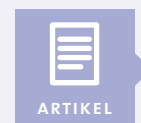


# Over ROM en benchmarken; reactie van SBG op rapport van de Algemene Rekenkamer en de actie ‘Stop ROM’

E. DE BEURS



De Algemene Rekenkamer (AR) heeft onderzoek gedaan naar de bekostiging van de ggz. Ze evalueerde prestatiebekostiging en onderzocht of uitkomstenbekostiging per 2020 een realistisch voornemen is. De AR heeft geadviseerd om voor de invoering van uitkomstenbekostiging meer tijd te nemen, omdat 1. er vooralsnog niet veel terecht is gekomen van marktwerking in de ggz, 2. het veld nog ervaring op moet doen met zorgstandaarden en 3. de uitkomstgegevens van Stichting Benchmark GGZ (SBG) nog onvoldoende volledig en betrouwbaar zijn voor dit doel (Algemene Rekenkamer 2017). De betrouwbaarheid van de informatie van SBG is overigens niet empirisch onderzocht door de AR (over de betrouwbaarheid zie TABEL 1).

Een deel van het ggz-veld heeft uit het rapport van de AR geconcludeerd dat uitkomstenbekostiging niet kan, uitkomst van zorg niets zegt over de kwaliteit van zorg en dat SBG beter kan stoppen met haar ambitie kwaliteit te meten ([www.stoprom.com](http://www.stoprom.com)). Die conclusie staat echter niet in het rapport van de AR, die juist lovend is over de stappen die in de ggz zijn gezet met ROM en de inmiddels bereikte transparantie over behandeluitkomsten en spreekt over ‘*één taal voor het meten van kwaliteit*’ (p. 19). De AR concludeert wel dat ROM nu nog niet geschikt is voor uitkomstenbekostiging, ze adviseert tegen overhaaste invoering van uitkomstenbekostiging op basis van ROM en verwijst naar het toetsingskader van Zorginstituut Nederland voor het eindoordeel over ROM (p. 46).

De huidige discussie over nut van ROM en benchmarken spitst zich toe op twee thema's: het onderscheid tussen ‘goede’ en ‘foute’ ROM en het gebruik van ROM-gegevens als inkoopinformatie voor verzekeraars en als keuze-informatie voor patiënten. De discussie zelf wordt gekenmerkt door begripsverwarring over wat met ROM en benchmarken bedoeld wordt en wat het doel is. ROM, benchmarken, verantwoord en afrekenen worden alle op één hoop

gegooid, maar het zijn echter zeer verschillende begrippen. Ook is er verwarring en verschil van mening over de rol die de verzekeraar zou moeten spelen en wat er met prestatie- of uitkomstenbekostiging in de zorg wordt bedoeld. Dat helpt de discussie niet verder. Voor de helderheid geef ik eerst een korte uitleg van de begrippen.

## Definitie begrippen en minimale ROM

ROM is regelmatig de uitkomst van de behandeling bepalen en is primair bedoeld om de behandeling bij de individuele patiënt te ondersteunen. Door klachten of functioneren frequent te meten houd je een vinger aan de pols bij de patiënt en kan je samen de verdere koers van de behandeling bepalen (gedeelde besluitvorming; Metz e.a. 2015). ROM kan helpen om mislukking van een behandeling te voorkomen (Lambert 2010) en kan ook helpen bij het realiseren van gepast gebruik (ministerie van vws 2015). *Benchmarken* is je uitkomsten vergelijken met andere – vergelijkbare – instellingen om er iets van te leren, bijvoorbeeld waar je verbeterpotentieel zit (Barendregt 2015). In de nieuwjaarsgroet van SBG legt Jeroen Muller, voorzitter raad van bestuur van Arkin, kort uit hoe benchmarken in zijn instelling heeft gewerkt (<https://youtube/y4ncou05NjA?t=79>).

*Verantwoord* is aan de buitenwereld laten zien wat de ggz oplevert. Prestaties van zorgaanbieders moeten voor iedereen beschikbaar zijn; dat hoort bij ons zorgstelsel en is bij wet zo geregeld ([www.zorginstituutnederland.nl](http://www.zorginstituutnederland.nl)). Met *afrekenen* bedoelt men uitkomstenbekostiging, ook wel bekend als *pay-for-performance* (P4P). Dit is overigens iets anders dan *prestatiebekostiging*; dat is het vergoeden per verrichting, dat met de systematiek van de diagnose-behandelcombinatie (DBC) in 2013 in de ggz is geïntroduceerd. P4P betekent zorgaanbieders belonen voor het leveren van goede kwaliteit (Eijkenaar & Schut 2015). Al

deze zaken - ROM, benchmarken, verantwoorden en uitkomstenbepaling - zijn erop gericht de kwaliteit van de zorg te verbeteren (de Beurs 2015).

Sinds 2010 is, ingevolge een afspraak tussen de zorgaanbieders en -verzekeraars, succesvol in het gehele land een minimale variant van ROM ingevoerd met generieke meetinstrumenten en een meting rondom een DBC. Op deze variant van ROM is veel kritiek van hulpverleners en die kritiek is terecht: de hulpverlener heeft weinig aan een jaarlijkse pre-postmeting om de behandeling van een individuele patiënt bij te sturen. Dat generieke meetinstrumenten ongeschikt zijn voor ROM is overigens niet juist: met instrumenten als de OQ-45 of de HONOS is bij de meeste patiënten goed te meten of zij baat hebben bij de behandeling (de Jong e.a. 2007; Mulder e.a. 2004). Voor goede ROM moeten we wel vaker meten, maar hier is de hulpverlener zelf aan zet. Die kan met de huidige infrastructuur voor ROM veel frequenter meten en kan desgewenst het generieke meetinstrument aanvullen met (stoornisspecifieke) schalen die optimaal zijn afgestemd op de problematiek van de patiënt.

### Inkoopinformatie en keuze-informatie voor patiënten

Een heikel punt in de huidige discussie is of het wenselijk is dat de ROM-informatie wordt gebruikt bij de bekostiging van de ggz of om patiënten te informeren en te helpen bij het kiezen van een zorgaanbieder. Wie zou niet vooraf willen weten hoe lang een behandeling gemiddeld duurt en welke uitkomst er te verwachten is bij de zorgaanbieder van zijn of haar keuze? Een gunstig neveneffect van zulke transparantie kan zijn dat zorgaanbieders extra gemotiveerd raken om goede resultaten te behalen (Greenhalgh e.a. 2017).

Tegenstanders stellen nu dat de ROM-gegevens een veel te beperkte kijk op kwaliteit van zorg bieden. Die omvat immers meer aspecten dan alleen medische effectiviteit of uitkomst. Dat is natuurlijk waar, maar ook staat vast dat de uitkomst van de behandeling op zijn minst een belangrijk onderdeel is van kwaliteit van zorg en uiterst relevant is voor patiënten in de ggz.

Onder invloed van het gedachtegoed van Michael Porter (2009; 2010) is de uitkomst van de behandeling centraal komen te staan in het denken over de kwaliteit van zorg. Volgens Porter is het grootste probleem in de gezondheidszorg dat de prikkels verkeerd gericht zijn. Een financieringssysteem waarbij betaald wordt per verrichting zet aan tot zo veel mogelijk doen, niet tot zo goed mogelijk doen. Als alternatief voor prestatiebekostiging stelt Porter uitkomstenbepaling voor. Het moet daarbij gaan om de behaalde gezondheidswinst afgezet tegen de gemaakte kosten: *value-based health care* (VBHC) op basis van wat voor

### AUTEUR

**EDWIN DE BEURS**, hoofd wetenschappelijk onderzoek SBG en hoogleraar ROM en Benchmarken, Universiteit Leiden.

### CORRESPONDENTIEADRES

Prof. dr. Edwin de Beurs, Faculteit Sociale Wetenschappen  
Universiteit Leiden, Wassenaarseweg 52, 2333 AK Leiden.  
E-mail: e.de.beurs@fsw.leidenuniv.nl

Geen strijdige belangen meegedeeld.

Het artikel werd voor publicatie geaccepteerd op 8-3-2017.

### TITLE IN ENGLISH

ROM and benchmarking; the reaction of SBG on the report of the Court of Audit and the action 'Stop ROM'

de patiënt van belang is. In de gezondheidszorg is een hele VBHC-beweging op gang gekomen en met talloze voorbeelden is aangetoond dat VBHC tot betere zorg en lagere kosten kan leiden (Porter e.a. 2017).

Voor beide zaken – patiënten informeren en uitkomstenbepaling – is betrouwbare informatie over de behandeluitkomst nodig. En daar zit een probleem (zie TABEL 1). De ggz is nog maar net aan het meten geslagen en dat kan nog veel beter en vollediger. Ook SBG is werk in uitvoering (Blijd-Hoogewys e.a. 2012) en werkt samen met haar wetenschappelijke raad en expertraden aan verbetering van de methodiek en de betrouwbaarheid van de informatie (de Beurs e.a. 2017). Het begint allemaal met goede ROM-gegevens. Wat dat betreft sluit ik me dus volmondig aan bij het pleidooi uit het veld voor een verbeterde variant van ROM. Dat kan echter beter bereikt worden door de bestaande ROM te hervormen dan door er geheel mee te stoppen.

### De zorgverzekeraar aan de haal met ROM?

Velen zijn nu bang dat de zorgverzekeraars met de ROM-gegevens aan de haal gaan en voorbarig en ondoordacht gaan inkopen op uitkomst: *pay-for-delta T* (P4ΔT). In de methodiek van SBG is ΔT de centrale prestatie-indicator; ΔT is een afgeleide maat van de effectgrootte van het verschil tussen een voor- en een nameting (de Beurs e.a. 2017). In de wetenschap zijn de meningen verdeeld over de vraag of uitkomstenbepaling tot betere zorg leidt (Eijkenaar & Schut 2015). Er zijn ook risico's. Uitkomstenbepaling kan gebruikt worden als middel om praktijkvariatie (Wennberg 2002) tegen te gaan en zo iedere patiënt een gelijke kans te bieden op even goede zorg.

Uitkomstenbekostiging kan echter praktijkvariatie ook juist vergroten (Eijkenaar & Schut 2015). Stel dat op het platteland de uitkomsten van zorg slechter zouden zijn dan in de stad. Nieuwe ontwikkelingen en nieuwe behandelingen die zijn ontwikkeld in universitaire centra hebben tijd nodig om verspreid te raken en dringen op het platteland moeizamer door. Met simpele uitkomstenbekostiging worden aanbieders in de stad beloond voor hun betere uitkomsten en aanbieders op het platteland krijgen minder geld. De verschillen zullen daardoor eerder toemenen dan afnemen.

Kortom, de informatie van SBG gebruiken in een P4 $\Delta$ T-model is een te simplistische benadering. Ten eerste moet je niet alleen kijken naar  $\Delta$ T, maar ook naar andere aspecten van kwaliteit van zorg, zoals veiligheid en patiëntervaring. Ten tweede moet je  $\Delta$ T bezien in de context waarin het resultaat behaald is (bij welke stoornis, met welke behandeling, tegen welke kosten en in welke omgeving). Dat vergt informatie over patiënt-, proces- en structuurfactoren. Een simplistische vorm van P4 $\Delta$ T kan ongewenste neveneffecten hebben, zoals ik met mijn denkbeeldig voorbeeld heb aangegeven.

Als zorgverzekeraars de uitkomst informatie gaan gebruiken, dan wil ik bij gebruik van inkoop informatie van SBG pleiten voor P4*improved*  $\Delta$ T: meerjarige afspraken maken met instellingen over verbetering van de zorg, en instellingen belonen met extra geld voor aantoonbaar betere resultaten dan in het verleden werden behaald.

### Tot slot: vertrouwen

Een voorwaarde voor een goed werkend uitkomstenbekostigingssysteem is onderling vertrouwen tussen de zorgaanbieders en -verzekeraars (Eijkenaar & Schut 2015). Betrouwbare informatie over de uitkomst van de behandeling afkomstig van een TTP (een derde vertrouwde partij) kan helpen om dit vertrouwen op te bouwen. En dat vertrouwen is noodzakelijk voor een financierbare en toekomstbestendige ggz, waarin we onze patiënten de best mogelijke behandeling tegen aanvaardbare kosten bieden.

**TABEL 1** De kwaliteit en validiteit van de gegevens van SBG

- Twijfel over de betrouwbaarheid van de informatie van Stichting Benchmark GGZ (SBG) berust vooralsnog alleen op theoretische overwegingen over de methodiek van benchmarken, en heeft betrekking op mogelijke bronnen van vertekening van de informatie (meetvariatie, confounding en onvoldoende representativiteit van de gegevens; van Os e.a. 2012).
- SBG heeft de afgelopen jaren veel onderzoek gedaan naar de invloed van meetinstrumenten en het meetmoment, naar casemixcorrectie en naar bias door selectieve inclusie of uitval (de Beurs e.a. 2017).
- Ons eigen onderzoek wijst uit dat boven een respons van 50% de gegevens vrij robuust zijn voor vertekening; de uitkomsten worden niet wezenlijk anders na statistische correctie voor casemixverschillen (Warmerdam e.a. 2017).
- Verder blijken de uitkomsten en de ranking van zorgaanbieders behoorlijk stabiel over de tijd, een klassiek teken van meetbetrouwbaarheid (intern rapport SBG).
- Uitkomsten van diverse meetinstrumenten en diverse manieren om uitkomsten te operationaliseren leiden tot dezelfde ranking, wat de validiteit van de methodiek ondersteunt (de Beurs e.a. 2016).
- Empirisch onderzoek naar de methodiek, de betrouwbaarheid en de validiteit van de uitkomstgegevens van SBG wordt bij voorkeur uitgevoerd door een onafhankelijke derde partij. Dat kan de huidige discussie over nut van benchmarken op een hoger plan brengen. SBG is vanzelfsprekend bereid ten volle mee te werken aan dergelijk onderzoek.

## LITERATUUR

- Algemene Rekenkamer. Bekostiging van de curatieve geestelijke gezondheidszorg. Den Haag: Algemene Rekenkamer; 2017. p. 58.
- Barendregt M. Benchmarken en andere functies van ROM: back to basics. Tijdschr Psychiatr 2015; 57: 517-25.
- Beurs E de, Barendregt M, Warmerdam L, red. Behandeluitkomsten: bron voor kwaliteitsbeleid in de GGZ. Boom, Amsterdam 2017.
- Beurs E de, Barendregt M, Heer A de, Duijn E van, Goeree B, Kloos M, e.a. Comparing methods to denote treatment outcome in clinical research and benchmarking mental health care. Clin Psychol Psychother 2016; 23: 308-18.
- Beurs E de, ROM en benchmarken, over meten, weten en wat dan? [oratie]. Leiden: Universiteit Leiden; 2015.
- Blijd-Hoogewys E, Dijk R van, Emmelkamp P, Mulder N, Oude Voshaar RC, Schippers G, e.a. Benchmarken is 'werk-in-uitvoering'. Tijdschr Psychiatr 2012; 54: 1031-8.
- Eijkenaar F, Schut E. Uitkomstbepaling in de zorg: een (on)begaanbare weg? Rotterdam: Erasmus Universiteit; 2015.
- Greenhalgh J, Dalkin S, Gooding K, Gibbons E, Wright J, Meads D, e.a. Functionality and feedback: a realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care. Health Services and Delivery Research 2017; 5.
- Jong K de, Nugter MA, Polak MG, Wagenborg JEA, Spinhoven P, Heiser WJ. The Outcome Questionnaire (OQ-45) in a Dutch population: A cross-cultural validation. Clin Psychol Psychother 2007; 14: 288-301.
- Lambert MJ. Prevention of treatment failure: the use of measuring, monitoring, and feedback in clinical practice. Washington: American Psychological Association; 2010.
- Metz MJ, Franx GC, Veerbeek MA, de Beurs E, van der Feltz-Cornelis CM, Beekman ATF. Shared decision making in mental health care using routine outcome monitoring as a source of information: a cluster randomised controlled trial. BMC Psychiatry 2015; 15: 1-10.
- Ministerie van VWS. Agenda ggz voor gepast gebruik en transparantie. Vergaderjaar 2015-2016, Tweede Kamer. Den Haag: Sdu; 2015. p. 20.
- Mulder C, Staring A, Loos J, Buwalda V, Kuijpers D, Sytema S, e.a. De Health of the Nation Outcome Scales (HONOS) als instrument voor 'routine outcome assessment'. Tijdschr Psychiatr 2004; 46: 273-84.
- Os J van, Kahn R, Denys D, Schoevers RA, Beekman AT, Hoogendijk WJ, e.a. ROM: gedragsnorm of dwangmaatregel? Overwegingen bij het themanummer over routine outcome monitoring. Tijdschr Psychiatr 2012; 54: 245-53.
- Porter ME. A strategy for health care reform - toward a value-based system. N Engl J Med 2009; 361: 109-12.
- Porter ME. What is value in health care? N Engl J Med 2010; 363: 2477-81.
- Porter ME, Kaplan RS, Frigo ML. Managing health care costs and value. Strategic Finance 2017; 98: 24.
- Warmerdam EH, Barendregt M, Beurs E de. Risk adjustment of self-reported clinical outcomes in Dutch mental health care. Journal of Public Health 2017; DOI:10.1007/s10389-017-0785-4.
- Wennberg JE. Unwarranted variations in healthcare delivery: implications for academic medical centres. BMJ 2002; 325: 961-4.