

ROM in de ggz: behandeluitkomsten met specifieke en generieke instrumenten bij patiënten met angststoornissen; een cohortonderzoek

M. A. NUGTER, G.P. OOSTERDIJK, A.A.P. VAN EMMERIK

- ACHTERGROND** Bij routine outcome monitoring (ROM) wordt veelal gebruikgemaakt van generieke vragenlijsten. Dat kan het klinisch gebruik van de vragenlijsten in de weg staan en een onderschatting betekenen van de feitelijke behandelresultaten.
- DOEL** Nagaan of er verschillen zijn in de mate van vooruitgang op specifieke vragenlijsten voor angstklachten en de mate van vooruitgang op de subschaal Symptomatische Distress van de Outcome Questionnaire-45 (OQ-45-SD).
- METHODE** Data van 604 patiënten met een angststoornis werden gebruikt. Uitkomsten op OQ-45-SD werden vergeleken met zeven veelgebruikte vragenlijsten voor angstklachten. Gemiddelde verschilscores, effectgroottes en percentages betrouwbare verbetering werden vergeleken.
- RESULTATEN** De effectgroottes van de OQ-45-SD en de verschillende stoornisspecifieke instrumenten waren 0,80 of hoger. De correlaties tussen de verschilscores waren matig, evenals de samenhang tussen de mate waarin patiënten betrouwbaar verbeterden. In een derde tot de helft van de gevallen was er sprake van een discrepantie.
- CONCLUSIE** Behandeluitkomsten op de OQ-45-SD en stoornisspecifieke vragenlijsten bij patiënten met angststoornissen zijn ongeveer even positief. Tegelijkertijd zijn discrepanties zichtbaar in de mate van vooruitgang. In overeenstemming met het advies van het International Consortium for Health Outcomes Measurement bevelen wij aan beide soorten vragenlijsten te gebruiken bij het evalueren van de voortgang van de behandeling.

TIJDSCHRIFT VOOR PSYCHIATRIE 60(2018)11, 741-749

TREFWOORDEN angststoornissen, ROM, routine outcome monitoring, stoornisspecifieke vragenlijsten



Bij het gebruik van *routine outcome monitoring* (ROM) in de ggz kan een indeling worden gemaakt in *outcome management* en *clinical management* (Andrews & Page 2005; Barendregt 2015). Outcome management betreft het gebruik van geaggregeerde data waarmee uitkomsten op groepsniveau worden geanalyseerd en geëvalueerd. Bij clinical management gaat het om het gebruik van ROM om de behandeling te sturen. Zowel voor als tijdens de behandeling vullen

patiënten meetinstrumenten in en het verloop van de scores wordt gebruikt als klinische feedback, op basis waarvan de behandeling zo nodig wordt aangepast. Verschillende systematische reviews en meta-analyses hebben laten zien dat het gebruik van ROM om de behandeling te sturen effectief kan zijn (Knaup e.a. 2009; Carlier e.a. 2010; Shimokawa e.a. 2010; Kendrick e.a. 2016). De effectiviteit lijkt vooral aanwezig bij patiënten die onvoldoende

voortuitgaan, de zogenaamde *not on track*(NOT)-gevallen, in korte behandelingen, en als er meer dan 1 keer feedback is gegeven.

In vrijwel alle ROM-systemen, zowel in Nederland als daarbuiten, wordt voor de ROM gebruikgemaakt van generieke instrumenten. Boswell e.a. (2015) raden ook aan om de instrumenten kort en simpel te houden. Andere auteurs adviseren een meer multidimensioneel systeem van ROM met zowel generieke als stoornisspecifieke instrumenten (Kelly & Bickman 2009; Delespaul 2015; Obbarius e.a. 2017).

Voor outcome management is de keuze voor een generiek instrument misschien logisch, voor clinical management zou de keuze van de instrumenten vooral bepaald moeten worden door de relevantie ervan voor de behandeling (Nugter & Buwalda 2012). In hun feedbacktheorie stellen Riemer en Bickman (2011) ook dat de waargenomen relevantie en de specificiteit van de ROM-feedbackboodschap belangrijke kenmerken zijn die kunnen bepalen of de clinicus de feedback ook kan gebruiken voor de behandeling. In de twee onderzoeken waarbij het gebruik is geregistreerd, blijkt dat niet alle behandelaars de ROM gebruikten in de behandeling (Bickman e.a. 2011; De Jong e.a. 2012). Dat is niet nieuw: ook voordat de ROM werd geïntroduceerd, was het gebruik van vragenlijsten in de ggz niet structureel ingebed in de evaluatie van de behandelvoortgang, terwijl er voor de verschillende stoornissen voldoende valide en betrouwbare instrumenten voorhanden zijn. Het gebruik van vragenlijsten is, mede door het huidige verplichte karakter ervan, veel gangbaarder geworden. De vraag is of het gebruik ervan kan worden verbeterd door de ROM-systeematiek meer te laten aansluiten op de specifieke problematiek waar de behandeling over gaat.

Ook voor outcome management is het de vraag of instrumenten zo veel mogelijk generiek moeten zijn. Het maakt de vergelijkbaarheid van de uitkomsten op groepsniveau makkelijker, maar een belangrijke beperking kan zijn dat de uitkomsten onvoldoende de feitelijke behandelresultaten reflecteren. Zo is aangetoond dat behandelresultaten in de praktijk minder positief zijn dan in gerandomiseerde gecontroleerde trials (RCT's; Hansen e.a. 2002; Barkham e.a. 2008). Een verklaring hiervoor is de doorgaans strikte controle in RCT's op het volgen van behandelprotocollen, dat in de praktijk kan worden bevorderd door het inzetten van de ROM. Een andere verklaring kan liggen in de keuze van de instrumenten: de primaire uitkomsten van RCT's betreffen vaak uitkomsten op de specifieke klachten die worden behandeld in de experimentele conditie (Aderka e.a. 2012). Hooguit in tweede instantie worden meer generieke uitkomsten gemeten.

AUTEURS

ANNET NUGTER, senior onderzoeker en hoofd afd.

Onderzoek en Monitoring, GGZ Noord-Holland-Noord.

GABRIELLE OOSTERDIJK, basispsycholoog, GGZ Noord-Holland-Noord en psychologiepraktijk Hollands Noorden.

ARNOLD VAN EMMERIK, onderzoeker, programmagroep Klinische Psychologie, Faculteit der Maatschappij- en Gedragwetenschappen, Universiteit van Amsterdam.

CORRESPONDENTIEADRES

M.A. Nugter, GGZ Noord-Holland-Noord, afd. Onderzoek en Monitoring, Postbus 18, 1850 BA Heiloo.

E-mail: a.nugter@ggz-nhn.nl

Geen strijdige belangen meegedeeld.

Het artikel werd voor publicatie geaccepteerd op 22-3-2018.

Er is een aantal onderzoeken gedaan waarin de voortuitgang op generieke instrumenten is vergeleken met die op specifieke instrumenten. We beperken ons tot vergelijkingen die in Nederland zijn gedaan met de *Brief Symptom Inventory* (BSI) en de *Outcome Questionnaire-45* (OQ-45). Aangetoond is dat bij patiënten met eetstoornissen de BSI minder sensitief is voor verandering dan een specifieke vragenlijst voor eetstoornissen (Dingemans & Van Furth 2017). Bij cliënten met depressieve stoornissen is vrijwel evenveel voortuitgang te zien op de BSI en op de subschaal Symptomatische Distress (SD) van de OQ-45 als op een depressieve klachtenlijst, terwijl de samenhang tussen de mate van voortuitgang op beide typen instrumenten gematigd was (Nugter e.a. 2017). Bij angststoornissen is de mate van voortuitgang op de BSI en verschillende specifieke vragenlijsten voor angststoornissen vergelijkbaar terwijl de totaalscore van de OQ-45 veel minder voortuitgang liet zien (Van der Mheen e.a. 2018). Een mogelijke verklaring voor deze laatste bevinding is dat met het gebruik van de OQ-45-totaalscore ook het functioneren van de patiënten in de vergelijking is meegenomen. Dat is ook de reden dat de Stichting Benchmark GGZ bij de benchmarking alleen de OQ-45-SD-schaal gebruikt.

Het doel van het huidige onderzoek is na te gaan of er verschillen zijn in mate van voortuitgang op specifieke vragenlijsten voor angstklachten en die op de meer generieke OQ-45-SD. Op basis van het onderzoek dat is gedaan, veronderstellen we dat de mate van verbetering vergelijkbaar zal zijn, maar dat de samenhang in voortuitgang op de OQ-45-SD en die op de stoornisspecifieke vragenlijsten gematigd zal zijn.

METHODE

Studieopzet

Wij verrichtten een prospectief cohortonderzoek. Bij de beschrijving zijn we zo veel mogelijk uitgegaan van de richtlijnen voor de rapportage van cohortonderzoek zoals die staan beschreven in het STROBE-statement (*Strengthening the Reporting of Observational Studies in Epidemiology*; von Elm e.a. 2008). Daarbij maakten we gebruik van data die bij GGZ Noord-Holland-Noord (GGZ-NHN) worden gegenereerd in het kader van de ROM, waarbij aan de behandelaar wordt overgelaten of deze, naast de landelijke verplichte generieke vragenlijsten, ook stoornisspecifieke vragenlijsten afneemt. In principe worden de vragenlijsten aan het begin en einde van de behandeling en verder eens per drie maanden door de patiënten ingevuld.

Deelnemers

Voor dit onderzoek werd gebruikgemaakt van anonieme data van 604 patiënten met een angststoornis volgens de *Diagnostic and Statistic Manual of Mental Disorders* (DSM-IV-TR; APA 2000) bij wie zowel een generieke vragenlijst als een stoornisspecifieke vragenlijst was gebruikt om het behandelresultaat te evalueren. De diagnose werd aan het begin

van de behandeling vastgesteld, aan de hand van een of meer intakegesprekken. In de organisatie wordt gewerkt met zorgprogramma's die gebaseerd zijn op de multidisciplinaire richtlijnen; bij angststoornissen wordt veelal cognitieve gedragstherapie aangeboden, al dan niet in combinatie met farmacotherapie.

Met het gebruik van stoornisspecifieke vragenlijsten werd begonnen in 2010; vanaf 2013 werden deze toenemend toegepast. De angststoornis kon een hoofddiagnose zijn, maar ook een nevendiagnose. Hiervoor werd gekozen, omdat de behandelaar de angststoornis, ook als deze comorbide was, kennelijk belangrijk genoeg vond om de behandeling ook met een stoornisspecifieke vragenlijst te evalueren.

De volgende diagnoses werden geïnccludeerd: agorafobie, gegeneraliseerde angststoornis, obsessieve-compulsieve stoornis, paniekstoornis, posttraumatische stressstoornis, sociale fobie, specifieke fobie en angststoornis NAO. Er werden alleen gegevens gebruikt van patiënten van 18 tot en met 65 jaar die hun behandeling hadden afgerond.

Vragenlijsten

Voor de generieke vragenlijst werd gebruikgemaakt van de OQ-45-SD (Lambert e.a. 1996; De Jong e.a. 2008). De SD-schaal

TABEL 1 Instrument, meetpretentie, aantal items en antwoordtype van de gebruikte stoornisspecifieke vragenlijsten

Instrument	Meetpretentie	Antwoordtype
AGO 20 items	Symptomen horend bij een agorafobie	0 = geheel geen angst tot 4 = zeer veel angst
PI-R 41 items	Symptomen horend bij een obsessief compulsieve stoornis	0 = nooit/helemaal niet tot 4 = zeer vaak
POL 45 items	Symptomen horend bij een paniekstoornis	0 = helemaal geen tot 100 = zeer veel*
PSWQ 16 items	Symptomen horend bij een gegeneraliseerde angststoornis	1 = helemaal niet kenmerkend tot 5 = erg kenmerkend
SFS 20 items	Symptomen horend bij een sociale fobie; angst in relatie tot observatie door anderen	0 = helemaal niet van toepassing tot 4 = helemaal van toepassing
SIAS 20 items	Symptomen horend bij een sociale fobie, angst in sociale interacties	0 = helemaal niet van toepassing tot 4 = helemaal van toepassing
SVL-22 22 items	Symptomen horend bij een posttraumatische stressstoornis	0 = helemaal niet tot 4 = erg vaak

*Bij de POL verschillen de antwoordmogelijkheden per subschaal.

AGO: *Agorafobie Schaal* (Ost 1990); PI-R: *Padua Inventory-Revised* (Sanavio 1988); POL: *Paniek Opinie Lijst* (Telch 1987); PSWQ: *Penn State Worry Questionnaire* (Meijer e.a. 1990); SFS: *Sociale Fobie Schaal* (Mattick & Clarke 1998); SIAS: *Social Interaction and Anxiety Scale* (Mattick & Clarke 1998); SVL-22: *Schokverwerkingslijst* (Horowitz e.a. 1979).

bestaat uit 25 vragen die beantwoord worden op een schaal die loopt van nooit (0) tot bijna altijd (4). Zowel de betrouwbaarheid als de validiteit van de Nederlandse versie wordt goed bevonden (De Jong e.a. 2007).

De volgende stoornisspecifieke vragenlijsten werden gebruikt: de *Agorafobie Schaal* (AGO; Ost 1990), de *Padua Inventory-Revised* (Sanavio 1988), de *Paniek Opinie Lijst* (POL; Telch 1987), de *Penn State Worry Questionnaire* (PSWQ; Meijer e.a. 1990), de *Sociale Fobie Schaal* (SFS; Mattick & Clarke 1998), de *Social Interaction and Anxiety Scale* (SIAS; Mattick & Clarke 1998) en de *Schokverwerkingslijst* (SVL-22; Horowitz e.a. 1979). Deze vragenlijsten en een aantal van hun eigenschappen staan weergegeven in **TABEL 1**. Voor alle vragenlijsten geldt dat hogere scores wijzen op meer klachten.

Procedure

Voor aanvang van de behandeling kregen patiënten schriftelijk informatie over de ROM. Ook werden patiënten op de hoogte gebracht van het anonieme gebruik van de gegevens voor wetenschappelijk onderzoek. Patiënten konden daartegen bezwaar maken. Geselecteerd werden die metingen die binnen drie maanden voor of na het eerste behandelcontact en binnen drie maanden voor of na het laatste behandelcontact waren afgenomen.

Data-analyse

Van de variabelen werd getoetst of er sprake was van een normaalverdeling. Vooruitgang van patiënten werd op meerdere manieren beoordeeld en getoetst.

1. De effectgrootte (Cohen 1988) van de verschillen op de OQ-45-SD-schaal werd vergeleken met die van de stoornisspecifieke lijsten. De effectgrootte werd berekend met de formule: $Cohens\ d = \frac{M1 - M2}{\sigma}$; waarbij $\sigma = \sqrt{SD1^2 + SD2^2 - 2 * r * SD1 * SD2}$. Deze laatste formule neemt de correlatie tussen de voor- en nameting mee, om zo te corrigeren voor de afhankelijkheid tussen beide metingen. Wanneer sprake was van een niet-normale verdeling van de verschillen werd Spearman's rangcorrelatie gebruikt in de formule.
2. Per diagnose werd met Pearson's correlatie getoetst of er een samenhang bestond tussen de verschillen op de OQ-45-SD-schaal en die op de betreffende stoornisspecifieke lijst. Bij een niet-normale verdeling werd Spearman's rangcorrelatie gebruikt.
3. Per vragenlijst werd de mate van vooruitgang op de OQ-45-SD-schaal en op de betreffende stoornisspecifieke vragenlijst vergeleken volgens de definitie van de *reliable change index* (RCI; Jacobson & Truax 1991). De RCI geeft aan in hoeverre een gevonden verandering ook een daadwerkelijke verandering representeert en niet slechts het gevolg is van een onbetrouwbaarheid van de meting. Bij een RCI groter dan 1,96 kan ervan worden

gegaan dat er echte verandering heeft plaatsgevonden, omdat de kans dat de gevonden verandering door toeval ontstaan is, kleiner is dan $p < 0,05$.

Van de OQ-45-SD was de RCI bekend (De Jong e.a. 2008), van de andere instrumenten niet. De RCI van deze instrumenten werd berekend met de volgende formule: $RCI = \frac{MI - M2}{Sdiff}$; waarbij MI en M2 staan voor de gemiddeldes op respectievelijk de voormeting en de nameting, waarbij $Sdiff = \sqrt{2(SE)^2}$ en waarbij $SE = SI * \sqrt{1 - rxx}$. Hierbij is SI de standaarddeviatie op de voormeting van de vragenlijsten. Voor de rxx gebruikten we de betrouwbaarheidscoëfficiënt Cronbach's alfa van de voormeting. Met deze RCI werd per vragenlijst bekeken hoeveel patiënten vooruitgingen, stagneerden of achteruitgingen en met de kappa-waarde (Ronk e.a. 2012) werd de overeenstemming vastgesteld tussen de verschillende specifieke vragenlijsten en de OQ-45-SD. Hierbij hanteerden we de volgende criteria: 0,01-0,20 = gering; 0,21-0,40 = matig; 0,41-0,60 = redelijk; 0,61-0,80 = substantieel; 0,81-1,00 = (bijna) perfect (Landis & Koch 1977).

RESULTATEN

Deelnemers

Van de 604 patiënten met voor- en nametingen op de OQ-45-SD en een specifieke vragenlijst was 211 (34,9%) man en 393 (65,1%) vrouw. De gemiddelde leeftijd was 34,72 (SD 11,99) jaar. Er waren 571 (94,5%) deelnemers met een angststoornis als hoofddiagnose. Bij de overige 33 deelnemers (5,5%) was de angststoornis comorbide, van wie bijna de helft ($n = 14$; 42,4%) met een stemmingsstoornis als hoofddiagnose. De gemiddelde behandelduur van de patiënten was 44,70 weken (SD 26,25).

Continue uitkomsten

In **TABEL 2** staat voor elke stoornisspecifieke vragenlijst de correlatie van de verschillen op de stoornisspecifieke vragenlijst en de OQ-45-SD weergegeven en de effectgrootte. Bij de stoornisspecifieke vragenlijsten ging het om de totaalscores, met uitzondering van de POL, waarvoor alleen subschaalscores beschikbaar waren: paniekverwachting, zorgen over de gevolgen van een paniekaanval, en gebrek aan zelfvertrouwen. De aantallen verschillen per vragenlijst.

De effectgrootte van de OQ-45-SD varieerde per subgroep. Het effect op deze schaal was over het algemeen groot ($> 0,80$), en was bij de groep bij wie ook de SFS was afgenomen met 0,79 het laagst.

Bij de stoornisspecifieke vragenlijsten waren de nametingen van de SVL, de SFS, de AGO en de twee van POL-subschalen niet normaal verdeeld en werd de effectgrootte berekend met Spearman's rangcorrelatie (zie **TABEL 2**). Ook bij de

TABEL 2 Correlaties tussen de verschillcores en effectgroottes op de stoornisspecifieke vragenlijsten en de subschaal Symptomatic Distress van de Outcome Questionnaire-45 (OQ-45-SD)

Stoornisspecifieke vragenlijst	Correlatie met verschillcores OQ-45-SD	Effectgrootte specifiek	Effectgrootte OQ-45-SD
AGO (n = 118)	0,51*	0,92*	0,99
PADUA (n = 61)	0,33*	1,19	1,21
POL (n = 186)			
POL-paniekverwachting	0,43	0,92*	1,09
POL-zorgen over gevolgen	0,45	0,78*	1,09
POL-gebrek zelfvertrouwen	0,51	1,02	1,09
PSWQ (n = 103)	0,59	0,81	1,14
SFS (n = 75)	0,62*	0,86*	0,79
SIAS (n = 73)	0,70	0,84	0,84
SVL-22 (n = 179)	0,63	1,17*	1,03

*Spearman's rangcorrelatie

Voor afkortingen: zie **TABEL 1**.

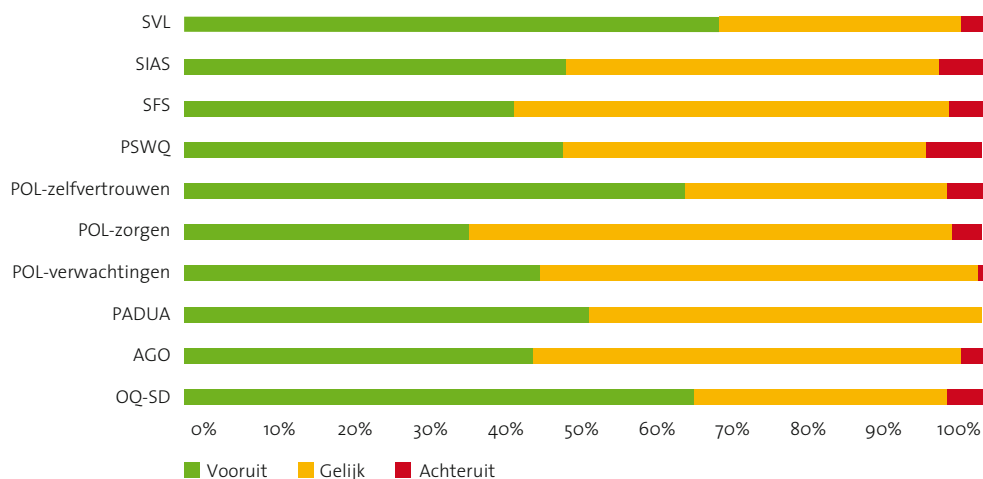
stoornisspecifieke vragenlijsten zagen we grote effecten, waarbij de effectgrootte van een van de subschalen van de POL (zorgen over de gevolgen van een paniekverwachting) met 0,78 het laagst was. Er waren zowel overeenkomsten als discrepanties te zien tussen de effectgroottes op de stoornisspecifieke lijsten en de OQ-45-SD. Dit gold vooral bij de POL, waar de effectgroottes op de subschalen varieerden, en bij de PSWQ, waar de grootte van het effect op OQ-45-SD beduidend in positieve zin afweek van dat van de PSWQ. De correlaties op de AGO, de SFS en de PADUA werden gebaseerd op Spearman's rangcorrelatie, omdat deze verschillcores niet normaal verdeeld waren. De tabel laat zien dat de correlaties tussen de verschillcores van de OQ-45-SD en

die van de stoornisspecifieke vragenlijsten varieerden van 0,33 voor de PADUA tot 0,70 voor de SIAS. Het hing dus van het type problematiek af in hoeverre de OQ-45-SD-schaal de mate van verandering op de specifieke vragenlijst reflecteerde. Dat lukte het beste bij de SIAS en de SFS.

Categorische uitkomsten

Per deelnemer berekenden we per vragenlijst de verschillscore en vervolgens gingen we na hoeveel deelnemers betrouwbaar vooruit waren gegaan, stabiel waren gebleven of betrouwbaar achteruit waren gegaan. De uitkomsten daarvan zijn te zien in **FIGUUR 1**. De verandering op de OQ-45-SD-subschaal was gebaseerd op de totale groep (n =

FIGUUR 1 Mate van vooruitgang per vragenlijst op basis van de reliable change index (RCI; Jacobson & Truax 1991)



Voor afkortingen: zie **TABEL 1**.

TABEL 3 Mate van overeenstemming in vooruitgang tussen de stoornisspecifieke lijsten en de subschaal Symptomatic Distress van de Outcome Questionnaire-45 (OQ-45-SD)

Vragenlijst	% overeenstemming	Kappa	p
AGO (n = 118)	61,9	0,29	< 0,001
PADUA (n = 61)	52,5	0,06	ns
POL (n = 186)			
POL-paniekverwachting	56,9	0,19	< 0,01
POL-zorgen over gevolgen	51,6	0,15	< 0,05
POL-gebrek zelfvertrouwen	63,4	0,25	< 0,001
PSWQ (n = 103)	55,3	0,18	< 0,05
SFS (n = 75)	62,7	0,35	< 0,001
SIAS (n = 73)	63,0	0,34	< 0,001
SVL-22 (n = 179)	72,1	0,41	< 0,001

Voor afkortingen: zie **TABEL 1**.

604), en die van de stoornisspecifieke instrumenten op subgroepen van verschillende grootte. In overeenstemming met de resultaten wat betreft de effectgrootte zagen we dat bijna twee derde van de patiënten was vooruitgegaan op de OQ-45-SD (64,2%). Bij de stoornisspecifieke instrumenten was meer dan 50% van de patiënten vooruitgegaan op de SVL (67,0%), de POL-subschaal voor gebrek aan zelfvertrouwen (62,9%), en de PADUA (50,8%). Op de andere stoornisspecifieke instrumenten was minder dan 50% vooruitgegaan. Veel patiënten bleven stabiel.

In **TABEL 3** staat het percentage overeenstemming in de mate van vooruitgang op elke stoornisspecifieke vragenlijst en de OQ-45-SD weergegeven, naast de kappawaarde. De gegevens laten zien dat de overeenstemming in de mate van vooruitgang volgens de RCI schommelde rond de 60% met een uitschieter van 72,1% tussen de SVL-22 en de OQ-45-SD. De kappawaarden lieten over het algemeen een matige overeenstemming zien tussen de vragenlijsten wat betreft vooruitgang, stagnatie en achteruitgang volgens de RCI, waarbij op de PADUA de mate van overeenstemming niet significant was.

DISCUSSIE

In dit onderzoek gingen wij na of er een verschil is in de mate van vooruitgang bij patiënten met een angststoornis wanneer de resultaten worden vastgesteld met een stoornisspecifieke vragenlijst of met een generieke vragenlijst. Belangrijke reden voor het onderzoek is de veronderstelling dat het gebruik van stoornisspecifieke vragenlijsten betekenisvoller is voor de behandeling en mogelijk ook andere resultaten laten zien. De mate van vooruitgang is op meerdere manieren geoperationaliseerd.

Allereerst blijkt dat er weliswaar een positieve samenhang bestaat tussen de verschilscores op de stoornisspecifieke vragenlijsten en die op de generieke vragenlijst, maar dat deze samenhang laag tot middelmatig is. De mate van vooruitgang volgens de RCI bevestigt deze matige samenhang: in zo'n 40% van de gevallen is er een discrepantie. Het komt dus voor dat patiënten op de generieke lijst vooruitgang laten zien en op de specifieke vragenlijsten niet, en vice versa. Dat kan erop duiden dat verschillende clusters van klachten worden vastgesteld met de verschillende instrumenten en dat bij sommige stoornisspecifieke klachten het gebruik van alleen een generieke vragenlijst niet altijd voldoende zicht geeft op de mate van de vooruitgang op deze specifieke klachten.

Een tweede bevinding is dat de grootte van de effecten op de verschillende vragenlijsten over het algemeen boven de 0,80 uitkwam. Op twee specifieke instrumenten, de SVL-22 en de SFS, was het effect groter dan op de OQ-45-SD. In alle andere gevallen was het effect op de OQ-45-SD groter. Deze bevinding komt niet overeen met het eerdere onderzoek van Van der Mheen e.a. (2018). In dat onderzoek is gebruikgemaakt van de totaalscore van de OQ-45 waarin het ook gaat om interpersoonlijke relaties en sociaal functioneren. In ons onderzoek hebben wij gebruikgemaakt van de SD-schaal van de OQ-45 die specifiek ingaat op symptomen en klachten, waaronder angstklachten.

Het verschil in resultaten van beide studies zou erop kunnen duiden dat uitkomsten op generieke lijsten meer overeenkomen met die van specifieke lijsten naarmate de meetdomeinen dichter bij elkaar liggen. Dat is bijvoorbeeld ook terug te zien in het onderzoek van Dingemans en Van Furth (2017), die betere uitkomsten op een vragen-

lijst voor eetstoornissen vaststelden dan op de BSI die geen eetstoornissen meet. Ander onderzoek heeft laten zien dat effecten op generieke instrumenten die kwaliteit van leven of de algemene gezondheid meten, minder groot zijn dan effecten op instrumenten die psychische klachten meten (Lee e.a. 2005; Ronk e.a. 2012).

Beperkingen

Ons onderzoek kent een aantal beperkingen. We hebben gebruikgemaakt van de gegevens van een cohort patiënten die niet zijn verzameld met als doel de uitkomsten op generieke en specifieke vragenlijsten te vergelijken. De afname van de instrumenten vond plaats op initiatief van de behandelaar, waardoor er waarschijnlijk sprake is van een selectie, zowel van patiënten als van behandelaars. Wellicht worden stoornisspecifieke instrumenten vooral gebruikt bij patiënten die daartoe gemotiveerd of in staat zijn; of juist bij patiënten bij wie het gaat om complexe problematiek. Ook kan het zijn dat behandelaars die specifieke instrumenten toepassen beter in staat of meer gemotiveerd zijn de ROM te gebruiken als feedback, wat de resultaten kan hebben beïnvloed.

De diagnoses van de patiënten zijn vastgesteld op basis van intakegesprekken en niet op basis van een gestructureerd klinisch interview, wat de betrouwbaarheid van de diagnoses verhoogd zou hebben.

Verder is in dit onderzoek geen rekening gehouden met eventuele comorbiditeit, die zeker in de specialistische ggz veelvuldig voorkomt. Het wel of niet hebben van één of meerdere stoornissen naast een angststoornis kan invloed hebben gehad op de gevonden resultaten. Een patiënt die alleen een angststoornis heeft, laat wellicht een ander beloop van klachten zien gedurende therapie dan iemand die daarnaast ook nog één of meerdere andere stoornissen heeft. Om rekening te kunnen houden met de eventuele invloed van comorbiditeit, zou in vervolgonderzoek hiervoor gecorrigeerd kunnen worden.

Bij dit onderzoek is alleen gebruikgemaakt van voor- en nametingen, waardoor we geen zicht hebben op eventuele verschillen in snelheid van verandering. Het is goed moge-

lijk dat veranderingen op generieke instrumenten een ander beloopspatroom laten zien dan op stoornisspecifieke instrumenten. Om dat te onderzoeken is het nodig metingen tijdens de behandeling te doen, bij voorkeur met een hoge frequentie, bijvoorbeeld per sessie. Onderzoek heeft uitgewezen dat de meeste verandering plaatsvindt aan het begin van de behandeling en er zijn aanwijzingen dat specifieke instrumenten die verandering mogelijk beter in kaart brengen (Schibbye e.a. 2014). Met de huidige dataset was het niet mogelijk dat te onderzoeken.

Voor de verschillende stoornisspecifieke vragenlijsten was geen RCI voorhanden, waardoor de gebruikte RCI-waarden op beperkte en uiteenlopende steekproeven zijn gebaseerd. Het kan zijn dat in andere steekproeven andere waarden worden gevonden en dat beperkt de generaliseerbaarheid.

Implicaties

Dit onderzoek heeft laten zien dat men verschillende informatie over een behandeluitkomst kan vinden bij gebruik van vragenlijsten die verschillen in specificiteit. Gezien de vele implicaties van ROM is het van belang er rekening mee te houden dat de generieke vragenlijsten die op dit moment verplicht zijn, mogelijk niet altijd het meest accurate en volledige beeld geven van de behandeluitkomsten. Vervolgonderzoek waarin frequenter wordt gemeten, zou moeten uitwijzen in hoeverre generieke en specifieke instrumenten ook verschillen in beloop tijdens de behandeling te zien geven.

Ondanks de beperkingen laat dit onderzoek zien dat uitkomsten op specifieke en generieke instrumenten kunnen verschillen en daarmee een ander inzicht kunnen geven in de voortgang van de behandeling. Mede op basis hiervan sluiten wij ons aan bij de conclusies van een recent consensusdocument van Obbarius e.a. (2017), die adviseren specifieke instrumenten frequent tijdens de behandeling af te nemen en te combineren met een minder frequente afname van meer generieke instrumenten. Daarmee zouden we rechtdoen aan de complexiteit van de problematiek van veel patiënten in de specialistische ggz.

LITERATUUR

- Aderka IM, Nickerson A, Bøe HJ, Hofmann SG. Sudden gains during psychological treatments of anxiety and depression: a meta-analysis. *J Consult Clin Psychol* 2012; 80: 93-101.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-IV-TR). Washington: American Psychiatric Association; 2000.
- Andrews G, Page AC. Outcome measurement, outcome management and monitoring. *Aust N Z J Psychiatry* 2005; 39: 649-51.
- Barendregt M. Benchmarking en andere functies van ROM: Back to basics. *Tijdschr Psychiatr* 2015; 57: 517-25.
- Barkham M, Stiles WB, Connell J, Twigg E, Leach C, Lucock M, e.a. Effects of psychological therapies in randomised trials and practice based studies. *Br J Clin Psychol* 2008; 47: 397-415.
- Bickman L, Kelley SD, Breda C, De Andrade AR, Riemer M. Effects of routine feedback to clinicians on youth mental health outcomes: A randomized cluster design. *Psychiatr Serv* 2011; 62:1423-9.

- Boswell JF, Kraus DR, Miller SD, Lambert MJ. Implementing routine outcome monitoring in clinical practice. Benefits, challenges, and solutions. *Psychother Res* 2015; 25: 6-19.
- Carlier IVE, Meuldijk D, Vliet IM van, Fenema E van, Wee NJA van der, Zitman FG. Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *J Eval Clin Pract* 2010; 18: 104-10.
- Cohen J. *Statistical power analysis for the behavioral sciences* (2e ed.). Hillsdale: Earlbaum; 1988.
- Delepaul PAEG. Routine outcome measurement in the Netherlands – A focus on Benchmarking. *Int Rev Psychiatry* 2015; 27: 320-8.
- Dingemans AE, Furth EF van. Het meten van verandering tijdens behandeling voor eetstoornissen: een vergelijking van algemene en specifieke vragenlijst. *Tijdschr Psychiatr* 2017; 59: 278-85.
- Hansen NB, Lambert MJ, Forman EM. The psychotherapy dose-response effect and its implications for treatment delivery services. *Clin Psychol Sci Prac* 2002; 9: 329-43.
- Horowitz M, Wilner N, Alvarez W. Impact of Event Scale: A measure of subjective stress. *Psychosom Med* 1979; 41: 209-18.
- 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 MA, Lambert MJ, Burlingame GM. Handleiding voor afname en scoring van de Outcome Questionnaire (OQ-45). Salt Lake City: OQ Measures LLC 2008.
- Jong K de, Sluis P van, Nugter MA, Heiser WJ, Spinhoven P. Understanding the differential impact of outcome monitoring: therapist variables that moderate feedback effects in a randomized clinical trial. *Psychother Res* 2012; 22: 464-74.
- Kendrick T, El-Gohary M, Stuart B, Gilbody S, Churchill R, Aiken L, e.a. Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. *Cochrane Database Syst Rev* 2016; 7: CD01119.
- Kelley SD, Bickman L. Beyond outcomes monitoring: measurement feedback systems in child and adolescent clinical practice. *Curr Opin Psychiatry* 2009; 22: 363-8.
- Knaup C, Koester M, Schoefer D, Becker T, Puschner B. Effect of feedback on treatment outcome in specialist mental healthcare: meta-analysis. *Br J Psychiatry* 2009; 195: 15-22.
- 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.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-74.
- Lee W, Jones L, Goodman R, Heyman I. Broad outcome measures may underestimate effectiveness: An instrument comparison study. *Child Adolesc Mental Health* 2005; 10: 143-4.
- Mattick RP, Clarke JC. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behav Res Ther* 1998; 36: 455-70.
- Mheen M van der, Mors IM ter, Hout MA van den, Cath DC. Routine outcome monitoring bij de behandeling van angststoornissen: diagnosespecifieke versus generieke meetinstrumenten. *Tijdschr Psychiatr* 2018; 60: 511-9.
- Meijer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. *Behav Res Ther* 1990; 28: 487-95.
- Nugter MA, Buwalda VJA. Achtergronden en gebruiksmogelijkheden van ROM in de ggz. *Tijdschr Psychiatr* 2012; 54: 111-20.
- Nugter MA, Hermens MLM, Robbers S, Van Son G, Theunissen J, Engelsbel F. Use of outcome measurements in clinical practice: How specific should one be? *Psychother Res* 2017; doi: 10.1080/10503307.
- Obbarius A, Van Maasackers L, Baer L, Clark DM, Crocker AG, De Beurs E, e.a. Standardization of health outcomes assessment for depression and anxiety: recommendations from the ICHOM Depression and Anxiety Working Group. *Qual Life Res* 2017; 26: 3211-25.
- Ost LG. The Agoraphobia Scale: An evaluation of its reliability and validity. *Behav Res Ther* 1990; 28: 323-9.
- Riemer M, Bickman L. Using program theory to link social psychology and program evaluation. In: Mark MM, Donaldson SI, Campbell B, editors. *Social Psychology and Evaluation*. New York: Guilford Press; 2011. p. 104-40.
- Ronk FR, Hooke GR, Page AC. How consistent are clinical significance classifications when calculation methods and outcome measures differ? *Clin Psychol Sci Prac* 2012; 19: 167-79.
- Sanavio E. Obsessions and compulsions: The Padua Inventory. *Behav Res Ther* 1988; 26: 167-77.
- Schibbye P, Ghaderi A, Ljotsson B, Hedman E, Lindefors N, Ruck C, e.a. Using early change to predict outcome in cognitive behaviour therapy: exploring timeframe, calculation method, and differences of disorder-specific versus general measures. *Plos One* 2014; 9(6): e100614.
- Shimokawa K, Lambert MJ, Smart DW. Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *J Consult Clin Psychol* 2010; 78: 298-311.
- Telch MJ. The Panic Appraisal Inventory. Unpublished Scale. University of Texas at Austin, 1987.
- Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008; 4: 344-9.

SUMMARY

ROM in mental health care: treatment outcomes of both generic and disorder specific instruments in patients with anxiety disorders; a cohort study

M.A. NUGTER, G.P. OOSTERDIJK, A.A.P. VAN EMMERIK

- BACKGROUND** In routine outcome monitoring (ROM) generic instruments are often used to measure outcomes. This could hamper the clinical use of these instruments and lead to a possible underestimation of treatment outcomes.
- AIM** To examine whether there are differences in the degree of progress between outcomes of disorder specific instruments measuring anxiety symptoms and the Symptomatic Distress subscale of the Outcome Questionnaire-45 (OQ-45-SD).
- METHOD** Data from 604 patients with an anxiety disorder were analyzed. The mean change scores, percentages of reliable change and effect sizes on the OQ-45-SD were compared to the outcomes of seven commonly used instruments that specifically measure the symptoms of anxiety disorders.
- RESULTS** The effect sizes on the OQ-45-SD and the disorder specific instruments were 0.80 or higher. The correlation between the change scores was moderate, as was the correspondence between the degree to which patients improved reliably. In one third to half of the cases changes on the OQ-45-SD and on the disorder specific instruments were not in agreement.
- CONCLUSION** The treatment outcomes of the OQ-45-SD are approximately as positive as those of the disorder specific instruments. However, discrepancies existed in the extent to which patients had improved. In accordance with the advice of the International Consortium for Health Outcomes Measurement, the use of both kinds of instruments is advised when evaluating the progress of the treatment.

TIJDSCHRIFT VOOR PSYCHIATRIE 60(2018)11, 741-749

KEY WORDS anxiety disorders, disorder specific instruments, ROM, routine outcome monitoring