

# ROM vroeg in de behandeling en specifiek; observationeel onderzoek naar generieke en specifieke vragenlijsten bij meten van vroege verandering bij depressiebehandeling

B.G. TIEMENS, M.I. KRAMER, M.W. KLOOS, J. SPIJKER

- ACHTERGROND** Ggz-instellingen maken gebruik van routine outcome monitoring (ROM) om vast te stellen of een cliënt goed reageert op de behandeling. Nog onbeslecht is of hiervoor het beste diagnosespecifieke of generieke meetinstrumenten gebruikt kunnen worden. Daarnaast is nog weinig bekend over wanneer men een eerste indicatie kan geven over de uitkomst van de behandeling.
- DOEL** Inzicht geven in de gevoeligheid voor verandering van twee vragenlijsten, een generieke en een specifieke, die beide gebruikt worden bij cliënten met een depressieve stoornis. Daarnaast inzicht geven over op welk moment een eerste indicatie kan worden gegeven over de uitkomst van de behandeling.
- METHODE** Retrospectieve cohortstudie met data van 518 cliënten met een depressieve stoornis. De Outcome Questionnaire (OQ-45.2) en de Inventory of Depressive Symptomatology (IDS-SR) werden gebruikt om het beloop van de behandeling te meten.
- RESULTATEN** De diagnosespecifieke IDS-SR bleek met name in het begin van de behandeling gevoeliger voor verandering dan de generieke OQ-45.2. Bij een meetfrequentie van eens in de vier weken lag het beste moment voor het verkrijgen van een eerste indicatie over de uitkomst van de behandeling tussen week 4 en week 8.
- CONCLUSIE** Voor cliënt en behandelaar die zich tot doel stellen depressieve klachten te reduceren is een diagnosespecifieke vragenlijst te verkiezen om dit te monitoren. De meting tussen 4 en 8 weken is belangrijk voor het zo nodig bijsturen van de behandeling en het verbeteren van het behandelresultaat.

TIJDSCHRIFT VOOR PSYCHIATRIE 62(2020)2, 121-130

**TREFWOORDEN** depressie, diagnosespecifiek of generiek meetinstrument, routine outcome monitoring, vroege verandering



In de ggz wordt gebruikgemaakt van vragenlijsten om voortgang in de behandeling te meten. Met routine outcome monitoring (ROM) krijgen de behandelaar en de cliënt feedback over het beloop van de klachten. Dit is nuttig omdat behandelaars die geen vragenlijsten gebruiken niet

goed in staat zijn te voorspellen welke van hun cliënten waarschijnlijk gaan verslechteren (Hannan e.a. 2005) en tevens slecht de resultaten van hun cliënten tijdens de behandeling kunnen beoordelen (Garb 2005; Sapyta e.a. 2005).

Meerdere studies laten zien dat de vroege respons op een behandeling (verandering bij de cliënt) de uitkomst van de behandeling voorspelt (Lutz e.a. 2006; 2009; Stulz e.a. 2007). Wanneer men ROM als feedbackinstrument in de behandeling inzet, zodat men die vroege respons (of het uitblijven ervan) kan opmerken, verbetert dit de behandelresultaten. Dit geldt vooral bij cliënten die niet goed vooruitgaan (Shimokawa e.a. 2010; Kendrick e.a. 2016). Het daadwerkelijk gebruik van ROM in de behandelpraktijk valt echter nog tegen, onder andere omdat behandelaars de gebruikte instrumenten vaak niet praktisch bruikbaar achten (Nuijen e.a. 2015).

### Generiek versus specifiek

Bij de keuze van de meetinstrumenten voor ROM maakt men onderscheid tussen generieke meetinstrumenten, die voor uiteenlopende groepen cliënten kunnen worden gebruikt, en diagnosespecifieke instrumenten. Obbarius e.a. (2017) adviseren om frequent herhaalde metingen gedurende de behandeling met diagnosespecifieke instrumenten te combineren met generieke instrumenten voor brede behandeluitkomsten die lager frequent worden afgenomen. Diagnosespecifieke instrumenten zijn van belang om verbeteringen in specifieke symptomen bij diagnostisch homogeenere subgroepen cliënten te kunnen volgen, bijvoorbeeld bij het toepassen van de richtlijn voor depressie (Nugter & Buwalda 2012).

Zowel bij cliënten met een depressie, als bij cliënten met een angststoornis of een eetstoornis lijken diagnosespecifieke instrumenten gevoeliger voor verandering dan een generiek instrument (Dingemans e.a. 2017; De Beurs e.a. 2018; Van der Mheen e.a. 2018). Nugter e.a. (2017) vinden daarentegen een vergelijkbare effectgrootte voor een generiek en een specifiek depressie-instrument bij cliënten met een depressie. Niettemin concluderen deze auteurs dat generieke en specifieke instrumenten niet inwisselbaar zijn, want het gebruik van beide lijkt de behandeluitkomst te verbeteren (Nugter e.a. 2017). Zij gebruikten in hun onderzoek echter alleen een begin- en eindmeting.

### Belang vroege prognose

Het huidige onderzoek werd uitgevoerd bij cliënten met een depressieve stoornis. De meeste mensen die behandeling krijgen voor een depressie herstellen binnen een periode van 6 tot 11 maanden. Een groot deel van de cliënten die een depressieve episode doormaken, krijgt echter te maken met terugval en 12% is niet volledig hersteld na 36 maanden. Hoe eerder de depressieve klachten worden behandeld, des te groter de kans op blijvend herstel, zonder terugval. Elke nieuwe depressieve episode verhoogt het risico op een chronisch verloop en dat is verontrustend (Spijker e.a. 2004; Ten Have e.a. 2017). Het is daarom

### AUTEURS

**BEA TIEMENS**, bijzonder hoogleraar Evidence based practice in mental health care, Radboud Universiteit Nijmegen, senior wetenschappelijk medewerker, Pro Persona Research, en programmaleider onderzoek, Indigo.

**MARE KRAMER**, gezondheidszorgpsycholoog in opleiding tot specialist, praktijkhouder Praktijk Kramer, Doetinchem.

**MARGOT KLOOS**, junior wetenschappelijk medewerker, Pro Persona Research.

**JAN SPIJKER**, psychiater, Pro Persona Expertisecentrum depressie, Nijmegen, bijzonder hoogleraar Chronische Depressie, Faculteit der Sociale Wetenschappen, Radboud Universiteit, Nijmegen.

### CORRESPONDENTIEADRES

Prof. dr. B. Tiemens, Pro Persona Research, de Bogen, Postbus 26, 6870 AA Renkum.  
E-mail: b.tiemens@propersona.nl

Geen strijdige belangen meegedeeld.

Het artikel werd voor publicatie geaccepteerd op 10-9-2019.

belangrijk dat een cliënt met een depressie zo snel mogelijk de juiste behandeling krijgt en dat de behandeling ook snel aanslaat om de kans op chroniciteit te verkleinen.

Bij voorkeur wordt zo vroeg mogelijk in de behandeling helder of er kans is op een goede behandeluitkomst, want hoe eerder men het eindresultaat kan voorspellen, des te therapeutischer de behandelaar kan ingrijpen bij stagnatie. De vraag is of men dan beter voor een generieke of een diagnosespecifieke vragenlijst kan kiezen om de voortgang in kaart te brengen en ook op welk moment van de behandeling met welk instrument men zo'n eerste indicatie kan geven over de uitkomst van de behandeling.

In dit artikel beschrijven we een onderzoek naar de gevoeligheid voor verandering van de generieke 00-45.2 en de diagnosespecifieke IDS-SR. Vervolgens onderzochten we wat het beste moment in de behandeling is om een eerste indicatie te geven over de uitkomst ervan.

### METHODE

#### Opzet, setting en cliënten

Het onderzoek betreft een retrospectief cohortonderzoek bij cliënten van Pro Persona, een grote ggz-instelling met meerdere vestigingen in Oost-Nederland. De zorg is georganiseerd volgens zorgprogramma's. Deze studie werd uitgevoerd binnen het zorgprogramma Depressie, met gegevens van cliënten bij wie de behandeling was gestart

**TABEL 1** DSM-IV-classificaties van 518 cliënten met een startmeting OQ-45.2 en IDS-SR

DSM-IV-classificatie	n (%)
Depressieve stoornis, ernstig, met psychotische kenmerken	20 (3,9)
Depressieve stoornis, ernstig, eenmalig	55 (10,6)
Depressieve stoornis, ernstig, recidiverend	91 (17,6)
Depressieve stoornis, licht/matig, eenmalig	108 (20,8)
Depressieve stoornis, licht/matig, recidiverend	180 (34,7)
Depressieve stoornis NAO	16 (3,1)
Dysthyme stoornis	33 (6,4)
Overige stoornis	15 (2,9)

tussen 2012 en 2015 en afgesloten voor 2016. 518 cliënten hadden een startmeting ingevuld op zowel de IDS-SR als de OQ-45.2. 90,7% van hen had een depressieve stoornis op as I van de DSM-IV, 6,4% een dysthyme stoornis en 2,9% een overige stoornis (TABEL 1). Bij 28,8% van de cliënten was er op as II van de DSM-IV ook een persoonlijkheidsstoornis vastgesteld. Cliënten ontvingen protocollaire cognitieve gedragstherapie (CGT), interpersoonlijke therapie (IPT) of sociaalpsychiatrische begeleiding, die in principe wekelijks werden aangeboden, al dan niet gecombineerd met antidepressieve medicatie.

### Procedure

Binnen het zorgprogramma Depressie wordt bij alle cliënten de behandeling gemonitord met vragenlijsten. Vanaf de start van de behandeling wordt de *Inventory of Depressive Symptomatology - Self-Report* (IDS-SR) om de 4 weken en de OQ-45.2 elke 12 weken afgenomen. Voor dit onderzoek werd gebruikgemaakt van geanonimiseerde gegevens uit de ROM en uit het EPD. Gegevens werden gecodeerd door de cliëntnummers te vervangen door een onderzoekscode en naast diagnose werden alleen leeftijd en geslacht meegenomen in de database. Cliënten konden bij de start van de behandeling bezwaar maken tegen gebruik van geanonimiseerde gegevens voor wetenschappelijk onderzoek. Twee cliënten deden dit in deze periode; van hen zijn de gegevens niet meegenomen in het onderzoek.

### Instrumenten

#### GENERIEK: OUTCOME QUESTIONNAIRE (OQ-45.2)

De OQ-45.2 (Lambert e.a. 2004) is een zelfrapportagevragenlijst en bestaat uit 45 vragen naar hoe de cliënt zich de afgelopen week heeft gevoeld. De OQ-45.2 meet klachten en functioneren en heeft drie subschalen: symptomatische

distress (SD), interpersoonlijke relaties (IR) en sociale rol (SR). Items worden beoordeeld op een 5-puntslikert-schaal van 0 'nooit' tot 4 'bijna altijd'.

#### SPECIFIEK: INVENTORY OF DEPRESSIVE SYMPTOMATOLOGY - SELF-REPORT (IDS-SR)

De IDS-SR bestaat uit 30 items en wordt ingezet om de ernst van depressieve klachten in kaart te brengen (Rush e.a. 1996). De vragenlijst richt zich op de DSM-IV-criteria van een depressie. Items worden beoordeeld op een 4-puntslikert-schaal (0 = helemaal niet; 3 = bijna altijd).

### Selectie van cliënten en metingen

Het totale cohort omvatte 1133 cliënten. Zij vulden in totaal 4821 keer de IDS-SR in, en 2484 keer de OQ-45.2 vanaf de start van de behandeling. Bij de start van de behandeling wordt een zorgpad depressie geopend in het EPD. Een zorgpad bestaat uit de beschrijving van een vooraf vastgesteld zorgproces (verplichte en optionele modules) voor een specifieke groep cliënten. Van cliënten die behandeld werden in meerdere zorgpaden binnen het zorgprogramma depressie, gebruikten we de gegevens vanaf het eerst gestarte zorgpad.

Omdat de IDS-SR het frequentst werd afgenomen, deelden we eerst de IDS-SR-metingen in meetperiodes. Vervolgens werden de metingen van de OQ-45.2 gekoppeld aan een IDS-SR-meting die in dezelfde week was afgerond (maximaal 7 dagen verschil in afronddatum). Op basis van de gekoppelde metingen werden de behandelingen in periodes ingedeeld (zie TABEL 2). De startmeting (meting I) was binnen 4 weken na de start van het zorgpad afgerond. Periode II van gekoppelde metingen IDS-SR/OQ-45.2 zou in de daarop volgende 12 weken moeten vallen, maar rond 12 weken waren er te weinig OQ-metingen. Deze periode II werd daarom langer dan de andere periodes, namelijk tussen 4 en 24 weken na start van het zorgpad, maar tussen de gekoppelde metingen zat ook hier maximaal 7 dagen. De periodes III en IV besloegen beide maximaal 12 weken. De gekoppelde metingen in periode I, II, III en IV gebruikten we voor het berekenen van de effectgrootte in de verschillende periodes (voor de leesbaarheid noemen we dit meetmoment, ten opzichte van de start van de behandeling). 518 cliënten hadden een startmeting van zowel de OQ-45 als de IDS-SR afgerond en 320 van hen hadden minimaal twee gekoppelde IDS-SR/OQ-45-metingen.

### Statistische analyses

#### EFFECTGROOTTE

Als eerste maat voor het verschil in gevoeligheid voor verandering gebruikten we de effectgrootte Cohens *d* (Cohen 1988). De effectgrootte werd berekend per meet-

**TABEL 2** Meetmomenten IDS-SR en OQ-45.2

weken	0-4	4-8	8-12	12-16	16-20	20-24	24-28	28-32	32-36	36-40	40-44	44-48
IDS-SR	1	2	3	4	5	6	7	8	9	10	11	12
OQ-45.2	I	II					III				IV	

moment voor de groep cliënten met zowel een IDS-SR- als OQ-45.2-meting op dat meetmoment. Het aantal cliënten in de analyse verschilde dus per meetmoment.

**REPEATED MEASURES**

Om ook de gevoeligheid in verandering tussen de IDS-SR en de OQ-45.2 bij eenzelfde groep cliënten te vergelijken, voerden we *general linear model (GLM) repeated measures* uit. Voor deze analyse werden cliënten geïncludeerd met minimaal twee gekoppelde IDS-SR/OQ-45.2-metingen (n = 320) en de missende metingen werden geïmputeerd. Om de scores van de IDS-SR en de OQ-45.2 vergelijkbaar te maken rekenden we de ruwe scores om naar z-scores.

Alvorens de missende metingen te imputeren werd gecontroleerd of de missende data at random ontbraken. Dit was het geval bij de IDS-SR (Littles MCAR-test was niet significant; p = 0,432), maar niet bij de OQ-45.2 (Littles MCAR-test; p = 0,049). Niettemin werd de analyse toch uitgevoerd, omdat deze vooral als extra check diende. Uiteraard moeten de resultaten met voorzichtigheid worden geïnterpreteerd. Voor de repeated measures werden missende waarden geïmputeerd volgens de maximum likelihoodmethode (Allison 2006).

**REGRESSIEANALYSE/VERKLAARDE VARIANTIE**

Tot slot voerden we verschillende lineaire regressies uit met steeds als onafhankelijke variabele de verschilscore tussen de IDS-SR-scores bij de startmeting en de verschillende vervolgmomenten en als afhankelijke variabele het uiteindelijk herstel oftewel de verschilscore tussen de start- en eindmeting op de IDS-SR. Dit herhaalden we nog een keer met als afhankelijke variabele de mate van verbe-

tering op de OQ-45.2 aan het eind van de behandeling.

Beide series analyses werden gecontroleerd voor leeftijd, geslacht, startscores en de behandelduur. Met de verklaarde variantie (R<sup>2</sup>) als *goodness-of-fit* maat bepaalden we op welk meetmoment de verandering op de IDS-SR in de eerste 24 weken van de behandeling het meest voorspellend leek te zijn voor het eindresultaat van de behandeling.

**RESULTATEN**

**Effectgrootte**

Voor het berekenen van de effectgrootte maakten we gebruik van de gekoppelde metingen op de verschillende meetmomenten. **TABEL 3** geeft weer van hoeveel cliënten op de verschillende meetmomenten er gegevens beschikbaar waren voor de OQ-45.2 en de IDS-SR afzonderlijk, alsook voor de gekoppelde metingen IDS-SR en OQ-45.2.

In **TABEL 4** worden de kenmerken van de 518 cliënten met een gekoppelde startmeting weergegeven ten opzichte van de cliënten die niet werden meegenomen in de analyses. Geslacht, leeftijd en de behandelduur in weken waren vergelijkbaar in beide groepen ( $\chi^2$ - en t-toets)

**FIGUUR 1** geeft de verandering op de IDS-SR en de OQ-45.2 weer tussen de eerste en latere metingen gedurende de behandeling. De mediaan van het aantal weken tussen de start van het zorgpad en de gekoppelde meting in periode I was 1,6 weken (n = 518), tussen periode I en II 13,3 weken (n = 145), tussen II en III 12,1 weken (n = 91) en tussen III en IV 12,4 weken (n = 68). De grootste verandering zagen we tussen de startmeting en meetperiode II. Dit gold voor beide vragenlijsten, waarbij de IDS-SR een hogere effectgrootte liet zien dan de OQ-45.2. Deze analyses werden ook apart uitge-

**TABEL 3** Cliënten per meetmoment na koppeling van vragenlijsten

	Startmeting I (0-4 weken)	Periode II (4-24 weken)	Periode III (24-36 weken)	Periode IV (36-48 weken)
IDS-SR	564	293	242	161
OQ-45.2	585	406	348	261
Gekoppeld	518	248	227	152

**TABEL 4** Kenmerken van cliënten met gekoppelde metingen vergeleken met overige cliënten

Kenmerken	Cliënten met gekoppelde startmetingen (n = 518)	Overige cliënten (n = 615)	p
Geslacht vrouw (%)	279 (53,9)	334 (54,3)	0,88
Gem. leeftijd, in j (SD)	39,9 (10,3)	40,3(10,4)	0,58
Gem. behandelduur, in w (SD)	39,5 (25,0)	41,9 (29,5)	0,14

**TABEL 5** Kenmerken van cliënten geselecteerd voor repeated measures vergeleken met de overige cliënten

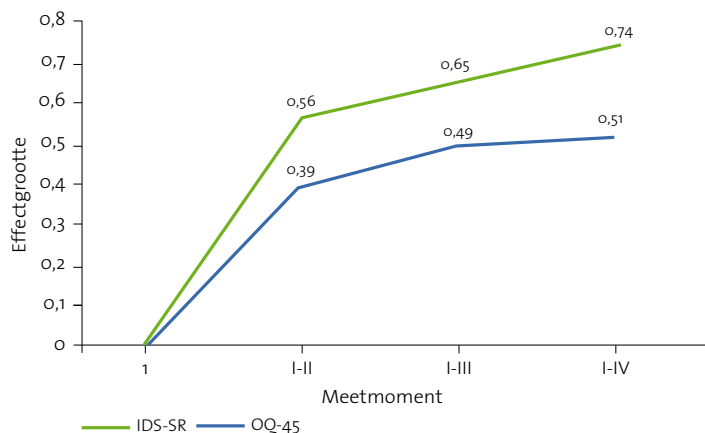
Kenmerken	Cliënten met $\geq 2$ metingen	n	Overige cliënten	n	p
Geslacht vrouw (%)	187 (58,4%)	320	426 (52,4%)	813	0,066
Leeftijd, in j (SD)	40,3 (10,1)	320	40,1 (10,4)	813	0,773
Behandelduur, in w (SD)	47,0 (24,0)	320	38,3 (28,5)	813	< 0,001
Startscore IDS-SR (SD)	38,4 (14,4)	262	37,2 (15,7)	256	0,394
Eindscore IDS-SR (SD)	24,4 (17,6)	248	30,4 (18,3)	244	< 0,001
Startscore OQ-45.2 (SD)	88,2 (24,6)	279	87,4 (26,7)	306	0,718
Eindscore OQ-45.2 (SD)	73,0 (31,1)	320	85,2 (28,52)	375	< 0,001

voerd voor de SD-schaal van de OQ-45.2 en daarmee werden vergelijkbare resultaten gevonden. Het verschil in effectgrootte tussen beide vragenlijsten op het eerste moment (effectgrootte tussen meetmoment I en II) was 0,17 en bleef aanwezig in de latere twee tijdsintervallen.

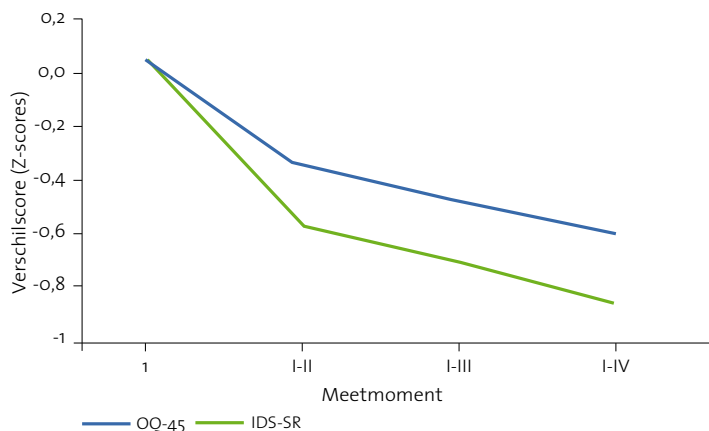
### Repeated measures

Vervolgens werden met GLM repeated measures de resultaten van 320 cliënten geanalyseerd die voldeden aan de eis dat er minimaal twee metingen van beide vragenlijsten

beschikbaar waren. **TABEL 5** geeft de kenmerken weer van deze 320 cliënten en de start- en eindscores (nog niet geïmputeerd), vergeleken met de groep cliënten die niet aan dit criterium voldeed. In de groep met minimaal twee gekoppelde metingen was de behandelduur langer (t-toets;  $p < 0,001$ ). De eindscores op zowel de IDS-SR als de OQ-45.2 waren hoger bij de groep overige cliënten (voor beide instrumenten  $p < 0,001$ ); deze cliënten waren dus minder verbeterd.

**FIGUUR 1** Effectgroottes van de IDS-SR en de OQ-45.2 ten opzichte van de startmeting\*

\* n(I-II) = 145; n(I-III) = 142; n(I-IV) = 91

**FIGUUR 2** GLM repeated measures van z-scores IDS-SR en OQ-45.2 (n = 320)

De repeated-measuresanalyse, met geïmputeerde scores voor de missende metingen, gaf hetzelfde patroon (zie **FIGUUR 2**) als de eerdere analyse op basis van de effectgrootte. De grootste afname van klachten zagen we tussen de startmeting en meetperiode II. Dit gold voor beide vragenlijsten waarbij de grootste afname van klachten zichtbaar was bij de IDS-SR. Het verschil tussen beide instrumenten was het grootst tussen meetmoment I en II, daarna bleef het verschil constant.

### Regressieanalyse

Met de IDS-SR, die gevoeliger leek voor verandering, gingen we na wat het meest voorspellende meetmoment was tijdens de behandeling voor het eindresultaat. Lineaire regressieanalyses werden uitgevoerd met als onafhankelijke variabelen leeftijd, geslacht, behandelduur in weken, startscore en de verandering op de IDS-SR tussen de eerste vier metingen van de IDS-SR en als afhankelijke variabele de verschilscore tussen de eerste en laatste meting van de

**TABEL 6** Lineaire regressie van vroege meetmomenten IDS-SR en verbetering op de IDS-SR aan het eind van de behandeling

Meetmoment	Aantal cliënten	R <sup>2*</sup>	R <sup>2**</sup>	B	p
1-2 (na 4 weken)	311	6,6%	26,5%	0,684	0,000
2-3 (na 8 weken)	191	12,0%	13,7%	0,244	0,058
3-4 (na 12 weken)	150	9,7%	10,9%	0,276	0,167
4-5 (na 16 weken)	145	9,9%	11,9%	0,322	0,075

\*Model met leeftijd, geslacht, behandelduur in weken en startscore IDS-SR als onafhankelijke variabelen.

\*\*Model met verschilscore IDS-SR en leeftijd, geslacht, behandelduur in weken en startscore IDS-SR als onafhankelijke variabelen.

**TABEL 7** Lineaire regressie van vroege meetmomenten IDS-SR en verbetering op de OQ-45.2 aan het eind van de behandeling

Meetmoment	Aantal cliënten	R <sup>2*</sup>	R <sup>2**</sup>	B	p
1-2 (na 4 weken)	254	6,8%	14,5%	0,766	0,000
2-3 (na 8 weken)	162	8,2%	10,1%	0,467	0,074
3-4 (na 12 weken)	131	13,0%	13,6%	0,333	0,341
4-5 (na 16 weken)	135	13,2%	14,2%	0,354	0,239

\*Model met leeftijd, geslacht, behandelduur in weken en startscore IDS-SR als onafhankelijke variabelen.

\*\*Model met verschilscore IDS-SR en leeftijd, geslacht, behandelduur in weken en startscore IDS-SR als onafhankelijke variabelen.

IDS-SR en OQ-45.2 (zie **TABEL 6 EN 7**). De  $R^2$  in de tabellen zijn niet cumulatief, ze betreffen steeds de afzonderlijke periode, met verschillend aantal cliënten doordat niet van alle cliënten gekoppelde metingen beschikbaar waren op alle meetmomenten.

In **TABEL 6 EN 7** wordt voor elk meetmoment eerst de  $R^2$  gegeven voor het model met alleen leeftijd, geslacht, behandelduur in weken en startscore op de IDS-SR en in de volgende kolom de  $R^2$  nadat de verschillscore op de IDS-SR in de betreffende meetperiode werd toegevoegd. De  $\beta$ - en  $p$ -waarde in de laatste twee kolommen betreffen de toegevoegde variabele verschillscore op de IDS-SR. Voor beide uitkomstmaten was de toegevoegde  $R^2$  het hoogst in het eerste model met het verschil tussen de IDS-SR-startmeting en de meting na 4 weken (meetmoment 1-2). Voor een klein deel was dit toe te schrijven aan de hogere  $R^2$  van de modellen zonder de verschillscore op de volgende meetmomenten, maar niet volledig.

## DISCUSSIE

### Belangrijkste resultaten

Het doel van deze studie was inzicht te krijgen in de gevoeligheid voor verandering van twee vragenlijsten die beide gebruikt worden om het behandelbeloop te meten bij cliënten met een depressieve stoornis, de generieke OQ-45.2 en de diagnosespecifieke IDS-SR. Daarnaast wilden we inzicht krijgen in het moment waarop een eerste indicatie kan worden gegeven over de uitkomst van de behandeling. Zowel de analyses op basis van de effectgrootte op de verschillende meetmomenten als de repeated measures laten zien dat de IDS-SR met name in het begin van de behandeling gevoeliger is voor verandering dan de OQ-45.2. De voorspellende waarde, uitgedrukt in verklaarde variantie van de verandering op de IDS-SR in het begin van de behandeling voor de totale behandeling, was het hoogst tussen het eerste en het tweede meetmoment voor zowel de IDS-SR als de OQ-45.2 als uitkomstmaat. Zoals verwacht, is de toegevoegde voorspellende waarde van de verschillscore van de IDS-SR hoger voor de IDS-SR als uitkomstmaat dan voor de OQ-45.2 als uitkomstmaat. Op basis van beide instrumenten gaf in dit cohort bij 4-wekelijks meten, het beloop op de IDS-SR 4 tot 8 weken na de start de meeste toegevoegde voorspellende waarde voor de behandeluitkomst.

### Gevoeligheid voor verandering

Onze bevinding dat het diagnosespecifieke instrument gevoeliger is dan het generieke, komt overeen met die van Van der Mheen e.a. (2018) en Dingemans e.a. (2017). Hoewel Nugter e.a. (2017) geen verschillen vonden tussen een diagnosespecifiek en generiek instrument, adviseerden ze toch beide te gebruiken omdat juist het gebruik van beide

instrumenten de behandeluitkomst lijkt te verbeteren.

Het is niet duidelijk waar de verschillende uitkomsten door worden bepaald. Voor zover we uit de publicaties begrijpen, zijn de cliënten in de betreffende studies voor hun specifieke psychiatrische stoornis behandeld. Interessant zou zijn om in vervolgonderzoek na te gaan of de mate waarin de behandeling al dan niet specifiek is ingezet op de specifieke aandoening, gerelateerd is aan de snelheid waarin die specifieke klachten en de comorbide klachten en functioneren veranderen. Wij vonden namelijk dat alleen in het begin van de behandeling (in deze studie tot 24 weken) een depressie-instrument de reactie op de behandeling beter meet dan een generiek instrument.

Het is heel goed mogelijk dat eventuele comorbide klachten en het functioneren pas later veranderen. Wat dan zeker ook voor een diagnosespecifiek instrument pleit, is dat het voor de cliënt motiverend kan zijn dat verbetering snel zichtbaar wordt of bij een verslechtering van de klachten juist vertrouwenwekkend als de behandelaar de voortgang zo goed in de gaten houdt (Solstad e.a. 2017).

### Voorspellende waarde van vroege verandering

Lutz e.a. (2009) onderzochten de patronen van vroege verandering bij depressiebehandelingen. Ze vonden dat niet het soort behandeling, maar de verschillende patronen van vroege reactie op de behandeling de uitkomst bij beëindiging van de behandeling en de periode erna voorspelden. Schibbye e.a. (2014) zochten naar het optimale moment en vonden dat verandering in week 4 de optimale voorspeller is voor het behaalde eindresultaat van een behandeling. Dit komt overeen met onze bevinding dat een tijdseenheid tussen 4 en 8 weken na de start het beste moment lijkt om vroege veranderingen bij het beloop van een behandeling van een depressie met de IDS-SR in beeld te brengen.

Hafkenscheid en Van Os (2018) zetten vraagtekens bij de studies naar de voorspellende waarde van vroege verandering, met name omdat in nagenoeg alle studies het instrument waarmee de voorspelling wordt gedaan hetzelfde is als het instrument waarmee de uitkomst wordt gemeten. Dit is inderdaad een zwak punt, omdat de metingen met hetzelfde instrument per definitie gecorreleerd zijn. Daarom hebben we dezelfde analyses ook uitgevoerd met de OQ-45.2 als uitkomstmaat. De toename in verklaarde variantie bij toevoeging van het beloop op de IDS-SR was duidelijk lager, maar nog steeds het sterkst tussen de eerste twee metingen.

### Meetfrequentie

Schibbye e.a. (2014) vonden hun resultaat in een afgebakende studie met korte behandelingen waarin bij elke sessie werd gemeten. In ons onderzoek hebben we gege-



vens geanalyseerd uit een naturalistische setting met veel onvoldedige metingen, wisselende behandelduur en een meetfrequentie eens per 4 weken, maar we vonden een vergelijkbaar resultaat. Hoewel gesuggereerd kan worden dat het vaker invullen van de IDS-SR mogelijk een effect op zich kan geven, en bovendien onmogelijk is in de dagelijkse praktijk, is een 4-wekelijkse afname in vergelijking tot buitenlandse studies nog steeds zeer laagfrequent. Sterker nog, Schiepek e.a. (2016) laten zien dat dit juist tot toevalsbevindingen zou kunnen leiden en dat daarom altijd elke sessie moet worden gemeten. Zowel de studie van Schiepek e.a. als het grote nationale Britse programma *Improving Access to Psychological Therapies* (IAPT) laat zien dat hoogfrequent meten wel degelijk haalbaar is (Clark 2018), en overwogen zou moeten worden.

### Beperkingen van het onderzoek

In deze studie werden gegevens van cliënten gebruikt die minimaal één keer een vragenlijst hebben ingevuld tijdens hun behandeling. Hiermee maakten we dus al een selectie omdat cliënten zonder metingen buiten beschouwing bleven. De behandeluitkomsten van de cliënten zonder metingen zijn duidelijk slechter, wat mogelijk verklaard kan worden door specifieke kenmerken van deze groep (zoals minder behandelmotivatie), maar dit kan ook een pleidooi vormen om nog meer in te zetten op gebruik van ROM bij alle cliënten. Duidelijk is wel dat onze resultaten alleen maar geldig zijn voor cliënten bij wie de ROM wordt afgenomen.

Naast deze selectie zijn er gedurende de behandeling van het cohort cliënten metingen uitgevallen. De uitval van metingen had onder andere te maken met de noodzaak de metingen van de twee instrumenten te koppelen. Daarnaast neemt in de loop van de tijd doorgaans het aantal cliënten, en daarmee het aantal metingen, af, simpelweg omdat behandelingen worden afgesloten. Hierdoor zal het

aantal cliënten dat een vragenlijst aan het begin van een behandeling invult altijd hoger zijn dan de groep cliënten die nog over is op een later moment.

Daarnaast includeerden we gegevens van cliënten waarbij de behandeling uiterlijk in december 2015 was afgesloten. Dit betekent dat de chronische cliënten bij wie de behandeling nog niet afgesloten was, buiten het onderzoek gehouden zijn. Dit heeft mogelijk gevolgen voor de representativiteit van het cohort.

Hoewel de database veel uitkomstdata bevat, is er relatief weinig aanvullende informatie over de cliënten meegenomen in dit onderzoek. Er zijn weinig tot geen achtergrondgegevens beschikbaar over de cliënten in de database. Ook het gegeven of de behandeling een eerste of een herhaling van een behandeling voor een depressieve episode betrof was niet altijd bekend. Als laatste zijn ook geen gegevens over de uitgevoerde behandeling meegenomen, zoals frequentie van de afspraken, en dus aantal afspraken tussen de metingen.

### Implicaties voor de praktijk

Afgelopen jaren heeft bij de ROM in Nederland de nadruk vooral gelegen op de voor- en nameting van een behandeling. Het belang van de metingen tijdens de behandeling, al te beginnen 4 weken (of eerder) na de start van de behandeling als indicatie voor succes, wordt nu steeds meer duidelijk. Als de cliënt en behandelaar zich tot doel stellen psychische klachten, waarop de diagnose is gebaseerd, te reduceren, dan is aan te raden om een diagnosespecifieke vragenlijst toe te voegen, zeker in het begin van de behandeling. Gezien de bevinding van Nugter e.a. (2017) en het internationale advies van Obbarius e.a. (2017) is het wel de vraag of men ROM moet beperken tot een specifiek instrument. Echter, we willen cliënten ook niet te veel belasten. Dit vraagt dus om zorgvuldige afwegingen bij het functioneel inrichten van elk ROM-proces.

### LITERATUUR

- Allison PD. Multiple Imputation of Categorical Variables Under the Multivariate Normal Model. Annual Meeting of the American Sociological Association, Montreal, 2006. Annual Meeting of SUGI (SAS User's Group International), Philadelphia, 2005.
- Beurs E de, Vissers E, Schoevers R, Carlier IVE, van Hemert AM, Meesters Y. Comparative responsiveness of generic versus disorder-specific instruments for depression: An assessment in three longitudinal data sets. *Depress Anxiety* 2018;1–10. <https://doi.org/10.1002/da.22809>
- Clark DM. Realising the mass public benefit of evidence-based psychological therapies: The IAPT Program. *Ann Rev Clin Psychology* 2018; 14: 159-83.
- Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale: Lawrence Erlbaum Associates; 1988.
- 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.



- Garb HN. Clinical judgment and decision making. *Ann Rev Clin Psychology* 2005; 1: 67-89.
- Hafkenscheid A, Os J van. Twee misvattingen over ROM. *De Psycholoog* 2018; april: 34-44.
- Hannan C, Lambert MJ, Harmon C, Nielsen SL, Smart DW, Shimokawa K, Sutton SW. A lab test and algorithms for identifying clients at risk for treatment failure. *J Clin Psychology* 2005; 61: 155-63.
- Hardy GE, Stiles WB, Leach C, Barkham M, Lucock M, Iveson S, Shapiro DA, Iveson MJ. Early sudden gains in psychotherapy under routine clinic conditions: practice-based evidence. *Consult Clin Psychol* 2003; 71: 14-21.
- Have M ter, Penninx BWJH, Tuithof M, Donselaer van S, Kleinjan M, Spijker J, e.a. Duration of major and minor depressive episodes and associated risk indicators in a psychiatric epidemiological cohort study of the general population. *Acta Psychiatr Scand* 2017; 136: 300-12.
- Hermans D, Putte J van de. Cognitieve gedragstherapie bij depressie. Houten: Bohn Stafleu Van Loghum; 2004.
- Judd LL. The clinical course of unipolar major depressive disorders. *Arch Gen Psychiatry* 1997; 54: 989-91.
- Judd LL, Akiskal HS, Paulus M.P. The role and clinical significance of subsyndromal depressive symptoms (SSD) in unipolar major depressive disorder. *J Affect Disord* 1997; 45: 5-17.
- Jong K de. De Nederlandse versie van de Outcome Questionnaire (OQ-45.2): een crossculturele validatie. *Psychologie & gezondheid* 2008; 36: 35-45.
- Keller MB. Depression: a long-term illness. *Br J Psychiatry* 1994; 165: 9-15.
- Keller MB, Boland RJ. Implications of failing to achieve successful long-term maintenance treatment of recurrent unipolar major depression. *Biological Psychiatry* 1998; 44: 348-60.
- 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: CD011119.
- Lambert MJ, Morton JJ, Hatfield DR, Harmon C, Hamilton S, Shimokawa K. Administration and scoring manual for the OQ-45.2 (Outcome Questionnaire). (3e ed.). Wilmington: American Professional Credentialing Services LLC; 2004.
- Lutz W, Lambert MJ, Harmon SC, Tschitsaz A, Schürch E, Stulz N. The probability of treatment success, failure and duration - what can be learned from empirical data to support decision making in clinical practice? *Clin Psychol Psychother* 2006; 13: 223-32.
- Lutz W, Stulz N, Köck K. Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *J Affect Disord* 2009; 118: 60-8.
- Meesters Y, Duijzer WB, Nolen WA, Schoevers RA, Ruhé HG. Inventory of Depressive Symptomatology en verkorte versie in routine outcome monitoring van Stichting Benchmark GGZ. *Tijdschr Psychiatr* 2016; 58: 48-54.
- Mheen M van der, Mors LM ter, Hout MA van den, Cath DC. Routine outcome monitoring bij de behandeling van angststoornissen: diagnosespecifieke versus generieke meetinstrumenten. *Tijdschr Psychiatr* 2018; 60: 11-9.
- Nugter MA, Buwalda VJA. Achtergronden en gebruiksmogelijkheden van ROM in de GGZ. *Tijdschr Psychiatr* 2012; 54: 111-8.
- Nugter MA, Hermens MLM, Robbers S, Son G Van, Theunissen J, Engelsbel F. Use of outcome measurements in clinical practice: How specific should one be? *Psychother Res* 2017; doi: 10.1080/10503307.2017.1408975.
- Nuijen J, Wijngaarden B van, Veerbeek M, Franx G, Meeuwissen J, Bonmartens M. Implementatie van ROM ter ondersteuning van de dagelijkse zorgpraktijk in de GGZ. Utrecht: Trimbos-instituut; 2015.
- Obbarius A, Maasackers L van, Baer L, Clark DM, Crocker AG, Beurs E de, 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.
- Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The Inventory of Depressive Symptomatology (IDS): psychometric properties. *Psychol Med* 1996; 26: 477-86.
- Sapyta J, Riemer M, Bickman L. Feedback to clinicians: theory, research, and practice. *J Clin Psychology* 2005; 61: 145-53.
- Schibbye P, Ghaderi A, Ljotsson B, Hedman E, Lindefors N, Christian R, 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: e100614.
- Schiepek G, Aichhorn W, Gruber M, Strunk G, Bachler E, Aas B. Real-time monitoring of psychotherapeutic processes: concept and compliance. *Front Psychol* 2016; 7: 604. doi: 10.3389/fpsyg.2016.00604
- 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 Cons Clin Psychol* 2010; 78: 298-311.
- Solstad SM, Castonguay LG, Moltu C. Patients' experiences with routine outcome monitoring and clinical feedback systems: a systematic review and synthesis of qualitative empirical literature. *Psychother Res* 2017; 27: 1-14.
- Spijker J, Bijl RV, Nolen WA. Duur van depressieve episoden en determinanten van beloop. *Tijdschr Psychiatr* 2004; 46: 229-35.
- Stulz N, Lutz W, Leach C, Lucock M, Barkham M. Shapes of early change in psychotherapy under routine outpatient conditions. *J Consult Clin Psychology* 2007; 75: 864-74.
- Tiemens BG, Sonsbeek M van. ROM als tomtom. *Tijdschr Psychiatr* 2017; 59: 489-93.

## SUMMARY

# Routine outcome monitoring (ROM) early in treatment and disorder specific. Observational study into generic and specific questionnaires to measure the early change in depression treatment

B.G. TIEMENS, M.I. KRAMER, M.W. KLOOS, J. SPIJKER

**BACKGROUND** Mental health care institutions use routine outcome monitoring (ROM) to determine whether a patient responds well to treatment. However, it has been still unanswered whether disorder specific or generic measurement instruments can best be used for this purpose. In addition, little is known about when a first indication can be given for the outcome of the treatment.

**AIM** To provide insight into the sensitivity to change of two questionnaires: a generic and a specific one, which were both used for clients with a depressive disorder. An additional objective is to provide insight into when a first indication can be given for the outcome of the treatment.

**METHOD** An observational cohort study with data from 518 patients with a depressive disorder. The Outcome Questionnaire (OQ-45.2) and the Inventory of Depressive Symptomatology (IDS-SR) were used to measure the course of treatment.

**RESULTS** The depression specific IDS-SR appeared to be more sensitive to change than the generic OQ-45.2, especially at the beginning of treatment. With a measurement frequency of once every four weeks, the best time to get a first indication on the outcome of the treatment was between week 4 and week 8.

**CONCLUSION** For clients and practitioners whose treatment focus is reducing depressive symptoms, a disorder specific questionnaire is preferable to monitor this. The measurement between 4 and 8 weeks is important for, if necessary, adjusting the treatment, and to improve the treatment outcome.

TIJDSCHRIFT VOOR PSYCHIATRIE 62(2020)2, 121-130

**KEY WORDS** depressive disorder, early chance, generic en specific questionnaires, routine outcome monitoring