

# Benchmarken en andere functies van ROM: back to basics

M. BARENDREGT

- ACHTERGROND** Sinds 2011 worden in de Nederlandse ggz op landelijke schaal uitkomstgegevens verzameld. Dit heeft tot verwarring geleid over de positie die benchmarken inneemt ten opzichte van routine outcome monitoring (ROM).
- DOEL** Inzicht verschaffen in verschillende doelstellingen en gebruik van geaggregeerde uitkomstgegevens.
- METHODE** Kwalitatieve review en analyse.
- RESULTATEN** Benchmarken is een strategie voor het vinden van best practices en het vergroten van de effectiviteit en valt in het domein van kwaliteitsmanagement. Bij benchmarken gebruikt men het vergelijken van uitkomstgegevens instrumenteel en het heeft relatief veel tolerantie wat betreft de validiteit van de data. Het moet onderscheiden worden van andere functies van ROM. Klinisch beleid, maatschappelijke verantwoording, onderzoeken, uitkomstbesteding, keuze-informatie voor patiënten en benchmarken zijn activiteiten die zowel een andere manier van terugkoppeling van data vergen als verschillende eisen stellen aan de validiteit van de onderliggende data.
- CONCLUSIE** Benchmarken wordt ten onrechte vaak te beperkt gezien als alleen het vergelijken van instellingen. Het is echter een veel meer omvattende methode van kwaliteitsverbetering met relatief soepelere criteria voor validiteit dan uitkomstbesteding en keuze-informatie voor patiënten. Voor benchmarken zijn de huidige beschikbare uitkomstgegevens goed te gebruiken.

TIJDSCHRIFT VOOR PSYCHIATRIE 57(2015)7, 517-525

**TREFWOORDEN** benchmarken, routine outcome monitoring, validiteit



In 2010 sloegen de zorgaanbieders en zorgverzekeraars de handen ineen voor het opzetten van een systeem om structureel uitkomsten te meten in de geestelijke gezondheidszorg (ggz). De ggz werd hiermee voorloper in het moderne kwaliteitsdenken, waarin de patiënt centraal staat en de waarde van de gezondheidszorg wordt uitgedrukt in termen van behandeluitkomsten (Franx 2014; Porter 2010).

Inmiddels is het op geaggregeerd niveau vergaren van uitkomstgegevens onderwerp van een sterk gepolariseerde discussie. Deels komt dit doordat verschillende betrokken partijen verschillende doelstellingen nastreefden met de dataverzameling. Binnen de (brede) gezondheidszorg wees de Gezondheidsraad (2006) al eerder op het risico dat het gebruik van behandeluitkomstgegevens voor publieke

verantwoording de intrinsieke motivatie om deze data te gebruiken voor kwaliteitsverbetering kan ondermijnen. Eenzelfde waarschuwing gaven Nugter en Buwalda (2012) wat betreft het gebruik van uitkomstdata in de ggz.

## Weerstand

Maar de weerstand in de ggz lijkt breder omdat de beoogde doelen van de verzameling breder zijn: financiering van de zorg, keuze-informatie voor patiënten, wetenschappelijk onderzoek én behandelen en begeleiden van patiënten. Dit heeft het risico dat één van de doelen de overhand krijgt en ander gebruik belemmert. Zo is de laatste jaren waar te nemen dat sommige instellingen louter meetinstrumenten kiezen om 'ROM-responspercentages' te halen voor de financier en niet vanwege inhoudelijk gebruik binnen de

behandeling. Delespaul (2011) pleit voor het *back to basics* terugbrengen van *routine outcome monitoring* (ROM) naar de oorspronkelijke context: een initiatief van medici om de zorg te optimaliseren.

Het gevolg van deze beweging is echter dat alle gebruik van uitkomstgegevens buiten een direct klinische setting op één hoop wordt gegooid. Er wordt zo binnen het pluriforme veld van ROM een kunstmatige tweedeling aangebracht tussen 'goede, klinische ROM' die omarmd moet worden en 'bureaucratische ROM', die we zouden moeten vermijden (Hafkenscheid & Van Os 2014, p. 27). Dit polariseert het debat alleen maar verder.

### Begripsverwarring

Er is sprake van onduidelijkheid omtrent het begrip 'benchmarken' en hoe benchmarken zich verhoudt tot ROM. Hoewel het debat zich vooral lijkt toe te spitsen op benchmarken, is het opvallend hoe weinig pogingen er zijn ondernomen om het begrip 'benchmarken' af te bakenen of te definiëren. Illustratief is dat onder de vier functies van ROM in het ROM-ggz-project (behandelen en begeleiden, leren, onderzoeken en verantwoorden) de term 'benchmarken' niet genoemd wordt (Van der Vlist 2011).

Daar waar er wel pogingen zijn ondernomen om het begrip te plaatsen gebeurt dit weinig consistent. Van der Vlist plaatst benchmarken onder 'verantwoorden', maar in hetzelfde boek worden benchmarks ook bij 'leren' genoemd (Van der Ham & Reitsma 2011). Vaak wordt benchmarken beperkt tot 'het vergelijken van behandelprestaties tussen aanbieders' (Blijd-Hoogewys e.a. 2012; Janssen e.a. 2014; Sytema & Van der Krieke 2013; Van Os e.a. 2012). Anderen beschouwen benchmarken als het bieden van transparantie, waarvoor het vergelijken slechts het middel vormt (Bruinsma e.a. 2012) of als het bieden van keuze-informatie (Laane & Luijk 2012). Benchmarken in de oorspronkelijke context van kwaliteitsmanagement daarentegen wordt omschreven als 'a technique for learning from others' successes in an area where the team is trying to make improvements' (Massoud e.a. 2001, p. 74).

Er is kortom wat conceptuele vaagheid over wat benchmarken precies is en hoe zich dat verhoudt tot ROM. Er is behoefte aan een zakelijke insteek waarbij de verschillende wijzen en doelen van gebruik van geaggregeerde uitkomstgegevens helder onderscheiden worden, zodat er een vruchtbare discussie kan plaatsvinden over de mogelijkheden en onmogelijkheden. In dit artikel breng ik het begrip 'benchmarken' terug naar de oorspronkelijke context en plaats het in de context van andere doelen en mogelijkheden van het gebruik van uitkomstdata.

### AUTEUR

**MARKO BARENDREGT**, theoretisch psycholoog, senior onderzoeker, Stichting Benchmark GGZ (SBG).

### CORRESPONDENTIEADRES

Dr. M. Barendregt, Stichting Benchmark GGZ, Rembrandtlaan 46, 3723 BK Bilthoven.

E-mail: Marko.barendregt@sbggz.nl

Geen strijdige belangen meegedeeld.

Het artikel werd voor publicatie geaccepteerd op 20-1-2015.

### De basis van benchmarken

Benchmarken behoort tot het domein van kwaliteitsmanagement en kan het beste gezien worden als een strategie voor kwaliteitsverbetering. Het is ontwikkeld door Richard Camp, wiens opdrachtgever Xerox kampte met een teruglopend marktaandeel. Vergeleken met andere strategieën voor kwaliteitsverbetering kenmerkt benchmarken zich door het leren van externe vergelijking. De vicepresident van Xerox beschreef de situatie aldus: *'we were horrified to find that Japanese were selling their machines at what it cost us to make ours... we had been benchmarking against ourselves. We weren't looking outside.'* (Omachonu & Ross 2004, p. 142).

Door een naar buiten gerichte oriëntatie profiteer je van kennis die elders is opgedaan. Reacties als 'Het kan niet, want dat hebben we nog nooit gedaan', worden met benchmarken omgedraaid in 'Het moet wel kunnen, want zij kunnen het ook'. Benchmarken is verankerd in *total quality management* en *continuous quality improvement* (Dale e.a. 2007). Kwaliteitsverbetering vindt plaats door een cyclisch proces van implementatie van verbeterplannen (*plan*), het uitvoeren daarvan (*do*), tussentijds monitoren of de gewenste verandering optreedt (*check*) en vervolgens analyse van de nieuwe spiegelinformatie (*act*) als input voor een nieuwe verbeteringsronde (de plan-do-check-act-cyclus). Het is een proces van het vinden van *best practices*, deze te onderzoeken op toepasbaarheid en ze toe te passen binnen de eigen organisatie (Wilson & Nathan 2003). In de checkfase worden uitkomstgegevens van verschillende instellingen met elkaar vergeleken. In de analysefase komen vragen aan bod als: is de ander beter en waarom is de ander beter? Voor benchmarken is de waaromvraag de belangrijkste, maar ook de moeilijkste om te beantwoorden. Geadviseerd wordt om in elk geval te kijken naar verschillen in procedures, managementsystemen, evaluatiesystemen, personeelsbeleid, cultuur en structuur van de organisatie (Meade 1998).

In de ggz betekent dit kijken naar toegepast behandelprotocol en protocoltrouw, timing van behandelplanbesprekingen, wachtlijstbeleid, organisatie van intervisie en opleiding etc. Benchmarken leidt zo tot meer inzicht in de onderliggende processen van de eigen en andermans organisatie tegen het licht van de waargenomen verschillen in uitkomsten (Inglis 2005; Zairi 1994).

Analyse van onderliggende processen zonder dat daar iets mee wordt gedaan, leidt niet tot kwaliteitsverbetering. In de planfase worden beloftevolle werkprocessen daarom onderzocht op toepasbaarheid in de eigen organisatie. Werkprocessen van instellingen met een betere behandeluitkomst kunnen niet zonder meer één op één overgenomen worden omdat deze processen altijd ontstaan in antwoord op specifieke vragen binnen een specifieke cultuur in een specifieke omgeving. Wat voor de ene zorgaanbieder werkt, hoeft voor de andere zorgaanbieder nog niet te werken, simpelweg omdat de ene organisatie de andere niet is. Men moet dus niet zonder meer bedrijfsprocessen overnemen, maar altijd eerst analyseren of deze op de eigen organisatie passend te maken zijn (Alstete 1995). Nadat de nieuwe werkprocessen zijn ingevoerd (do), kijkt men of dit effect heeft op de uitkomsten (check), waarna een nieuwe ronde van analyse en adaptatie begint.

Vaak wordt benchmarken gezien als het vergelijken van instellingen of het creëren van prestatiestandaarden. Feitelijk is dit dus slechts één van de fases van benchmarken. Vergelijkingsinformatie is binnen benchmarken puur instrumenteel en ondergeschikt aan het proces van kwaliteitsverbetering (De Beurs 2011). Constaten dat de een beter is dan de ander, zonder je af te vragen waarom en hoe je dat kunt veranderen, is géén benchmarken. Wanneer de betekenis van benchmarken wordt verengd tot uitsluitend genereren van vergelijkingsinformatie komt benchmarken niet los van het opstellen van ranglijstjes van beste en slechtste instellingen zoals die wel eens in de media verschijnen (Van Hoorn e.a. 2006). Dergelijke ranglijstjes zijn een terecht doelwit van methodologische criticasters, die aangeven dat de wetenschappelijke basis voor de rangorde op z'n minst discutabel is. Het verwarren van benchmarken als kwaliteitsmanagement met het opstellen van prestatieranglijsten leidt tot *'naming and shaming'* die het draagvlak voor het gebruiken van vergelijkingsinformatie voor verbetering juist ondermijnt.

#### PSYCHOMETRISCHE KWALITEIT VERSUS INFORMATIEKWALITEIT

Eén van de kritiekpunten op benchmarken is dat het vergelijken van behandeluitkomsten niet wetenschappelijk verantwoord is (o.a. Sytema & Van der Krieke 2013; Van Os e.a. 2012). Voor benchmarken als kwaliteitsmanagement zijn de psychometrische vereisten aan de data echter niet zo streng als die men aan data stelt in zuiver wetenschap-

pelijk onderzoek. In het recent vastgestelde toetsingskader beschrijft het Zorginstituut Nederland (2014) de criteria die aan meetinstrumenten gesteld worden. Hierbij worden expliciet psychometrische vereisten gesteld aan meetinstrumenten indien deze gebruikt worden voor financiering of keuze-informatie (o.a. betrouwbaarheid, validiteit, populatievergelijkbaarheid en registratievergelijkbaarheid).

Voor kwaliteitsverbetering geldt dat discussies over praktijkvariatie in uitkomsten niet per se dood hoeven te slaan bij enige twijfel over de validiteit van de benchmarkinformatie, omdat andere gebruikersfactoren van minstens even groot belang zijn. De vergelijkingsinformatie dient immers niet voor waarheidsvinding, maar wordt pragmatisch en instrumenteel ingezet. Enige onzekerheid over de juistheid van vergelijkingsinformatie op instellingsniveau is acceptabel, zolang de informatie desondanks aanleiding biedt om naar de eigen en andermans processen te kijken. Bij benchmarken is de relevante vraag niet of de rangorde een perfect juiste weergave van de werkelijke kwaliteitsverschillen geeft, maar of de variatie in uitkomsten erin slaagt gebruikers tot actie aan te zetten om de kwaliteit te verbeteren. De bruikbaarheid van data voor een gegeven doel wordt in de informatica *information quality* genoemd (Kahn e.a. 2002).

Om zinvol te zijn zal de benchmarkinformatie ook aan andere criteria moeten voldoen: accuraatheid, volledigheid, consistentie, tijdigheid, uniekheid (Cykana e.a. 1996). Bovendien zal de informatie ook moeten appelleren aan de behoeftes van de gebruiker zoals begrijpelijkheid, interpreteerbaarheid, objectiviteit, geloofwaardigheid, beschikbaarheid, reputatie en gebruiksgemak (Lee e.a. 2002). Ook moet ze aansluiten bij het taalgebruik en de beleving van de gebruikers.

Hoewel benchmarken tolerantie vertoont wat betreft een mindere validiteit, is het geen *'anything goes'*. Gebruikers die met tegenvallende resultaten worden geconfronteerd, zullen vaak eerst twifelen aan de data zelf (Oudejans e.a. 2011). Binnen de ggz wordt vooral gewezen op vertekening door confounding en tekortschietende uniformiteit in meetmethodiek (Noom e.a. 2012). Naarmate de resultaten vaker worden gerepliceerd en zijn gecorrigeerd voor patiënt- en meetkenmerken, is men eerder geneigd de data als uitgangspunt te nemen om over kwaliteitsverbetering na te denken. Als ondergrens mag men eisen dat de vergelijkingsinformatie dus ten minste voldoende valide moet zijn om gebruikers ervan te overtuigen dat de informatie aanleiding is naar de eigen processen te kijken.

#### Ander gebruik van uitkomstdata

Er zijn drie manieren waarop ROM-data gebruikt kunnen worden: op individueel patiëntniveau, op geaggregeerd

groepsniveau en ten slotte als vergelijkingsinformatie tussen instellingen (zie ook Nugter & Buwalda 2012). Verschillende functies van ROM vereisen vaak een andere manier van teruggeven van data. Benchmarken is kwaliteitsmanagement en is een heel specifieke manier van gebruik van uitkomstdata, als check in een PDCA-cyclus. Het vereist dat behandeluitkomsten worden geaggregeerd op instellingsniveau en dat deze instellingsgemiddelden onderling vergeleken worden. Er zijn meer functies van ROM die vereisen dat instellingen direct met elkaar vergeleken worden, bijvoorbeeld uitkomstenfinanciering (selectieve inkoop) en keuze-informatie voor patiënten.

Andere functies van ROM maken weliswaar gebruik van geaggregeerde gegevens, maar vereisen niet per se een onderlinge vergelijking tussen instellingen (maatschappelijke verantwoording en wetenschappelijk onderzoek). Ook kunnen ROM-data op patiëntniveau teruggegeven worden aan de gebruiker binnen de behandeling. Schematisch is dit weergegeven in FIGUUR 1. In deze paragraaf worden deze andere functies van ROM beschreven. Het doel is niet een volledige beschrijving van deze functies te geven, maar om een context te bieden voor benchmarken en benchmarken af te bakenen van andere gerelateerde concepten.

#### BEHANDELEN EN BEGELEIDEN

Uitkomstdata die vergaard zijn binnen een behandeling kunnen ook op patiëntniveau worden teruggegeven ter ondersteuning van het klinische oordeel van de behande-

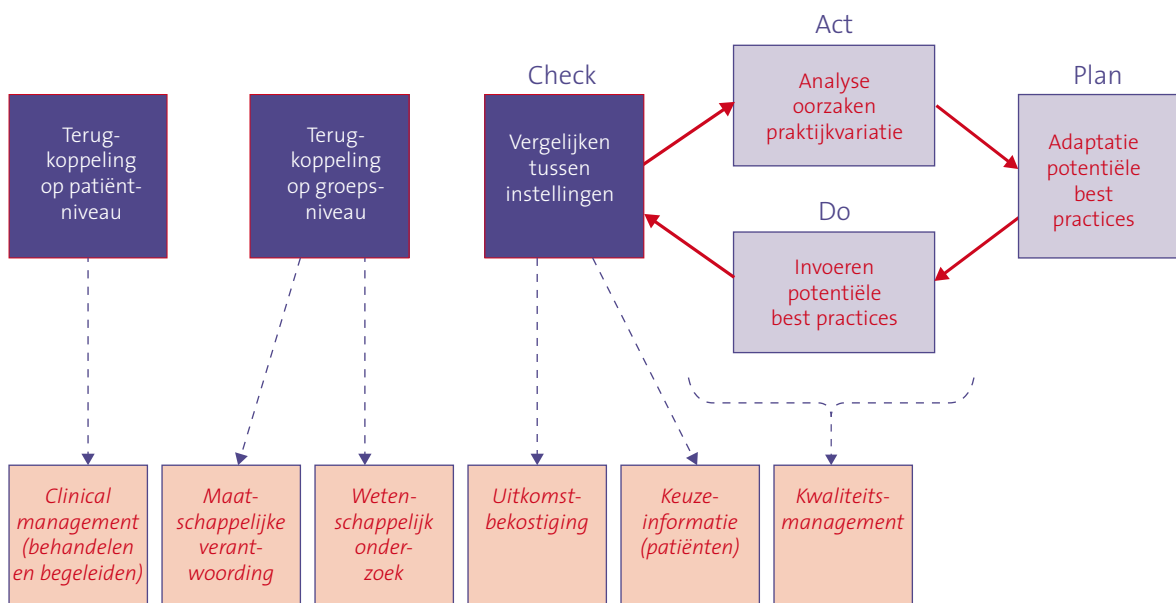
laar. Dit is klinisch beleid (*clinical management*), oftewel behandelen en begeleiden (Andrews & Page 2005). Bij de intake zijn ROM-gegevens dienstig bij de diagnostiek en gaandeweg de behandeling kan het monitoren van de voortgang op basis van uitkomstgegevens bijdragen aan een beter behandelresultaat, in het bijzonder door snelle detectie wanneer de behandeling niet aanslaat (Bickman e.a. 2012; Carlier e.a. 2012; De Jong e.a. 2014; Knaup e.a. 2009). Dit vergt dus een actieve rol van de behandelaar, die de uitkomstgegevens niet alleen moet vergelijken, maar deze ook moet monitoren. Uitkomstgegevens gelden voor de behandelaar als één, maar niet de enige, informatiebron bij het nemen klinische beslissingen (Andrews & Page 2005).

Het gebruik van uitkomstgegevens voor klinisch beleid is in de ggz vrij algemeen omarmd en anno 2014 weinig controversieel. Hafkenscheid en Van Os (2014), die dit 'sturende ROM' noemen, geven aan dat de uitkomstdata weliswaar multi-interpretabel zijn, maar dat dit geen probleem is wanneer de data louter gebruikt worden om 'een dialoog en vruchtbaar onderhandelingsproces tussen patiënt en behandelaar te bevorderen' (p. 22).

#### MAATSCHAPPELIJKE VERANTWOORDING

De samenleving investeert jaarlijks miljarden in de ggz. Transparant maken wat deze zorg oplevert aan gezondheidswinst is essentieel om het maatschappelijk draagvlak ervoor te behouden. Keer op keer blijkt de ggz kwetsbaar voor een bezuinigende overheid: aanpassingsstoornissen

FIGUUR 1 Terugkoppeling van uitkomstgegevens en functies van ROM



uit het basispakket, invoering eigen bijdrage, transitie van de jeugd-ggz. Het aantonen van pre-postverschillen geeft invulling aan maatschappelijke verantwoording. Voor de ggz als branche kan een landelijk gemiddelde worden berekend dat laat zien in welke mate de patiënten zijn verbeterd aan het eind van de behandeling ten opzichte van het begin van de behandeling. Op een lager niveau van aggregatie kan dit ook op instellingsniveau. Als de ggz bijvoorbeeld aantoont dat bij 60% van de patiënten na een half jaar behandelen de klachten aanzienlijk zijn afgenomen, kan de ggz daarmee haar bestaansrecht onderstrepen. Een belangrijk verschil tussen maatschappelijke verantwoording en benchmarken is dat bij verantwoord niet per se een vergelijking plaatsvindt tussen instellingen. Een geaggregeerd gemiddeld behandelingseffect geeft, ook zonder directe vergelijking tussen instellingen, een beeld van wat de behandeling heeft opgebracht.

Als ROM ten behoeve van de klinische behandeling voldoende is geïmplementeerd, ontstaat er van nature een dataset om te gebruiken voor geaggregeerde uitspraken. Daarmee kan men een begin maken met maatschappelijke verantwoording. Hoewel methodologisch niet uitgesloten kan worden dat (een deel van) het effect ook aan andere factoren dan de behandeling toegeschreven kan worden (spontane remissie, 'hello-goodbye'-effect), worden het definiëren en verzamelen van uitkomstendata gezien als een belangrijke eerste stap in het inzichtelijk maken van de maatschappelijke opbrengst (Nicholls e.a. 2012).

Voorzichtigheid in de interpretatie van behandelingseffecten is wel aan de orde. Voor zinvolle uitspraken over een patiëntenpopulatie is het nodig dat de groep waarvoor uitkomstmetingen beschikbaar zijn representatief is voor de totale populatie. Helaas zijn er altijd patiënten die zich niet goed laten meten of die geen vragenlijsten willen invullen. Als deze patiënten anders zijn dan patiënten van wie wel uitkomstmetingen beschikbaar zijn, ontstaat er vertekening (selectieve uitval). Om uitspraken op groepsniveau goed te kunnen duiden, zal er altijd informatie beschikbaar moeten zijn in welke mate de groep patiënten met beschikbare uitkomstgegevens op relevante kenmerken representatief is voor de totale groep. Met name selectieve uitval vanwege het ontbreken van metingen van patiënten die voortijdig de behandeling staken, is een probleem, omdat het leidt tot overschatting van het behandelingseffect.

Ook voor het uitblijven van behandelingseffect kunnen er goede verklaringen zijn. Een instelling die zich actief inzet om zorgmijders in zorg te houden middels bemoeizorg zal bijvoorbeeld vermoedelijk weinig verbetering laten zien. Het maatschappelijk nut van de behandeling is bij deze subgroep waarschijnlijk meer zichtbaar op andere domeinen dan klachtenreductie.

Selectieve uitval vormt niet alleen bij naturalistisch vergaarde data een uitdaging, maar ook bij de door Hafkenscheid en Van Os (2014) bepleite steekproeftrekkingen. Bij steekproeven is selectieve uitval mogelijk nog uitdagender doordat patiënten wellicht wel willen meewerken aan uitkomstmetingen wanneer deze worden gebruikt in de behandeling, maar niet als de gegevens alleen voor onderzoek worden verzameld en zij er geen feedback over ontvangen.

#### WETENSCHAPPELIJK ONDERZOEK

Vergaarde uitkomstmetingen zijn bruikbaar voor wetenschappelijk onderzoek naar de effectiviteit van behandelingen. De gouden standaard voor effectstudies is de zogenaamde gerandomiseerde gecontroleerde trial (RCT), waarbij patiënten willekeurig worden toegewezen aan een experimentele behandelgroep of een controlegroep. Voor het aantonen van de werkzaamheid (*efficacy*) van een behandeling gelden strenge methodologische vereisten. Zo worden in- en exclusiecriteria voor de studie nauwkeurig gedefinieerd en moet men de behandeling conform een protocol uitvoeren. Hierdoor is de interne validiteit van een RCT erg hoog, maar staat de externe validiteit onder druk. De gemiddelde patiënt in de klinische praktijk voldoet vaak niet aan in- en exclusiecriteria en ook zal de behandeling in de praktijk vaak niet zo zorgvuldig het protocol volgen als binnen een RCT. Het is daardoor nog maar de vraag of de onder ideale omstandigheden werkzame behandeling ook in werkelijkheid effectief is (Marley 2000).

RCT's worden daarom vaak aangevuld met observationeel onderzoek. Hierbij kijkt men of de behandelingen die in de gecontroleerde situatie werkzaam (*efficacious*) zijn, ook in de klinische praktijk effectief blijken (*effectiveness*). Net als bij een RCT zijn hierbij voor- en nametingen nodig. De uitkomstmetingen die gebruikt worden binnen een behandeling vormen hiervoor een nuttige bron (De Beurs & Barendregt 2010).

#### UITKOMSTBEKOSTIGING VAN DE ZORG

De financiering van de zorg vindt momenteel vooral plaats op basis van aantallen verrichtingen ('prestatiebekostiging'). Er is behoefte aan zorgfinanciering die zich meer richt op de kwaliteit van de geleverde zorg. Bekostiging van instellingen op basis van de uitkomst van de behandeling wordt 'uitkomstbekostiging' genoemd (Hayen e.a. 2013). Uitkomsten zijn, naast proces- en structuurkenmerken, belangrijke kwaliteitsindicatoren (Donabedian 1988) op basis waarvan verzekeraars zorg kunnen inkopen. Volgens velen is de uitkomst van de behandeling zelfs de belangrijkste kwaliteitsindicator (Porter 2010). Het idee is daarbij

dat instellingen onderling gaan concurreren op de geleverde kwaliteit (Laane & Luijk 2012). Er zijn verschillende vormen van uitkomstbepaling zoals de bonus-malusregeling op basis van het al dan niet behalen van een vooraf vastgestelde uitkomstdoelstelling, en selectieve inkoop waarbij de verzekeraar op basis van uitkomstindicatoren aanpassingen doet in het volume of de prijs van de ingekochte zorg.

Bij uitkomstbepaling zijn de behaalde score op de uitkomstmaat en de plaats op de rangorde niet vrijblijvend, maar hebben financiële consequenties. Dit vergt dat behandeluitkomsten nauwkeurig vastgesteld worden voor een juist antwoord op de vraag wie goede en wie minder goede zorg levert. In verhouding tot benchmarken is het hier dus uitermate belangrijk dat de rangordering op basis van ROM-data overeenkomt met de werkelijkheid. Om die reden gelden er strenge methodologische criteria voor uitkomstbepaling. Het is de vraag of naturalistische data ooit aan dergelijke criteria kunnen voldoen.

Daar staat tegenover dat de zorgverzekeraar de wettelijke taak heeft in te kopen op basis van kwaliteit en hiervoor de geschiktste instrumenten zal uitzoeken. De vraag is niet alleen of behandeluitkomsten als indicator van kwaliteit valide genoeg is, maar of deze meer valide is dan de alternatieven. Het is immers beter om het zorgbudget te verdelen op basis van een instrument met gebrekkige validiteit dan om het zorgbudget te verdelen op volstrekt niet-inhoudelijke of willekeurige informatie. De Algemene Rekenkamer (2013) heeft vastgesteld dat 0% van de alternatieve prestatie-indicatoren goed scoort op methodologische criteria. Het zou daarom best kunnen blijken dat de via outcome measurement verkregen uitkomstdata, ondanks alle beperkingen, als een van de accuratere en daarmee als bruikbaarste kwaliteitsindicator uit de bus komen.

### KEUZE-INFORMATIE VOOR PATIËNTEN

In het huidige model van gereguleerde marktwerking, waarin vraagsturing centraal staat, wordt ook aan patiënten een sturende kracht toebedeeld. De achterliggende rationale is dat vrije keuze van de patiënt, zorgaanbieders en zorgverzekeraars stimuleert om op goede en efficiënte zorg te concurreren. Transparantie over de kwaliteit is een voorwaarde om dit deel van de marktwerking te laten slagen (Damman e.a. 2010; Sauter 2009). Het is belangrijk dat informatie waarop de patiënt zijn of haar keuze baseert goed is; anders is het gevaar dat de patiënt niet de meest passende behandeling krijgt. Keuzegedrag door patiënten kan bovendien, net als uitkomstbepaling, financiële consequenties hebben voor zorgaanbieders en is dus niet vrijblijvend. Dat vereist dus, net als bij uitkomstbepaling, dat de informatie op basis waarvan de patiënt een

keuze maakt aan veel strengere wetenschappelijke vereisten voldoet dan nodig is voor kwaliteitsmanagement.

Goede keuze-informatie voor patiënten is nog complexer dan informatie bieden voor uitkomstbepaling. Voor uitkomstbepaling wil de financier van de zorg een betrouwbare schatting van de gemiddelde kwaliteit van een instelling. Een individuele patiënt die een behandeling zoekt, heeft echter weinig aan een instellingsgemiddelde en wil een betrouwbare schatting van de verwachte behandeluitkomst voor zijn specifieke klachten. Statistiek leent zich slecht voor het doen van individuele voorspellingen. Bovendien ontstaan door het opnemen van individuele achtergrondkenmerken erg kleine subgroepen waarop een dergelijke voorspelling kan worden gebaseerd. Hierdoor is het opstellen van echt goede keuze-informatie voor patiënten erg ingewikkeld en staat de ontwikkeling ervan nog in de kinderschoenen.

De mate waarin keuzevrijheid voor patiënten zal gaan werken als sturend naar efficiëntere zorg lijkt overigens beperkt (Delnoij 2008), omdat patiënten, ook bij voldoende aanbod en goede keuze-informatie, niet altijd rationele keuzes maken op basis van kwaliteit (Sauter 2009; Schaufli & Mordavsky 2001).

### CONCLUSIE

Het was dienstig geweest voor een breder draagvlak als bij de opzet van de landelijke uitkomstendatabase de doelstellingen door de betrokken partijen explicieter gedefinieerd en duidelijker afgebakend waren. Daarmee had wellicht veel onrust, verwarring en controverses voorkomen kunnen worden. Te vaak nog worden het vergelijken tussen instellingen, benchmarken, verantwoorden en 'afrekenen' over één kam geschoren en als methodologisch onverantwoord of onwetenschappelijk terzijde geschoven. Als we benchmarken terugbrengen naar waar het ooit vandaan kwam, is er weinig reden voor controverse. Benchmarken is een strategie voor het ontdekken van best practices en kwaliteitsverbetering (de Beurs 2011).

Voor uitkomstbepaling gelden andere methodologische vereisten waaraan vooralsnog niet voldaan wordt. Het feit dat de huidige dataverzameling de wetenschappelijke toetssteen niet kan doorstaan, maakt uitkomstbepaling echter niet onmogelijk. Uitkomstgegevens zijn de belangrijkste vorm van kwaliteitsinformatie; uiteindelijk gaat het voor de patiënt om de uitkomst van de zorg en om de vraag of die er beter van geworden is of niet. Op dit moment zijn de via outcome measurement verkregen data de enige landelijk beschikbare vergelijkingsinformatie op het gebied van uitkomsten in de ggz. Gezien de nog heersende twijfels aan de validiteit is het verstandig veel slagen om de arm te houden bij het gebruik van deze informatie voor bekostiging en dit slechts toe te passen op een beperkt

deel van het zorgbudget. Maar het zou evenzeer onverantwoord zijn om de beschikbare uitkomstgegevens volledig terzijde te schuiven bij de verdeling van de middelen.

Een belangrijke vraag is of uitkomstbepaling en benchmarken überhaupt wel samengaan, of dat het risico van financiële (of publieke) afrekening de intrinsieke prikkel om te benchmarken in het veld ondermijnt, zoals gesteld door de Gezondheidsraad (2006). Hier gelden niet zozeer methodologische vraagstukken, maar vooral sociale. Vult men de vragenlijst wel eerlijk in als er naast kwaliteitsverbetering ook euro's tegenover staan en zo niet, wat betekent dat voor de betrouwbaarheid van de uitkomstgegevens?

De boodschap is dat kwaliteitsverbetering altijd op de eerste plaats moet staan: dit kan nu al met de verzamelde gegevens, is goed voor de patiënt en voor de doelmatigheid

van de ggz en dus ook voor de risicodragende financier. Benchmarken mag meer gestimuleerd worden, omdat op dit moment de uitkomstgegevens nog onvoldoende binnen kwaliteitscycli gebruikt worden. Verdergaande vormen van uitkomstbepaling zijn vanwege de hogere methodologische criteria vooralsnog niet aan te raden. De taak van de wetenschap is om de methodiek steeds verder door te ontwikkelen of betere alternatieven te vinden. Het is te hopen dat alle betrokken partijen hiervoor de ruimte kunnen scheppen.

✍ Edwin de Beurs las het manuscript kritisch door en deed suggesties voor relevante rom-literatuur; de Wetenschappelijke Raad van sbg las het manuscript kritisch door en deed waardevolle suggesties.

## LITERATUUR

- Alstete JW. Benchmarking in higher education: adapting best practices to improve quality. ASHE-ERIC Higher Education Report No. 5. Washington: ERIC; 1995
- Andrews G, Page AC. Outcome measurement, outcome management and monitoring. *Aust N Z J Psychiatry* 2005; 39: 649-51.
- Beurs E de. Benchmarken, kansen en valkuilen. In: van Hees S, van der Vlist P, Mulder N, red. *Van meten naar weten; ROM in de ggz*. Amsterdam: Boom; 2011. p. 227-39.
- Beurs E de, Barendregt M. De evidence base van zorgprogramma's in de tbs; een visie op therapie-effectonderzoek. *Proces* 2010; 89: 331-43.
- Bickman L, Kelley SD, Athay M. The technology of measurement feedback systems. *Couple Family Psychol* 2012; 1: 274-84.
- Blijd-Hoogewys E, van Dijk R, Emmelkamp P, Mulder N, Oude Voshaar R, Schippers G, e.a. Benchmarken is 'werk-in-uitvoering'. *Tijdschr Psychiatr* 2012; 54: 1031-8.
- Bruinisma CL, Verbraak MJPM, Verbraak P. Transparantie in ggz gebaat bij ROM en benchmarking. *Tijdschr Psychiatr* 2012; 54: 254-6.
- Carlier IV, Meuldijk D, van Vliet IM, van Fenema EM, van der Wee NJ, Zitman FG. Empirische evidence voor de effectiviteit van routine outcome monitoring; een literatuuronderzoek. *Tijdschr Psychiatr* 2012; 54: 121-8.
- Cykana P, Paul A, Stern M. DoD Guidelines on Data Quality Management. *IQ* 1996; 154-71.
- Dale BC, van der Wiele T, van Iwaarden J van. *Managing quality*. Malden: John Wiley; 2013.
- Damman OC, Hendriks M, Delnoij DM. Keuze-informatie op basis van patiëntenervaringen: aanbevelingen en dilemma's. *TSG* 2010; 88: 396-405.
- Jong, K de, Timman R, Hakkaart-Van Roijen L, Vermeulen P, Kooiman K, Passchier J e.a. The effect of outcome monitoring feedback to clinicians and patients in short and long-term psychotherapy: A randomized controlled trial. *Psychother Res* 2014; 24: 629-39.
- Delespaul P. Hulpverlening – 'Back to basics'. In: van Hees S, van der Vlist P, Mulder N, red., *Van meten naar weten; ROM in de ggz*. Amsterdam: Boom; 2011. p. 241-9.
- Delnoij D. Zicht op kwaliteit; transparantie in de zorg vanuit patiëntenperspectief. Rede, uitgesproken op 15 mei 2008.
- Donabedian A. The quality of care: How can it be assessed? *JAMA* 1988; 260: 1743-8.
- Franx G. Sturen op resultaten; ROM in de USA. *MGv* 2014; 69: 12-8.
- Gezondheidsraad/RVZ. *Vertrouwen in verantwoorde zorg? Effecten van en morele vragen bij het gebruik van prestatie-indicatoren*. Den Haag: Centrum voor ethiek en gezondheid; 2006.
- Hafkenscheid A, van Os J. Naar een deugdelijke ROM. *MGv* 2014; 69: 20-8.
- Ham M van, Reitsma. Context – taal en vorm. In: van Hees S, van der Vlist P, Mulder N, red. *Van meten naar weten; ROM in de ggz*. Amsterdam: Boom; 2011. p. 27-31.
- Hayen AP, Bekker PJGM de, Ouwens MMTJ, Westert GP, Jeurissen PPT. No cure, no pay? Onderweg naar uitkomstbepaling in de Nederlandse zorg; Huidige en toekomstige mogelijkheden. Nijmegen: Celsus; 2013.
- Inglis A. Quality improvement, quality assurance, and benchmarking: comparing two frameworks for managing quality processes in open and distance learning. *International Review of Research in Open & Distance Learning* 2005; 6: 1-13.
- Janssen M, Wensing M, van Deursen P, Cornelissen I, van der Gaag RJ, Buitelaar J. Niet rammen maar ROMmen; dwingende ROM -responspercentages en methodieken: kan het anders? *MGv* 2014; 69: 29-32.

- Kahn BK, Strong DM, Wang RY. Information quality benchmarks: product and service performance. *Communications of the ACM* 2002; 45: 184-92.
- Knaup C, Koesters M, Schoefer D, Becker T, Puschner B. Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. *Br J Psychiatry* 2009; 195: 15-22.
- Laane R, Luijk R. ROM en positie van de zorgverzekeraars. *Tijdschr Psychiatr* 2012; 54: 135-9.
- Lee YW, Strong DM, Kahn BK, Wang RY. AIMQ: a methodology for information quality assessment. *Information & management* 2002; 40: 133-46.
- Marley J. Efficacy, effectiveness, efficiency. *Australian Prescriber* 2000; 23: 114-5.
- Massoud R, Askov K, Reinke J, Franco LM, Bornstein T, Knebel E, e.a. A modern paradigm for improving healthcare quality. Bethesda: Quality Assurance Project; 2001.
- Algemene Rekenkamer. Indicatoren voor kwaliteit in de zorg. Tweede Kamer, vergaderjaar 2012-2013, 33 585, nr. 22013. Den Haag: Sdu; 2013.
- Meade P. A guide to benchmarking. Dunedin: University of Otago; 1998.
- Nicholls J, Lawlor E, Neitzert E, Goodspeed T. A guide to social return on investment. London: Office of the Third Sector, The Cabinet Office; 2012.
- 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 A, Buwalda V. Achtergronden en gebruiksmogelijkheden van ROM in de ggz. *Tijdschr Psychiatr* 2012; 54: 111-20.
- Omachonu VK, Ross JE. Principles of total quality. Boca Raton: CRC Press; 2004.
- 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.
- Oudejans S, Spits M, Schippers G. Leren van uitkomsten – van taseki naar jiseki. In: van Hees S, van der Vlist P, Mulder N, red. *Van meten naar weten; ROM in de ggz*. Amsterdam: Boom; 2011. p. 35-43.
- Porter ME. What is value in health care? *N Engl J Med* 2010; 363: 2477-81.
- Sauter W. Marktwerving in de zorg – Toezicht: met oog op de consument. Rede, uitgesproken op 6 februari 2009.
- Schaffler HH, Mordavsky JK. Consumer reports in health care: do they make a difference? *Annu Rev Public Health* 2001; 22: 69-89.
- Sytema S, Kriek L van der. Routine outcome monitoring: A tool to improve the quality of mental health care? In: Thornicroft G, Ruggeri M, Goldberg D, red. *Improving mental health care: the global challenge*. Chichester: John Wiley; 2013. p. 246-263.
- Hoorn A van, Houdenhoven M van, Wullink G, Hans E, Kazemier G. Een nieuw stappenplan voor benchmarking. *Management Executive* 2006; 1-15.
- Vlist P van der. Indeling en leeswijzer. In: van Hees S, van der Vlist P, Mulder N, red. *Van meten naar weten; ROM in de ggz*. Amsterdam: Boom; 2011. p. 17-24.
- Wilson A, Nathan L. Understanding benchmarks. *Home Healthc Nurse*, 2003; 21: 102-7.
- Zairi M. Measuring performance for business results. Londen: Chapman & Hall; 1994.
- Zorginstituut Nederland. Toetsingskader kwaliteitsstandaarden en meetinstrumenten. <http://www.zorginstituutnederland.nl/kwaliteit/toetsingskader+en+register/toetsingskader> (geraadpleegd op 6 oktober 2014); 2014.



## SUMMARY

# Benchmarking and other functions of ROM: back to basics

M. BARENDREGT

**BACKGROUND** Since 2011 outcome data in the Dutch mental health care have been collected on a national scale. This has led to confusion about the position of benchmarking in the system known as routine outcome monitoring (ROM).

**AIM** To provide insight into the various objectives and uses of aggregated outcome data.

**METHOD** A qualitative review was performed and the findings were analysed.

**RESULTS** Benchmarking is a strategy for finding best practices and for improving efficacy and it belongs to the domain of quality management. Benchmarking involves comparing outcome data by means of instrumentation and is relatively tolerant with regard to the validity of the data. Although benchmarking is a function of ROM, it must be differentiated from other functions of ROM. Clinical management, public accountability, research, payment for performance and information for patients are all functions of ROM which require different ways of data feedback and which make different demands on the validity of the underlying data.

**CONCLUSION** Benchmarking is often wrongly regarded as being simply a synonym for 'comparing institutions'. It is, however, a method which includes many more factors; it can be used to improve quality and has a more flexible approach to the validity of outcome data and is less concerned than other ROM functions about funding and the amount of information given to patients. Benchmarking can make good use of currently available outcome data.

TIJDSCHRIFT VOOR PSYCHIATRIE 57(2015)7, 517-525

**KEY WORDS** benchmarking, quality improvement, routine outcome monitoring (ROM)