

## Benchmarken is ‘werk-in-uitvoering’

E. BLIJD-HOOGEWYS, R. VAN DYCK, P. EMMELKAMP, N. MULDER,  
R. OUDE VOSHAAR, G. SCHIPPERS, P. SPINHOVEN, R. VERMEIREN

**ACHTERGROND** In het maartnummer van dit tijdschrift uitte een groep hoogleraren kritiek op de methodiek van de Stichting Benchmark GGZ (SBG). Deze kritiek betreft de vermeende doelstellingen (prestatiebekostiging in de ggz) en methodologische kritiekpunten, zoals confounding, bias en lage sensitiviteit van meetinstrumenten.

**DOEL** Nadere toelichting geven over SBG en de rol van de Wetenschappelijke Raad (WR) en over de maatschappelijke context waarbinnen ROM en benchmarken zich ontwikkelen en onze visie geven op de methodologische kritiekpunten.

**METHODE** Verhelderen van de SBG-methodiek en de positie van de WR en de huidige stand van zaken betreffende ROM en benchmarken.

**RESULTATEN** Wij onderkennen dat er methodologische problemen zijn. Ook de WR waarschuwt voor het trekken van voorbarige conclusies op basis van gegevens waarvan de validiteit nog ter discussie staat. Ervaring op het vlak van ROM en benchmarken voor de ggz ontbreekt nagenoeg en moet ontwikkeld worden. Daar staat tegenover dat ROM-gegevens en onderlinge vergelijking tot verbetering van zorg kunnen leiden, mits deze goed wordt uitgevoerd. In de SBG-methodiek worden de gesignaleerde knelpunten op zijn minst ten dele ondervangen. Zo wordt er niet alleen gekeken naar symptoomreductie, maar ook naar functioneren en kwaliteit van leven.

**CONCLUSIE** De WR bepleit een wetenschappelijk gevalideerd benchmarksysteem voor de ggz. De mate waarin benchmarkgegevens wat zeggen over kwaliteit en/of de vergelijking tussen instellingen valide is, is nauwelijks onderzocht. Een landelijke ROM-systematiek is noodzakelijk en de huidige initiatieven bieden mogelijkheden om hiernaar onderzoek te doen, bij voorkeur in samenwerking van wetenschappers en professionals.

[TIJDSCHRIFT VOOR PSYCHIATRIE 54(2012)12, 1031-1038]

**TREFWOORDEN** benchmarken, routine outcome monitoring, Stichting Benchmark GGZ

Stichting Benchmark GGZ (SBG) is in 2010 opgericht om als onafhankelijke partij uitvoering te geven aan benchmarking in de ggz. Dit volgde uit het bestuurlijk akkoord tussen GGZ Nederland en Zorgverzekeraars Nederland, dat als doel stelde routine outcome monitoring (ROM) op meerdere niveaus als hulpmiddel te gebruiken voor het bevorderen van de kwaliteit van de zorg en de transparantie van de geestelijke gezondheidszorg

in Nederland. ROM houdt in dat periodiek de toestand van de patiënt wordt gemeten om inzicht te krijgen in de voortgang van de behandeling (van Hees e.a. 2011). Het bestuurlijk akkoord stelt dat hulpverleners ROM-gegevens gaan gebruiken om waar nodig de behandeling en de begeleiding bij te sturen. De gegevens die zijn verzameld, dienen ook ter beschikking te worden gesteld aan een derde partij, SBG, die deze in geaggregeerde vorm

zal gebruiken om te benchmarken. Zorgverzekeraars, zorgaanbieders en patiënten zien dit laatste als een manier om door onderlinge vergelijking beter zicht te krijgen op wie bij welke patiëntengroep bovengemiddelde prestaties weet te realiseren.

De Wetenschappelijke Raad (WR) van SBG is conform de recent vastgelegde statuten belast met het toezicht op het wetenschappelijke kader waarbinnen SBG opereert. De WR ziet erop toe dat SBG voor zover mogelijk gevalideerde en/of bewezen methoden en technieken gebruikt voor onderzoek naar therapie-uitkomsten en benchmarking. Hiertoe mag de WR proactief en reactief, gevraagd en ongevraagd, (bindend) advies uitbrengen aan het bestuur en de directie van SBG. De WR komt vier keer per jaar bijeen en daarnaast wordt de WR tussentijds geconsulteerd wanneer daar aanleiding toe is vanwege vragen die bij SBG leven.

#### KRITIEK OP SBG

In een recent essay (van Os e.a. 2012) veroordeelden hoogleraren en hoofddocenten psychiatrie de activiteiten van SBG als misbruik van ROM voor pseudowetenschap. Ze veggen de verschillende doelstellingen van ROM, benchmarken en prestatiebekostiging op één hoop en ze zetten benchmarking weg als een bezuinigingsmaatregel. Hun stelling is dat benchmarking op basis van ROM tussen afdelingen en instellingen geen zin heeft, vanwege vertekening (*confounding*), bias, lage sensitiviteit en belangenconflicten. De eerste drie punten van deze stelling zijn te beschouwen als belangrijke methodologische problemen, die ook de WR van SBG ziet. Het vierde bezwaar is van een andere orde.

In hun essay roepen de auteurs op tot krachtig verzet tegen SBG, het gezamenlijke initiatief van de zorgverzekeraars en GGZ Nederland. Als dit verzet succesvol zou zijn, 'ontstaat de ruimte om een wetenschappelijk gevalideerd benchmarkingsysteem te ontwerpen' in de ggz. Deze stellingname impliceert dat Van Os e.a. de eerder gesignaleerde methodologische problemen oplosbaar achten, maar SBG

niet als een neutraal en onafhankelijk instituut beschouwen.

#### WETENSCHAPPELIJKE RAAD SBG

De oorspronkelijke leden van de WR van SBG (Emmelkamp, Spinhoven en Van Dyck) hebben in de fase dat benchmarken binnen de ggz nog uitsluitend een project van het vroegere Kenniscentrum Zorg Nederland (KZN) was, erop aangedrongen dat dit omgezet moest worden naar een gezamenlijk project van verzekeraars en aanbieders. Dit was ook het streven van de samenwerkende partijen binnen het project ROM ggz, een initiatief van GGZ Nederland, in samenwerking met het Nederlands Instituut van Psychologen (NIP), het Landelijk Platform GGz (LPGGZ) en de Nederlandse Vereniging voor Psychiatrie (NVVP). Dit heeft in 2009 geresulteerd in het bestuurlijk akkoord tussen GGZ Nederland en Zorgverzekeraars Nederland en de oprichting van SBG in 2010.

Na de oprichting van SBG is vanuit de WR SBG en de NVVP sterk gepleit voor een slagvaardige WR met een voor het veld representatieve samenstelling. SBG heeft dit gehonoreerd en heeft in januari 2012 een uitgebreide WR geïnstalleerd die meer expertisegebieden vertegenwoordigt en bindende adviezen kan geven. Dit maakt dat de NVVP het streven ondersteunt van SBG om te komen tot transparantie in de zorg door het ontwikkelen van een valide manier van benchmarking. Dit heeft het bestuur van de NVVP samen met de kernhoogleraren Psychiatrie in een brief (d.d. 25 mei 2012) aan al haar leden kenbaar gemaakt en is nog eens bevestigd in de nadruk op het stimuleren van transparantie in de ggz in het convenant tussen zorgaanbieders, zorgverzekeraars, patiënten en de overheid, dat 16 juni jl. getekend werd. De leden van de WR willen aan het benchmarken bijdragen door vanuit een wetenschappelijk kader de ontwikkelingen te sturen. Het statutair vastgelegde mandaat laat toe om onafhankelijk te adviseren over de wetenschappelijke mogelijkheden en onmogelijkheden van kwaliteitsmeting.

## GEDEELDE ZORGEN

Als leden van de WR van SBG delen wij een aantal zorgen van Van Os e.a. (2012). Meerdere van de door hen opgevoerde argumenten zien wij zelf ook als methodologische knelpunten. Tegelijkertijd zijn we ervan overtuigd dat gegevens afkomstig uit ROM ons in staat kunnen stellen om tot verbetering van zorg te komen door onderlinge vergelijking (tussen en binnen instellingen), mits deze goed wordt uitgevoerd. Deze ontwikkeling past in een internationale trend (zie voor het Verenigd Koninkrijk: Roland 2004; voor de VS: Institute of Medicine 2006; voor Australië: Trauer 2010). Ook in Nederland zijn er verschillende voorbeeldprojecten uit de somatische gezondheidszorg die veelbelovende resultaten laten zien, bijvoorbeeld Nationale Intensive Care Evaluatie (NICE; van der Veer e.a. 2011); Dutch Institute for Clinical Auditing (DICA 2012). Het streven naar transparantie over behandeluitkomst in de ggz is ook een logische vervolgstap in de ontwikkeling van prestatie-indicatoren, zoals die door Zichtbare Zorg (ZiZo) ter hand is genomen. Zie bijvoorbeeld Berg e.a. (2005) over prestatie-indicatoren voor ziekenhuiszorg.

We zijn ons ervan bewust dat de uitrol van ROM en benchmarken in Nederland een unieke ontwikkeling is, alleen al vanwege de ambitie om op landelijke schaal gegevens te gaan verzamelen. Dit vraagt om uniformiteit, waarbij keuzes gemaakt moeten worden betreffende ROM (o.a. meetinstrumenten en -momenten) en benchmarken (o.a. operationaliseren van behandeluitkomst en verzamelen van relevante casemixvariabelen). Deze onderneming vergt nader onderzoek naar de onderliggende methodiek van ROM en benchmarken. SBG weet dat ze dit niet alleen kan doen en de WR hoopt dit samen met de academische centra en de schrijvers van het essay te doen. We reageren in het navolgende op verschillende onderdelen van het essay.

## METHODOLOGISCHE BEZWAREN

*Bias of vertekening*

De auteurs waarschuwen voor vertekening van de resultaten als het aan de zorgaanbieders zelf wordt overgelaten om ROM-gegevens te verzamelen, gekoppeld aan de vrees die daar bestaat door zorgverzekeraars op resultaat afgerekend te worden. Ook wij zien dit gevaar, internationaal bekend als *gaming*: het bespelen van de methodologie om betere resultaten te verkrijgen (Bevan & Hood 2006; Bilsker & Goldner 2002). We merken op dat het gevaar van selectief includeren of selectief nameten beperkt wordt door de ambitie van zorgaanbieders om een zo hoog mogelijke ROM-respons te behalen. Mogelijke overschatting van de bereikte behandeluitkomsten wordt verder tegengegaan door het gebruik van gestandaardiseerde instrumenten, die de patiënt zelf invult. Echter, de WR erkent ook dat het onvoldoende duidelijk is wat het effect is van eventuele beloning/straf (door de zorgverzekeraar) gekoppeld aan benchmarking op de ROM-uitkomsten. Amerikaanse ervaringen met *pay-for-performance* programma's geven aan dat beloning niet automatisch tot een verhoging van de kwaliteit van zorg leidt (Rosenthal & Dudley 2007).

Het is voorts niet ondenkbaar dat instellingen bijvoorbeeld geen ROM meer zouden toepassen bij cliënten van wie ze vermoeden dat de uitkomst van de behandeling tegenvalt. De WR is er zich bewust van dat er daarnaast nog tal van andere manieren bestaan om de resultaten te bespelen. Daarom ziet de WR de huidige fase ook als een opstartfase, waarin scherp op bronnen van vertekening moet worden gelet. Immers, een valide vergelijking van instellingen is pas mogelijk als dergelijke bias in de resultaten voldoende valt uit te sluiten. De WR pleit dan ook voor grote terughoudendheid bij onderlinge vergelijking van uitkomsten van instellingen: het verhaal achter de data is in deze opstartfase het belangrijkste. Vertrouwen in de zeggingskracht van de gegevens die SBG teruggeeft aan het veld is er niet op voor-

hand. Dit dient met de tijd te groeien. De WR ziet het ook als zijn taak om te adviseren over de zeggingskracht van de gegevens, en dus de beperkingen wat betreft de financiële consequenties die aan de uitkomst van benchmarken verbonden kunnen worden, zoals het afrekenen op behandelresultaat.

#### Confounding of vergelijken zonder casemixcorrectie

ROM is te beschouwen als observationeel onderzoek en onderscheidt zich daarmee van experimenteel onderzoek naar therapie-effect dat in de geneeskunde doorgaans uitgevoerd wordt in de vorm van een gerandomiseerde gecontroleerde trial (RCT). In een RCT onderzoekt men de werkzaamheid van een behandeling. Vanwege de daarbij toegepaste aselecte toewijzing van patiënten aan een experimentele conditie en een controleconditie kunnen we ervan uitgaan dat een verschil in behandelingsuitkomst tussen beiden groepen is toe te schrijven aan de experimentele behandeling en niet aan een (eventueel ongemeten) verschil tussen de groepen dat al op voorhand bestond. Invloed van versturende variabelen noemt men *confounding* van het resultaat. In observationeel onderzoek wordt niet aselect toegewezen, mist men dus deze belangrijke controle en kan een alternatieve verklaring voor verschillen in therapie-effect niet uitgesloten worden.

Observationeel onderzoek is hierdoor minder geschikt voor het beantwoorden van fundamentele wetenschappelijke vragen, maar dit onderzoek heeft toch veel waarde, vooral omdat het inzicht kan bieden in de effectiviteit van behandelingen in de klinische praktijk, waarvan de werkzaamheid in RCT's is aangetoond (*efficacy* versus *effectiveness*, bijv. Seligman 1995). Zo laat recent Nederlands onderzoek op basis van ROM-data zien dat de *effectiveness* van behandeling van lichte tot matige depressie in de dagelijkse praktijk geringer is dan de *efficacy* in RCT's (van der Lem e.a. 2012). Ook bleek dat – hoewel de meerderheid van patiënten in de dagelijkse praktijk niet in aanmerking zou komen voor een RCT – geschiktheid voor deelname aan een RCT hoegenaamd

geen voorspellende waarde had voor de effectiviteit van behandeling in de dagelijkse praktijk (van der Lem e.a. 2011).

Van een fundamenteel probleem bij gebruik van ROM-gegevens voor vergelijking op groepsniveau, waar Van Os e.a. (2012) gewag van maken, is geen sprake. Zulk gebruik van ROM-gegevens is observationeel onderzoek. Wel moeten we bij de interpretatie van resultaten van observationeel onderzoek extra voorzichtig zijn en een open oog houden voor *confounding*.

Vergelijken van de gemiddelde uitkomst van de behandeling zonder rekening te houden met potentiële *confounders* kan tot onjuiste conclusies leiden, doordat dergelijke versturende variabelen, zoals systematische verschillen in kenmerken van de patiënten van verschillende zorgaanbieders een uitkomst kunnen verzwakken of juist versterken. Op welke patiëntvariabelen er systematische verschillen zijn en in welke mate die variabelen voorspellende waarde hebben voor de behandeluitkomst is voor de ggz grotendeels onbekend (Garfield 1994; Kemp e.a. 2008; Riedel e.a. 2011). Verdere dataverzameling is nodig om in kaart te brengen wat in de praktijk de belangrijkste *confounders* zijn. Om dit te onderzoeken zijn grote databestanden nodig. Dit is een grote opgave, maar zou een aanzienlijke wetenschappelijke winst betekenen. Om tegemoet te komen aan eventuele verschillen in de casemix kunnen uitkomsten per subgroep van patiënten worden vergeleken (subgroepselectie). Op termijn moeten we nagaan of statistische casemixcorrectie haalbaar en wenselijk is (Hermann e.a. 2007). Of dit tot valide en betekenisvolle vergelijkingen zal leiden, is een van de vragen die volgens de WR ook nader onderzocht moeten worden.

#### Sensitiviteit

Van Os e.a. stellen dat de meetinstrumenten die door de expertgroepen worden aanbevolen voor ROM een onvolledig beeld geven en te weinig sensitief zijn om de uitkomst in kaart te brengen. Het is een misverstand te denken dat SBG klach-

tenlijsten en symptoomreductie automatisch als de beste maten beschouwt om voor alle probleemgebieden verbetering te meten. Ook is het niet juist dat SBG uitsluitend oog heeft voor verbetering of afname van symptomen. Meerdere stoornissen zijn immers chronisch, waarbij stabilisatie het hoogst haalbare is. In de methodiek van SBG wordt daar rekening mee gehouden en is er ook aandacht voor uitkomsten op andere gebieden, zoals functioneren en kwaliteit van leven.

Overigens beschouwt de WR de huidige methodiek van de SBG als een eerste stap en niet als definitief: 'SBG is werk in uitvoering'. Nader onderzoek naar de meest sensitieve en responsieve uitkomstmaten, waarmee we optimaal kunnen differentiëren tussen geslaagde en mislukte behandelingen, is essentieel voor de doelstellingen van SBG en dergelijk onderzoek wordt door SBG uitgevoerd (de Beurs 2010; de Beurs e.a. 2012). De ontwikkeling van een herstelmaat voor patiënten met ernstige psychiatrische aandoeningen wordt nauwlettend gevolgd.

De recente installatie van een uitgebreide WR (januari 2012) heeft onder andere tot doel om de expertise op de diverse meetgebieden te vergroten. Zo is bij veel patiënten de mate van functioneren in het dagelijks leven ook een belangrijke uitkomstmaat. Het ligt daarom voor de hand dat voor uiteenlopende patiëntencategorieën verschillende maten gekozen moeten worden. De door SBG gekozen methodiek voorziet hierin. Tevens is nog onvoldoende duidelijk of alle patiëntencategorieën zich lenen voor de ROM-praktijk of dat er uitzonderingen gemaakt dienen te worden (denk bijvoorbeeld aan de crisisdienst, de forensische zorg of de ziekenhuispsychiatrie). Ook dergelijke inzichten kunnen wetenschappelijke en klinische winst opleveren.

Daarbij is het niet zo dat de beleidsmakers de instrumenten voorschrijven. De huidige selectie van instrumenten is een afspiegeling van de huidige ROM-praktijk en is bepaald door het veld, in de expertgroepen van professionals ingesteld door GGZ Nederland. De selectie van meetinstrumenten loopt in de pas met internationale ontwikke-

lingen op het gebied van ROM en kwaliteitsmeting in de zorg. Zo maakt men in het Verenigd Koninkrijk en Australië gebruik van de 'Health of the Nation Outcome Scales' (HONOS) en in de VS wordt vaak de 'Outcome Questionnaire' (OQ-45) gebruikt bij ambulante curatieve zorg. Ten slotte kunnen instellingen alternatieve instrumenten voordragen. SBG zal deze opnemen als geaccepteerd instrument op basis van het kwaliteitsoordeel dat de WR hierover velt na advies te hebben ingewonnen van de expertraden voor ROM en benchmarken.

#### BELANGENCONFLICT?

De WR is het eens met Van Os e.a. dat het ongewenst is dat commerciële uitgevers geld verdienen aan het gebruik van meetinstrumenten die met publieke middelen zijn ontwikkeld. Zo komt het voor dat uitgevers beslag weten te leggen op de rechten van vragenlijsten terwijl dat niet de bedoeling is geweest van ontwikkelaars. Dit heeft er onder andere mee te maken dat er geen degelijk, niet-commercieel distributiesysteem van vragenlijsten bestaat. Het zou wellicht wenselijk zijn dat op te zetten, zeker nu ROM zo'n vlucht neemt in de ggz in Nederland.

Een goed initiatief op dit punt is dat de site van het Tijdschrift voor Psychiatrie een sectie Meetinstrumenten heeft (zie [www.tijdschriftvoorpsychiatrie.nl](http://www.tijdschriftvoorpsychiatrie.nl), onder meetinstrumenten). Hier wordt per test aangegeven of er een Nederlandstalige versie beschikbaar is en of deze gratis te downloaden is. Ook geeft de site links naar bron en vertaling.

Dat er vragenlijsten voorkomen op de lijst van geaccepteerde instrumenten van SBG waarvoor wel auteursrechten betaald moeten worden, is niet de keuze van de WR. Het zijn instellingen die erop hebben aangedrongen om instrumenten te accepteren waar zij ervaring mee hebben en die zij liever niet wilden inruilen voor een ander, gratis instrument. De WR wil stimuleren dat de dataverzameling er op termijn toe zal bijdragen dat een nieuwe batterij van instrumenten zonder auteursrechten ontwikkeld en gevalideerd kan

worden. Dit is bij uitstek een onderwerp waarin wij samenwerking willen met de academische centra, omdat dit wellicht al met huidig beschikbare databestanden verwezenlijkt kan worden.

#### KRACHTENVELD

De Nederlandse politiek heeft ervoor gekozen om veel verantwoordelijkheid voor de uitvoering van de zorg neer te leggen bij de zorgverzekeraars. Dit is een ontwikkeling waar de meningen verdeeld over zijn. Het belast de zorgverzekeraars met financiële verantwoordelijkheden, die zij tot op zekere hoogte zelf kunnen invullen. Om de stijgende zorgkosten te kunnen beheersen, zoekt men manieren om de prijs-kwaliteitverhouding van zorg in beeld te brengen. Wat betreft benchmarken zijn wij bekend met de berichten dat individuele zorgverzekeraars instellingen financieel onder druk zetten om ROM-gegevens hiervoor aan te leveren. Hoewel we enige druk vanuit de verzekeraars geenszins onwenselijk vinden, hebben wij daar zorgen over. Te grote druk en financiële afrekening dreigen immers ten koste te gaan van de kwaliteit van de ROM-gegevens en daardoor komt ook de kwaliteit van het benchmarken in het gedrang.

De hele ggz is sterk in beweging, niet enkel omwille van prestatiebekostiging, maar ook door bezuinigingen, de wens tot meer ambulant werken, et cetera. In deze ontwikkelingen is het nog niet duidelijk hoe kwaliteit van zorg het beste in beeld gebracht kan worden: door outcomemetingen, prestatie-, proces- of structuurindicatoren, of een combinatie van deze mogelijkheden. Stellige standpunten van beleidsmakers lopen vaak vooruit op feitenkennis. Ook op het gebied van ROM en benchmarken wordt maar al te vaak simplistisch gedacht en worden voordelen verondersteld die (nog) niet aangetoond zijn. Zo verkondigde minister Schippers onlangs dat ROM-gegevens op termijn een veel betere basis voor de financiering van de ggz bieden dan DBC's.

Er is in de gezondheidszorg een toenemende vraag naar transparantie over behaalde resultaten

(Porter & Teisberg 2006). Wij zijn ervan overtuigd dat we de ontwikkeling van kwaliteitssystemen voor de ggz, zoals benchmarken, niet aan politici en zorgverzekeraars mogen overlaten. Hun vraag dit te ontwikkelen is begrijpelijk. Het is vervolgens aan wetenschappers en professionals om dit op een wetenschappelijk verantwoorde wijze te laten plaatsvinden. Daarom verwelkomen we landelijke samenwerking met academische centra, teneinde vanuit gezamenlijke inspanningen en gedeelde visie te kunnen bijdragen.

#### CONCLUSIE

Evenals Van Os e.a. spreekt de WR zich uit voor het ontwikkelen van een wetenschappelijk gevalideerd benchmarksysteem voor de ggz. Gezien alle geschetste problemen en openstaande vragen delen we de overtuiging dat er nog een lange weg te gaan is. In de eerste plaats moeten basale methodologische knelpunten overwonnen worden, om te komen tot een deugdelijke ROM-methodiek als basis voor vergelijking en benchmarking. Pas in een latere fase kan geoordeeld worden of, in welke mate en op welke wijze benchmarken sturend kan zijn voor financiële besluitvorming (zie ook Berg e.a. 2005).

Op bestuurlijk niveau is besloten om SBG als onafhankelijk instituut voor de uitvoering van benchmarken voor de ggz op te richten. De mate waarin benchmarkgegevens wat zeggen over kwaliteit en/of de vergelijking tussen instellingen valide is, is niet bekend en nauwelijks onderzocht. Het ontwikkelen van een dergelijke landelijke systematiek biedt een uitgelezen mogelijkheid om dit te onderzoeken. SBG kan en wil graag bijdragen aan dit onderzoek en we hopen van harte dit in een brede samenwerking te kunnen doen, waarbij de kritische bijdrage van Van Os e.a. een aanzet kan zijn om hiertoe de handen ineen te slaan.

## LITERATUUR

- Berg M, Meijerink Y, Gras M, Goossensen A, Schellekens W, Haecck J, e.a. Feasibility first: developing public performance indicators on patient safety and clinical effectiveness for Dutch hospitals. *Health Policy* 2005; 75: 59-73.
- Bevan GWYN, Hood CHR. What is measured is what matters: targets and gaming in the English public health care system. *Public Administration* 2006; 84: 517-38.
- Bilsker D, Goldner EM. Routine outcome measurement by mental health-care providers: is it worth doing? *Lancet* 2002; 360: 1689-90.
- Beurs E de. De genormaliseerde T-score, een 'euro' voor testuitslagen. *Maandblad Geestelijke Volksgezondheid* 2010; 65: 685-96.
- Beurs E de, Barendregt M, Flens G, van Dijk E, Huijbrechts I, Meerding WJ. Equivalentie in responsiviteit van veel gebruikte zelfrapportage meetinstrumenten in de geestelijke gezondheidszorg. *Maandblad Geestelijke Volksgezondheid* 2012; 67: 259-64.
- Dutch Institute for Clinical Audit (DICA). Rapportages 2011. *Transparantie, keuzes en verbetering van zorg*. DICA; 2012. <http://www.clinicalaudit.nl/images/documenten/2012/jaar-rapportage%202011.pdf>
- Garfield SL. Research on client variables in psychotherapy. In: Bergin AE, Garfield SL, red. *Handbook of psychotherapy and behavior change* (4de druk), Oxford: John Wiley & Sons; 1994. p. 190-228.
- Hees S van, van der Vlist P, Mulder N, red. *Van weten naar meten: ROM in de GGZ*. Amsterdam: Boom; 2011
- Hermann RC, Rollins CK, Chan JA. Risk-adjusting outcomes of mental health and substance-related care: a review of the literature. *Harv Rev Psychiatry* 2007; 15: 52-69.
- Institute of Medicine. Committee on Crossing the Quality Chasm: *Adaptation to mental health and addictive disorders. Improving the quality of health care for mental and substance-use conditions: Quality Chasm Series*. Washington: National Academic Press; 2006.
- Kemp AH, Gordon E, Rush AJ, Williams LM. Improving the prediction of treatment response in depression: integration of clinical, cognitive, psychophysiological, neuroimaging, and genetic measures. *CNS Spectrums* 2008; 13: 1066-86.
- Lem R van der, van der Wee NJA, van Veen T, Zitman FG. The generalizability of antidepressant efficacy trials to routine psychiatric out-patient practice. *Psychol Med* 2011; 41: 1353-63.
- Lem R van der, van der Wee NJA, van Veen T, 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.
- Os J van, Kahn R, Denys D, Schoevers RA, Beekman AT, Hoogendijk WJ, e.a. ROM: gedragsnorm of dwangmaatregel? Overwegingen bij het themanummer over routine outcome monitoring. *Tijdschr Psychiatr* 2012; 54: 245-53.
- Porter M, Teisberg EO. *Redefining health care*. Cambridge: Harvard Business School Publishing; 2006.
- Riedel M, Maller HJ, Obermeier M, Adli M, Bauer M, Kronmuller K, e.a. Clinical predictors of response and remission in inpatients with depressive syndromes. *J Affect Disord* 2011; 133: 137-49.
- Roland M. Linking physicians' pay to the quality of care - A major experiment in the United Kingdom. *N Engl J Med* 2004; 351: 1448-54.
- Rosenthal MB, Dudley R. Pay-for-performance: will the latest payment trend improve care? *JAMA* 2007; 297: 740-4.
- Seligman MEP. (1995). The effectiveness of psychotherapy: the Consumer Reports study. *American Psychologist* 1995; 30: 965-74.
- Trauer T. *Outcome measurement in mental health*. Cambridge: Cambridge University Press; 2010.
- Veer SN van der, de Vos ML, Jager KJ, van der Voort PH, Peek N, Westert GP, e.a. Evaluating the effectiveness of a tailored multifaceted performance feedback intervention to improve the quality of care: protocol for a cluster randomized trial in intensive care. *Implement Sci* 2011; 6: 119.

## AUTEURS

ELS BLIJD-HOOGEWYS, psycholoog en teamleider Inhoudelijke Zaken Autismeteam Noord-Nederland, Lentis, Groningen.

RICHARD VAN DYCK, emeritus hoogleraar Psychiatrie VUmc, Amsterdam, EMGO+ Institute for Health and Care Research, GGZ Altrecht, visitatiecommissie Stichting TOPGGZ.

PAUL EMMELKAMP, hoogleraar Klinische Psychologie, Programmagroep Klinische Psychologie, Universiteit van Amsterdam.

NIELS MULDER, hoogleraar Openbare Geestelijke Gezondheidszorg, Erasmus MC, afd. Psychiatrie, Rotterdam en psychiater en onderzoeker bij BavoEuropoort.

RICHARD OUDE VOSHAAR, hoogleraar Ouderenpsychiatrie, Universitair Centrum Psychiatrie, UMC Groningen.

GERARD SCHIPPERS, hoogleraar Verslavingszorg en Zorg-evaluatie, afd. Psychiatrie, Amsterdam Institute for Addiction

Research Academisch Medisch Centrum Universiteit van Amsterdam, tevens Arkin GGZ.

PHILIP SPINHOVEN, hoogleraar Klinische Psychologie, afd. Klinische Psychologie, Universiteit Leiden.

ROBERT VERMEIREN, hoogleraar Kinder- en Jeugdpsychiatrie, Curium, Leiden en hoogleraar Forensische Jeugdpsychiatrie, VUmc, Amsterdam.

Correspondentieadres: Edwin de Beurs, inhoudelijk directeur Stichting Benchmark GGZ, Rembrandtlaan 46, 3723 BK Bilthoven.

E-mail: edwin.debeurs@sbggz.nl

Strijdige belangen: de auteurs zijn lid van de Wetenschappelijke Raad van de Stichting Benchmark GGZ.

Het artikel werd voor publicatie geaccepteerd op 25-9-2012.

## SUMMARY

Benchmarken is 'work in progress' – E. Blijd-Hoogewys, R. van Dyck, P. Emmelkamp, N. Mulder, R. Oude Voshaar, G. Schippers, P. Spinhoven, R. Vermeiren –

**BACKGROUND** In the March edition of this journal a group of professors criticised the methods adopted by the SBG (Benchmark foundation of the Dutch Mental Health Services). The critical comments relate to the alleged objectives ('pay-for-performance' in the Dutch Mental Health Services) and the methods involving confounding, bias and the use of insufficiently sensitive instruments.

**AIM** To provide more detailed information about the SBG and the role of the scientific council (WR) and about the social context within which ROM and benchmarking are developing, and to give our views on the methodological criticisms.

**METHOD** We elucidate the methods of the SBG and the role of the WR and we clarify the ways in which ROM and benchmarking are currently being used.

**RESULTS** We recognise that there are methodological problems. The WR too warns against jumping to conclusions on the basis of data whose validity is still under discussion. The Dutch Mental Health Services have practically no experience of ROM and benchmarking. The Dutch Mental Health Services need to gain more experience in this field. On the other hand, it is true that ROM data and comparisons between institutions can lead to improvements in the standard of care, provided these procedures are carried out carefully. In the SGB methods the aforementioned problems have been removed, at least partially. The SBG is concerned not only with symptom reduction but also with the functioning of the individual and with his or her quality of life.

**CONCLUSION** The WR believes that the Dutch Mental Health Services should operate a scientifically validated benchmark system. So far, there has been hardly any research into whether benchmark data really say anything about quality and/or whether comparisons between institutions are valid. A national ROM-system is needed. Current initiatives are creating ways in which these matters can be investigated, preferably in conjunction with scientists and professionals.

[TIJDSCHRIFT VOOR PSYCHIATRIE 54(2012)12, 1031-1038]

**KEY WORDS** benchmark, routine outcome monitoring, Netherlands