



www.DirectieveTherapie.nl

Dank voor uw download

U kunt er natuurlijk uit citeren, graag zelfs, maar dan wel met bronvermelding. U mag dit artikel ook ruimhartig verspreiden mits het niet voor commerciële doeleinden is. In die gevallen pas na onze schriftelijke toestemming.

Opname in bloemlezingen en readers moedigen wij aan, maar wel graag eerst even overleggen.

Alle rechten van de artikelen liggen bij
de Stichting Cognitie en Psychose.

Voor alle vragen:
info@gedachtenuitpluizen.nl

ROM en benchmarken

Een voorbeeld van een geïntegreerde aanpak

Routine Outcome Monitoring (ROM) is een methodiek in de ggz om standaard meetinstrumenten (zelfrapportagevragenlijsten of beoordelingsschalen) te gebruiken als hulpmiddel bij de behandeling. Meetinstrumenten worden bij de intake ingezet om zicht te krijgen op de aard en ernst van de klachten en zo tot een behandelplan te komen. Met vervolgmetingen tijdens of na de behandeling kunnen we de uitkomst vaststellen. De meetgegevens voorzien de behandelaar en de patiënt periodiek van feedback over de voortgang die geboekt wordt met de behandeling (de Beurs et al., 2011). Het betrekken van meetgegevens bij het bepalen van het behandelplan en het evalueren van de resultaten past prima binnen het zogenaamde *shared decision making* (gezamenlijke besluitvorming) van patiënt en behandelaar. ROM kan zo een van de manieren zijn om tijdig te detecteren dat een behandeling op koers ligt of dreigt te mislukken. Vooral potentiële mislukking detecteren is relevant, aangezien behandelaars over het algemeen een te zonnige kijk hebben op hun eigen effectiviteit (Lambert, 2012). Zo vonden Hannan en collega's (Hannan et al., 2005) dat van een groep van 550 patiënten maar bij 3 van hen door de behandelaars een negatieve uitkomst voorspeld werd, terwijl het daadwerkelijke aantal patiënten met een negatieve uitkomst uiteindelijk 40 bedroeg. Dit komt neer op 7,5% correct voorspelde negatieve uitkomsten. Een feedbackalgoritme daarentegen voorspelde in 77% van deze gevallen de negatieve uitkomst correct.

Meten is op zichzelf niets nieuws in de klinische psychologie en de psychiatrie. Sinds jaar en dag worden er zelfrapportagevragenlijsten (zoals de SCL-90; Arrindell & Ettema, 2003) gebruikt om de ernst of de aard van de klachten of symptomen op een gestandaardiseerde manier vast te stellen. Het gebruik ervan wordt dan ook als noodzakelijk gezien voor het kunnen toepassen van onder andere de verschillende multidisciplinaire behandelrichtlijnen (zie www.ggzrichtlijnen.nl). De brede toepassing van ROM in de ggz past in een tendens om te professionaliseren, en de klinische blik van de intaker of behandelaar aan te vullen en te verrijken met gegevens van gestandaardiseerde meetinstrumenten, teneinde passend bij de patiëntengroep en de specifiek te behalen behandeldoelen psy-

chische klachten, sociaal en maatschappelijk functioneren en/of kwaliteit van leven te kwantificeren.

Recent is de invoering van ROM in de ggz in een stroomversneling geraakt en staat zij sterk in de belangstelling, getuige een drietal boeken (Buwalda, Nugter, Swinkels, & Mulder, 2011; Buwalda, Nugter, van Tilburg, & Beekman, 2013; van Hees, van der Vlist, & Mulder, 2011) en een themanummer in het *Tijdschrift voor Psychiatrie* (februari 2012). Deze belangstelling is mede te verklaren uit een andere ontwikkeling: de plannen om geaggregeerde ROM-gegevens te gebruiken om de kwaliteit van de geboden zorg transparant te maken bij groepen patiënten. Zo'n groep kan bestaan uit alle patiënten van een instelling, de patiënten van een afdeling of de caseload van een behandelaar. De geaggregeerde uitkomstgegevens kunnen vergeleken worden met een norm, bijvoorbeeld het landelijk gemiddelde. We noemen dit 'benchmarken' (de Beurs, 2011). Instellingen krijgen zo zicht op hun sterke en zwakke kanten, en kunnen zien waar extra inspanningen nodig zijn. Zorgverzekeraars en zorggebruikers (patiënten) krijgen zicht op de kwaliteit van de zorg die een instelling biedt (van der Meeren, 2011). Mits aan bepaalde voorwaarden is voldaan, kan aggregatie van gegevens die primair ten behoeve van het klinisch handelen zijn verzameld op deze manier waardevolle informatie opleveren en transparantie scheppen over de effectiviteit van een zorgsector die toch vooral als een *black box* bekend stond: het was onbekend wat je daar als patiënt kon verwachten en wat het opleverde. Voorwaarden voor waarachtige transparantie en een eerlijke onderlinge vergelijking zijn wel (1) dat we beschikken over accurate, betrouwbare, valide en vergelijkbare gegevens, en (2) dat er rekening wordt gehouden met op voorhand bestaande verschillen tussen zorgaanbieders wat betreft hun patiëntenpopulaties.

De plannen om te gaan benchmarken in de ggz hebben de implementatie van ROM een forse impuls gegeven. Het benchmarken is echter ook aanleiding geweest tot discussie over de wenselijkheid van deze ontwikkeling (Mulder, 2010; Keet, Nugter, Hafkenscheid, Eede, & Thunnissen, 2012). Vanuit wetenschappelijke hoek is twijfel geuit over de bruikbaarheid van ROM-gegevens voor verantwoording of over de haalbaarheid van benchmarken op basis van ROM-gegevens. Daarbij wordt gewezen op problemen met betrekking tot de vergelijkbaarheid van meetinstrumenten, behandeluitkomsten, patiëntengroepen, samen te vatten als bias in de gegevens vanwege selectie, en instrumentatie en *confounding* (van Os et al., 2012). Maar vanuit wetenschappelijke hoek wordt ook om enig geduld gevraagd: benchmarken is 'werk in uitvoering', en er moet nog het nodige uitgezocht en onderzocht worden. Er wordt dan ook gewaarschuwd voor het trekken van voorbarige conclusies op basis van gegevens waarvan de validiteit nog ter discussie staat, maar tevens

wordt gesteld dat onderlinge vergelijking op basis van ROM-gegevens wel degelijk tot verbetering van zorg kan leiden (Blijd-Hoogewys et al., 2012).

Ook door beroepsbeoefenaars in het veld worden er vraagtekens gezet bij het gebruik van geaggregeerde ROM-gegevens voor kwaliteitsmanagement en transparantie over wat de ggz vermag. Zo is er de vrees dat ROM-gegevens, wanneer ze gebruikt worden als prestatie-indicator van de kwaliteit van zorg, vroeg of laat gecorrumpeerd raken. Volgens deze redenering zal ROM averij oplopen wanneer de gegevens gebruikt worden voor het verantwoorden van bestede gelden in de ggz. Een veelgehoorde stelling is: 'Als we erop afgerekend worden, ontstaat een perverse prikkel om de ROM-gegevens te vertekenen, of om bijvoorbeeld alleen nog bij succesvol verloopende behandelingen ROM toe te passen.' In deze redenering dient benchmarken slechts een afrekencultuur ('prestatiebekostiging'), leidt het tot politiek geïnspireerde en ongewenste ontwikkelingen als 'marktwerking in de zorg', en moeten we er verre van blijven, of op zijn minst zeer omzichtig met geaggregeerde gegevens omgaan. De druk om ROM in het veld te implementeren en de bonus/malus-regeling die door sommige verzekeraars aan de toepassing van ROM wordt verbonden, verhogen nog eens de achterdocht bij de beroepsgroep. ROM dreigt zo in een kwaad daglicht te komen. Dat staat succesvolle implementatie op de werkvloer in de weg. Er dreigt nu een situatie te ontstaan dat ROM niet wordt toegepast ter ondersteuning van de behandeling, maar omdat het moet van de zorgverzekeraar. Begrijpelijkerwijs neemt dat veel enthousiasme weg bij de beroepsbeoefenaars in de ggz en kalft het draagvlak af voor invoering van ROM.

De discussie wordt ook gekenmerkt door verwarring over de verschillende doelen en betekenissen van ROM en benchmarken. Voor de behandelaar in de ggz is ROM vooral een middel om de behandeling te ondersteunen (*Routine Outcome Monitoring*, ROMo; zie ook het naschrift bij dit artikel). Voor een manager levert ze gegevens op om te sturen en zich te verantwoorden over behaalde resultaten (benchmarken; *Routine Outcome Management*, ROMa). Ten slotte is er nog het doel om te leren van de uitkomstgegevens en ze te gebruiken voor wetenschappelijk onderzoek naar de effectiviteit van verschillende behandelingen (*Routine Outcome Measurement*, ROMe). Essentieel voor een goede implementatie is dat de nadruk ligt op de ondersteunende functie van ROM voor de behandeling; de behandelaar is immers de sleutel tot succes.

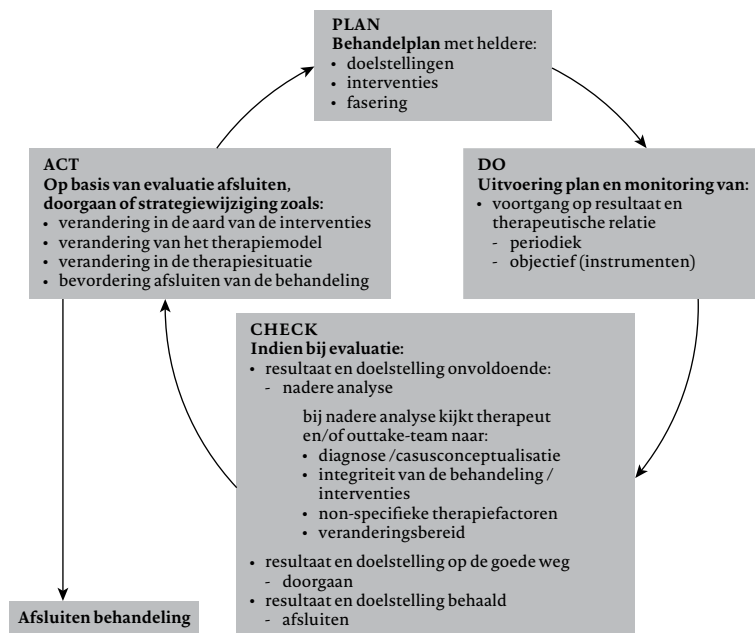
Bij een goede implementatie liggen de doelen van ROM en benchmarken in elkaars verlengde, maar zoals het nu gaat, lijken deze doelen tegenstrijdig en elkaar tegen te werken. Maar zo hoeft het niet te gaan. ROM en benchmarken kunnen ook hand in hand

gaan en gezamenlijk beide doelen dienen (ondersteunen van de individuele behandeling en kwaliteitsmanagement). Als illustratie van een succesvolle integratie van ROM in de behandeling, in combinatie met benchmarken, willen we de *practice* bij een middelgrote, landelijke ggz-aanbieder beschrijven. Het dubbele gebruik van ROM-gegevens (behandeling ondersteunen en benchmarken) wordt door de werkwijze bij deze organisatie geïllustreerd.

ROUTINE OUTCOME MONITORING BIJ HSK

De HSK Groep is een landelijk opererende, private ggz-organisatie voor zowel generalistische basis-ggz als gespecialiseerde ggz, die sinds 2004 standaard bij al haar patiënten voortgangsmetingen toepast. De aldus verkregen gegevens worden tevens gebruikt voor het benchmarken van vestigingen en behandelaars. Er wordt *evidence-based* gewerkt met klachtspecifieke behandelprotocollen volgens de cognitieve gedragstherapie, toegepast bij ambulante patiënten met de zogenaamde *common mental disorders*: depressieve stoornis, dysthyme stoornis, alle angststoornissen, somatoforme stoornissen, middelenstoornissen, stoornissen in de impulscontrole, en stressgerelateerde stoornissen (burn-out) met en zonder comorbiditeit (zie ook: Dijkgraaf-Hartland, Brals, Verbraak, & van Dijk, 2013). Gemiddeld beslaat een behandeling bij HSK 18 zittingen, maar onder het motto 'zo kort als mogelijk, zo lang als nodig' kunnen het er ook 12 of 35 zijn.

Binnen iedere individuele behandeling wordt om de vijf behandelingen, conform het binnen de organisatie gehanteerde kwaliteitsmanagementsysteem voor de behandeling van de individuele patiënt (zie figuur 1; Verbraak & Hoogduin, 2013), routinematig gewerkt met evaluatiemomenten tussen behandelaar en patiënt. Tijdens deze evaluatiemomenten wordt gebruikgemaakt van een meting van de ernst van de problematiek op dat moment, op basis van de SCL-90 als algemene ernstmaat (Arrindell & Ettema, 2003) en stoornisspecifieke klachtenvragenlijsten, zoals de Hamilton Depressie Schaal (HDS; Bech, Kastrup, & Rafaelsen, 1989) bij depressieve patiënten, of de Utrechtse Burnout Schaal (UBOS; Schaufeli & van Dierendonck, 2000) bij patiënten met burn-out. Daarnaast wordt aan de hand van de WerkAlliantievragenlijst (WAV; Stinkens, Ulburghs, & Claes, 2009) navraag gedaan naar de kwaliteit van de therapeutische relatie zoals ervaren door de patiënt. Deze vragenlijsten zijn voorafgaand aan de evaluatie afgenomen en de uitslag is beschikbaar bij de zitting. Op het evaluatiemoment worden de uitkomsten samen met de patiënt nauwlettend bestudeerd om te zien of de behandeling op de goede weg zit of dat er eventueel



FIGUUR 1
Kwaliteitsmanagementsysteem voor de behandeling van de individuele patiënt (bron: Verbraak & Hoogduin, 2013)

bijgestuurd moet worden. De ernst van de problematiek op het moment van de evaluatie wordt beschouwd en vergeleken met die bij intake of het onderzoek voorafgaande aan de behandeling. De vooruitgang, stagnatie of achteruitgang wordt besproken met patiënt, evenals de redenen die daar mogelijk aan ten grondslag liggen. Geven de scores de werkelijkheid juist weer? Zijn volgens de patiënt de klachten af- of juist toegenomen, zoals de scores uitwijzen? Als dat zo is, zijn we dan nog steeds op de goede weg? Deze evaluatie met de patiënt kan zo aanleiding geven tot continuering, aanpassing of afronding van de behandeling. Over de uitkomst van de evaluatie wordt telkens op een gestandaardiseerde wijze gerapporteerd aan de verwijzer. In de correspondentie met de verwijzer wordt de ernst van de problematiek op dat moment weergegeven en wordt de score afgezet tegen de ernst bij intake. Met andere woorden: iedere therapeut gebruikt de uitkomsten op de ernstmaten, waaronder standaard de SCL-90, niet alleen om te evalueren, maar ook om in al zijn inhoudelijke correspondentie te rapporteren over de stand van zaken rondom zijn patiënt. Het is de behandelaar die bij de patiënt de 'psychologische thermometer' afleest en terugkoppelt aan zowel patiënt als verwijzer. Dit betekent dus dat beide partijen hiermee frequent en actief aan de gang zijn; binnen een

behandeling van 18 zittingen wordt zo'n vijf maal een SCL-90 afgenomen.

Maar ook tussen de behandelaar en het team waarin hij werkzaam is vindt periodiek een bespreking plaats over de voortgang van de behandelingen van de patiënten in zijn caseload (*outtake-team*; zie ook Verbraak & Hoogduin, 2013). Ook in dit evaluatiegesprek wordt standaard gebruikgemaakt van de uitkomsten op de vragenlijsten, teneinde een overzicht te krijgen van de voortgang bij individuele behandelingen over de tijd heen.

BENCHMARKEN BIJ HSK

De uitkomstmetingen van individuele patiënten komen ook terug in de werkbegeleiding, in teambesprekingen en in beleidsvergaderingen, waarin wordt stilgestaan bij resultaten van behandelaars, teams en vestigingen, met als doel de zorg te verbeteren door te leren van vergelijkingen. Uitdrukkelijk uitgangspunt hierbij is dat iedereen de wens heeft om beter te worden als behandelaar. En dat iedereen bereid is te leren. Maar daarvoor moet je wel weten waar je als behandelaar, team of vestiging staat, ook (of juist) in vergelijking met anderen. Benchmarken is een middel om leren van elkaar mogelijk te maken. Met deze wijze van benchmarken wordt permanente kwaliteitsverbetering nagestreefd. Dit is een werkwijze die zich over een langere periode heeft moeten ontwikkelen en die feitelijk nog steeds in ontwikkeling is (zie kader 1).

KADER 1 ONTWIKKELEN VAN DE ACCEPTATIE VAN BENCHMARKEN

De invoering van een benchmarksysteem kan gevoelens van angst, onveiligheid en weerstand oproepen bij de behandelaars. Immers, resultaten worden transparant gemaakt. Binnen HSK werden na de introductie van de benchmark vooral twijfels geuit over de gevolgde methodiek. Deze zou onvolledig zijn en geen recht doen aan de complexiteit van de therapiesituatie. De data kloppen niet. Hier zien we het eerste stadium terug, beschreven door Don Berwick, hoofd van het Amerikaanse Institute for Healthcare Improvement, van een proces waar professionals doorheen gaan als ze geconfronteerd worden met (tegenvallende) prestaties van zichzelf of hun organisatie (Berwick, 2003a, 2003b, in Oudejans, Spits & Schippers, 2011). Berwick beschrijft dit proces als een reis van *taseki* naar *jiseki*. Dit zijn twee Japanse termen, waarbij *taseki* zoveel betekent als 'de bal ligt bij jou, het is niet mijn pro-

bleem', terwijl jiseki zo veel wil zeggen als 'de verantwoordelijkheid ligt bij mij, ik pak de handschoen op'. Om van taseki naar jiseki te komen, dienen professionals vier min of meer opeenvolgende stadia van omgaan met gegevens over (hun) prestaties te doorlopen. In stadium 1, waarin de data niet kloppen, is er sprake van ongelooft. De verzamelde gegevens geven de werkelijkheid niet weer. Er worden dan ook vraagtekens gezet bij de wijze waarop de data zijn verzameld en hoe de berekeningen zijn uitgevoerd (Oudejans et al., 2011). Zo ook binnen HSK. Omdat bekend is dat het werken met professionals vraagt om een heldere en herhaalde uitleg over het hoe en waarom van doelstellingen, taken en inhoud, hebben we daarvan ook gebruikgemaakt. De doelstelling van de benchmark, de gekozen indicatoren, berekeningswijze en hun uiteindelijke norm zijn op verschillende manieren en bij herhaling onder de aandacht gebracht en besproken. Maar belangrijk was ook dat er geluisterd werd naar de bezwaren en vraagtekens van de behandelaars. Dit leidde uiteindelijk tot enige aanpassingen in de systematiek, waarmee voor de behandelaars een meer betrouwbare weergave van de werkelijkheid bereikt werd. Vestigingsmanagers en behandelaars werden ook uitgenodigd om mee te denken over mogelijke andere geschikte kwaliteitsparameters waarmee verbetermogelijkheden te achterhalen zouden zijn. Dit heeft ertoe geleid dat vanaf 2011 ook gebenchmarkt wordt op patiënttevredenheid, eerst nog met de GGZ Cliëntthermometer (van Wijngaarden & Kok, 2008), later met de CQ-index (van Wijngaarden, Kok, Meije, & Fotiadis, 2007), op het niveau van vestigingen en behandelaars. Daarnaast wordt gewerkt aan het toevoegen van uitkomstmaten op het niveau van functioneren (onder andere in welke mate de patiënt arbeid verricht aan het begin en aan het einde van de behandeling) en het welbevinden van patiënten volgens de SF-36 (Aarons et al., 1998).

Na stadium 1 volgt voor de professional stadium 2, waarin de data dan misschien wel kloppen, maar er geen probleem is. De gegevens worden gezien als onvermijdelijk. Ze geven de praktijk van alledag weer en maken actie onnodig (Oudejans et al., 2011). Zo ook binnen HSK. Er werd bij herhaling genoemd dat er nu eenmaal verschillen waren in de casemix en caseload tussen therapeuten en tussen vestigingen. De trainee zag alleen maar makkelijke patiënten, terwijl de oude, ervaren rot in het vak alle moeilijke patiënten kreeg. De ene vestiging kreeg veel meer verwijzingen van huisartsen, terwijl de andere veel meer verwijzingen van bedrijfsartsen kreeg. Analyses van der-

gelijke casemix- en caseload-variabelen op verzoek van de behandelaars lieten vaak zien dat de verschillen wel meevielen, terwijl er toch verschillen bleven tussen therapeuten en vestigingen in uitkomst. Dergelijke kwalitatieve analyses lieten bijvoorbeeld zien dat het feit dat iemand een trainee of ervaren therapeut was geen of onvoldoende verklaring was voor geconstateerde verschillen. Analyses wezen uit dat therapeuten die vooral 'moeilijke' patiënten behandelden vaak hogere verbeterpercentages behaalden. Ook bleken bij patiënten die verwezen waren door een huisarts vergelijkbare, soms zelfs betere resultaten behaald te kunnen worden als bij patiënten die verwezen waren door een bedrijfsarts. Nog steeds komt bij een constatering van verschillen met enige regelmaat de vraag nog naar voren of deze verschillen maar niet gewoon geaccepteerd moeten worden, omdat ze samenhangen met specifieke kenmerken. Dergelijke vragen worden dan beantwoord door voor, en bij voorkeur samen met, de behandelaar of vestiging een nadere analyse te maken.

Door nadrukkelijk te blijven communiceren dat de benchmark een kwaliteitssysteem is, dat in die zin een belangrijke bijdrage kan leveren aan het leren en de professionele ontwikkeling van de behandelaar, werd getracht door stadium 3 heen te komen. Stadium 3 wordt beschreven als het stadium waarin een sfeer ontstaan is dat de data kloppen, en dat er wel een probleem is, maar dat het niet ons probleem is (Oudejans et al., 2011). De oplossing van het probleem ligt niet bij de behandelaar zelf, maar elders. Zo werd binnen HSK in dit kader de nadruk gelegd op de te hoge werkdruk, beperkte faciliteiten en veeleisende opdrachtgevers. Oorzaken werden buiten de eigen invloedssfeer gelegd. Door de wens tot professionele ontwikkeling aan te spreken bij behandelaars en de daarin gedane beloften ook gestand te doen werd dit stadium 'overwonnen'. Deze beloften hielden namelijk ook in dat er tijd beschikbaar kwam om te leren, waarmee de werkdruk enigszins omlaagging en duidelijk werd dat ontwikkelen gefaciliteerd werd. Ook het omgaan met veeleisende opdrachtgevers werd ter hand genomen, alsmede het managen van de verwachtingen van diezelfde opdrachtgevers. Bij dit stadium hoorde binnen HSK ook een zorgvuldige en genuanceerde benadering van vestigingen en behandelaars die onder de gewenste target bleven. De resultaten van individuele medewerkers werden alleen verstrekt aan de vestigingsmanager. Dit voorkwam dat er een sfeer van onveiligheid of ongezonde competitie ontstond tussen behandelaars onderling.

Terwijl de behandelaars bij de invoering van de benchmark enige weerstand en 'koudwatervrees' vertoonden, is er inmiddels een houding ontstaan waarin men uitziet naar de nieuwe resultaten. Daarmee zijn we in stadium 4 terechtgekomen, waarin het uitgangspunt is dat de data betrouwbaar zijn, dat er een probleem uit kan blijken, en dat het dan het probleem van de behandelaar is, die er wat aan gaat doen (Oudejans et al., 2011). Dit bleek binnen HSK onder andere bij de invoering van een nieuw digitaal cliëntvolgsysteem in 2008. Er is toen tijdelijk een periode geweest waarin er geen benchmarkgegevens beschikbaar waren. Herhaaldelijk is toen door behandelaars de vraag gesteld wanneer er weer gebenchmarkt zou gaan worden. Ze misten het zicht op hun uitkomsten van behandeling. Nu ze dat zicht wel hebben, vragen ze zelf (individueel of als vestiging) om meer supervisie, specifieke intervisies of workshops over inhoudelijke thema's, die in potentie de kwaliteit van hun werk kunnen vergroten.

Uiteraard zien we binnen HSK nog weleens bewegingen van jiseki naar taseki. Fasemodellen zoals dat van Berwick zijn zelden lineair. Met name bij veranderingen in de organisatie of personeelssamenstelling blijkt dat. Dan valt men nog wel eens terug in een eerder stadium. Er wil dan nog wel eens een smeulend vuurtje opflakkeren, waarbij de betrouwbaarheid en het nut van de data in twijfel worden getrokken. Goed uitleggen, maar ook luisteren, blijft van belang. Maar vooral laten zien dat je doet wat je beloofd hebt: benchmark is een hulpmiddel, ingezet om leren te bevorderen en daarmee de kwaliteit te verbeteren. Wat dat betreft is de implementatie van benchmarking nooit echt klaar.

Vanaf januari 2004 is de HSK Groep met een zogenaamde 'interne kwalitatieve benchmark' gaan werken, die een onderlinge vergelijking maakt van de kwaliteit van de behandelingen, afgemeten aan geaggregeerde gegevens op een tweetal indicatoren:

- 1 Het percentage patiënten met complete scores op de SCL-90 (SCL-90-score bij intake en bij ontslag).
- 2 Het gemiddelde percentage verbetering op de SCL-90 bij ontslag.

De eerste indicator is een procesvariabele. Hij geeft zicht en greep op het percentage patiënten bij wie de voortgang is gemonitord. Daarop is de tweede indicator gebaseerd. Het percentage gemonitorde patiënten dient uiteraard zo hoog mogelijk te zijn om een representatief beeld te krijgen van de gemiddelde uitkomst. De

tweede indicator, de feitelijke uitkomstvariabele, is geoperationaliseerd als het percentage verbetering op de totaalscore van de SCL-90-klachtenlijst.

Het percentage verbetering wordt berekend door de prescore (score bij onderzoek/intake of aanvang van de behandeling) te vergelijken met de postscore (de score bij ontslag of afsluiting van de behandeling). Om de gemiddelde klachtenreductie in een percentage uit te drukken is een nulpunt noodzakelijk. Hiervoor is een score van 118 op de SCL-90 genomen. De rekenregel voor percentage verbetering luidt daarmee:

$$((SCL-90 \text{ pre-score}) - (SCL-90 \text{ post-score}) / (SCL-90 \text{ pre-score} - 118)) \times 100\%$$

De waarde 118 is de gemiddelde score van de gemiddelde Nederlander volgens de normering van de SCL-90 (Arrindell & Ettema, 2003; voor een nadere uitleg over het percentage verbetering, zie ook kader 2). Het streven in de behandeling is om wat betreft uitkomst in de buurt van de gemiddelde Nederlander uit te komen, en het percentage verbetering geeft aan in welke mate dit is gelukt. Een verandering van pre 198 naar post 158 is halverwege de verschuiving 198 naar 118 en levert dus een uitkomst op van 50% verbetering.

KADER 2 BEREKENING VAN PERCENTAGE VERBETERING

Op de berekeningswijze van percentage verbetering is statistisch/methodologisch gezien het een en ander aan te merken (de Beurs et al., 2015). Ze is niet de meest zuivere beoordeling van verbetering, maar ze is wel intuïtief aansprekend en zeer praktisch in het gebruik. Voor behandelaars is ze begrijpelijk, en als voor iedereen de formule op dezelfde manier wordt toegepast, is onderlinge vergelijking mogelijk. Dat laatste is het doel, de gebruikte formule het middel. Maar er zijn om het een hanteerbaar middel te laten zijn wel een aantal andere aanpassingen/voorwaarden gesteld aan de berekening van deze formule. Volgens deze formule kan de verslechtering namelijk ver boven de 100% uitkomen (dat gebeurt zelfs exponentieel naarmate het resultaat slechter is), maar kan de verbetering nooit boven de 100% uitkomen. Als de rekenregel niet aangepast zou zijn, dan zou een verslechtering onevenredig zwaar de gemiddelde score van behandelaar of vestiging gaan beïnvloeden. Daarom is besloten om van alle patiënten die een verslechtering laten zien die boven de 100% ligt de score vast te zetten op -100%. Hiermee wordt op het percentage verbetering of verslechtering een maximum van + of -100% toegepast.

Daarnaast zijn er ook patiënten die een prescore hebben onder de 118. Daarmee krijgt de deler in de formule een negatieve waarde, waarmee bij een lagere postscore de verbetering toch berekend wordt als een verslechtering. Daarom is de regel ingesteld dat eindscores op de SCL-90 gelijk aan of lager dan die van de gemiddelde Nederlander (een score van 118) onafhankelijk van de beginscore als 100% verbetering genoteerd worden.

De vestigingen en de behandelaars onderling worden vergeleken op deze indicatoren. Uit de scores van de vestigingen wordt tevens een gemiddelde score op de indicatoren voor de gehele organisatie berekend.

Na het starten met de benchmark in 2004 werd per vestiging en per behandelaar ook een aantal na te streven targets vastgesteld. Het ging niet alleen om het onderlinge vergelijken, maar bij gebrek aan extern vergelijkingsmateriaal ook om het als organisatie gaan vergelijken met een wenselijk geacht streven. Zo werd, nadat ook gebleken was dat dit in principe haalbaar was, het streefpercentage complete scores op minimaal 80% en het gemiddelde streefpercentage verbetering op de SCL-90 op 85% vastgesteld.

Een belangrijke rol in het gebruik van de benchmarkresultaten ligt van het begin af aan bij de vestigingsmanager. Het is de vestigingsmanager die de resultaten van de eigen vestiging vergelijkt met die van de andere vestigingen en beoordeelt of de gestelde targets gehaald zijn. De resultaten op vestigingsniveau zijn transparant en worden met de medewerkers van de vestiging gedeeld en besproken. Wanneer de resultaten voldoen aan de vastgestelde targets, is verdere actie niet nodig. Wanneer de verbeterpercentages op vestigingsniveau achterblijven, zorgt de vestigingsmanager voor een analyse van de resultaten. Waar is de vestiging goed in en waar is de vestiging minder goed in? Waar kan verbetering van het behandelresultaat gerealiseerd worden? Vervolgens wordt zo nodig een meer of minder intensief begeleidingstraject gestart. Dit kan aanvullende supervisie op vestigingsniveau zijn. Hierbij neemt een senior behandelaar van een andere vestiging, of desgevraagd de inhoudelijk directeur, de betreffende vestiging onder zijn hoede. Maar dit kan ook uitwisseling zijn van werkwijzen met een andere goed scorende vestiging, of het vragen om, of het zelf uitwerken van, nieuwe of aanvullende inhoudelijke instrumenten, richtlijnen of procedures, die bij voorkeur organisatiebreed toegepast en dus geïmplementeerd kunnen worden, door het binnen de organisatie bestaande ontwikkelteam.

De resultaten van individuele medewerkers worden in beginsel niet alom gedeeld, maar door de vestigingsmanager met elke behandelaar persoonlijk besproken. Wanneer een behandelaar onder de gestelde targets presteert, nemen vestigingsmanager en behandelaar het gehele patiëntenbestand van de behandelaar door. Er wordt onder andere gekeken naar de verschillende diagnoses van de patiënten in de caseload van de behandelaar, het al dan niet volgens protocollen behandelen, de aanwezigheid van door andere behandelaren overgedragen patiënten en de effectiviteit van de supervisie. Bij dat laatste wordt ook de supervisor betrokken. Deze analyse leidt zo nodig tot een specifiek plan van aanpak. Er worden verbeterpunten met de behandelaar geformuleerd waaraan in de supervisie aandacht dient te worden besteed. Er kan ook worden besloten dat de behandelaar een cursus gaat volgen in specifieke interventiemethoden of in het behandelen van specifieke patiëntgroepen. Er vindt aanvullende supervisie of scholing plaats tot het gemiddelde resultaat op niveau van de benchmark ligt en de target dus is gehaald. De supervisie wordt op dit moment in de regel gegeven door medewerkers van de HSK Groep, maar er kan ook voor gekozen worden een externe supervisor aan te trekken. Tot slot: de benchmark laat zien dat er behandelaren zijn die steeds een percentage verbetering van 100% hebben op de SCL-90. Dit kan betekenen dat het 'topbehandelaren' betreft, maar het is tevens een reden om met hen in gesprek te gaan. Wellicht is de patiëntenpopulatie te licht en is er behoefte om meer ingewikkelde of andersoortige problematiek te behandelen.

Vanaf de start in januari 2004 is een overwegend opwaartse lijn waar te nemen in zowel percentage verbetering op de SCL-90 als percentage complete scores. Tegelijkertijd blijven kleine maar soms ook grotere verschillen tussen behandelaren en vestigingen waarneembaar. Opmerkelijk is de constatering dat beleidswijzigingen die de afgelopen jaren in veelvoud over de ggz heen zijn gekomen hun weerslag hebben op de benchmarkcijfers. Daarmee zijn deze cijfers zeker niet stabiel over de tijd heen. Er zit duidelijk beweging in. Benchmarks in binnen de organisatie inmiddels geaccepteerd als een van de middelen om kwaliteit te bewaken en waar nodig te verbeteren.

DISCUSSIE

Het werken volgens een benchmarkstelsel vraagt om een zorgvuldige implementatie in de organisatie. Wanneer medewerkers zich gecontroleerd voelen in plaats van gestimuleerd om hun prestaties te verbeteren loopt het stelsel zijn doel mis. Om dat laatste

te voorkomen is het belangrijk goed te benadrukken dat het om een kwaliteitssysteem gaat, waarbij herhaald en op vaste momenten volgens afspraak de uitkomsten worden gevolgd. Het gaat steeds weer om het laten zien dat benchmarken gebruikt wordt waarvoor het oorspronkelijk bedoeld is vanuit een context van kwaliteitsmanagement: 'A technique for learning from others' successes in an area where the team is trying to make improvements' (Massoud et al., 2001, p. 74, in Barendregt, 2015). Daarbij gaat het ook om consequent, volgens afspraak en gepast actie te ondernemen wanneer de uitkomsten achterblijven bij die van anderen of bij de verwachtingen (de target). Medewerkers zien en ervaren dan ontwikkeling van zichzelf en hun kwaliteit.

Inmiddels wordt de benchmark door behandelaars overwegend als positief ervaren. Iets soortgelijks speelt overigens bij het toepassen van ROM bij patiënten. Herhaaldelijk worden behandelaars die het interval van evalueren om de 5 zittingen niet zo in acht nemen geconfronteerd door hun patiënten met de vraag of het niet weer eens tijd wordt voor hun 'psychologische thermometer'. Meten wordt door patiënten gewaardeerd omdat er serieus naar gekeken wordt, er teruggerapporteerd wordt en er actie wordt ondernomen op basis van de informatie die met ROM is verkregen.

Wat geleerd is bij de casus van HSK is dat ROM en benchmarken als in elkaars verlengde liggende instrumenten kwaliteitsverbetering mogelijk maken en een positief effect hebben op het kwaliteitsdenken. Een bekend kwaliteitsmanagementmotto is hier van toepassing: 'What gets measured, gets managed.' Maar ook hebben we bij HSK geleerd dat ROM en benchmarken geen doelen op zich zijn, maar hulpmiddelen om een doel te bereiken. Het zijn geen afrekeninstrumenten, zoals vooraf wel werd gevreesd door behandelaars en ondernemingsraad. Daarover zijn goede en transparante afspraken gemaakt. Deze afspraken zijn geëvalueerd en nagekomen, hetgeen het vertrouwen in ROM en benchmarken als kwaliteitsinstrumenten heeft versterkt. Maar uitdrukkelijk is ook geleerd dat er daarnaast andere middelen (nodig) zijn die zicht geven op kwaliteit, zoals audits van het klinisch proces: is er volgens protocol behandeld en heeft de behandelaar zich goed aan het protocol gehouden? Dergelijke informatie is ook belangrijk om goed te begrijpen hoe het behandelresultaat tot stand is gekomen en om eventuele verschillen tussen vestigingen of behandelaars te duiden. Dergelijke kwalitatieve analyses geven de kwantitatieve data hun echte betekenis. Feitelijk gaat het bij ROM en benchmarken steeds om streven naar transparantie en leren van uitkomsten. Benchmarken is dienstbaar aan het streven naar transparantie over kwaliteit, maar ook benchmarken zelf is gediend bij transparantie over wat het systeem vermag en hoe het verbeterd kan worden. Een van de dingen die we

namelijk nog niet met zekerheid kunnen zeggen (omdat de wetenschappelijke evidentie ervoor ontbreekt), is of de behandelingen in uitkomsten er gemiddeld genomen echt veel beter van worden. De ervaring tot nu toe laat vooral zien dat de spreiding in behandeluitkomsten minder groot wordt. Dat is toch ook zeker winst voor de patiënten.

En dan toch die weerstand in de ggz tegen met name benchmarken. Die is te verklaren vanuit de verschillende doelen die met benchmarken door verschillende partijen worden voorgestaan (Barendregt, 2015). Een mooi voorbeeld daarvan levert de Raad voor de Volksgezondheid en Zorg, een adviesorgaan van parlement en regering, in zijn rapport *Sturen op gezondheidsdoelen* (Raad voor de Volksgezondheid en Zorg, 2011). Hij stelt dat de zorg meer gericht moet zijn op het bereiken van concrete gezondheidsdoelen, en spreekt de verwachting uit dat dit zal leiden tot 'een betere kwaliteit van de gezondheidszorg en betere zorg voor de patiënt. Bovendien vergroot het de effectiviteit van de ingezette middelen, zowel in menskracht als in geld.' De raad beveelt onder andere aan om het doel van de behandeling met de patiënt te expliciteren, uitkomstindicatoren vast te stellen en uitkomstgegevens van de behandeling standaard te registreren in het elektronisch patiëntendossier. De ROM in de ggz wordt hierbij als voorbeeld gepresenteerd. In het rapport wordt de overtuiging uitgesproken dat meer oog voor de effectiviteit van zorg op zichzelf al tot een kwaliteitsverbetering kan leiden, zowel op het niveau van de individuele patiënt (microniveau), op het niveau van behandelingen binnen een instelling (mesoniveau), als op het niveau van de ggz als onderdeel van de gehele gezondheidszorg (macroniveau).

Maar in het hierboven aangehaalde rapport vinden we nog een aanbeveling. Op termijn zou de bekostiging van de zorg omgeboogen moeten worden van een financiering op basis van verrichtingen naar financiering op basis van behaalde resultaten. Dit is een echo van het pleidooi van Porter en Teisberg (2005) voor *outcome based financing* om marktwerking in de zorg te stimuleren. Ook de minister van vws heeft een toekomst geschetst van monitoren in het primaire proces en de verzamelde gegevens gebruiken voor *performance management* en financiering van de zorg baseren op behandeluitkomsten. Hieruit blijkt dat de beoogde doelen van benchmarken breed zijn: behandelen van patiënten, wetenschappelijk onderzoek, keuze-informatie voor patiënten en financiering van de zorg. Het risico is dat een van deze doelen de overhand krijgt en het gebruik voor andere doelen belemmert (Barendregt, 2015). En dit is ook precies wat er gebeurt. ROM wordt toegepast omdat responspercentages gehaald dienen te worden, wil men als zorgaanbieder geen zorgbudget mislopen. Dit is echt iets anders dan benchmarken

om te kunnen leren van elkaar door uitkomsten te vergelijken, om vervolgens in gesprek te gaan over onderliggende processen.

Benchmarking als instrument voor bekostiging van zorg als stip aan de horizon is mooi, maar voor het zover is, moet het vertrouwen in outcome monitoring en de zeggingskracht van verzamelde uitkomstgegevens nog groeien. De geschiedenis leert dat het vertrouwen toeneemt naarmate we meer ervaring en kennis opdoen met de ROM-methodiek in de behandelpraktijk en met de benchmarkmethodiek als vorm van kwaliteitsmonitoring. Tot die tijd is het opstellen van ranglijstjes van de best en slechtst presterende zorgaanbieders uit den boze. Constateren dat de een het beter doet dan de ander zonder je af te vragen waarom dit zo is en hoe je het kunt veranderen, is geen benchmarken (Barendregt, 2015). Sterker nog, het is voer voor methodologische criticasters die betogen dat de methodologische basis voor een dergelijk gebruik van ROM op dit moment nog discutabel is (Hafkenscheid & van Os, 2014). Goede keuze-informatie voor patiënten is daarbij complexer dan informatie bieden voor uitkomstbekostiging.

TOT BESLUIT

We willen afsluiten met een pleidooi voor implementatie van ROM eerst en vooral ten bate van de individuele behandeling van de patiënt, en aansluitend bij een intrinsieke motivatie tot ROM bij behandelaars (*bottom-up* in plaats van *top-down*). Bij een juiste uitvoering van ROM kunnen alle doelen goed gediend worden (feedback voor de behandeling (ROMO), geaggregeerde uitkomstgegevens ten behoeve van benchmarken (ROMa) en observationeel onderzoek (ROME)). Voorwaarde is wel een prudente omgang met geaggregeerde ROM-resultaten door zorgaanbieders en zorgverzekeraars. Voorbarige conclusies over de kwaliteit van geleverde zorg op basis van onvolledige of onbegrepen uitkomstgegevens zal deze poging tot transparantie over wat de ggz vermag voortijdig smoren.

ROMO, ROMa en ROME zijn zo genoemd in navolging van collega Jochanan Huyser van Arkin, die tijdens een werkconferentie van Stichting Benchmark GGZ (SBG), Zorgverzekeraars Nederland en GGZ Nederland op 13 maart 2013 sprak over ROMO (Routine Outcome Monitoring) en ROMa (Routine Outcome Management).

PROF. DR. MARC VERBRAAK is klinisch psycholoog en cognitief gedragstherapeut, bijzonder hoogleraar gezondheidszorgpsychologie aan de Radboud Universiteit, Behavioural Science Institute, en hoofdopleider gz-psychologen voor de gz-opleiding van het Radboud Centrum Sociale Wetenschappen te Nijmegen.

Hij was tot voor kort werkzaam bij de HSK Groep. Inmiddels is hij werkzaam bij Pro Persona.

DRS. SUUS THEUWS is gz-psycholoog en cognitief gedragstherapeut bij de HSK Groep te Breda.

DR. CARA VERDELLEN is klinisch psycholoog en cognitief gedragstherapeut en eveneens tot voor kort als manager Expertise Centra werkzaam bij de HSK Groep. Zij is momenteel werkzaam bij PsyQ Zwolle.

Referenties

- Aaronson, N. K., Muller, M., Cohen, P. D., Essink-Bot, M. L., Fekkes, M., Sanderman, R. ... Verrips, E. (1998). Translation, validation and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *Journal of Clinical Epidemiology*, 51, 1055-1068.
- Arrindell, W. A., & Ettema, J. H. M. (2003). *SCL-90: Herziene handleiding bij een multidimensionele psychopathologie-indicator*. Lisse: Swets & Zeitlinger.
- Barendregt, M. (2015). Benchmarken en andere functies van ROM: Back to basics. *Tijdschrift voor Psychiatrie*, 57 (in druk).
- Bech, P., Kasrup, M., & Rafaelsen, O. (1989). *Mini-compendium van beoordelingsschalen in de psychiatrie* (Nederlandse vertaling: H. D'Haenen en W. Verhoeven). Brussel: VUB-Press.
- Blijd-Hoogewys, E., van Dijck, R., Emmelkamp, P., Mulder, N., Oude Voshaar, R., Schippers, G., ... Vermeiren, R. (2012). Benchmarken is 'werk-in-uitvoering'. *Tijdschrift voor Psychiatrie*, 54, 1031-1038.
- Buwalda, V. J. A., Nugter, M. A., Swinkels, J. A., & Mulder, C. L. (2011). *Praktijkboek ROM in de ggz: Een leidraad voor gebruik en implementatie van meetinstrumenten*. Utrecht: De Tijdstroom.
- Buwalda, V. J. A., Nugter, M. A., van Tilburg, W., & Beekman, A. T. F. (2013). *Praktijkboek ROM in de ggz II: Implementatie en gebruik bij verschillende doelgroepen*. Utrecht: De Tijdstroom.
- de Beurs, E. (2011). Benchmarken: Kansen en valkuilen. In S. van Hees, P. van der Vlist, & N. Mulder (Red.), *Van meten naar weten: ROM in de GGZ* (pp. 219-226). Amsterdam: Uitgeverij Boom.
- de Beurs, E., Barendregt, M., de Heer, A., van Duijn, E., Goeree, B., Kloos, M. ... Merks, A. (2015). Comparing methods to denote treatment outcome in clinical research and benchmarking mental health care. *Clinical Psychology and Psychotherapy*. Epub ahead of print. Doi: 10.1002/cpp.1954
- de Beurs, E., den Hollander-Gijsman, M. E., van Rood, Y. R., van der Wee, N. J. A., Giltay, E. J., van Noorden, M. S. ... Zitman, F. G. (2011). Routine outcome monitoring in the Netherlands: Practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clinical Psychology & Psychotherapy*, 18, 1-12.
- Dijkgraaf-Hartland, W., Brals, C., Verbraak, M., & van Dijk, M. (2013). Cormorbide persoonlijkheidsproblematiek en de effectiviteit van protocollaire behandeling. *Directieve Therapie*, 33, 14-30.
- Hafkenscheid, A., & van Os, J. (2014). Naar een deugdelijke ROM. *Maandblad voor de Geestelijke Volksgezondheid*, 69, 20-28.
- Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology*, 61, 155-163.
- Keet, I. P. M., Nugter, A. M., Hafkenscheid, A., Eede, V. D., & Thunnissen, M. M. (2012). ROM: Primair een

- instrument voor de behandeling. *Tijdschrift voor Psychiatrie*, 54, 107-109.
- Lambert, M. J. (2012). Helping clinicians to use and learn from research-based systems: The OQ-analyst. *Psychotherapy*, 49, 109-114.
- Mulder, N. (2010). De genormaliseerde T-score (1): Reactie op Edwin de Beurs, 'De genormaliseerde T-score'. *Maandblad voor de Geestelijke Volksgezondheid*, 65, 840-841.
- Oudejans, S., Spits, M., & Schippers, G. (2011). Leren van uitkomsten – van taseki naar jiseki. In S. van Hees, P. van der Vlist, & N. Mulder (Red.), *Van meten naar weten; ROM in de GGZ* (pp. 35-43). Amsterdam: Uitgeverij Boom.
- Porter, M. E., & Teisberg, E. O. (2006). *Redefining health care: Creating value-based competition on results*. Harvard Business Press.
- Raad voor de Volksgezondheid en Zorg (2011). *Sturen op gezondheidsdoelen*. Den Haag: Raad voor de Volksgezondheid en Zorg.
- Schaufeli, W. B., & van Dierendonck, D. (2000). *Utrechtse Burnout Schaal (UBOS): Testhandleiding*. Amsterdam: Harcourt Test Services.
- Stinckens, N., Ulburghs, A., & Claes, L. (2009). De WerkAlliantie Vragenlijst als sleutelement in therapiegebeuren. Meting met behulp van de wav-12, de Nederlandstalige, verkorte versie van de Working Alliance Inventory. *Tijdschrift voor Klinische Psychologie*, 39, 44-60.
- van der Meeren, W. (2011). Verzekeraar - transparantie van zorguitkomsten. In S. van Hees, P. van der Vlist, & N. Mulder (Red.), *Van meten naar weten; ROM in de GGZ* (pp. 219-226). Amsterdam: Uitgeverij Boom.
- van Hees, S., van der Vlist, P., & Mulder, N. (2011). *Van meten naar weten; ROM in de GGZ*. Amsterdam: Uitgeverij Boom.
- van Os, J., Kahn, R., Denys, D., Schoevers, R. A., Beekman, A. T., Hoogendijk, W. J., ... Leentjens, A. F. G. (2012). ROM: gedragsnorm of dwangmaatregel? Overwegingen bij het themanummer over routine outcome monitoring. *Tijdschrift voor Psychiatrie*, 54, 245-253.
- van Wijngaarden, B., & Kok, I. (2008). Prestatie-indicatoren en meetinstrumenten: Prestaties door de ogen van de client. In J. Havenaar, P. van Splunteren & H. J. Wennink (Red.), *Koersen op kwaliteit in de GGZ*. Assen: Koninklijke Van Gorcum.
- van Wijngaarden, B., Kok, I., Meije, D., & Fotiadis, L. (2007). *Een Consumer Quality Index voor kortdurende ambulante ggz: De ontwikkeling en psychometrische kwaliteiten van een vragenlijst voor het meten van cliëntervaringen*. Utrecht: Trimbos-instituut.
- Verbraak, M., & Hoogduin, K. (2013). Management van de behandeling. In P. Emmelkamp & K. Hoogduin (Red.), *Van mislukking naar succes in de psychotherapie* (pp. 93-112). Amsterdam: Boom.