

Cijfers over prestatie-indicatoren: een wetenschappelijke bijsluiter¹

Tom van Yperen & Mariska van der Steege
Nederlands Jeugdinstituut / NJi
September 2008

Inleiding

Het bereiken van de resultaten van de hulp vormt in het primaire proces een cruciale toetssteen om al dan niet van geslaagde hulp te spreken. Het meten van die resultaten is dan ook een belangrijk onderdeel van de kwaliteitsbepaling van de hulp. Bijzonder veel instellingen in de jeugdzorg hebben de laatste jaren een start gemaakt met het meten van die resultaten.

Het gebruik van de meetgegevens kent echter verschillende valkuilen. In de notitie ‘Trefzeker tellen’ wees Van Yperen al op waarschuwingen van uiteenlopende auteurs².

- Verzameling van gegevens garandeert nog geen gebruik van gegevens in het kader van kwaliteitsbeleid.
- Een cultuur van meten en afrekenen de ondernemingszin en innovatie kan remmen. Innoveren is immers risico nemen. Als een mislukking negatief doortikt in de prestatiecijfers, kan dit de lust bederven om vernieuwend te werken.
- Het gebruik van prestatiemetingen kan professionals kan vervreemden van hun vak, omdat het alleen nog maar om bureaucratische cijfertjes lijkt te gaan.
- Het werken met prestatie-indicatoren kan perverterend werken: waarom zou een instelling bijvoorbeeld nog moeite doen voor doelgroepen die moeilijk te behandelen zijn als dat een negatieve invloed heeft op de prestatiecijfers?

Met het oog op dit soort waarschuwingen is het van belang dat prestatie-indicatoren moeten passen bij de ambitie om de kwaliteit van de zorg (verder) te verbeteren. Daarom is in hoofdstuk 2 van het raamwerk gesteld dat de indicatoren betrekking moeten hebben op zaken waarin partijen (de professionals, financiers, bestuurders) hun verantwoordelijkheden en bevoegdheden kennen om – voor zover nodig – verbeteringen mogelijk te maken en tot stand te brengen.

De inmiddels gegroeide praktijk met betrekking tot de hier verzamelde *resultaatgegevens* laat zien dat het interpreteren van de gegevens ook een aantal meer specifieke valkuilen kent. Deze hebben te maken met drie kwesties:

- A priori zeggingskracht van de gegevens over het effect van de hulp;
- Psychometrische kwesties, zoals de betrouwbaarheid en de validiteit;
- Complicaties bij het bundelen van gegevens naar een hoger aggregatieniveau (het zogeheten aggregatievraagstuk).

In deze notitie passeren in een notendop de belangrijkste valkuilen de revue en geven we handreikingen om ermee om te gaan.

¹ Dit is een bewerking van een eerder gepresenteerde bijsluiter specifiek over Doelrealisatie. Deze notitie is in belangrijke mate gebaseerd op Yperen, T.A. van, Kroes, G. & Bijl, B. (2008). Meten van doelrealisatie. In: T.A. van Yperen & J.W. Veerman (Red.). *Zicht op effectiviteit. Handboek voor praktijkgestuurd effectonderzoek in de jeugdzorg* (p. 271-292). Delft: Eburon.

² Yperen, T.A. van (2005). *Trefzeker tellen. Samenvatting bijeenkomsten VWS over prestatie-indicatoren over de jeugdzorg*. Utrecht: NIZW Jeugd. Te downloaden van www.nji.nl.

A priori zeggingskracht

Valkuilen

De zeggingskracht van cijfers over de resultaten van de hulp is op dit moment principieel beperkt van aard. Dat heeft te maken met een combinatie van drie belangrijke factoren.

- a. *Kwaliteit van het object van de metingen.* In de eerste plaats is de kwaliteit van het werk waarop de metingen betrekking hebben nog sterk in ontwikkeling is. Veel hulpverleners moeten bijvoorbeeld het stellen van goede doelen nog voldoende in de vingers krijgen. De kwaliteit van de prestatie-indicator 'Doelrealisatie' is van die stand van ontwikkeling sterk afhankelijk. Aan die ontwikkeling wordt in de sector hard gewerkt, maar dit proces is nog niet afgerond.
- b. *Determinanten van uitkomsten hulp.* De gegevens over de prestatie-indicatoren zeggen niet alles over de effectiviteit van de hulp. De uitkomsten van de hulp zijn immers slechts voor een deel toe te schrijven aan effectief methodisch handelen. Ook bijvoorbeeld cliëntfactoren en de kwaliteit van de relatie tussen hulpverlener en cliënt spelen een rol. Welk deel van de resultaten dan wel door de hulp wordt bepaald is – gegeven de stand van kennis – moeilijk te zeggen. Om een goede schatting te maken is strengwetenschappelijk onderzoek nodig. 'Platte' cijfers over de indicatoren werpen daar onvoldoende licht op.
- c. *Normen.* Gegevens hebben weinig zeggingskracht als normen ontbreken voor wat goed, matig of slecht is te noemen. Die normen zijn er op dit moment nog niet. Die zullen in discussies over de eerste gegevens over de prestatie-indicatoren moeten uitkristalliseren.

Hoe hier mee om te gaan?

Cijfers met betrekking tot de prestatie-indicatoren zijn te beschouwen als de eerste indicator voor de resultaten van hulp. Daarbij is het vooral van belang *dat* hulpverleners en instellingen cijfers over de resultaten verzamelen en rapporteren. Het stimuleert de kritische reflectie op de kwaliteit van het werk. Ook prikkelt het om initiatieven te nemen om de vragen die de cijfers oproepen te gaan beantwoorden. Stel dat een hulpvorm van een instelling laat zien dat slechts 45 procent van de cliënten de einddoelen realiseert. Is het dan niet nuttig om een uitgebreid gesprek te gaan voeren over wat hier aan de hand is? Is het bij zo'n lage score zinvol om een duur en intensief wetenschappelijk effectonderzoek te starten, of kijkt men liever eerst wat deze lage successcore kan verklaren en hoe de resultaten te verbeteren zijn? Zijn er instellingen die bij een soortgelijke hulpvorm veel betere resultaten halen? Het zijn nuttige vragen.

Van belang daarbij is de cijfers over de prestatie-indicatoren niet geïsoleerd te beoordelen, maar in relatie tot bijvoorbeeld gegevens over

- de zwaarte van de doelgroep (bij een gemakkelijke doelgroep is het gemakkelijk succes te hebben),
- de hulpvorm waar het hier om gaat (bijvoorbeeld geen lichte ambulante hulp, maar intensieve 24-uurs daghulp);
- de mate waarin het is gesteld met scholing van medewerkers in bijvoorbeeld het stellen van goede doelen en het effectief handelen bij bepaalde problemen van cliënten;
- de indicatoren voor de resultaten van de hulp in relatie tot elkaar (zoals de mate van doelrealisatie gecombineerd met de afname van de problematiek, de cliënttevredenheid, de mate waarin al dan niet vervolghulp nodig is etc.).

Psychometrische problemen

Valkuilen

Waar gemeten wordt, maakt men meetfouten. Met het stapelen van de gegevens, stapelt men ook de meetfouten op elkaar. Cijfers over de prestatie-indicatoren kunnen verstoord raken door bijvoorbeeld:

- *Instabiliteit van de metingen.* De cliënt of de hulpverlener willen gedurende het hulpverleningsproces nog wel eens de kwaliteit van de hulp wisselend waarderen. Op het ene moment zijn ze bijvoorbeeld positiever over de mate waarin de doelen zijn gerealiseerd dan op het andere moment, zonder dat er een daadwerkelijk verschil is de mate van realisatie is. En de ene keer vullen ze een tevredenheidsvragenlijst vrolijker in dan een andere keer. Dat zorgt ervoor dat de metingen teveel een momentopname kunnen zijn.
- *Gebrek aan consensus.* Resultaatgegevens zijn alleen geldig als de betrokkenen het eens zijn over wat het probleem was, waar de hulp op was gericht en de mate waarin de hulp daar een goed antwoord op was. Bij gebrek aan consensus moet helder zijn wie het geldige eindoordeel mag vellen. Niet iedereen hanteert daarbij dezelfde regels. Dit alles tast de betrouwbaarheid en validiteit van de cijfers aan.
- *Koppeling indicatiebesluit en uitvoering.* De geregistreerde resultaten moet een graadmeter zijn voor wat er met de hulp daadwerkelijk is bereikt. Voorwaarde is dan dat de hulp ook daadwerkelijk is gericht op het realiseren van het indicatiebesluit. De praktijk laat zien dat hier nog geregeld ‘mismatches’ in optreden, met name als het gaat om de doelen in het indicatiebesluit van bureau jeugdzorg en de bij de zorgaanbieder uitgevoerde hulp.

Hoe hier mee om te gaan?

Vanwege de meetfouten zijn cijfers over de resultaten slechts met een ruime foutenmarge te gebruiken. Bijvoorbeeld, als in een overzicht staat dat bij instelling A 70 procent van de doelen geheel zijn gerealiseerd, bij instelling B 60 procent en bij instelling C 45 procent, is het verstandig om rond deze getallen een marge te denken. Hoe groot die marge precies moet zijn is nu nog moeilijk te zeggen. Een veilige vuistregel zou kunnen zijn dat bij elk percentage er 10 procentpunten worden afgetrokken en opgeteld. Figuur 1 geeft een voorbeeld.

Figuur 1. Cijfers doelrealisatie met een foutenmarge

Instelling	Percentage doelen geheel gerealiseerd	Foutenmarges	Opmerking
A	70 %	60 – 80 %	Verschil met C
B	60 %	50 – 70 %	
C	45 %	35 – 55 %	Verschil met A

Omdat de marges van instelling A en B elkaar overlappen, kunnen we niet zeggen dat deze instellingen echt van elkaar verschillen. Anders gezegd, er moet tussen twee cijfers minstens een verschil zijn van 20 punten, wil er een betekenisvol verschil zijn. Bij vergelijking van instelling A en B is dat niet het geval. Dan is het ook niet de moeite waard om aan het verschil aandacht te besteden. Het verschil tussen A en C vraagt wel om nadere bespreking: hoe zou dit verschil te verklaren zijn? Een zelfde aanpak moet worden gehanteerd bij de andere prestatie-indicatoren: door foutenmarges te schatten voorkomen we dat aan kleine verschillen teveel waarde wordt gehecht.

Aggregatieproblemen

Valkuilen

Bij het opschalen van de resultaatgegevens van cliëntniveau naar een algemener niveau (bijvoorbeeld de module, de instelling, de regio) is het verstandig rekening te houden met twee bekende valkuilen.

- *Percentages over kleine aantallen zeggen weinig.* Deze kleine aantallen kunnen bijvoorbeeld voorkomen bij cijfers over instellingen die maar weinig cliënten per jaar opnemen. Als bij de ene instelling 40 procent van de cliënten alle doelen zijn gerealiseerd, en bij de andere 80 procent, dan zegt dat nog niet zoveel als die cijfers betrekking hebben op 12 respectievelijk 15 cliënten. Immers, hoe kleiner de groep, hoe gevoeliger de cijfers voor allerlei toevalsinvloeden.
- *Gemiddelen zeggen niet zoveel.* Sommige prestatie-indicatoren, zoals doelrealisatie, werken met eenvoudige schalen die verleiden tot het berekenen van gemiddelden. Wetenschappelijk gesproken gaat het bij dit soort schalen om een zogeheten 'laag meetniveau'. Het berekenen van gemiddelden kan dan een schijn van exactheid wekken, terwijl die exactheid feitelijk niet bestaat. Ook kan het voorkomen dat een gemiddelde is gebaseerd op veel extreme scores: veel cliënten hebben bijvoorbeeld een hele hoge score op de FJ-schaal van de STEP, de meeste andere juist heel laag; daartussen zit er weinig. Als de spreiding van de scores zo slecht is, is het gemiddelde geen goede representant van de scores waarop deze is gebaseerd.

Hoe hier mee om te gaan?

Twee adviezen zijn hier op zijn plaats:

- *Gebruik alleen percentages bij groepen van meer dan 20 cliënten.* Noem bij kleinere groepen bij voorkeur geen percentages of geef duidelijk aan (bijvoorbeeld door de cijfers tussen haakjes te zetten) dat de percentages geen betrouwbaar beeld leveren.
- *Gebruik alleen gemiddelden als het type schaal en de spreiding van scores dat toelaten.* Vermijd hiermee schijn-exactheid.

Conclusie

Cijfers over de prestatie-indicatoren zijn zinvol. Ze geven een *eerste indicatie* van de resultaten van de hulp. Voor een goed gebruik is ten eerste van belang dat ze ingezet worden in het kader van de ambitie om de kwaliteit van de zorg (verder) te verbeteren. Daarom hebben de indicatoren betrekking op zaken waarin partijen (de professionals, financiers, bestuurders) hun verantwoordelijkheden en bevoegdheden kennen om – voor zover nodig – verbeteringen mogelijk te maken en tot stand te brengen.

Voorts is van belang te beseffen dat de cijfers een lage zekerheidsgraad hebben. Ze geven daarom niet zo gauw *uitsluitel* over de effectiviteit van de hulp. De cijfers zijn dan ook alleen op een nuttige manier te gebruiken als men de valkuilen bij het gebruik incalculeert:

- Van belang is de cijfers niet geïsoleerd te beoordelen, maar in relatie tot andere gegevens te interpreteren.
- De cijfers zijn alleen met een ruime foutenmarge te gebruiken. Bij een percentage moet bijvoorbeeld altijd minstens 10 procentpunten opgeteld en afgetrokken worden.
- Sommige cijfers hebben een laag meetniveau. Bij groepen van 20 of minder cliënten bieden percentages geen geldige cijfers. Het berekenen van gemiddelden is niet altijd aan te raden.

Gezien het bovenstaande geldt als algemene moraal: *cijfers over de prestatie-indicatoren spreken niet voor zich, er moet altijd over gesproken worden.*