

ROM: gedragsnorm of dwangmaatregel?

Overwegingen bij het themanummer over routine outcome monitoring

J. VAN OS, R. KAHN, D. DENYS, R.A. SCHOEVERS, A.T.F. BEEKMAN, W.J.G. HOOGENDIJK, A.M. VAN HEMERT, P.P.G. HODIAMONT, F. SCHEEPERS, PH.A.E.G. DELESPAUL, A.F.G. LEENTJENS

ACHTERGROND De kosten van de ggz zijn mogelijk meer gestegen dan die van andere sectoren in de gezondheidszorg. In een poging deze kosten te reduceren, proberen beleidsmakers enerzijds een selectieve financiële drempel op te werpen voor mensen die psychische zorg vragen, en anderzijds een systeem van prestatiebekostiging in te voeren op basis van data verkregen met routine outcome monitoring (ROM). Er dreigt misbruik van klinische data die door hulpverleners zijn verzameld, nu zorgverzekeraars en GGZ Nederland instellingen willen verplichten om data te leveren aan de Stichting Benchmark GGZ (SBG). De SBG moet het mogelijk maken om 'prestaties' te vergelijken tussen instellingen.

DOEL Nadere bestudering van het SBG-project met een focus op wetenschappelijke validiteit.

METHODE Kwalitatieve review en analyse.

RESULTATEN Beschouwing leert dat benchmarken op basis van ROM-data in Nederland, zoals voorgesteld door SBG en zorgverzekeraars, wetenschappelijke noch medisch-ethische toetsing kan doorstaan. De voornaamste problemen zijn: bias, confounding, mix van gebruikte instrumenten, lage sensitiviteit en het vergroten van de kosten in de ggz door onnodige en omstreden copyrightuitgaven aan commerciële partijen.

CONCLUSIE Wij ontraden invoering van deze dure en onwetenschappelijke bureaucratische maatregel in de gestaag uitdijende management-registratie-controlecyclus van de ggz, zeker daar nog niet alle betrokken partijen hun fiat hebben gegeven. Dan ontstaat de ruimte om een wetenschappelijk gevalideerd benchmarkingsysteem te ontwerpen dat kan bijdragen aan efficiënter gebruik van schaarse middelen.

[TIJDSCHRIFT VOOR PSYCHIATRIE 54(2012)3, 245-253]

TREFWOORDEN benchmarking, biomedische ethiek, routine outcome monitoring/measurement (ROM)

In het boeiende themanummer over routine outcome monitoring (ROM) van het Tijdschrift voor Psychiatrie wordt terecht gesteld dat ROM primair een instrument voor de behandeling is en dat het gebruik ervan voor benchmarking (gebruik van ROM voor vergelijkend onderzoek naar behandelresultaten) nog veel discussie oproept (Keet e.a.

2012). Noom e.a. (2012) bespreken een aantal van de problemen bij benchmarking, zoals casemix en de keuze van instrumenten. Nugter en Buwalda (2012), alsmede Mulder en Kortrijk (2012) bespreken het probleem van mogelijke bias. Alle auteurs stellen dat er oplossingen gevonden moeten worden voor de aanzienlijke problemen rond bench-

marking en wijzen erop dat er nog sprake is van een ontwikkelingsproces.

Na het lezen van het themanummer blijft de lezer zitten met een prangende vraag. Waarom is eigenlijk opdracht gegeven tot een kostbare, grootscheepse, nationale dataverzameling, in de vorm van de activiteiten van Stichting Benchmark GGZ, om instellingen te vergelijken op basis van data die blijkbaar niet vergelijkbaar zijn? Had een goede medisch-ethische commissie dit ooit laten passeren? Deze vragen zijn des te klemmender doordat de zorgverzekeraars openlijk hun intentie uitspreken in het themanummer (Laane & Luijk 2012) om benchmarkdata te gebruiken voor kwaliteitsconcurrentie met potentieel belangrijke consequenties voor zorg en werkgelegenheid.

ROM ALS KLINISCHE ACTIVITEIT HEEFT ZIN

Het is een goede zaak dat psychiaters en psychologen het ziektebeloop en het behandelresultaat van individuele patiënten kwantitatief vastleggen; dit past bij de huidige *state-of-the-art* benadering van de moderne psychiatrische zorg. Als iedereen dit bij alle patiënten doet, kunnen we inderdaad spreken van ROM. Het is uiteraard aan de beroepsgroep om de instrumenten te kiezen die het beste aansluiten bij de evaluatie van hun klinische werk. Men kan zich ook voorstellen dat de ROM-data bij onderzoek, zonder dat er door zorgverzekeraars economische consequenties aan verbonden worden, op groepsniveau *exploratief* kunnen worden gebruikt in een *pre-post* vergelijking om bijvoorbeeld de zorg van een welomschreven patiëntengroep te evalueren (Bak e.a. 2007).

Het wordt echter problematisch als (1) de beleidsmaker de beroepsgroep wil voorschrijven hoe zij deze activiteiten moet uitvoeren en met welke instrumenten, (2) de data door de beleidsmaker worden gebruikt voor niet valide (en dus onethische) vergelijkingen van 'prestaties', (3) de zorgverzekeraar hiermee beoordeelt of behandelingen door de beroepsgroep met 'kwaliteit' worden uitgevoerd, (4) deze vergelijkingen geheel of

gedeeltelijk de basis vormen van de financiering van de zorg, en (5) de beleidsmaker en de zorgverzekeraar hiermee voorbijgaan aan het stelsel van gedifferentieerde en getrapte zorg zoals gebruikelijk is in Nederland.

HET PROBLEEM: MISBRUIK VAN ROM

Het is verdedigbaar de kwaliteit van de geleverde zorg te willen verbeteren. Het is eveneens verdedigbaar de uitkomst van de geleverde zorg op een wetenschappelijke manier te willen evalueren. Dit heeft echter geleid tot een vreemd huwelijk tussen zorgverzekeraars en GGZ Nederland, die besluiten tot een verplichte dataverzameling waarvan de analyse tot nuttige inzichten moet gaan leiden. Miljoenen euro's worden uit de patiëntenzorg gehaald om een nationaal benchmarksysteem op te zetten, ondergebracht bij Stichting Benchmark GGZ (SBG), in de veronderstelling dat beleidsmakers kwaliteitsvergelijkingen tussen instellingen kunnen uitvoeren. Maar de doelstellingen van dit 'huwelijk' zijn niet met elkaar te verzoenen.

Het verzamelen, analyseren en interpreteren van kwantitatieve data behoren tot het domein van de wetenschap, die zeer specifieke kwaliteitseisen stelt, vooral als het gaat over meten en vergelijken van behandelresultaten. Het is in Nederland goed gebruik om deze eisen in eerste instantie te laten toetsen door een medisch-ethische commissie, om misstanden te voorkomen. Het ethische argument hier is niet gering: de zorgverzekeraars immers geven aan dat de data gebruikt zullen worden voor 'concurrentie op kwaliteit' met potentieel grote consequenties voor patiënten, hulpverleners en instellingen (Laane & Luijk 2012).

De SBG stelt op haar website (<http://www.sbggz.nl/Benchmarken.aspx>): 'Benchmarken is voor zorgaanbieders een middel om de eigen prestaties te vergelijken met de prestaties van gelijksoortige aanbieders'. Als we verder lezen, blijkt 'prestatie' te refereren aan 'behandelresultaat'. Laane en Luijk (2012) stellen namens de zorgverzekeraars in het thema-

nummer: 'ROM biedt voor behandelaren een kwaliteitsverbetering en voor verzekeraars een middel om uitspraken te kunnen doen over kwaliteit van de ggz op geaggregeerd niveau'. Zijn al deze claims wel juist, en zouden ze een medisch-ethische toetsing hebben kunnen doorstaan?

We zullen in dit artikel aantonen dat benchmarking op basis van ROM, met andere woorden, het vergelijken van de behandelresultaten tussen instellingen of afdelingen binnen instellingen, geen zin heeft. Meer nog, het is wetenschappelijk niet gefundeerd, niet gevalideerd en niet ethisch door een aantal onoverkomelijke methodologische problemen: confounding, bias, instrumentmix en lage sensitiviteit. Ook versterkt deze vorm van ROM 'copyrightwoeker'.

METHODOLOGISCHE PROBLEMEN

Confounding

Een van de belangrijkste problemen met benchmarking is de invloed van de *casemix*: de vertroebeling (*confounding*) van vergelijking van uitkomsten, prestaties of behandelresultaten door variabelen zoals diagnose, somatische aandoeningen, comorbiditeit, placebo-effecten, ervaring van de therapeut, aard van de therapie, leeftijd, etnische groep, geslacht, sociaaleconomische status, regionale karakteristieken, voorgeschiedenis, duur van de behandeling, effecten van de bredere sociale omgeving, et cetera. Dit is niet onbelangrijk: in de ggz zijn demografische variabelen even belangrijk voor de uitkomst als de diagnose – of zelfs belangrijker dan die laatste (Kisely e.a. 2000); zie hierover ook Mulder en Kortrijk (2012) in het themanummer ROM. Het doen van vergelijkende uitspraken over effecten die verder gaan dan hypothese genererende exploratie (Bak e.a. 2007) heeft alleen zin indien men deze 'confounding' variabelen kan uitsluiten, bijvoorbeeld door randomisatie. Zoals aangegeven in het themanummer, is hier nog geen oplossing voor (Noom e.a. 2012).

Het is niet rationeel om een grootscheepse, doorlopende dataverzameling te beginnen zonder idee hoe om te gaan met confounding (die speelt zowel binnen als tussen instellingen). Sterker nog, enige wetenschappelijke reflectie leert dat er geen oplossing is voor het casemixprobleem omdat we met de huidige ROM niet eens een fractie van alle relevante vertroebelende factoren op valide wijze kunnen verzamelen. Ook kunnen onbekende confounders bij uitkomstvergelijkingen een grotere rol spelen dan bekende (Clayton & Hills 1993; Hennekens & Buring 1987). Bovendien zijn statistische correcties bij vergelijkingen tussen personen, bij gebruik van een model dat veel variabelen bevat die in meer of mindere mate fundamenteel met elkaar samenhangen, zeer problematisch en niet goed interpreteerbaar (Miller & Chapman 2001).

Bias

Bias duidt op vertekening van de resultaten door een onnatuurlijke methodologische 'kortsluiting' tussen blootstelling en uitkomsten, in dit geval de behandeling en de uitkomst van die behandeling, waarvoor niet kan worden gecorrigeerd in de analyses (Hennekens & Buring 1987), zie ook Mulder en Kortrijk (2012) hierover in het themanummer ROM. Als een observationeel onderzoek ook maar een klein risico loopt op besmetting door bias, is het raadzaam de uitvoering uit dan wel af te stellen tot het biasprobleem is opgelost – zo dat überhaupt mogelijk is.

De potentiële bias bij SBG-benchmarking is evident: degenen die de data verzamelen, zijn ook degenen die door de verzekeraar financieel kunnen worden gekort als de resultaten achterblijven bij het landelijk gemiddelde. Een uitnodiging tot vertekening dus. De vraag doet zich wederom voor: waarom een kostbare dataverzameling beginnen waarvan van tevoren al vaststaat dat het risico op bias onacceptabel groot is, en die dus in wetenschappelijke termen niet valide is?

Instrumentenmix

Delespaul (2010) wees reeds op het risico van het gebruik van zeer verschillende instrumenten met zeer verschillende sensitiviteit en kwaliteit bij verschillende instellingen voor verschillende patiëntengroepen van de ROM in Nederland. Indien deze gegevens gebruikt worden voor onderlinge vergelijkingen, ontstaan psychometrische problemen die niet kunnen worden opgelost met een simpele T-score. Men heeft voor deze exercitie immers niet-vertekend referentiemateriaal nodig om T-scores te bepalen, dat wil zeggen randpopulatiegegevens, niet de eigen SBG-data.

Om te kunnen vergelijken is een klassiek psychometrische 'ijking' van het instrument noodzakelijk. Lokale ijking is voldoende voor lokale beslissingen, maar niet voor externe vergelijkingen. Men kan opteren om in ROM-consortia de data te poolen, maar een voorwaarde blijft dat men dezelfde instrumenten in verschillende instellingen moet gebruiken. Ook moeten de instellingen een representatieve verdeling vormen van de ggz in Nederland. Ijking gebeurt steeds in relatie tot een referentiegroep. Een willekeurige steekproef van de Nederlandse bevolking, maar ook een aselecte steekproef van mensen die met een angstprobleem in de ggz zorg zoeken in Nederland, zijn legitieme referentiegroepen. De eigen caseload of instellingsdatabase is dat niet.

We zijn ons ervan bewust dat men in de praktijk niet te puristisch moet zijn. Maar benchmarking is een serieuze zaak, met potentieel gevolgen voor economie en gezondheidszorg, en moet dus wetenschappelijk exact zijn.

Lage sensitiviteit

Psychische problemen zijn onlosmakelijk verbonden met lichamelijke aandoeningen omdat mentale stoornissen vaak het gevolg zijn van somatische ziekten. In de ziekenhuispsychiatrie hangen psychische behandeluitkomsten bijna zonder uitzondering samen met de ernst en de prognose van de lichamelijke aandoening. Zonder

informatie over de uitkomsten van specifieke lichamelijke aandoeningen en de behandeling daarvan zijn veranderingen in psychische uitkomsten niet goed interpreteerbaar. Bij benchmarken zoals beschreven door de SBG tracht men psychiatrische diagnose en het behandelresultaat vast te leggen, maar niet de somatische diagnose en het behandelresultaat.

Hieruit lijkt een visie te spreken dat psychisch lijden een afzonderlijke beleving is die in isolatie kan worden beschouwd van de lichamelijke component. Dit leidt tot een fundamenteel wetenschappelijk meetprobleem: men wil een gezondheidsuitkomst beschrijven, maar gebruikt hiervoor een incompleet instrument, waardoor de uitkomst niet-sensitief in kaart wordt gebracht. Dit bezwaar geldt in het bijzonder voor de ziekenhuispsychiatrie.

De SBG stelt op de website, bij het zorgdomein 'volwassenen kort': *'primair doel van behandeling is reductie van klachten/symptomen'*. Teneinde deze symptoomreductie te kunnen meten stelt de stichting een aantal monitorinstrumenten voor. Dit is een naïeve gedachte die niet zonder meer valide is en in ieder geval nuancering behoeft.

Ten eerste toont Hafkenscheid (2012) in het themanummer op treffende wijze aan dat interpretatie van ROM-data een subjectieve aangelegenheid is en dat *'hogere scores op de monitorinstrumenten niet noodzakelijk positief en lagere scores niet per se negatief zijn'*.

Een tweede, daaraan gerelateerd, argument is dat het in de praktijk bij de behandeling van psychische klachten vaak meer gaat om symptoomacceptatie (die in de huidige opzet van de ROM niet gemeten wordt) dan om symptoomreductie. Somberheid kan persisteren, maar acceptatie verlicht het lijden; fobische angst blijft aanwezig, maar de persoon laat zich er minder door beïnvloeden. Bovendien kan symptoomreductie een toename in klachten induceren; zo kan een afname van wanen en hallucinaties tijdens behandeling met klassieke antipsychotica leiden tot extrapiramidale verschijnselen en een toename van dysforie of apathie.

Ook kan het zijn dat verbetering niet door de patiënt, maar wel door de omgeving wordt waargenomen. In dit geval is zelfrapportage van symptoomreductie niet sensitief. Hoewel men kan beargumenteren dat symptoomacceptatie zonder symptoomreductie kan worden geïdentificeerd als er gelijktijdige verbetering in de kwaliteit van leven is, kan men dit niet altijd aannemen. De zelfrapportage van kwaliteit van leven, zoals voorgesteld bij ROM, vertoont immers een grote en complexe overlap met de symptomen, zoals depressie (de Leval 1999; Moore e.a. 2005); deze kan dus niet als onafhankelijke maat voor kwaliteit van leven gelden.

De grote afwezigheid in de huidige opzet van ROM is een valide maat van herstel, want dit is de uitkomst die door patiënten als de meest valide wordt beschouwd.

Copyright versus open source

In de wetenschap is er al jaren aandacht voor de rol van de grote commerciële uitgeverij en gerelateerde partijen, die zich vooral toespit op de vraag hoe het verdienmodel van deze partijen interfereert met de vrije disseminatie van wetenschappelijke kennis en het vrije gebruik van wetenschappelijke instrumenten (Newman & Feldman 2011). Een belangrijk thema in dit debat zijn de prijzen die commerciële partijen berekenen voor het gebruik van symptoomlijstjes en andere instrumenten die tot 40 jaar geleden zijn ontwikkeld, vrijwel zonder uitzondering met publiek geld. Velen zijn van mening dat de praktijk om tot 3 euro per afname te moeten betalen voor een dergelijk lijstje kenmerken heeft van woeker met als gevolg dat geld onterecht wordt weggesluisd uit de gezondheidszorg en de wetenschap.

De vaak gehoorde verdediging van commerciële partijen dat deze inkomsten noodzakelijk zijn voor 'vertaling', 'normering' en 'ontwikkeling' is niet valide, omdat de kosten hiervan niet in verhouding staan tot de cumulatieve inkomsten. Gezondheidszorg- en wetenschapsinstellingen wer-

ken daarom steeds vaker intensief samen om gratis instrumenten te ontwikkelen.

Gegeven deze ontwikkelingen is het van kardinaal belang om bij ROM en benchmarking uitsluitend gratis instrumenten aan te bieden, mede om instellingen erop attent te maken dat zij soms al jaren ten onrechte geld uitkeren aan commerciële partijen. Van de 6 ROM-instrumenten die Buwalda e.a. (2012) in het themanummer bespreken, zijn er 4 niet gratis. Het gebruik hiervan draagt bij aan woekerwinsten en sluist nodeloos middelen weg uit de ggz.

ARGUMENTEN PRO EN CONTRA

De combinatie zorgverzekeraars-GGZ Nederland is zich wel bewust van de kritiek. Het argument dat men ter verdediging opvoert, is dat men voorlopig vooral uit is op het uitvoeren van ROM *per se*, omdat dit sowieso goed zou zijn voor de patiëntenzorg. Er ontstaat zo een situatie dat de zorgverzekeraar blijkbaar degene is die de beroepsgroep de richtlijnen gaat voorschrijven waarop het klinisch handelen gebaseerd moet zijn. Het resultaat is duidelijk: een monstrueuze *one-size-fits-all* beleidsmaatregel die voorbij gaat aan de complexiteit van de geestelijke gezondheidszorg, en waarvoor geen sluitend bewijs is. Een valide meta-analyse dat een dergelijke *one-size-fits-all* benadering de ROM-uitkomsten in de ggz verbetert, dwars door alle settingen heen, bestaat niet. Carlier e.a. (2012) constateren in hun kwalitatieve overzicht in het themanummer: 'studies naar het effect van ROM op het gedrag van de patiënt (bijvoorbeeld therapietrouw) en diens (psychische) gezondheid of herstel zijn schaarser, met niet-eenduidige resultaten'.

Bovendien zijn de door hen beschouwde onderzoeken zeer heterogeen en weinig representatief voor de praktijk van de ggz in Nederland: het betreft activiteiten die niet vergelijkbaar zijn met ROM in het kader van SBG. Het is bijvoorbeeld onduidelijk in welke mate en hoe verplichte ROM in de Nederlandse ggz ook gebruikt zal worden voor systematische feedback aan behandelaars – voorlopig lijken instellingen veel meer bezig met

het leveren aan SBG *per se* en komt verantwoord gebruik van ROM in de praktijk op de tweede plaats.

Dat er wellicht positieve effecten te verwachten zijn van ROM in de Nederlandse ggz is aannemelijk – reden waarom beroepsgroepen in de ggz steeds meer ROM-activiteiten ontwikkelen met door hen gekozen instrumenten, die geschikt zijn voor de specifieke setting en patiëntengroep. Voor patiënten met ernstige psychiatrische aandoeningen bijvoorbeeld werd onlangs de Nederlandse consensus gepubliceerd (Drukker e.a. 2010; Mulder e.a. 2010). De eerste resultaten zijn bemoedigend (Drukker e.a. 2010; 2011) maar zeker niet sluitend.

Gegeven deze ontwikkelingen is een bureaucratische *one-size-fits-all* verplichting van de zorgverzekeraar op dit moment nodig noch gewenst, zeker niet als de onderliggende agenda er een is van versterking van de voortdurend uitdijende management-, registratie- en controlebureaucratie in de ggz (Laane & Luijk 2012). Meta-analysen zijn blijkbaar voor de richtlijnen van de wetenschappelijke vereniging – de beleidsmaker hoeft alleen op indirecte kwalitatieve argumenten te varen. Daar waar klinici al jaren gewend waren hun specifieke doelgroep in hun specifieke setting kwantitatief te evalueren, met door hen zorgvuldig geselecteerde ROM-instrumenten, komen bezorgde managers ons nu vertellen dat het verplichte eenheidsproduct van de beleidsmaker voortaan de scepter zwaait: de zorgverzekeraar immers dreigt met korting.

Echter, nog afgezien van de vraag of ROM in de voorgestelde vorm een effect oplevert, ligt het grootste mysterie in de volgende observatie: als de zorgverzekeraar alleen wil dat ROM wordt toegepast en erkent dat benchmarken voorlopig niet mogelijk is, waarom dan een peperdure SBG inrichten voor een verplichte, voor instellingen kostbare, *one-size-fits-all*-, nationale dataverzameling die geen ander redelijk doel kan dienen dan benchmarken?

IS BENCHMARKING ÜBERHAUPT MOGELIJK?

Het is duidelijk dat benchmarken op zich nuttig is: systematische feedback op basis van routinematige uitkomstvergelijkingen kan de kwaliteit naar een hoger plan tillen. Wetenschappelijke beroepsverenigingen in vele landen werken aan de ontwikkeling van een valide vorm van benchmarken in verschillende takken van gezondheidszorg. De ROM-benchmarking die nu in Nederland dreigt te worden gelanceerd, is echter gebaseerd op het fundamentele misverstand dat de beleidsmaker objectieve vergelijkingen kan uitvoeren tussen groepen op basis van data die de behandelaar in de klinische praktijk gebruikt om het beloop van de individuele patiënt te beoordelen.

Voor de behandelaar zijn de ‘objectieve’ ROM-data slechts één van de vele informatiebronnen die de basis voor de behandeling van de individuele patiënt vormen. De behandelaar heeft ook informatie over lichamelijke aandoeningen en de behandeling daarvan, het niveau van acceptatie van de klachten, de veranderlijkheid van de klachten over de dagen, weken en maanden, het specifieke demografische profiel van de patiënt en de sociale context die de klachten veroorzaakt, onderhoudt of verlicht. Binnen de context van individuele behandeling door een en dezelfde behandelaar is ROM dus nuttig; de behandelaar kan de behandeling benchmarken met zichzelf over de tijd – omdat het over vergelijkingen bij dezelfde patiënt gaat speelt confounding geen rol.

Het uitgangspunt dat men geïsoleerde symptoomlijstjes voor vergelijkingen bij dezelfde patiënt kan verheffen tot instrument voor het vergelijken van prestaties tussen instellingen is gebaseerd op een wetenschappelijke denkfout. Routine-uitkomsten kunnen niet gelijk worden gesteld met ‘prestaties’ of behandelresultaten. Valide ROM-activiteiten, waarvan voorbeelden bestaan in het buitenland, zijn veel diepgaander en vooral gericht op *routine process monitoring*; zonder *process monitoring* verschaalt ROM tot een controle-instrument van beleidsmakers (Hafken-

scheid 2010). Bovendien kunnen we stellen dat indien men vanuit beleidsperspectief in vergelijkende zin iets zinnigs over de behandeluitkomst wil vaststellen, dit vooral op het niveau van de beleeftde uitkomst zal moeten zijn en niet de technisch-klinimetriche uitkomst.

HOOP OP AFWIJZING SBG

Nu de zorgverzekers veel geld hebben geïnvesteerd in het optuigen van SBG, is de kans gering dat de combinatie zorgverzekers-GGZ Nederland zal erkennen dat er een fundamentele vergissing aan de basis ligt van de ROM-benchmarkgedachte. Het is wellicht een teken van de tijd dat wetenschappelijke resultaten worden gezien als 'maakbaar' en als secundair aan beleidsbelangen; wat nu rest is een prestigekwestie. Onze hoop rust daarom bij de brancheorganisaties zoals de Nederlandse Federatie van UMC's (NFU) en de Nederlandse Vereniging van Ziekenhuizen (NVZ) die, in tegenstelling tot GGZ Nederland (in elk geval bij het ter perse gaan van deze bijdrage), nog niet hebben ingestemd met het gebruik van ROM-benchmarking in ziekenhuizen. Zij worden hierin gesteund door de Orde van Medisch Specialisten (OMS).

Zolang al deze partijen geen toezegging hebben gedaan, lijkt er geen juridische basis waarop zorgverzekers algemene en universitaire ziekenhuizen kunnen verplichten tot het invoeren van ROM-benchmarking, net zo min als de zorgverzekeraar de chirurg kan opleggen welk scalpel te gebruiken.

In het belang van de patiënten is verzet nodig: een systeem van economische concurrentie invoeren op basis van ondeugdelijke data is onethisch en het gijzelen van hulpverleners met het zoveelste voorbeeld van kostbare, ondoordachte en onwetenschappelijke bureaucratie dient een krachtig halt toe te worden geroepen. Pas dan ontstaat de ruimte, of beter: de vrijheid, om een wetenschappelijk gevalideerd benchmarkingsysteem te ontwerpen voor gebruik in de ggz.

Vooralsnog blijft de ultieme vraag (Delespaul 2010): wat betekent een SBG-benchmarkscore van 7,5 in instelling X in vergelijking met een 7,1 in instelling Y, of een T-scoreverbetering van 14 ten opzichte van 25? Kan het zijn dat de 7,1 betere zorg is dan de 7,5, en dat verbetering van 14 beter is dan 25? Het antwoord is een ontvullend: ja.

LITERATUUR

- Bak M, van Os J, Delespaul P, de Bie A, á Campo J, Poddighe Ge.a. An observational, 'real life' trial of the introduction of assertive community treatment in a geographically defined area using clinical rather than service use outcome criteria. *Soc Psychiatry Psychiatr Epidemiol* 2007; 42: 125-30.
- Buwalda VJA, Nugter MA, Van de brug SY, Draisma S, Smit JH, Swinkels JA, e.a. ROM in de volwassenenpsychiatrie: een evaluatie van meetinstrumenten. *Tijdschr Psychiatr* 2012; 54: 147-52.
- Carlier IVE, Meuldijk D, Van Vliet I, van Fenema EM, van der Wee NJA, Zitman FG. Empirische evidence voor de effectiviteit van routine outcome monitoring; een literatuuronderzoek. *Tijdschr Psychiatr* 2012; 54: 121-8.
- Clayton D, Hills M. Testing hypotheses. In: Clayton D, Hills M, red. *Statistical models in epidemiology*. Oxford: Oxford Science Publications; 1993. p. 237-49.
- Delespaul AEG. De genormaliseerde T-score (2). Reactie op Edwin de Beurs, 'De genormaliseerde T-score'. *Maandblad Geestelijke Volksgezondheid* 2010; 65: 841-3.
- Drukker M, Bak M, J AC, Driessen G, Van Os J, Delespaul P. The cumulative needs for care monitor: a unique monitoring system in the south of the Netherlands. *Soc Psychiatry Psychiatr Epidemiol* 2010; 45: 475-85.
- Drukker M, van Os J, Bak M, á Campo J, Delespaul P. Systematic monitoring of needs for care and global outcomes in patients with severe mental illness. *BMC Psychiatry* 2010; 10: 36.
- Drukker M, van Os J, Dietvorst M, Sytema S, Driessen G, Delespaul P. Does monitoring need for care in patients diagnosed with severe mental illness impact on psychiatric service use? Comparison of monitored patients with matched controls. *BMC Psychiatry* 2011; 11: 45.
- Hafkenscheid A. Rammelende ROM in de ggz: geen ROM zonder routine process monitoring. *GZ-Psychologie* 2010; 8: 12-7.

- Hafkenscheid A. Subjectiviteit bij de interpretatie van het grafisch scoreverloop op monitorinstrumenten. *Tijdschr Psychiatr* 2012; 54: 129-34.
- Hennekens CH, Buring JE. Analysis of epidemiologic studies: evaluating the role of bias. In: Hennekens CH, Buring JE, red. *Epidemiology in medicine*. Toronto: Little, Brown and Company; 1987. p. 272-87.
- Hennekens CH, Buring JE. *Epidemiology in medicine*. Toronto: Little, Brown and Company; 1987.
- Keet IPM, Nugter AM, Hafkenscheid A, Van Den Eede F, Thunnissen MM. ROM: primair een instrument voor de behandeling. *Tijdschr Psychiatr* 2012; 54: 107-9.
- Kisely S, Preston N, Rooney M. Pathways and outcomes of psychiatric care: does it depend on who you are, or what you've got? *Aust N Z J Psychiatry* 2000; 34: 1009-14.
- Laane R, Luijk R. ROM en positie van de zorgverzekeraars. *Tijdschr Psychiatr* 2012; 54: 135-9.
- Leval N de. Quality of life and depression: symmetry concepts. *Qual Life Res* 1999; 8: 283-91.
- Noom MJ, de Jong K, Tiemens B, Kamsteeg F, Markus MT, Pot AM, e.a. Routine outcome monitoring en benchmarking: hoe kunnen we behandelresultaten op een zorgvuldige manier vergelijken? *Tijdschr Psychiatr* 2012; 54: 141-5.
- Nugter MA, Buwalda VJA. Achtergronden en gebruiksmogelijkheden van ROM in de ggz. *Tijdschr Psychiatr* 2012; 54: 111-20.
- Miller GA, Chapman JP. Misunderstanding analysis of covariance. *J Abnorm Psychol* 2001; 110: 40-8.
- Moore M, Hofer S, McGee H, Ring L. Can the concepts of depression and quality of life be integrated using a time perspective? *Health Qual Life Outcomes* 2005; 3: 1.
- Mulder CL, Kortrijk HE. De invloed van de duur van behandeling op het interpreteren van ROM-metingen bij ACT. *Tijdschr Psychiatr* 2012; 54: 191-6.
- Mulder CL, van der Gaag M, Bruggeman R, Cahn W, Delespaul PA, Dries P, e.a. Routine outcome monitoring voor patiënten met ernstige psychiatrische stoornissen; een consensusdocument. *Tijdschr Psychiatr* 2010; 52: 169-79.
- Newman JC, Feldman R. Copyright and open access at the bedside. *N Engl J Med* 2011; 365: 2447-9.

AUTEURS

JIM VAN OS is hoogleraar Psychiatrische Epidemiologie, Maastricht UMC.

RENÉ S. KAHN is hoogleraar Psychiatrie, UMC Utrecht.

DAMIAAN DENYS is hoogleraar aan de Universiteit van Amsterdam (UvA) en afdelingshoofd Psychiatrie, AMC Amsterdam.

ROBERT A. SCHOEVERS is hoogleraar Psychiatrie en afdelingshoofd Universitair Centrum Psychiatrie Groningen.

AARTJAN T. BEEKMAN is hoogleraar en hoofd afdeling Psychiatrie, VUmc.

WITTE J. HOOGENDIJK is hoogleraar Psychiatrie, Erasmus UMC.

BERT M. VAN HEMERT is hoogleraar Psychiatrie, Leids UMC.

PAUL P. HODIAMONT is hoogleraar Psychiatrie, UMC St. Radboud.

FLOOR SCHEEPERS is hoofd afdeling Kinder- en Jeugdpsychiatrie, UMC Utrecht.

PHILIPPE A.E.G. DELESPAUL is klinisch psycholoog, universitair hoofddocent, Maastricht UMC en programmaleider Integrale Zorg en Psychotische Stoornissen in Mondriaan, Zuid-Limburg.

ALBERT F.G. LEENTJENS is psychiater en universitair hoofddocent, Maastricht UMC.

Correspondentieadres: prof. dr. Jim van Os, Afdeling Psychiatrie en Psychologie, Maastricht Universitair Medisch Centrum, Postbus 616, 6200 MD (DRT 12), Maastricht.

E-mail: j.vanos@maastrichtuniversity.nl

Geen strijdige belangen meegedeeld.

Het artikel werd voor publicatie geaccepteerd op 31-1-2012.

SUMMARY

Behavioural standard or coercive measure? Some considerations regarding the special issue on ROM – J. van Os, R. Kahn, D. Denys, R.A. Schoevers, A.T.F. Beekman, W.J.G. Hoogendijk, A.M. van Hemert, P.P.G. Hodiament, F. Scheepers, Ph.A.E.G. Delespaul, A.F.G. Leentjens –

BACKGROUND The cost of mental health care has possibly risen more than costs in other sectors of health care in the Netherlands. In an attempt to control the rising costs, new policies have been implemented that include the introduction of selective financial penalties for those in need of mental health care as well as the start of performance-based mental health care reimbursement. In order to achieve the latter goal, a nation-wide large-scale data collection was introduced based on clinical routine outcome monitoring (ROM) data, with a view to using these data for benchmarking.

AIM Closer inspection of the benchmarking efforts in terms of scientific validity.

METHOD Qualitative review and analysis.

RESULTS Analysis shows that the type of ROM data that is collected in the Netherlands is valid for tracking the outcomes of individual patients, but unsuitable for performance comparisons between institutions for reasons of case-mix, instrument-mix, bias and lack of sensitivity.

CONCLUSION Attempts to introduce benchmarking based on ROM will probably have a negative impact on the practice of mental health care in the Netherlands. More input from mental health professionals and scientists is required in order to identify more rational and efficient ways of spending scarce resources.

[TIJDSCHRIFT VOOR PSYCHIATRIE 54(2012)3, 245-253]

KEY WORDS benchmarking, biomedical ethics, routine outcome monitoring/ measurement (ROM)