlijnontwikkeling in de gezondheidszorg, AMC, Amsterdam.

WILLEM VAN TILBURG is psychiater en emeritus hoogleraar Klinische psychiatrie, VUmc/GGZ inGeest.

Correspondentieadres: Victor Buwalda, Stadionweg 128, 1077 SW Amsterdam.

E-mail: imsep@xs4all.nl

Geen strijdige belangen meegedeeld.

Het artikel werd voor publicatie geaccepteerd op 8-9-2011.

SUMMARY

ROM in adult psychiatry: an evaluation of measurement instruments – V.J.A. Buwalda, M.A. Nugter, S.Y. van de Brug, S. Draisma, J.H. Smit, J.A. Swinkels, W. van Tilburg –

BACKGROUND An evaluation of the most commonly used ROM measures in Dutch psychiatry is lacking, both for severe mental illnesses and for common psychiatric disorders.

AIM To provide an overview of the characteristics and quality of outcome measures.

METHOD A literature study yielded six outcome measures. The psychometrical, clinical and practical aspects of these scales are described.

RESULTS The measures are suitable and are of adequate quality.

DISCUSSION It remains to be seen if any of the outcome measures are suitable for both serious and less serious mental illnesses. The use of a combination of a self-rating scale and an observer-rating scale that measure symptoms and domains of functioning may be the most promising choice.

[TIJDSCHRIFT VOOR PSYCHIATRIE 54(2012)2, 147-152]

KEY WORDS outcome measures, psychometric quality, ROM rating scales