

# Generaliseerbaarheid van depressietrials naar de dagelijkse praktijk

R. VAN DER LEM, N.J.A. VAN DER WEE, T. VAN VEEN, F.G. ZITMAN

- ACHTERGROND** Bewezen effectieve behandelingen voor depressie, zoals beschreven in de richtlijnen, zijn gebaseerd op resultaten uit gerandomiseerde gecontroleerde trials (RCT's). Het is echter onbekend in hoeverre resultaten uit RCT's voor depressie generaliseerbaar zijn naar de dagelijkse praktijk.
- DOEL** Het vergelijken van de resultaten voor depressie uit RCT's met die uit de dagelijkse praktijk. Daarnaast vaststellen in hoeverre patiëntselectie in RCT's de behandeluitkomsten in de dagelijkse praktijk beïnvloeden.
- METHODE** In een groep van 1653 patiënten uit de dagelijkse praktijk die volgens de MINIplus een depressieve stoornis hadden en van wie gegevens uit routine outcome monitoring (ROM) beschikbaar waren (de baselinegroep) werd onderzocht in welke mate zij geëxcludeerd zouden worden voor RCT's voor antidepressiva en psychotherapie. Tevens werd onderzocht in hoeverre patiënten uit een ROM-bestand verschilden van patiënten uit RCT's in sociodemografisch en socio-economisch opzicht. Bij een groep van 626 patiënten met ten minste 1 vervolg-ROM-meting werd onderzocht in hoeverre de behandelresultaten, gemeten met de MADRS en de BDI-II, overeenkwamen met die van 15 meta-analyses van RCT's. Daarnaast werd onderzocht in hoeverre de exclusiecriteria en de sociodemografische en socio-economische variabelen invloed hadden op het behandelresultaat.
- RESULTATEN** Percentages remissie (21-27% in ROM versus 34-58% in RCT's) en effectgrootte (0,85 in ROM versus 1,71 in RCT's, within-groupdata) lagen in de dagelijkse praktijk lager dan in RCT's. Patiënten uit de ROM verschilden in veel klinische en sociodemografische kenmerken van patiënten uit RCT's. De invloed van exclusiecriteria en sociodemografische kenmerken op behandeluitkomst bleek echter gering (verklaarde variantie 1-11%).
- CONCLUSIE** Het behandelresultaat bij depressie is in de dagelijkse praktijk minder dan in RCT's en patiënten uit de dagelijkse praktijk verschillen in veel kenmerken van die uit RCT's. Echter, deze resultaten kunnen het verschil in behandel succes niet verklaren. De generaliseerbaarheid van uitkomsten uit depressietrials naar de dagelijkse praktijk wordt minder beïnvloed door gebruik van exclusiecriteria en sociodemografische verschillen dan in eerder onderzoek werd gesuggereerd.

TIJDSCHRIFT VOOR PSYCHIATRIE 57(2015)8, 579-587

**TREFWOORDEN** depressie, dagelijkse praktijk, effectiviteit, gerandomiseerde gecontroleerde trial (RCT), routine outcome monitoring (ROM)



*U kent hem wel: dr. X, een gewaardeerde collega, 60 jaar oud, en werkzaam in een polikliniek van een grote ggz-instelling. (We hebben gekozen voor een mannelijke collega, we hadden evengoed een vrouwelijke collega als voorbeeld kunnen kiezen.) Hij ziet iedere dag veel patiënten. In de loop van zijn carrière heeft dr. X veel ontwikkelingen in de psychiatrie meegemaakt: nieuwe*

*psychofarmaca, de opkomst van geprotocolleerde behandelingen en de voortschrijdende inzichten op gebied van biologische aspecten van de psychiatrie. In de ggz-instelling waar dr. X werkt, zijn de richtlijnen voor de behandeling van psychiatrische stoornissen ingevoerd in de dagelijkse patiëntenzorg. Dr. X was geïnteresseerd, maar ook wat sceptisch, en bang dat alle creati-*

viteit uit zijn vak zou verdwijnen. Toch committeerde hij zich aan de behandelalgoritmes. Dr. X leest de veelbelovende resultaten van gerandomiseerde gecontroleerde trials (RCT's). Ondertussen zijn de behandelresultaten in zijn eigen ggz-instelling vaak wat teleurstellend. Dr. X krijgt de indruk dat de behandeling van depressie in RCT's veel succesvoller is dan in 'werkelijkheid'. Hij vraagt zich af: lijken mijn patiënten wel op deelnemers aan RCT's? Vertellen RCT's ons eigenlijk wel iets over de dagelijkse psychiatrische praktijk?

In dit overzichtsartikel proberen we de vragen van dr. X te beantwoorden op basis van ons onderzoeksproject, dat resulteerde in het proefschrift: Are depression trials generalizable to clinical practice? Something clinicians always wanted to know about RCT's but were afraid to ask (Van der Lem 2013).

## Achtergrond

Voor de jaren negentig van de vorige eeuw was de behandeling van psychiatrische stoornissen gebaseerd op de persoonlijke ervaring van de clinicus. Tegenwoordig is ook in de psychiatrie *evidence-based medicine* leidend. In hoeverre de resultaten van RCT's echter generaliseerbaar zijn naar de dagelijkse praktijk is niet goed bekend. Wij onderzochten dit voor depressie. In de behandelrichtlijnen geldt een systematische review van ten minste twee dubbelblinde RCT's van goede kwaliteit en met een grote onderzoekspopulatie als betrouwbaarste bewijs van effectiviteit. Om te begrijpen waar de generaliseerbaarheid van resultaten uit RCT's in het gedrang kan komen, is het van belang te weten hoe RCT's zijn opgezet. In RCT's streeft men naar een maximale interne validiteit: resultaten uit RCT's moeten replicerbaar zijn en slechts terug te voeren zijn tot de onderzochte therapievorm. Om andere invloeden te beperken, maakt men gebruik van een placebogroep, randomisatie, dubbelblinde evaluatie en exclusiecriteria voor patiëntselectie. Het behandel-effect wordt (meestal) gedefinieerd als de mate van reductie van de symptomen. Als in een RCT is aangetoond dat fluoxetine bij depressie effectief is, dan is bij een groep patiënten tussen de 18-55 jaar met een depressieve stoornis zonder comorbiditeit, protocolgetrouwe inname van fluoxetine effectiever dan placebo binnen een vastgestelde periode. De interne validiteit is cruciaal om de effectiviteit van een therapievorm vast te stellen.

Voor een inschatting van de werkzaamheid van de behandeling in de dagelijkse praktijk is de externe validiteit belangrijk. De externe validiteit bepaalt de mate waarin een resultaat generaliseerbaar is naar een grotere patiëntenpopulatie met heterogenere kenmerken, zoals in de dagelijkse praktijk. De externe validiteit neemt toe als er in RCT's minder risico is op selectiebias (oververtegenwoordiging van patiënten met bepaalde kenmerken).

## AUTEURS

**ROSALIND VAN DER LEM**, psychiater, Kijvelanden/Het Dok, forensische psychiatrische polikliniek, tot 2011 werkzaam als psychiater-promovendus bij Rivierduinen/Leids Universitair Medisch Centrum.

**NIC VAN DER WEE**, hoogleraar Psychiatrie, afd. Psychiatrie, Leids Universitair Medisch Centrum.

**TINEKE VAN VEEN**, epidemioloog, afd. Psychiatrie, Leids Universitair Medisch Centrum.

**FRANS ZITMAN**, emeritus hoogleraar Psychiatrie, Leids Universitair Medisch Centrum.

## CORRESPONDENTIEADRES

Dr. Rosalind van der Lem, Kijvelanden/Het Dok, Diergaardesingel 69-73, 3014 AE Rotterdam.

E-mail: Rosalind.van.der.lem@hetdok.nl

Geen strijdige belangen meegegeedeld.

Het artikel werd voor publicatie geaccepteerd op 2-2-2015.

In- en externe validiteit staan op gespannen voet met elkaar. Rothwell (2005) geeft een overzicht van de aspecten van de onderzoeksopzet van een RCT die de externe validiteit kunnen verlagen. Hij geeft hierin aan dat door het gebruik van exclusiecriteria in RCT's het merendeel van de patiënten in de dagelijkse praktijk niet in aanmerking komt voor RCT-deelname. Daarnaast kan de wijze waarop men bij RCT's patiënten rekruteert de selectiebias verder vergroten. Verschillende onderzoekers hebben vanwege deze factoren vraagtekens gezet bij de generaliseerbaarheid van de resultaten van RCT's naar 'echte' patiënten (Goodwin e.a. 1999; Ten Have e.a. 2003; Leucht 2006; Zimmerman e.a. 2002).

In ons onderzoeksproject onderzochten we of de klinische indruk van dr. X klopt: is de behandeling volgens de richtlijnen in de dagelijkse praktijk minder effectief dan in RCT's? Daarna onderzochten we of verschillen tussen 'patiënten uit de dagelijkse praktijk' en patiënten uit RCT's, die door patiëntselectie ontstaan, een verklaring kunnen vormen voor verschil in behandelresultaat. In dit overzichtsartikel beschrijven wij de methodologie van ons project en geven wij de belangrijkste resultaten weer. We noemen de beperkingen van ons onderzoek en doen aanbevelingen voor toekomstig onderzoek. De uitgebreide resultaten van ons onderzoek zijn gepubliceerd in vijf artikelen en beschreven in een proefschrift (Van der Lem 2013; Van der Lem e.a. 2011a; b; 2012a; b; 2013).

## METHODEN

### Routine outcome monitoring

In dit onderzoeksproject gebruikten wij data uit *routine outcome monitoring* (ROM). In Rivierduinen, een ggz-instelling met o.a. vijf regionale poliklinieken in de Randstad ([www.rivierduinen.nl](http://www.rivierduinen.nl)) wordt ROM door onafhankelijke testverpleegkundigen bij alle patiënten afgenomen als standaardonderdeel van de intakeprocedure en bij de overgang naar een volgende behandelstap. Alleen als de patiënt onvoldoende Nederlands spreekt of te ziek is om vragenlijsten af te nemen, wordt de patiënt geëxcludeerd voor ROM. Het wetenschappelijk gebruik van geanonimiseerde ROM-data is goedgekeurd door de medisch-ethische commissie van het LUMC. De volledige lijst van instrumenten die gebruikt worden bij ROM in Rivierduinen is te vinden op: <http://www.lumc.nl/psychiatry/ROM-instruments>. Voor ons onderzoeksproject gebruikten we data uit de volgende vragenlijsten: MINIplus (Sheehan e.a. 1998); DAPP SF (de Beurs e.a. 2009); *Montgomery Asberg Depression Rating Scale* (MADRS; Asberg e.a. 1978); *Beck Depression Inventory-II* (BDI-II, Beck e.a. 1961); DEMOG: sociodemografische en socio-economische gegevens van een patiënt.

### Patiënten

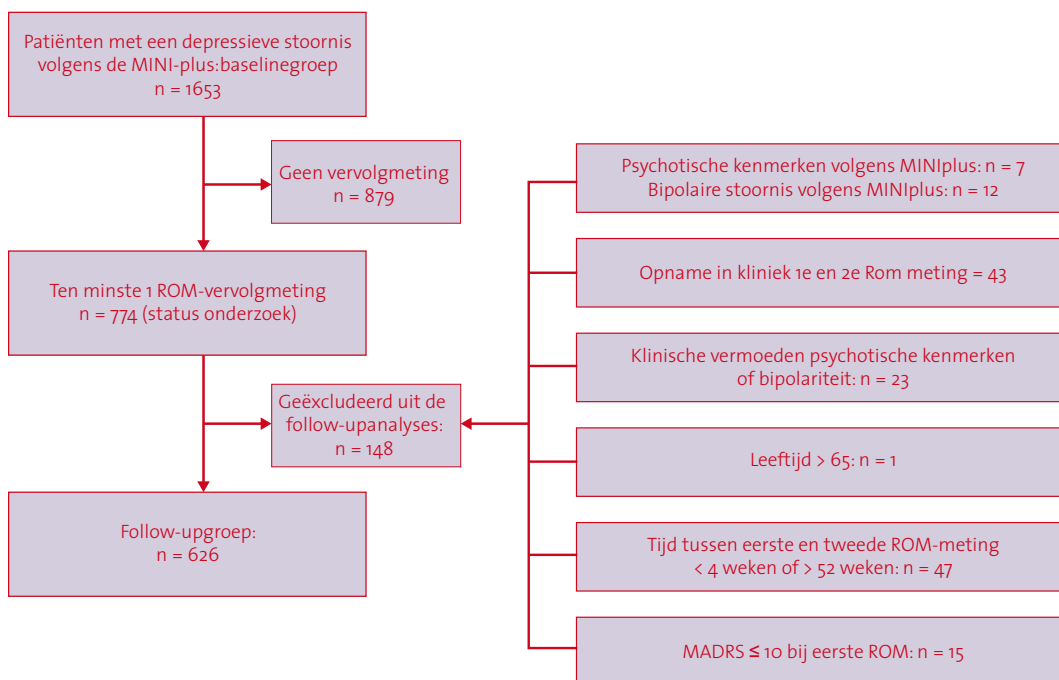
Voor dit onderzoeksproject selecteerden wij twee groepen patiënten. De eerste groep bestond uit volwassen patiënten

(18-65 jaar) die tussen 2002 en 2006 behandeling zochten bij Rivierduinen en een eerste ROM-meting hadden (circa 5000 patiënten). Om geïnccludeerd te worden moesten de patiënten voldoen aan de criteria van een depressieve stoornis volgens de MINIplus. Deze patiënten vormden de 'baselinegroep'. Vervolgens selecteerden wij hieruit de patiënten die ten minste één ROM-vervolgmeting hadden. Bij deze patiënten verrichtten wij statusonderzoek om informatie te krijgen over de behandeling en de primaire diagnose die door de behandelaar werd gesteld. Uit de laatste groep selecteerden wij de 'follow-upgroep'. In **FIGUUR 1** staat de patiëntselectie weergegeven.

### Groep uitvallers

Aangezien een grote groep patiënten geen ROM-vervolgmetingen bleek te hebben (> 50%, vergelijkbaar met STAR\*D), vergeleken wij onze 'follow-upgroep' met deze patiënten. De helft van de patiënten zonder vervolgmetingen heeft de behandeling niet afgemaakt; de andere helft bleef in behandeling, maar had geen vervolgmetingen. Patiënten zonder vervolgmeting waren iets jonger en iets vaker mannen. Daarnaast kwam PTSS vaker voor in deze groep. Echter, wat betreft de meeste comorbide stoornissen en sociodemografische variabelen, waren de groepen vergelijkbaar. Nader onderzoek naar risicofactoren voor uitval uit ROM en behandeling viel buiten het bestek van ons onderzoeksproject, maar is zeker aanbevelenswaardig.

**FIGUUR 1** Selectie van patiënten met ROM-gegevens



## Behandelsucces

Om de effectiviteit van depressiebehandelingen in RCT's te kunnen vergelijken met de effectiviteit van behandeling in de dagelijkse praktijk zoals gemeten met ROM selecteerden wij 15 meta-analyses over de effectiviteit van therapie bij een unipolaire depressie, over antidepressiva, individuele psychotherapie (cognitieve gedragstherapie (CGT) en interpersoonlijke therapie (IPT)) en over combinatietherapie. We includeerden alleen meta-analyses die zogenoemde *within-groupeffectiviteitscijfers* geven. Dit houdt in dat in de meta-analyse de effectiviteit van een bepaalde therapievorm in een patiëntengroep wordt weergegeven ten opzichte van de uitgangssituatie van diezelfde groep, in plaats van de effectiviteit van deze therapievorm ten opzichte van een andere therapievorm of placebo (*tussen-groupeffectiviteit*). De effectiviteit in de meta-analyse moest gemeten zijn met de *Hamilton Depression Rating Scale* (HAM-D; Hamilton 1967), de MADRS of de BDI-II. Er was geen overlap tussen de RCT's die geïncludeerd waren in de 15 geselecteerde meta-analyses. Om vergelijking mogelijk te maken hanteerden wij in onze ROM-follow-upgroep dezelfde maten voor effectiviteit als in de meta-analyses. Aangezien in de meta-analyses vaker de HAM-D werd gebruikt, rekenden wij de MADRS-scores om (Carmody e.a. 2006; Zimmerman e.a. 2004b).

Om de ROM-behandeluitkomsten te kunnen vergelijken met die in de meta-analyses, definieerden we behandeluitkomst op dezelfde manier als in de meta-analyses gedaan werd:

- Remissiepercentage: percentage patiënten met een  $MADRS \leq 10$  ( $\approx$  HAM-D 6-7) na behandeling.
- Effectgrootte:  $(\mu_{post} - \mu_{pre}) / \sigma$  (mate van verbetering gecorrigeerd voor groeps grootte,  $\sigma$  = standaarddeviatie bij aanvang, gemeten op de BDI-II om vergelijking met de geïncludeerde meta-analyse die deze definitie gebruikt mogelijk te maken (Minami e.a. 2007)).

Wij vergeleken onze behandelresultaten ook met de uitkomsten van de STAR\*D-trial (Rush e.a. 2004), een bekende grote 'pragmatische' trial waarin men de dagelijkse praktijk zoveel mogelijk trachtte te benaderen.

## Patiëntselectie op basis van exclusiecriteria in RCT's

Voor onderzoek naar de invloed van patiëntselectie op basis van exclusiecriteria gebruikten wij het model van Zimmerman (Zimmerman e.a. 2004a), waarin de exclusiecriteria worden beschreven die consistent in antidepressiva-efficacytrials (AET's) worden gebruikt. Deze exclusiecriteria zijn: (hypo)manische episodes in de voorgeschiedenis, psychotische kenmerken, suïcidaliteit, alcohol en/of middelenmisbruik of -afhankelijkheid, niet voldoen aan een minimum ernstcriterium (HAM-D > 18), comorbide

dysthyme stoornis, ziekte duur korter dan 4 weken of meer dan 2 jaar, comorbide as I-stoornissen en comorbide borderlinepersoonlijkheidsstoornis. 'In aanmerking komen voor AET's' definieerden we als voldoen aan alle criteria.

Voor de psychotherapie-efficacytrials (PET's) bestond een dergelijk model niet. Daarom deden wij een literatuuronderzoek naar PET's, analoog aan de methode van de cochrane-reviews. Wij includeerden uiteindelijk 15 PET's. Hieruit bleek dat in PET's exclusiecriteria minder consistent werden gebruikt. Het maken van een model analoog aan dat van Zimmerman was daardoor niet mogelijk. De drie meest gebruikte exclusiecriteria (> 75%) waren: (hypo)manische episodes in de voorgeschiedenis, psychotische kenmerken, alcohol of middelenmisbruik of -afhankelijkheid en niet voldoen aan een minimum ernstcriterium (HAM-D > 14). Vanwege de klinische relevantie namen wij in onze analyses ook het volgende exclusie criterium mee: gebruik van antidepressiva voorafgaand aan de psychotherapietrial (gebruikt in 70% van de onderzochte PET's). We berekenden de invloed van de verschillende exclusiecriteria op therapie-uitkomst in univariate analyses, en ook in 'joint analyses'. In joint analyses werd de invloed van de afzonderlijke exclusiecriteria bepaald terwijl er rekening werd gehouden met de invloed van de andere exclusiecriteria. Wij controleerden daarnaast voor mogelijke confounding door leeftijd, geslacht, primaire klinische diagnose en behandeltype.

## Sociodemografische/socio-economische verschillen tussen patiënten in RCT versus dagelijkse praktijk

Om de sociodemografische en socio-economische status (SES) van patiënten uit RCT's te onderzoeken gebruikten we dezelfde AET's en PET's die gebruikt waren voor de identificatie van de exclusiecriteria. De volgende kenmerken werden consistent beschreven in de RCT's: leeftijd, geslacht, etniciteit, huwelijkse staat en 'arbeidsstatus'. Wij definieerden arbeidsstatus als 'aan het werk zijn ten tijde van behandeling'. Patiënten zonder betaald werk en patiënten die een ziekte wettuitkering kregen werden beschouwd als 'niet aan het werk'. We vergeleken de kenmerken van patiënten uit RCT's met die van onze ROM-patiënten. We berekenden de invloed van de verschillende SES-kenmerken op therapie-uitkomst afzonderlijk en ook in de eerder genoemde joint analyses. Wij corrigeerden hierbij voor mogelijke confounding door de ernst van de depressie bij start van de behandeling, behandeltype en opleidingsniveau.

## Statistische analyses

Voor de vergelijking tussen *efficacy* en *effectiveness*, gebruikten we de methode voor de vergelijking van twee onafhan-

kelijke proporties (Altman 1991). Voor onze predictiemodellen over de invloed van de exclusiecriteria en SES-kenmerken op behandeluitkomst maakten wij gebruik van uni- en multivariate logistische- en lineaire-regressieanalyses (Altman 1991). De verklaarde variantie geeft aan in hoeverre het gekozen predictiemodel verklaard kan worden door de geïncludeerde variabelen, en is daarmee een maat voor de validiteit van predictie. Wij hebben, waar geïndiceerd, gecorrigeerd met de MICE-methode (Royston 2005) voor ontbrekende data ( $n = 103$  voor de MADRS in de baselinegroep,  $n = 28$  voor type behandeling en  $n = 82$  voor SES-kenmerken in de follow-upgroep).

## RESULTATEN

### Behandelsucces

We vonden in ROM een remissiepercentage van 22,6% en een effectgrootte van 0,85. De effectiviteit van behandeling met antidepressiva, individuele psychotherapie en combinatietherapie in de dagelijkse praktijk (ROM) was lager dan in RCT's/meta-analyses (remissiepercentages 34-63%; effectgrootte 1,71; 95%-BI: 1,60-1,82). Er werd geen verschil gevonden met de behandeluitkomsten in de STAR\*D-trial.

### Verschillen tussen patiënten in RCT's en in ROM door patiëntselectie

Op basis van de exclusiecriteria voor RCT's, kwam slechts een minderheid (< 25%) van de ROM-populatie in aanmerking voor deelname aan RCT's. De meest voorkomende redenen voor exclusie waren:

- De ernst van de depressie bij patiënten uit de dagelijkse praktijk lag onder de minimumernst bij aanvang bij AET (gold voor 42%) en bij PET (22%).
- Patiënten uit de dagelijkse praktijk rapporteerden suïcidaliteit (15%).
- Patiënten uit de dagelijkse praktijk rapporteerden middelengebruik en/of -afhankelijkheid (9%), hadden vaak andere comorbide as I-stoornissen (63%) en soms een comorbide borderlinepersoonlijkheidsstoornis (7%).
- Patiënten uit de dagelijkse praktijk hadden vaak antidepressiva gebruikt voorafgaand aan hun psychotherapie (44%).

We vonden dat patiënten uit RCT's ouder waren, iets vaker van het mannelijk geslacht (verschil 4%;  $p = 0,05$ ), iets vaker blank (verschil 4%;  $p < 0,001$ ) en wat vaker alleenstaand (verschil 7%;  $p = 0,0003$ ) dan patiënten uit de ROM. De invloed van exclusiecriteria en SES-kenmerken op behandeluitkomst staan in **TABEL 1**. De verklaarde variantie van onze predictiemodellen was zowel voor exclusiecriteria (1-4% voor de AET-criteria; 4-11% voor de PET-criteria) als voor SES-kenmerken (3-7%) erg laag. Dit betekende dat

er andere factoren, die buiten het bestek van ons onderzoeksproject lagen, invloed hadden op de therapie-uitkomst en dat de invloed van de door ons onderzochte variabelen zeer bescheiden was.

## DISCUSSIE

Clinici hebben gelijk wanneer ze stellen dat resultaten van depressiebehandeling in de praktijk minder gunstig zijn dan die in RCT's. Wij hebben vastgesteld dat als we in de dagelijkse praktijk alleen die patiënten behandelen die in aanmerking komen voor RCT-deelname, het behandel succes nauwelijks verbetert. Ook verschillen patiënten uit RCT's erg van die uit de dagelijkse praktijk door selectieprocedures. We vonden dat de exclusie van patiënten met een lichte depressie en patiënten die niet aan het werk zijn, leidt tot toename van behandel succes in de dagelijkse praktijk. Echter, de mate waarin het verschil tussen efficacy en effectiveness verklaard kan worden door exclusiecriteria is erg klein. Wij denken dat andere verschillen tussen RCT's en de dagelijkse praktijk bijdragen aan het verschil in behandel succes en hebben hier een aantal gedachten over. *Dr. X voert aan dat de teleurstellende behandeluitkomsten typerend zijn voor Rivierduinen, de Leidse regio of de Nederlandse psychiatrie.*

Hier moeten we opmerken dat de therapie-uitkomsten die wij vonden niet verschillen van die van de STAR\*D-trial. Dit wijst erop dat, ook in de Verenigde Staten, behandel effecten in de praktijk kleiner zijn dan bij RCT's, zelfs als er veel wordt geïnvesteerd in therapietrouw bij patiënten en therapeuten (zoals bij STAR\*D het geval was).

### Verschil in patiëntpopulatie

*De eerste patiënt van de dag is mw. Y. Mw. Y is een vrouw van 55 jaar, die recent haar baan is kwijtgeraakt. Gesprekken bij een psycholoog in de eerste lijn en antidepressiva van haar huisarts hebben haar stemming niet doen opklaren. Ze is wat sceptisch over de verwijzing naar dr. X, maar is vast van plan hem al haar problemen te vertellen. Hoe groot is de kans dat mw. Y mee wil doen aan een trial voor een nieuw antidepressivum in een academisch ziekenhuis?*

Patiënten in een RCT verschillen waarschijnlijk in veel opzichten van die in de dagelijkse praktijk. Mogelijk zijn de eersten gemotiveerder voor behandeling. Wellicht sturen klinici patiënten met een slechte prognose niet in voor deelname of zijn juist deze patiënten niet gemotiveerd voor deelname? Belangrijker nog is de invloed van het Nederlandse verwijssysteem. Alleen patiënten die onvoldoende baat hebben bij een behandeling in de eerste lijn worden verwezen naar de tweede lijn. Dit zijn waarschijnlijk patiënten met depressieve stoornis en een slechtere prognose. Voor RCT's rekruteert men vaak uit de populatie

**TABEL 1** De invloed van exclusiecriteria en sociaaleconomische status op behandeluitkomst; gecorrigeerde associaties uit multivariate regressieanalyse (joint analysis) en bij enkele variabelen de ongecorrigeerde associatie

Patiëntkenmerken	Invloed op remissiepercentage	Invloed op mate van verbetering (effectmaat, alleen gebruikt in PET's)
Exclusiecriteria		
AET: suicidaliteit	NS <sup>1</sup>	-
AET: bij aanvang voldoen aan ernst	OR 2,0 (95%-Bl: 1,3-3,1)	-
HAMD > 18		
PET: bij aanvang voldoen aan ernst	OR 0,5	B = 7,2
HAMD > 14	(95%-Bl: 0,33-0,81)	(95%-Bl: 5,3-9,1)
AET en PET: alcohol/middelenmisbruik	NS	NS
AET: comorbide dysthymie	NS	-
AET: comorbide angststoornis	NS	-
AET: comorbide somatoforme stoornis	NS	-
AET: comorbide overige as I-stoornissen	NS	-
PET: eerder gebruik van antidepressiva/	NS	-
ECT	ongecontroleerde analyse: OR 1,5 (95%-Bl: 1,0-2,3)	
In aanmerking komen voor deelname aan een AET	NS (OR 1,0 (95%-Bl: 0,5-2,0))	-
sociaaleconomische status		
Leeftijd	NS	-
Geslacht	NS	-
Etniciteit	NS	-
Huwelijkse staat	NS	-
Arbeidsstatus	OR 1,9 (95%-Bl: 1,2-2,8) ongecontroleerde analyse: OR 2,3 (95%-Bl: 1,6-3,4)	-

AET = antidepressiva-efficacytrials; PET = psychotherapie-efficacytrials; NS = niet significant.

van zowel de eerste als de tweede lijn en RCT's hebben daardoor een groep met een 'gemengde' prognose.

### Verskil tussen klinici bij RCT en in de dagelijkse praktijk

Het is vrijdagavond en dr. X verlaat zijn polikliniek. Het is een drukke week geweest: twintig patiënten per dag, besprekingen, supervisie, crisisbeoordeling, invallen voor een langdurig zieke collega, etc. Dr. X is moe en kijkt uit naar zijn pensioen. In zijn tas zit een brochure van een veelbelovende trial. Op de cover staat een glimlachende collega in een smetteloos witte jas, die ongeveer twee keer zo jong is als dr. X.

Clinici in de dagelijkse praktijk verschillen mogelijk van klinici die participeren in trials: de eersten hebben waarschijnlijk minder tijd per patiënt beschikbaar. Mogelijk volgen klinici uit de dagelijkse praktijk de richtlijnen minder strikt dan hun collega's betrokken bij RCT's en die

laatst zijn mogelijk extra geschoold of gemotiveerd, wat kan bijdragen aan het behandel succes in RCT's.

### Placebo-effect

Dr. X meldt dhr. A. aan voor deelname. Dhr. Z is een 45-jarige patiënt met een depressieve stoornis. Hij is wat sceptisch over medicatie, maar de trialtherapeut overtuigt hem ervan dat het nieuwe medicijn veelbelovend is. Hij legt uit dat dhr. Z mogelijk in de placebogroep terecht zal komen. Zonder het te weten krijgt dhr. Z. inderdaad de placebo en hij knapt op. Na drie maanden wordt het behandel effect gemeten en neemt de therapeut afscheid van dhr. Z.

Iedere behandeling kent een placebo-effect (behandel succes dat niet toe te schrijven is aan de verondersteld werkzame onderdelen van een behandeling). Waarschijnlijk is de invloed van het placebo-effect in trials groter dan in de dagelijkse praktijk, aangezien in RCT's de patiënten die

spontaan genezen, worden meegenomen in het behandel-effect. In de dagelijkse praktijk komen deze patiënten niet meer opdagen of wordt, als het herstel aan het begin van de behandeling optreedt, de behandeling afgesloten. Verder hebben patiënten in RCT's het gevoel dat ze aan een nieuwe veelbelovende behandeling zijn begonnen. Dit gevoel draagt bij aan het behandelresultaat en staat bekend als het Hawthorne-effect (Kirsch 1985; Leonard 2008). In RCT's ligt de frequentie van de afspraken met de therapeut waarschijnlijk hoger dan in de dagelijkse praktijk en een hogere frequentie draagt bij aan een hoger behandelingsucces (Posternak & Zimmerman 2007).

### Publicatiebias

*Therapeut A. voerde een trial uit over een nieuwe vorm van psychotherapie. Helaas vallen de resultaten tegen. Hoe groot is de kans dat haar negatieve resultaten in een toonaangevend wetenschappelijk tijdschrift worden gepubliceerd?*

Er is sprake van publicatiebias: positieve resultaten in trials worden sneller en in belangrijker tijdschriften gepubliceerd dan negatieve resultaten (Fanelli 2010). Hierdoor kan de effectiviteit zoals die in gepubliceerde RCT's (en de meta-analyses daarvan) wordt weergegeven, te gunstig lijken. Er zijn verschillende maatregelen genomen om publicatiebias te beperken en er zijn mogelijkheden om er statistisch voor te corrigeren in meta-analyses, maar desondanks zal publicatiebias nog steeds invloed hebben op de gevonden effectiviteit. Hoe groot deze invloed op dit moment is, is onbekend.

### Beperkingen van ons onderzoeksproject

In de onderzochte RCT's werden verschillende meetinstrumenten, verschillende definities van behandelingsucces en verschillende sets exclusiecriteria gebruikt. Hierdoor zou de validiteit van de door ons gebruikte behandeluitkomsten uit de meta-analyses beperkt kunnen worden. Of een patiënt voldoet aan de exclusiecriteria van een RCT bepalen onderzoekers waarschijnlijk mede op basis van een klinische inschatting en die kan tussen RCT's variëren. In ons project baseerden wij ons op een semigestructureerd interview. Het is mogelijk dat het percentage patiënten dat in werkelijkheid in aanmerking komt voor RCT-deelname in ons project hierdoor onder- of overschat is.

Er was bij de ROM-metingen sprake van ontbrekende variabelen en een grote groep patiënten zonder vervolgmetingen. Dit laatste is deels te verklaren doordat patiënten hun behandeling niet afmaken (50%). Anderen blijven wel in behandeling, maar krijgen geen vervolgmeting. Wij weten niet wat de redenen hiervoor zijn. Misschien heeft de patiënt geen behoefte aan vervolgmetingen, of onderschrijft de behandelaar de noodzaak tot vervolgmetingen

niet. Het is ook mogelijk dat patiënten om logistieke redenen niet komen (bereikbaarheid patiënten, roosters etc.). Hoewel wij getracht hebben het gebrek aan vervolgmetingen zoveel mogelijk methodologisch te ondervangen, kan dit toch invloed hebben gehad op de validiteit van onze resultaten. Als vooral patiënten die goed opknappen geen vervolgmetingen meer krijgen, heeft dit geleid tot een onderschatting van het behandelingsucces in de praktijk. Als het vooral patiënten zijn die niet beter worden, dan is het behandelingsucces in dagelijkse praktijk in werkelijkheid mogelijk nog lager dan in ons onderzoeksproject.

Tot slot is er een beperking in de generaliseerbaarheid van onze resultaten. Hoewel de kenmerken van onze ROM-populatie representatief lijken voor patiënten met een depressieve stoornis in de dagelijkse tweedelijns ggz, zijn zij niet generaliseerbaar naar de eerste of de derde lijn. Daarnaast is het mogelijk dat de effectiviteit van behandeling in andere ggz-instellingen anders is dan in Rivierduinen, door verschil in behandelaanbod, scholing van personeel, personeelsopbouw etc. Hierdoor zijn de behandelresultaten die wij vonden in onze ROM-populatie niet zonder meer generaliseerbaar naar andere tweedelijns ggz-instellingen.

### Aanbevelingen voor toekomstig onderzoek

Uit ons project blijkt het belang van verder onderzoek naar de behandelresultaten bij patiënten uit de dagelijkse praktijk. Meer kennis over deze grote groep patiënten zou kunnen leiden tot aanpassingen in onze behandeling en in onze ROM. Om vragen uit de klinische praktijk wetenschappelijk te kunnen beantwoorden, zijn grote databases met gegevens over verschillende aspecten van psychiatrische stoornissen en patiënten noodzakelijk. Op die manier verkrijgen we meer kennis over patiënten uit de dagelijkse praktijk en kunnen we uiteindelijk een wetenschappelijk fundament leggen voor 'behandeling op maat'. Daarnaast kan deze kennis bijdragen aan een grotere werkzaamheid van de bestaande behandelmethodes. We vinden onze onderzoeksresultaten daarom geen aanleiding voor somberheid, maar juist een stimulans voor verder onderzoek.

## LITERATUUR

- Altman DG. *Practical Statistics for Medical Research*. 1991. p. 229-76 + p. 325-51.
- Asberg M, Montgomery SA, Perris C, Schalling D, Sedvall G. A comprehensive psychopathological rating scale. *Acta Psychiatr Scand Suppl* 1978; 271: 5-27.
- Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry* 1961; 4: 561-71.
- Beurs E de, Rinne T, van KD, Verheul R, Andrea H. Reliability and validity of the Dutch Dimensional Assessment of Personality Pathology-Short Form (DAPP-SF), a shortened version of the DAPP-Basic Questionnaire. *J Pers Disord* 2009; 23: 308-26.
- Carmody TJ, Rush AJ, Bernstein I e.a.. The Montgomery Asberg and the Hamilton ratings of depression: a comparison of measures. *Eur Neuropsychopharmacol* 2006; 16: 601-11.
- Fanelli D. 'Positive' results increase down the Hierarchy of the Sciences. *PLoS one* 2010; 5 (4): e10068.
- Goodwin PJ, Pritchard KI, Spiegel D. The Fox guarding the clinical trial: internal vs. external validity in randomized studies. *Psychooncology* 1999; 8: 275.
- Hamilton M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 1967; 6: 278-96.
- Kirsch I. Response expectancy of experience and behavior. *American Psychologist* 1985; 40: 1189-202.
- Lem R van der. Are depression trials generalizable to clinical practice? Something clinicians always wanted to know about RCT's but were afraid to ask. Leiden: LUMC; 2013.
- Lem R van der, Stamsnieder P, Wee NJA van der, Veen T van, Zitman FG. Socio-demographic features in randomized controlled trials for major depression: generalizability and individualization. *Int J Pers Cent Med* 2011a, 1: 268-78.
- Lem R van der, Stamsnieder PM, Wee NJA van der, Veen T van, Zitman FG. Influence of sociodemographic and socioeconomic features on treatment outcome in RCT's versus daily psychiatric practice. *Soc Psychiatry Psychiatr Epidemiol* 2013; 48: 975-84.
- Lem R van der, Wee NJA van der, Veen T van, Zitman FG. The generalizability of antidepressant efficacy trials to routine psychiatric out-patient practice. *Psychol Med* 2011b; 41: 1353-63.
- Lem R van der, Wee NJA van der, Veen T van, Zitman FG. Efficacy versus effectiveness: a direct comparison of the outcome of treatment for mild to moderate depression in randomized controlled trials and daily practice. *Psychother Psychosom* 2012; 81: 226-34.
- Lem R van der, Wever WW de, Wee NJA van der, Veen T van, Cuijpers P, Zitman FG. The generalizability of psychotherapy efficacy trials in major depressive disorder: An analysis of the influence of patient selection in efficacy trials on symptom outcome in daily practice. *BMC Psychiatry* 2012, 12:192. DOI: 10.1186/1471-24X-12-192.
- Leonard KL. Is patient satisfaction sensitive to changes in the quality of care? An exploitation of the Hawthorne effect. *J Health Econ* 2008; 27: 444-59.
- Leucht S. Translating research into clinical practice: critical interpretation of clinical trials in schizophrenia. *Int Clin Psychopharmacol* 2006; 21 Suppl 2: S1-10.
- Minami T, Wampold BE, Serlin RC, Kircher JC, Brown GS. Benchmarks for psychotherapy efficacy in adult major depression. *J Consult Clin Psychol* 2007; 75: 232-43.
- Posternak MA, Zimmerman M. Therapeutic effect of follow-up assessments on antidepressant and placebo response rates in antidepressant efficacy trials: meta-analysis. *Br J Psychiatry* 2007; 190: 287-92.
- Rothwell PM. External validity of randomised controlled trials: 'to whom do the results of this trial apply?'. *Lancet* 2005; 365: 82-93.
- Royston P. Multiple imputation of missing values: update. *Stata Journal* 2005; 5: 188-201.
- Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA, e.a. Sequenced treatment alternatives to relieve depression (STAR\*D): rationale and design. *Control Clin Trials* 2004; 25: 119-42.
- Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, e.a. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998; 59 (Suppl 20): 22-33.
- TenHave TR, Coyne J, Salzer M, Katz I. Research to improve the quality of care for depression: alternatives to the simple randomized clinical trial. *Gen Hosp Psychiatry* 2003; 25: 115-23.
- Zimmerman M, Chelminski I, Posternak MA. Exclusion criteria used in antidepressant efficacy trials: consistency across studies and representativeness of samples included. *J Nerv Ment Dis* 2004a; 192: 87-94.
- Zimmerman M, Mattia JI, Posternak MA. Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? *Am J Psychiatry* 2002; 159: 469-73.
- Zimmerman M, Posternak MA, Chelminski I. Implications of using different cut-offs on symptom severity scales to define remission from depression. *Int Clin Psychopharmacol* 2004b; 19: 215-20.



## SUMMARY

# The generalisability of depression trials to clinical practice

## Something clinicians always wanted to know about RCTs but were afraid to ask

R. VAN DER LEM, N.J.A. VAN DER WEE, T. VAN VEEN, F.G. ZITMAN

**BACKGROUND** Evidence-based therapies for major depression, as described in the clinical guidelines, are based on results from randomised controlled trials (RCTs). So far, it is not known to what extent results of RCTs on major depression can be generalised to 'real life' clinical practice.

**AIM** To compare treatment results for major depression from RCTs (efficacy) and results from daily practice (effectiveness); furthermore, to assess to what extent eligibility criteria and (un)intended selection by recruitment procedures influences treatment outcomes in daily practice.

**METHOD** In a 'real life' patient population (n = 1653) suffering from major depression (established by the MINIplus) and assessed in routine outcome monitoring at baseline, we explored how many patients met the eligibility criteria for antidepressant and psychotherapy efficacy trials. Furthermore we explored to what extent RCT participants differed in socio-demographic and socio-economic status from 'daily practice' patients. 626 of the ROM patients had at least one follow-up assessment. In this follow-up group we compared the treatment outcome (assessed by the MADRS and BDI-II) to the results of 15 meta-analyses of RCTs. We also explored to what extent patient selection based on eligibility criteria and socio-demographic/socio-economic status influenced treatment outcome.

**RESULTS** Remission percentages (21-27% in ROM versus 34-58% in RCTs) and effect sizes (0.85 in ROM versus 1.71 in RCTs, within-group data) were lower in daily practice than in RCTs. ROM patients differed from RCT participants in many disease-specific and socio-economic features. These differences are due to patient selection in RCTs. However, the influence of patient selection based on eligibility criteria and socio-demographic differences in treatment outcome were very modest (explained variances 1-11%).

**CONCLUSION** Treatment success for major depression is lower in daily practice than in RCTs and 'real life' patients differ in many features from RCT participants. However, these differences cannot explain the difference between efficacy and effectiveness. The generalisability of the results of depression trials to daily practice might not be jeopardised by the use of eligibility criteria and recruitment procedures to the extent suggested in earlier research.

TIJDSCHRIFT VOOR PSYCHIATRIE 57(2015)8, 579-587

**KEY WORDS** depression, daily practice, efficacy, effectiveness, randomised controlled trial (RCT), routine outcome monitoring (ROM)