

Sector : Marktordening  
Afdeling/Project : Project AWBZ  
Samensteller(s) : Ben Vollaard  
Nummer : 121  
Datum : 2 mei 2005

## Doelmatigheidswinsten in de AWBZ

SAMENVATTING. Recente benchmark studies van verzorg- en verpleeghuizen en de thuiszorg (Prismant, 2005, Arcares, 2004, PWC, 2002, SCP, 1996, Kooreman, 1994) geven om vier redenen geen goed beeld van de omvang van te behalen doelmatigheidswinsten:

- *Onvergelijkbaarheid doelmatigheidsscores door verschillen in kwaliteit.*  
Bij de doelmatigheidsmeting blijven vele productdimensies zoals deskundigheid van de zorgverlening en bejegening van cliënten buiten beschouwing. Een lage doelmatigheidsscore kan betekenen dat de kosten hoog zijn door hoge kwaliteit van de zorg.
- *Onvergelijkbaarheid doelmatigheidsscores door verschillen in zorgzwaarte.*  
Grote verschillen in zorgzwaarte – en daarmee kosten per cliënt – tussen instellingen maakt het doen van sectorbrede uitspraken onmogelijk. De studies beperken zich tot uitspraken over doelmatigheid binnen clusters van instellingen met soortgelijke zorgzwaarte.
- *Onvergelijkbaarheid doelmatigheidsscores door verschillen in omgevingsfactoren.*  
Omdat geen goed beeld bestaat van de invloed van omgevingsfactoren op de doelmatigheid (bijvoorbeeld arbeidsmarktomstandigheden), kunnen we scores van instellingen uit verschillende regio's niet zomaar met elkaar te vergelijken.
- *Steekproef levert geen representatief en betrouwbaar beeld op.*  
Voor verpleeg- en verzorgingshuizen staat een kleine, niet representatieve steekproef betrouwbare uitspraken op sectorniveau in de weg.

Voor een betere doelmatigheidsmeting is het nodig om (1) dezelfde gegevens te verzamelen voor een groter aantal verpleeg- en verzorgingshuizen, (2) meer productdimensies in de doelmatigheidsmeting mee te nemen, (3) beter rekening te houden met omgevingsfactoren en (4) gebruik te maken van innovaties in methoden van doelmatigheidsmeting.



# 1 Introductie

## 1.1 Aanleiding

Zijn er in de AWBZ-gefinancierde zorg nog doelmatigheidswinsten te halen? Een eerste blik op recente benchmark studies suggereert een positief antwoord op deze vraag. Niet alle instellingen functioneren even doelmatig. Zo geeft tabel 1.1 een aantal gemiddelde doelmatigheidsscores weer voor de ouderenzorg die ruim afliggen van het best haalbare (een score van 100%). Als de benedengemiddeld presterende instellingen zijn op te trekken naar een hoger niveau, dan zouden doelmatigheidswinsten te behalen zijn.<sup>1</sup> Bij een één op één vertaling van doelmatigheidsscores naar te behalen doelmatigheidswinsten, dan zouden op basis van deze resultaten besparingen van grofweg 1 á 2 miljard haalbaar zijn.

	Verpleeghuizen	Verzorgingshuizen	Thuiszorg
	84	73	85

<sup>a</sup> Een score van bijvoorbeeld 84 procent betekent dat dezelfde zorg gemiddeld genomen ook tegen 16 procent minder kosten kan worden aangeboden.

Bron: Arcares (2004) voor verpleeg- en verzorgingshuizen. PWC (2002) voor thuiszorg.

### Steeds verder 'uitknijpen'?

We kunnen doelmatigheidsscores alleen direct doorvertalen naar te behalen besparingen als we er van uitgaan dat het totale zorgbudget voldoende is om in de vraag te voorzien. Het kan zijn dat ook de best presterende instelling niet het maatschappelijk gewenste niveau van kwaliteit kan leveren, simpelweg omdat het budget ontoereikend is. Door op basis van de huidige spreiding in doelmatigheidsscores instellingen steeds verder 'uit te knijpen', kan de kwaliteit van de zorg in de knel komen. Naast het meten van doelmatigheid van instellingen, is het dus ook van belang om na te gaan in hoeverre het aanbod ook kan aansluiten bij de vraag. Alleen wanneer dit ook duidelijk is, kunnen we concluderen welke besparingen mogelijk zijn.

### Betrouwbaarheid benchmarkresultaten

Twijfel bestaat er ook over de betrouwbaarheid van de cijfers waarmee instellingen worden vergeleken. Zo stelt een journalist na een gesprek met een directeur van een tiental verpleeghuizen: *'De hoogste cijfers kregen (...) zijn slechtere verpleeghuizen, terwijl de goede lager werden gewaardeerd omdat ze de papieren verantwoording niet op orde hadden.'* (NRC Handelsblad, 12-03-2005). Onderzoekers bevestigen dit beeld. Zo stelt een (anoniem te blijven) zorgonderzoeker: *'Ik ben erg geschrokken van het gebrek aan kwaliteit van de data voor*

<sup>1</sup> Zo vertaalt Blank (1998) doelmatigheidsscores in één adem door naar te behalen kostenbesparingen.

*verpleeghuizen. Bij alle extreme waarden ben ik nagegaan op basis van de jaarrekeningen of onze gegevens kloppen. In alle gevallen waren er fouten gemaakt. Meestal hangen die samen met fusies en samenwerkingsverbanden. De productie en de werkgelegenheid hebben dan niet betrekking op dezelfde entiteit.'*

In dit memorandum gaan we dieper in op de vraag wat we op basis van benchmark studies kunnen concluderen over te behalen doelmatigheidswinsten binnen de AWBZ-gefinancierde zorg. We richten ons daarbij op de verzorging en verpleging van ouderen: dit vormt een groot deel van de AWBZ.

## **1.2 Onderzoeksvraag**

De volgende onderzoeksvraag staat centraal in dit memorandum::

*Wat kunnen we op basis van benchmark studies concluderen over het potentieel voor doelmatigheidswinsten op sectorniveau in de AWBZ-gefinancierde zorg voor ouderen?*

## **1.3 Selectie van studies**

Belangrijkste bronnen vormen Arcares (2004) voor verpleeg- en verzorgingshuizen en PWC (2002) voor de thuiszorg. We vergelijken Arcares (2004) ook met een recente studie naar arbeidsproductiviteit van verpleeghuizen, Prismant (2005). Daarnaast analyseren we omwille van de gebruikte methode twee studies die een wat gedateerder beeld van de AWBZ-gefinancierde zorg geven: SCP (1996) en Kooreman (1994). De in deze studies gebruikte cijfers zijn meer dan 10 jaar oud. We zijn de heer Postma (PWC) en mevrouw Neuteboom (Arcares) erkentelijk voor het verschaffen van nadere informatie over de studie van Arcares (2004). We danken Evelien Eggink (SCP) voor haar inbreng op het methodologische vlak.

## **1.4 Structuur**

Eerst gaan we in op het type diensten dat binnen dit onderdeel van de AWBZ wordt geleverd (hoofdstuk 2). Vervolgens discussiëren we de meting van doelmatigheid op instellingsniveau (hoofdstuk 3 en 4). We bespreken de vertaling van benchmark resultaten naar sectorniveau in hoofdstuk 5. In hoofdstuk 6 trekken we conclusies. Daarbij gaan we expliciet in op de vraag wat met betere gegevens en methoden mogelijk zou zijn.

## **2 Doelstellingen van de ouderenzorg**

In dit hoofdstuk gaan we kort in op de aard van de zorg die in verpleeghuizen, verzorgingstehuizen en door de thuiszorg geleverd wordt. We volgen deze 'oude' indeling van de AWBZ zorg en niet de nieuwe indeling naar functies, omdat de oude indeling nog terugkomt in de te bespreken benchmark studies. Later in dit memorandum analyseren we in hoeverre deze doelstellingen tot uitdrukking komen in de meting van efficiëntie. Alle zorg is gericht op het verbeteren van de levenskwaliteit van patiënten, maar de specifieke doelstelling loopt natuurlijk uiteen tussen zorginstellingen.

### **2.1 Thuiszorg**

Het doel van de thuiszorg is mensen die niet alles zelf meer kunnen in het huishouden en/of de persoonlijke verzorging in staat te stellen zelfstandig te blijven wonen. Om deze doelstelling te realiseren biedt de thuiszorg een groot aantal diensten waaronder, huishoudelijke hulp ('Alpha-hulp'), (gespecialiseerde) verzorging (hulp bij het op orde brengen van het leven thuis), verpleging, alarmopvolging, nachtzorg en uitleen- en maaltijdservice. De thuiszorg omvat ook kraamzorg, consultatiebureaus en de uitleen van hulpmiddelen. In 2002 waren er in totaal circa 160 AWBZ-gefinancierde thuiszorginstellingen (CBS, Jaarboek Thuiszorg).

### **2.2 Verzorgingshuiszorg**

Het doel van verzorgingshuiszorg is mensen die niet langer thuis kunnen wonen een begeleide woonomgeving te bieden, waarin zij hun zelfstandigheid zo veel mogelijk kunnen handhaven en mogelijk verbeteren. De zorg betreft begeleiding en niet-complexe verpleging. Verzorgingshuizen zijn actief op het gebied van wonen, zorg en welzijn van ouderen. Mensen kunnen ook tijdelijk worden opgenomen in verzorgingshuizen. In 2002 waren er in totaal circa 1340 verzorgingshuizen (CBS).

### **2.3 Verpleeghuiszorg**

Verpleeghuiszorg omvat naast medische zorg en verpleging ook revalidatie die erop is gericht de patiënt in een aangepast tempo voor te bereiden op een terugkeer naar zijn eigen thuissituatie. In 2002 waren er in totaal circa 330 verpleeghuizen (Prismant). Er zijn twee soorten verpleeghuizen: somatische en psychogeriatrische (de meeste verpleeghuizen kennen een combinatie van beide soorten zorg).

Mensen met een lichamelijke ziekte of handicap kunnen terecht in een somatisch verpleeghuis. De zorg is gericht op terugkeer naar de oude woonomgeving. Is dat niet meer mogelijk, dan wordt er langdurige zorg geboden. Een aantal verpleeghuizen is gespecialiseerd in specifieke patiëntengroepen. Denk aan comapatiënten, mensen met AIDS of mensen met een aandoening aan de luchtwegen. Een revalidatiecentrum kan onderdeel uitmaken van een verpleeghuis. Revalidatiecentra richten zich alléén op herstel. Terugkeer naar huis of naar een aangepaste woonomgeving is dan het doel.

Psychogeriatrische verpleeghuizen richten zich op mensen met geheugenstoornissen, verwardheid, onrust of andere geestelijke (ouderdoms)aandoeningen. Bijvoorbeeld door vormen van dementie zoals de ziekte van Alzheimer. Ook jonge mensen, bijvoorbeeld met de ziekte van Korsakov, kunnen in een psychogeriatrisch verpleeghuis terecht.

## **2.4 Conclusie**

Het werk binnen de AWBZ-gefinancierde zorg voor ouderen heeft vaak een ‘open einde’ karakter. Alleen somatische verpleeghuizen zitten wat dichterbij het leveren van een vooraf goed in te schatten dienst. Bij de meeste vormen van (ouderen)zorg bestaat vaak onzekerheid over de precieze zorg en de duur er van. Dit betekent dat de indicatiestelling door het Centrum Indicatiestelling Zorg (CIZ) ook een belangrijke determinant vormt van de doelmatigheid van de zorgverlening. Het CIZ heeft direct invloed op efficiënte toewijzing van zorg waaraan behoefte is. Dit doet het CIZ niet alleen met de initiële indicatiestelling, maar ook met herzieningen daarvan op basis van nieuwe informatie. Daarbij speelt ook de ‘verkeerde bedden’ problematiek: het al dan niet efficiënt toewijzen van cliënten aan verschillende vormen van zorg. We gaan hier niet verder op in: wij richten ons op de efficiëntie van zorginstellingen gegeven de huidige invulling van het ‘opdrachtgeverschap’.

### **3 Gemeten doelmatigheid op instellingsniveau**

Welke methoden gebruiken de vijf studies om te bepalen in hoeverre zorginstellingen de doelstellingen uit het vorige hoofdstuk doelmatig weten te bereiken? Hier gaan we in dit hoofdstuk op in.

#### **3.1 Definitie doelmatigheid**

Efficiëntie geeft aan wat de verhouding is tussen geleverde 'productie' en de inzet van middelen. Bij productie gaat het niet alleen om bijvoorbeeld verpleegdagen, maar natuurlijk ook om de kwaliteit van de zorg. Een efficiënte zorginstelling weet het gewenste niveau van productie te behalen met minimale middelen tegen de laagste prijs. Zo weet een efficiënt somatisch verpleeghuis tegen minimale kosten een patiënt weer terug te brengen naar de oude thuissituatie.

Zoals we hieronder laten zien, gebruiken de benchmark studies een enge definitie van doelmatigheid. Veel productdimensies, kortweg 'kwaliteit van de zorg', blijven buiten beschouwing. Gemeten doelmatigheid is dus vaak heel wat anders dan werkelijke doelmatigheid. In het volgende hoofdstuk gaan we hier specifiek op in.

Hieronder gaan we in op de drie stappen die nodig zijn om efficiëntie te meten: het meten van de productie (1), kosten (2) en het toepassen van een methode om de relatieve doelmatigheid te bepalen (3). Vervolgens gaan we afzonderlijk in op de wijze waarop de studies rekening houden met verschillende karakteristieken van cliënten en verschillende regionale omstandigheden.

#### **3.2 Stap 1: Meet productie**

##### **De ideale wereld**

De productie van een zorginstelling is de bijdrage aan de doelstelling van de zorg. Idealiter zouden we een goed meetbare maatstaf van kwaliteit van leven hebben en vervolgens de bijdrage van een instelling aan deze maatstaf bepalen. Stel dat we een goede indicator hebben van kwaliteit van leven, dan is het nog moeilijk om de bijdrage van de zorgverlening hieraan te meten. De invloed van velerlei andere factoren op de kwaliteit van leven is moeilijk weg te filteren. Daarnaast is het belangrijk om te controleren voor waarneembare en niet-waarneembare karakteristieken die de levenskwaliteit en het effect van zorg voor individuele cliënten beïnvloeden.

### **De praktijk**

De benchmark studies nemen het aantal verpleeguren of het aantal cliënten als maatstaven van productie. Het idee is dat de uiteindelijke doelstelling van de zorg, betere kwaliteit van leven, sterk samenhangt met het aantal verpleeguren dan wel cliënten. Dat is te verwachten, omdat de indicatie van het CIZ deze 'productie' voorschrijft. De indicatie geeft de aard van de zorg aan (verpleging, begeleiding, etc.), de omvang (uren per dag/week), de frequentie (aantal malen per dag/week) en de duur.

Een instelling heeft veel vrijheid om de zorg verschillend in te vullen. In hoofdstuk 2 zagen we, dat het gevraagde product noodzakelijkerwijs vrij 'open' wordt gespecificeerd. De indicatie luidt in functionele termen en die zijn breed gedefinieerd. Bijvoorbeeld: de functie verpleging kan worden uitgevoerd door een HBO of MBO verpleegkundige.

Veel keuzes in de zorgverlening blijven dus buiten de productiemeting. Voor cliënten is het ook belangrijk of de pijnbestrijding niet te kort schiet, of ze persoonlijk aandacht krijgen, of ze de juiste (combinatie van) medicijnen krijgen voorgeschreven, enzovoort. In het volgende hoofdstuk gaan we dieper in op het 'kwaliteitsvraagstuk' bij de doelmatigheidsmeting.

Om rekening te houden met een verschillend cliëntenbestand tussen instellingen, wordt onderscheid gemaakt tussen verschillende typen verpleegdagen dan wel cliënten. Hier gaan we later in dit hoofdstuk nader op in. Het is belangrijk om het kwaliteitsvraagstuk te onderscheiden van de 'cliëntenmix'. Bejegening kan immers voor iedere cliënt van belang zijn.

De tabel op de volgende pagina geeft een overzicht van de definitie van productie zoals gebruikt door de vijf studies waar we in dit memorandum naar kijken.

### **Niet meegenomen producten**

Niet alleen blijven veel productdimensies buiten beschouwing, ook komen niet alle producten terug in de meting van productie. Zo merkt Kooreman (1994: 313) op, dat de instellingen deels ook een onderwijkstraak hebben. Het aannemen van leerling-verpleegkundigen leidt tot lagere efficiëntie, maar draagt ook bij aan hun opleiding. PWC (2002) geeft aan dat zij binnen de thuiszorg voedingsvoorlichting en dieetadvisering, gezondheidsvoorlichting en opvoeding en uitleen van verpleegartikelen niet meenemen. Deze producten worden niet door alle instellingen aangeboden en/of er zijn onvoldoende gegevens over beschikbaar. Zij proberen hier wel zo goed mogelijk voor te corrigeren door te achterhalen welke kosten zijn toe te rekenen aan deze producten (ibid., p. 43).



**Tabel 3.1 Van doelstelling naar 'productie' van zorginstellingen**

	Verpleeghuizen	Verzorgingshuizen	Thuiszorg
<b>Doelstelling</b>	<i>Somatische patiënten:</i> Terugkeer naar oude woon-omgeving; langdurige zorg als dat niet meer kan. <i>Psychogeriatrische patiënten:</i> langdurige zorg.	Een begeleide woonomgeving bieden aan mensen die niet langer thuis kunnen wonen	Mensen die niet alles zelf meer kunnen in staat te stellen zelfstandig te blijven wonen

**Definitie productie**

Prismant, 2005	Aantal verpleegdagen <sup>a</sup>	–	–
Arcares, 2004	Aantal zorgminuten <sup>b</sup>	Aantal zorgminuten <sup>b</sup>	–
PWC, 2002	–	–	Aantal uren thuiszorg <sup>c</sup>
SCP, 1996	Aantal cliënten <sup>d</sup>	–	–
Kooreman, 1994	Aantal cliënten <sup>e</sup>	–	–

<sup>a</sup> Vier typen zorgdagen: dagbehandeling, somatisch, standaard, psychogeriatrisch. De dagen worden gewogen met zorgzwaarte uitgedrukt in een index van de hulpbehoefendheid van cliënten (Algemene Dagelijkse Levensverrichtingen, ADL). <sup>b</sup> Vijf typen zorgminuten: behandeling, activerende begeleiding, ondersteunende begeleiding, verpleging, persoonlijke verzorging, huishoudelijke verzorging. <sup>c</sup> Het aantal productiesoorten is wat onduidelijk, mogelijk gaat het om alphahulp, (standaard/huishoudelijke/ gespecialiseerde) verzorging, (standaard/gespecialiseerde) verpleging, advies/instructie/ voorlichting, en dagverzorging voor ouderen. <sup>d</sup> Vier typen cliënten: somatische en psychogeriatrische die óf klinisch zijn óf dagpatiënt. De cliënten worden gewogen met zorgzwaarte (a.h.v. leeftijd en behandelingscategorie). <sup>e</sup> Vier typen cliënten, net als SCP (1996), maar zonder (initieel) weging met zorgzwaarte.

### 3.3 Stap 2: Bepaal kosten

De volgende stap bij het bepalen van de efficiëntie is het bepalen van de kosten. Het is gebruikelijk om niet direct toewijsbare kosten (zoals voor algemeen en administratief personeel) uniform aan de producten toe te kennen. Tabel 3.2 geeft een overzicht van de kosten die in de verschillende studies worden meegenomen.

**Meting doelmatigheid vertekend door weglaten kapitaalinvesteringen**

De meeste benchmark studies richten zich vooral op de doelmatigheid van de productiefactor arbeid. Productie wordt vaak gemeten in aantal zorgdagen, bij kosten gaat het vaak om personeelskosten. Hierdoor kan doelmatigheid overschat worden: het personeel is bijzonder productief dankzij kapitaalinvesteringen, waarvan de kosten niet worden meegenomen (denk bijvoorbeeld aan een tilhulp). Ook kan de doelmatigheid onderschat worden, wanneer kapitaalinvesteringen leiden tot een grotere zelfredzaamheid van cliënten. De gevoeligheid voor vertekening is groter wanneer productie wordt gemeten in zorgdagen in plaats van het aantal cliënten. Wel is het zo dat het aantal zorgdagen een wat preciezer beeld geeft van de productie dan het aantal cliënten.

Door verschillende typen personeel te onderscheiden houden Prismant (2005), SCP (1996) en Kooreman (1994) rekening met verschillen in cliëntenbestand tussen instellingen. We gaan hier later in dit hoofdstuk nader op in.

Prismant (2005), Arcares (2004) en Kooreman (1994) laten de kapitaalkosten buiten beschouwing. Het management heeft weinig invloed op dit type kosten. Alleen PWC (2002) neemt kapitaalkosten voor huisvesting mee (voor analyse van de thuiszorg). SCP (1996) neemt kapitaal niet als variabel middel mee zoals personeel, maar als vast ingezet middel.

**Tabel 3.2 Meting van kosten ouderenzorg**

Prismant (2005)	Arcares (2004)	PWC (2002)	SCP (1996)	Kooreman (1994)
Aantal FTEs <sup>a</sup>	<b>Zorgfunctie</b>	Personeel	Personeelskosten	Aantal FTEs <sup>c</sup>
	Loonkosten directe zorg; sociale lasten en overige personeelskosten; opslagen <sup>b</sup>	Huisvesting	gediplomeerd verpleegkundigen,	
		Automatisering	idem overig	
		Overig	personeel,	
	<b>Verblijffunctie</b>		materiaalkosten	
	Personeel hotel- en terreinfuncties; sociale lasten loonkosten; voedingskosten; hotelmatige kosten; belastingen en verzekeringen; terrein- en gebouwgebonden kosten			

<sup>a</sup> Efficiëntiemaatstaf is op niet nader toegelichte wijze geschoond voor samenstellingseffecten van de factor arbeid. <sup>b</sup> Opslagen voor: verlof, studiedagen, ziekteverzuim, niet direct zorggebonden tijd, management en staf, algemeen en administratief personeel, algemene kosten, cliëntgebonden kosten. <sup>c</sup> De analyse onderscheidt zes typen personeel.

### 3.4 Stap 3: Bepaal de relatieve doelmatigheid van een instelling

Na het bepalen van de productie en de middelen is de uitdaging om op basis van een onderlinge vergelijking te bepalen hoe goed elke instelling presteert. DEA is de meest gebruikte methode, hieronder gaan we in op de reden voor het gebruik hiervan. Tenslotte geven we een overzicht van de wijze waarop DEA is toegepast binnen de studies.

#### Waarom DEA en niet een eenvoudig input/output verhoudingsgetal?

Data Envelopment Analysis (DEA) is een methode om de (relatieve) efficiëntie van een instelling te bepalen wanneer meerdere producten en typen middelen een directe vergelijking op basis van een eenvoudige input/output ratio bemoeilijken. Als we met verhoudingscijfers willen werken, dan zien we al snel dat de ene instelling goed scoort op het ene cijfer en de andere instelling op het andere cijfer. Conclusies trekken is dan moeilijk (zie Blank, 1998 voor een discussie). Zo ‘produceren’ zorginstellingen verpleegdagen voor verschillende typen cliënten en zetten ze hiervoor verschillende typen personeel in. Door deze ongelijksoortigheid in producten en middelen kunnen we niet makkelijk de efficiëntie van een instelling uitdrukken in het aantal zorgminuten per euro.

Om met meerdere *inputs* and *outputs* toch tot een eenvoudige *input/output* ratio te komen, zouden we de gewogen som van de producten en de gewogen som van de middelen kunnen

nemen. Maar daar hebben we gewichten voor nodig. We moeten precies weten wat de waarde is van het ene product ten opzichte van het andere, en hetzelfde voor de middelen. Voor de hele verzameling instellingen zouden we dezelfde gewichten moeten toepassen. Om deze gewichten te bepalen, zouden we veronderstellingen moeten maken over de functionele vorm van de relatie tussen *inputs* and *outputs*, veronderstellingen die wellicht niet met de realiteit overeenkomen en weinig ruimte laten voor een verschillende waardering van *inputs* en *outputs* door zorginstellingen.

Met DEA is dat allemaal niet nodig. Met DEA bepalen we de efficiëntie van een instelling door naar lineaire combinaties van productieniveaus en middelen van andere instellingen te kijken. Is met dezelfde middelen een hogere productie te behalen? De gewichten van de producten en middelen worden zo gekozen dat de behaalde efficiëntie van een instelling in vergelijking met een andere instelling zo hoog mogelijk is. Als een instelling inefficiënt blijkt, zelfs met de meest gunstige keuze van gewichten, dan is dat natuurlijk een sterke uitspraak. De stelling dat het aan de keuze van gewichten ligt, is dan niet houdbaar.

De meest efficiënte instelling krijgt een score van 100%. Deze instelling levert, gegeven een bepaald kostenniveau, het grootste aantal zorgminuten. Een instelling met een doelmatigheidsscore van bijvoorbeeld 80% realiseert dus voor dezelfde kosten als de meest efficiënte instelling slechts 80% van het aantal minuten zorg van wat mogelijk is.

#### DEA of SFA?

Alle studies behalve Prismant (2005) gebruiken Data Envelopment Analysis (DEA) om de doelmatigheidsscores te bepalen.<sup>1</sup> DEA is een goed te begrijpen methode die relatief makkelijk is toe te passen. Parametrische methoden, zoals Stochastic Frontier Analysis (SFA), vragen vanwege de complexiteit veel meer van de onderzoekers en ook van de gebruikers. Een onderzoeker merkt hierover op: '*SFA is niet aan beleidsmakers uit te leggen; het wordt daarmee een black box die scores genereert*'. Daarnaast vereist SFA een groot aantal waarnemingen; een voorwaarde waar lang niet altijd aan voldaan wordt. Op puur theoretische gronden is geen duidelijke voorkeur aan te geven voor één bepaalde methode. Elke methode heeft haar sterke en zwakke punten (hieronder meer over de zwakke punten van DEA). Het SCP heeft de gewoonte om de resultaten van beide methoden te laten zien: zo worden de uitkomsten genuanceerd. Als beide methoden dezelfde kant opwijzen, dan is het empirisch bewijs sterker dan wanneer dat niet het geval is.

<sup>1</sup> Prismant gebruikt arbeidsproductiviteit als efficiëntie maatstaf: het productievolume (dan wel de toegevoegde waarde) gedeeld door het aantal FTEs.

#### Beperkingen DEA

Naast het feit dat DEA alleen instellingen onderling vergelijkt en dus geen beeld geeft van het 'maximaal haalbare' efficiëntie in de sector zijn er drie beperkingen van deze methode:

- Een instelling wordt gescoord ten opzichte van een ‘extreem’; de best presterende instelling. De methode is daarmee **gevoelig voor ‘uitbijters’**.
- Daarnaast kan ruis door **meetfouten** een significante vertekening opleveren. Ruis heeft niet alleen invloed op de eigen score, maar ook – door de vergelijking – op de score van anderen.
- Hoe minder vergelijkbaar de instellingen, hoe eerder ze allemaal als efficiënt worden aangemerkt. Met  $T$  *outputs* and  $M$  *inputs* is het te verwachten dat we  $T^*M$  efficiënte instellingen vinden. Het **aantal instellingen** moet dus veel groter zijn dan  $T^*M$  om tussen de instellingen te kunnen discrimineren.

### Toepassing DEA binnen de benchmark studies

Tabel 3.3 geeft een overzicht van de wijze van toepassing van DEA in de vier benchmark studies. Wat opvalt is dat het kleine aantal waarnemingen in Arcares (2004).

Tabel 3.3		Toepassing DEA in verschillende studies			
	Arcares (2004)	PWC (2002)	SCP (1996)	Kooreman (1994)	
Aantal productsoorten	5 (aantal minuten zorg per zorgfunctie) <sup>a</sup>	Onduidelijk <sup>b</sup>	4 (verpleeg- en dagbehandelingsdagen voor somatische en psychogeriatrische patiënten)	4 (idem als SCP, 1996)	
Aantal middelensoorten	1 (kosten van zorg en behandeling)	Onduidelijk <sup>b</sup>	3 (kosten gediplomeerd verplegend personeel, overig personeel, materiaal)	6 (doktoren, gewone en leerling-verpleegkundigen, therapeuten, algemene staf, overig)	
Aantal waarnemingen	21 (vph), 54 (vzh)	96	307 <sup>c</sup>	292	

<sup>a</sup> Zorgfuncties: persoonlijke verzorging, verpleging, behandeling, ondersteunende begeleiding, huishoudelijke verzorging, activerende begeleiding. <sup>b</sup> PWC (2002) geeft weinig technische details over de analyse. <sup>c</sup> SCP gebruikt in totaal 1939 cases uit 8 jaren, dus gemiddeld 240 per jaar (307 in het laatste jaar).

### Meetfouten

Zowel Arcares (2004) als PWC (2002) proberen meetfouten in productie- en kostengegevens te beperken door dubbele controles. Zo heeft de vragenlijst voor kosten van Arcares (2004) een ingebouwde foutcontrole en hebben accountants de antwoorden na retourontvangst gecontroleerd en aangevuld. De software applicaties van PWC (2002) kennen ook foutcontroles. Arcares (2004, p. 17) vermeldt wel dat de productieregistratie niet perfect is. Een week lang hebben betrokken medewerkers via een handheldcomputer geregistreerd welke handelingen zij tijdens hun werktijd hebben verricht. Bij een klein aantal instellingen was de tijdregistratie niet volledig of bevatte de week waarin de zorgtijd gemeten is feestdagen.

Onduidelijk is hoe ernstig de resterende meetfouten zijn, wel is duidelijk dat de betrouwbaarheid hoger is dan bij gegevens gebruikt door bijvoorbeeld SCP (1996).

### 3.5 Houd rekening met verschillende karakteristieken van cliënten

Bij de doelmatigheidsmeting is het belangrijk om rekening te houden met verschillen in het cliëntenbestand van instellingen. Sommige cliënten zijn immers veel ‘duurder’ om voor te zorgen dan andere: ze hebben een ander type behandeling nodig met duurder materiaal dan wel personeel. Omdat dit zo’n belangrijke zaak is, gaan we hier in deze paragraaf afzonderlijk op in. De studies gebruiken verschillende strategieën om hiermee rekening te houden:

- *Meerdere productsoorten.* Arcares (2004) onderscheidt vijf verschillende zorgfuncties (verpleging, huishoudelijke verzorging, etc.), voor PWC (2002) is het niet duidelijk, SCP (1996) en Kooreman (1994) onderscheiden vier verschillende typen cliënten.
- *Meerdere kostensoorten.* SCP (1996) en Kooreman (1994) onderscheiden daarnaast ook verschillende typen personeel.
- *Clustering van instellingen.* Arcares (2004) en PWC (2002) clusteren de doelmatigheidsscores op basis van zorgzwaarte (twee clusters verpleeghuizen, vier clusters verzorgingshuizen, zes clusters thuiszorginstellingen). Het idee is dat de doelmatigheidsscores alleen *binnen* deze clusters zijn te vergelijken. PWC (2002, p. 51) laat ook zien dat de gemiddelde doelmatigheid per cluster verschilt. Een gemiddelde doelmatigheidsscore voor de hele sector is daarmee weinig informatief. Binnen Arcares (2004) laat de clusterindeling nog vrij veel ruimte in de zorgzwaarte tussen instellingen.<sup>2</sup>
- *Regressieanalyse.* Kooreman (1994) en SCP (1996) voeren als een tweede stap een regressieanalyse op basis van de doelmatigheidsscores uit om te corrigeren voor verschillen in zorgzwaarte (die overblijven na het onderscheiden van verschillende productie- en kostensoorten). Voordeel van deze aanpak is dat zorgzwaarte een continue variabele is, terwijl clusters altijd een enigszins willekeurige indeling vormen van instellingen. Gegeven het aantal waarnemingen, stond deze optie ook open voor PWC (2002), niet voor Arcares (2004). Overigens blijken de twee door Kooreman (1994) gebruikte variabelen (deel patiënten >85jaar, gemiddelde verblijfsduur patiënten) verschillen in doelmatigheidsscores niet te kunnen verklaren.

<sup>2</sup> Om een idee te geven: tussen verpleeghuizen zit maximaal een speling van 11 procent in de kosten als gevolg van verschillen in zorgzwaarte (uitgaande van de 16 zorgbehoeftegroepen, zie Arcares, 2004, tabel 2-9, p. 34). De spreiding in doelmatigheidsscores van verpleeghuizen is 75% (100 vs. 57) en kan dus voor maximaal 15 procent het resultaat zijn van verschillen in de cliëntenmix. Voor verzorgingshuizen is de mogelijke invloed nog veel groter. Binnen de vier clusters zit een speling van maximaal 47 procent in zorgzwaarte. De spreiding in doelmatigheidsscores van 117% (100 vs. 46) kan dus voor 40% het resultaat zijn van verschillen in de cliëntenmix.

Kooreman (1994) en SCP (1996) geven de indruk dat met een groot aantal waarnemingen en verschillende product- en kostensoorten in de DEA voor verschillen in zorgzwaarte is te corrigeren. Arcares (2004) kent te weinig waarnemingen en te weinig kostensoorten om goed met verschillen in zorgzwaarte rekening te houden. De indeling naar clusters met soortgelijke zorgzwaarte in Arcares (2004) en PWC (2002) blijft altijd enigszins willekeurig, biedt beperkt soelaas in het geval van Arcares (2004), en maakt sectorbrede uitspraken over doelmatigheid onmogelijk.

### **3.6 Houd rekening met verschillende regionale omstandigheden**

PWC (2002, hoofdstuk 7) laat zien dat verschillen in regionale omstandigheden invloed hebben op de doelmatigheidsscores.<sup>3</sup> De doelmatigheidsscores liggen relatief laag in de Randstad en in enkele noordelijke provincies. Instellingen in de Randstad huren relatief veel personeel in, wat een kostenverhogend effect heeft. Verder blijkt dat in verstedelijkte gebieden meer kosten worden gemaakt voor de werving en voor het behoud van personeel. Het verloop onder het personeel is groter dan gemiddeld en er wordt relatief veel laaggeschoold personeel aangenomen, in functiegroepen waar het ziekteverzuim hoger is. Bovendien heeft een groter aantal medewerkers een contract voor slechts een beperkt aantal uren. Door al deze oorzaken moet er veel tijd geïnvesteerd worden in het inwerken, de begeleiding en de opleiding van personeel. Daardoor is het aandeel van de indirecte kosten in de totale personeelskosten relatief hoog en de productiviteit relatief laag. Weliswaar zijn de kosten per betaalluur voor het product huishoudelijke verzorging lager (door het grote aandeel laaggeschoold personeel), maar dit effect wordt teniet gedaan door de lage productiviteit en het hoge aandeel indirecte kosten.

Kooreman (1994) ondersteunt het idee dat stedelijkheid de doelmatigheid buiten de invloed van het management om kan beïnvloeden. Hij vindt een duidelijk significant 'grote stad'-effect. De regiovariabelen (noord, oost, zuid) zijn overigens niet significant.

Arcares (2004) laat de invloed van omgevingsfactoren buiten beschouwing. In hoeverre verschillende scores zijn veroorzaakt door factoren die buiten de invloedssfeer van het management liggen, is dus onduidelijk.

<sup>3</sup> Vanwege een financieringsmethode binnen de thuiszorg die ondoelmatigheid in de hand kan werken, onderzoekt PWC (2002) ook in hoeverre de aangeboden productmix overeenkomt met de benodigde productmix. Hiertoe wordt een vraagvoorspelmodel gebruikt op regioniveau. We gaan hier niet verder op in.

### 3.7 Conclusie

Over de meting van doelmatigheid in de betreffende studies concluderen we het volgende:

- Bij de doelmatigheidsmeting blijven **vele productdimensies** zoals deskundigheid van de zorgverlening en bejegening van cliënten **buiten beschouwing**. Een lage doelmatigheidsscore betekent daarom niet altijd dat een instelling ondoelmatig presteert. In het volgende hoofdstuk gaan we hier nader op in.
- Doelmatigheidsverbeteringen door **vervanging van arbeid door kapitaal** (bijvoorbeeld tillift) komen onvoldoende tot uitdrukking in de huidige doelmatigheidsscores.
- De meest gebruikte methode om de relatieve doelmatigheid van een instelling te bepalen, **DEA, is gevoelig voor ‘uitbijters’ en meetfouten**.
- Grote verschillen in **zorgzwaarte** – en daarmee kosten per cliënt – tussen instellingen maakt het doen van sectorbrede uitspraken onmogelijk. Het onderscheiden van clusters met soortgelijke zorgzwaarte biedt beperkt soelaas.
- Tenslotte bestaat **geen goed beeld van de invloed van externe factoren** (zoals regionale arbeidsmarktomstandigheden) op de doelmatigheid. Scores van instellingen in verschillende zorgkantoorregio’s zijn dus niet zomaar met elkaar te vergelijken.





## 4 Kosten en kwaliteit op instellingsniveau

Bij de doelmatigheidsmeting blijven vele productdimensies zoals deskundigheid van de zorgverlening buiten beschouwing. Dit is een probleem als prestaties op niet-meegenomen productdimensies niet gelijk op gaan met de doelmatigheidsscores. Een lage doelmatigheidsscore betekent dan niet dat een instelling ondoelmatig presteert.

De benchmark studies scharen alle productdimensies die niet meegenomen zijn in de doelmatigheidsmeting onder de noemer 'kwaliteit'. De vraag is dus of 'kwaliteit' gelijk op gaat met de doelmatigheidsscores. Eerst bespreken we meting van kwaliteit, dan gaan we in op het verband tussen kwaliteit en de doelmatigheidsscores.

### 4.1 Meting van 'kwaliteit'

In lijn met het woordgebruik in de benchmark studies is kwaliteit te definiëren als *alle niet in de doelmatigheidsmeting meegenomen productdimensies*. De vijf benchmark studies verschillen sterk in de wijze waarop kwaliteit van de zorg wordt gemeten. Prismant (2005) laat kwaliteitsvraagstukken buiten beschouwing. Dankzij een specifiek op kwaliteit gerichte enquête hebben Arcares (2004) en PWC (2002) veel meer informatie over allerlei kwaliteitsaspecten van de zorg dan SCP (1996) en Kooreman (1994). Hieronder gaan we in op de verschillende bronnen voor het meten van kwaliteit. De tabel geeft een overzicht.<sup>4</sup>

**Tabel 4.1 Meting kwaliteit van de zorg<sup>a</sup>**

Prismant (2005)	Arcares (2004)	PWC (2002)	SCP (1996)	Kooreman (1994)
– Enquêtes onder cliënten en contactpersonen over tevredenheid met zorg <sup>b</sup> + onder verzorgenden over problemen waar patiënten mee te maken hebben (valpartijen, depressie, enz.)	Schriftelijke enquête onder cliënten (organisatie en inhoud zorg)	Enquêtes onder instellingen: aandeel leerlingverplk., kwaliteitsprijzen (gedipl.) verpleegkundig en overig personeel, inzet van kapitaal per bed, aanwezigheid apotheek	Enquêtes onder instellingen: aanwezigheid patiëntenraad, raad voor familie van patiënten, klachtenprocedure, restricties op bezoeken	

<sup>a</sup> We laten de medewerkersraadpleging over het werkklimaat buiten beschouwing. <sup>b</sup> Zorg betreft hier: introductie, wonen, maaltijden, activiteiten, zorg, (para)medische diensten, en familie.

<sup>4</sup> We zien het medewerkersoordeel over het werkklimaat (PWC, 2002 en Arcares, 2004) niet als onderdeel van de kwaliteit van een instelling, maar als een *determinant* daarvan (voorbeelden zijn 'zelfstandigheid' en 'afstemming werk-privé'). Een motiverend werkklimaat werkt immers prestatieverhogend. Denk aan ziekteverzuim, verloop en betrokkenheid van het personeel.

### **(1) Enquête onder verzorgend personeel over niet meegenomen productdimensies**

Arcares (2004) enquêteert verzorgend personeel over een aantal zorginhoudelijke kwaliteitsindicatoren. Het gaat om de volgende kwaliteitsindicatoren: valpartijen; depressiesymptomen met/zonder antidepressivagebruik; gebruik van negen of meer verschillende medicijnen; ongewild gewichtsverlies; lastveroorzakend gedrag hoogrisicogroep/laagrisicogroep; drie maal of vaker per week slaapmiddelen; dagelijkse toegepaste lichaamsfixatie; doorligwonden hoogrisicogroep/laagrisicogroep; tekortschietende pijnbestrijding; aandacht.

Een directe vergelijking tussen instellingen op deze kwaliteitsindicatoren is moeilijk, omdat de zorgzwaarte van de cliënten tussen instellingen verschilt. Het percentage depressieve patiënten binnen een instelling kan gewoon hoger liggen dan elders door een gemiddeld hogere zorgzwaarte. Dit vergelijkingsprobleem geldt niet voor presteren onder de gezamenlijk afgesproken ondergrens (IGZ, 2004), maar wel voor de kwaliteit die daar bovenuit gaat. Arcares (2004) houdt hier gedeeltelijk rekening mee, omdat de *best practice* instelling per cluster verpleeg- en verzorgingshuizen met soortgelijke zorgzwaarte wordt bepaald.

Conclusie is dat een enquête onder verzorgenden over zorginhoudelijke factoren een goede informatiebron vormt voor de kwaliteit van de zorg<sup>5</sup>, maar dat voor een goede vergelijking wel correctie voor de cliëntenmix nodig is. Daarnaast is natuurlijk van belang dat het personeel betrouwbare informatie geeft over de toestand van de cliënten.

### **(2) Enquête onder cliënten en contactpersonen over kwaliteitsoordeel**

Arcares (2004) voert een enquête uit onder cliënten en contactpersonen naar de tevredenheid met de thema's wonen, maaltijden, activiteiten en zorg. Met behulp van cliëntenpanels is nagegaan welke factoren het zwaarste wegen. PWC (2002) voert een soortgelijke enquête uit onder cliënten. In de vragenlijst wordt onderscheid gemaakt naar de organisatie van de zorg (continuïteit, toegankelijkheid, communicatie) en de inhoud van de zorg (bejegening, deskundigheid, autonomie). Tegelijk wordt gevraagd hoe belangrijk cliënten verschillende kwaliteitsaspecten vinden: zo blijkt 'bejegening' belangrijker te zijn dan 'autonomie' (PWC, 2002, figuur 3-5, p. 65).

Vergeleken met indicatoren als het percentage cliënten met doorligwonden, komt de toegevoegde waarde van de zorg *theoretisch gezien* beter in beeld door naar het cliëntenoordeel

<sup>5</sup> De soms schokkende cijfers (24% van de hoogrisicogroep in verpleeghuizen heeft doorligwonden; 21% in verzorgingshuizen) zijn niet volledig toe te schrijven aan tekortschietende zorg binnen verpleeg- en verzorgingshuizen. Zij hebben immers geen invloed op het traject vóór intake, waar een deel van deze problemen kan zijn ontstaan.

te vragen. Cliënten kunnen immers ook rekening houden met het feit dat ze bijvoorbeeld de doorligwonden al in het ziekenhuis hebben opgelopen. Een directe vergelijking tussen instellingen op basis van het cliëntenoordeel lijkt makkelijker te maken dan met de hierboven genoemde kwaliteitsindicatoren.

Toch zijn er twee belangrijke redenen waarom een dergelijke vergelijking problematisch is. Allereerst kan de cliëntenmix toch een rol spelen, omdat sommige cliënten – al krijgen ze dezelfde zorg – toch anders oordelen. PWC (2002: 69) vindt aanwijzingen dat mannelijke, oudere en lager opgeleide cliënten positiever oordelen dan vrouwelijke, jonge, en hoog opgeleide cliënten. Zowel PWC (2002) als Arcares (2004) houden hier deels rekening mee door clusters van instellingen te onderscheiden met een soortgelijke cliëntenmix.

#### **Geeft het cliëntenoordeel een goed beeld van de kwaliteit van de zorg?**

IGZ (2004) laat zien dat algemene vragen naar tevredenheid van cliënten een beperkt beeld geeft van de kwaliteit van de zorg. Ondanks het feit dat veel verpleeghuizen niet eens de minimale zorg kunnen bieden, zijn de geïnterviewde bewoners bijna alle tevreden over de zorg die zij krijgen. Dit wordt grotendeels bepaald door de grote waardering die men heeft voor de inzet en betrokkenheid van de zorgverleners: "*Zij zetten zich maximaal in en kunnen niet meer doen dan mogelijk is.*" (ibid., p. 15). Het idee is dat cliënten zich niet voldoende kunnen distantiëren van de persoon van de zorgverlener om een goed oordeel te vellen.

Daarnaast kan het cliëntenoordeel positief gekleurd zijn door 'meevoelen met het personeel' (zie het tekstkader). Hoe verhoudt zich dat tot de tevredenheidsscores in Arcares (2004) en PWC (2002)? Allereerst kunnen ondanks een positieve vertekening in het oordeel van cliënten *verschillen* in het oordeel tussen instellingen wellicht wel informatief zijn, vooral op specifieke onderdelen. Algemene tevredenheidscijfers lijken weinig informatief. Op specifieke onderdelen kan wellicht gedeeltelijk worden gecorrigeerd voor de positieve vertekening door instellingen alleen te vergelijken op basis van het deel van de cliënten dat bijzonder tevreden is. Gegeven de vertekening in het cliëntenoordeel is 'enigszins tevreden' waarschijnlijk nog altijd onder de maat.

Tenslotte is er een groep respondenten die niet of minder in hun oordeel wordt beïnvloedt door contact met het personeel: contactpersonen. Arcares (2004) maakt hier veel gebruik van, omdat veel cliënten (vooral psychogeriatrische) niet in staat zijn een eigen oordeel te geven. Contactpersonen blijken aanzienlijk minder positief over de kwaliteit van de zorg dan cliënten.

Conclusie is dat een vergelijking van instellingen op basis van een cliëntenoordeel informatief is als rekening wordt gehouden met een positieve vertekening daarin (alles onder 'bijzonder tevreden' is indicatie van onvoldoende kwaliteit) en als de vragen voldoende specifiek zijn (algemene oordelen over 'de kwaliteit van de zorg' zeggen niet zo veel).

### (3) Indicatoren van 'kwaliteitsbeleid'

SCP (1996) en Kooreman (1994) beperken zich tot enkele indicatoren die met het kwaliteitsbeleid van instellingen kunnen samenhangen. SCP (1996) kijkt naar het personeelsbeleid, ingezet kapitaal per bed (denk aan grotere kamers en minder mensen op één kamer) en de aanwezigheid van een apotheek. Kooreman (1994) kijkt naar inspraak van patiënt en familie, klachtenprocedures en restricties op bezoeken. Van beleidsindicatoren is het natuurlijk maar de vraag of ze ook daadwerkelijk leiden tot hogere kwaliteit.

### Conclusie

Arcares (2004) en PWC (2002) meten kwaliteit van de zorg door enquêtes onder zorgpersoneel en cliënten uit te voeren. Deze meting van kwaliteit vormt een verbetering op de indicatoren van een 'kwaliteitsbeleid' zoals gebruikt in SCP (1996) en Kooreman (1994). Voor een vergelijking tussen instellingen op basis van de 'toestand van patiënten' en de tevredenheid van cliënten is correctie voor verschillen in zorgzwaarte nodig, al is deze correctie van kleiner belang in het geval van cliëntentevredenheid. Het cliëntenoordeel is informatief als rekening wordt gehouden met een positieve vertekening daarin en als de vragen voldoende specifiek zijn.

## 4.2 Doelmatigheid zoals in de praktijk gemeten vs. 'kwaliteit'

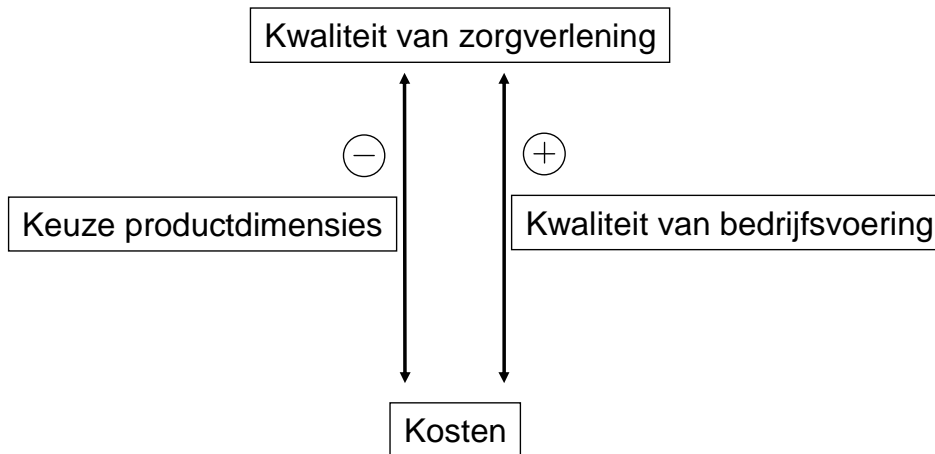
### Relatie theoretisch gezien ambigu

De relatie tussen de *gemeten* doelmatigheid en kwaliteit – ofwel prestaties op niet-meegenomen productdimensies – kan zowel positief als negatief zijn:

- **Positief verband.** Het kan zijn dat door een goede *bedrijfsvoering* een instelling zowel in termen van doelmatigheid als kwaliteit goed scoort. Als bijvoorbeeld veel medewerkers tevreden zijn, dan kunnen door een hogere motivatie en een lager ziekteverzuim zowel de doelmatigheid als de kwaliteit toenemen.
- **Negatief verband.** Daarnaast bestaat er een *afweging* tussen kwaliteit en gemeten doelmatigheid. Het zo goedkoop mogelijk produceren van verpleegdagen kan ten kosten gaan van kwaliteit, ofwel niet meetbare productdimensies (zie het tekstkader over 'De Egmontshof'). Denk bijvoorbeeld aan het inzetten van goedkoper maar minder deskundig personeel zonder het benodigde toezicht: een hogere doelmatigheid kan dan ten koste gaan van de kwaliteit van de zorg. Gegeven de behoefte van cliënten en de budgettering weegt het management af welk niveau van kwaliteit de instelling biedt. Instellingen kunnen daarin natuurlijk verschillen, omdat ook de wensen van cliënten uiteen kunnen lopen. Ook de waardering van verschillende elementen van kwaliteit kan tussen instellingen uiteenlopen.

Als we kunnen controleren voor de kwaliteit van de bedrijfsvoering is een negatief verband tussen kwaliteit en doelmatigheidsscores te identificeren.

**Figuur 4.1** Het verband tussen kosten en kwaliteit



Noot: Bij keuze van productdimensies gaat het om het kwaliteitsniveau van de zorgverlening, bijvoorbeeld de tijd die besteed wordt aan toezicht van patiënten.

#### **Kwaliteit en gemeten doelmatigheid: ervaringen in verpleeghuis 'De Egmontshof'**

Eind 2003 kondigt het management van verpleeghuis 'De Egmontshof' in Oud-Beijerland aan dat door bezuinigingen patiënten bij toerbeurt in bed moeten blijven ('pyjamadagen') en nog maar één keer in de drie weken een douchebeurt krijgen. Na een bezoek aan het verpleeghuis concludeert de Inspectie Gezondheidszorg november 2003 dat de zorg onder de maat is. De persoonlijke hygiëne, toiletgang en dagindeling verkeren in een 'risicogebied'. De rapportage geeft aan dat de problemen al langer spelen en zowel financieel als zorginhoudelijk van aard zijn.

Het benchmark rapport over dit verpleeghuis van Arcares (2002) – een test voor de in deze studie besproken studie van Arcares – laat zien dat de instelling goed scoort op doelmatigheid, maar slecht op cliëntentevredenheid. De kosten per geleverd uur zorg liggen op goed niveau, maar de kwaliteit blijft dus achter. Wanneer veel productdimensies buiten beschouwing blijven kunnen doelmatigheidsscores alleen dus een misleidend beeld geven van de prestaties van een instelling.

#### **Bevindingen**

Arcares (2004) en PWC (2002) voeren een kwalitatieve analyse uit naar de relatie tussen de doelmatigheidsscores en kwaliteit van zorginstellingen.

Arcares (2004) geeft elke instelling een kwaliteitsranking. Allereerst is voor elke instelling de positie in de rangorde berekend. Dit is gebeurd voor alle drie kwaliteitsinstrumenten: cliënttevredenheid, zorginhoudelijke kwaliteitsindicatoren en ook medewerkertevredenheid. Vervolgens zijn de scores opgeteld. Zowel voor verpleeghuizen als verzorgingshuizen vertoont deze kwaliteitsranking geen verband met de doelmatigheidsscores.

PWC (2002) creëert een cliëntscoor om de kwaliteit van een instelling aan te geven (we laten de medewerkersscore buiten beschouwing). Er lijkt een licht positief verband te bestaan tussen cliëntscoor en de doelmatigheidsscoor, maar het verband is niet sterk (ibid., p. 92). SCP (1996) en Kooreman (1994) nemen indicatoren van kwaliteitsbeleid mee in de regressieanalyse van de doelmatigheidsscoor (tobit en probit model). Kooreman (1994) vindt negatieve tekens voor alle vier kwaliteitsindicatoren, maar de statistische significantie is niet hoog (rond de 10 procent). SCP (1996) vindt ook aanwijzingen voor een negatief verband tussen doelmatigheid en kwaliteit. De inzet van beter opgeleid en meer ervaren verplegend personeel verlaagt de doelmatigheid, omdat de hogere loonkosten niet volledig door een hogere arbeidsproductiviteit worden gecompenseerd. Meer kapitaal per bed betekent ook minder doelmatigheid.

#### **Verband tussen kwaliteit en kosten blijkt soms positief, soms negatief**

Soms is het verband tussen kwaliteit en kosten positief, soms negatief, dat verschilt met de geboden kwaliteit van de zorgverlening. Dit blijkt uit een overzicht van Amerikaanse studies naar kosten en kwaliteit in verpleeghuizen (Weech-Maldonado, Mor en Oluwole, 2004). Zo blijkt het omlaagbrengen van doorligwonden vanaf een hoog niveau eerst heel veel extra kosten mee te brengen, daarna minder, en tenslotte zelfs geld op te leveren. Verpleeghuizen met een lage kwaliteit kunnen slechts tegen relatief hoge kosten de kwaliteit verbeteren. Denk aan het inhuren van voldoende personeel per cliënt. Na een zekere drempel nemen de kosten voor kwaliteitsverbetering sterk af. Verpleeghuizen met hoge kwaliteit kunnen tegelijkertijd de kwaliteit verbeteren en de kosten verlagen. De kosten voor behandeling van doorligwonden nemen af, het marktaandeel en daarmee de bezettingsgraad nemen toe.

De precieze relatie tussen kwaliteit en kosten verschilt tussen productdimensies. Zo ligt het verband voor depressiviteit weer anders: daar zijn de kosten voor kwaliteitsverbetering bij een laag niveau van kwaliteit laag en nemen ze toe na een zekere drempelwaarde.

### **4.3 Conclusie**

- Arcares (2004) en PWC (2002) meten de prestatie op niet in de doelmatigheidsmeting meegenomen productdimensies, kwaliteit dus, door enquêtes onder verzorgend personeel en cliënten uit te voeren. Deze meting van kwaliteit vormt een verbetering op de indicatoren van een 'kwaliteitsbeleid' zoals gebruikt in SCP (1996) en Kooreman (1994).
- Zowel Arcares (2004) als PWC (2002) laten geen duidelijk positief verband zien tussen de doelmatigheidsscoor en de prestatie op niet in de doelmatigheidsmeting meegenomen productdimensies.
- De gemeten doelmatigheid geeft dus een beperkte indicatie van de daadwerkelijke doelmatigheid.

## 5 Vertaling naar sectorniveau

In dit hoofdstuk gaan we in op de vertaling van de doelmatigheidsmeting naar sectorniveau. We gaan per AWBZ-zorgsector na in hoeverre de afzonderlijke studies voldoen aan de eisen van een dergelijke vertaling. We laten de problemen met de doelmatigheidsmeting op instellingsniveau uit hoofdstuk 3 en 4 even terzijde, in hoofdstuk 6 brengen we alle punten met elkaar in verband.

### 5.1 Steekproef representatief en voldoende groot?

Gemiddelde doelmatigheidsscores geven een goed beeld voor de hele sector als aan twee vereisten is voldaan. Allereerst moet de steekproef van instellingen representatief voor de sector zijn. Daarnaast is ook een minimum aantal waarnemingen nodig om voldoende betrouwbare uitspraken te doen op sectorniveau.

### 5.2 Verpleeghuiszorg (Arcares, 2004)

De steekproef telt 21 verpleeghuizen, wat neerkomt op 6% van de populatie (daarnaast zijn 17 gecombineerde verpleeg- en verzorgingshuizen in de steekproef opgenomen). De steekproef is feitelijk nog iets kleiner: zo blijken een aantal instellingen die tot één en dezelfde rechtspersoon horen identieke antwoorden te geven op vragen over de bedrijfsvoering (Arcares, 2004, p. 52). De deelname aan de benchmark is op vrijwillige basis.

Door de kleine steekproef bestaat een ruim betrouwbaarheidsinterval rond de gemiddelde doelmatigheidsscore. Een berekening op de achterkant van een sigarendoos laat bijvoorbeeld zien dat voor verpleeghuizen het 95% betrouwbaarheidsinterval voor de gemiddelde doelmatigheidsscore loopt van 78 tot 90 procent (uitgaande van een aselechte steekproef).

De steekproef is niet representatief; deelname is op vrijwillige basis. Navraag leert dat instellingen verschillende redenen hebben om deel te nemen. Het zijn niet alleen instellingen die hoge doelmatigheid en kwaliteit al als prioriteit hebben. Het is daarom niet duidelijk of door de steekproef een overschatting of een onderschatting van de gemiddelde doelmatigheidsscore ontstaat.

Conclusie is dat de steekproef niet representatief is en bovendien te klein om betrouwbare uitspraken op sectorniveau te doen.

De drie andere benchmark studies van verpleeghuizen hebben een veel grotere steekproef, waardoor vertekening door de steekproef kleiner en de statistische betrouwbaarheid van de resultaten groter is. Prismant (2005) gebruikt gegevens uit een enquête onder alle verpleeghuizen. Een respons van 66% levert een steekproef van 219 instellingen op (gegevens voor 2003). Ook SCP (1996) maakt gebruik van een enquête onder de gehele populatie. Een respons van 94% levert een steekproef op van 307 instellingen (gegevens voor 1993 en ook voor 1984 en 1987-1992). Tenslotte baseert ook Kooreman (1994) zich op een enquête onder de gehele populatie. Een respons van 91% levert 292 verpleeghuizen op (gegevens over 1989).

### **5.3 Verzorgingshuizen (Arcares, 2004)**

Ook hier geldt dat de steekproef niet toelaat betrouwbare en representatieve uitspraken te doen op sectorniveau. Het aantal verzorgingshuizen in de steekproef is 54, wat neerkomt op 4 procent van de populatie (daarnaast zijn 17 gecombineerde verpleeg- en verzorgingshuizen in de steekproef opgenomen). Deelname is op vrijwillige basis.

### **5.4 Thuiszorg (PWC, 2002)**

Bijna alle thuiszorginstellingen zijn onderdeel van de steekproef van PWC (2002). Het aantal instellingen is 96, wat neerkomt op 95 procent van de populatie. De steekproef is representatief en levert betrouwbare resultaten op.

### **5.5 Conclusie**

Vanwege een kleine en niet representatieve steekproef is vertaling naar sectorniveau voor Arcares (2004) niet mogelijk. In vergelijking met eerdere studies over verpleeg- en verzorgingshuizen is de steekproef veel kleiner, maar de bron van gegevens – vooral op het gebied van kwaliteit – veel rijker. PWC (2002) laat zien dat een meer uitgebreide gegevensverzameling niet noodzakelijk een kleine steekproef hoeft te betekenen.



## 6 Conclusies

In dit laatste hoofdstuk trekken we conclusies en geven we aan wat manieren zijn om de doelmatigheidsmeting in de toekomst te verbeteren.

### 6.1 Geen zicht op te behalen doelmatigheidswinsten

Recente benchmark studies lenen zich niet voor het bepalen van het potentieel voor doelmatigheidswinsten in de ouderenzorg, om de volgende redenen:

1. *Onvergelijkbaarheid door verschillen in kwaliteit*  
Bij de doelmatigheidsmeting blijven vele productdimensies zoals deskundigheid van de zorgverlening en bejegening van cliënten buiten beschouwing. Een lage doelmatigheidsscore kan betekenen dat de kosten hoog zijn door hoge kwaliteit van de zorg.
2. *Onvergelijkbaarheid door verschillen in zorgzwaarte*  
Grote verschillen in zorgzwaarte – en daarmee kosten per cliënt – tussen instellingen maakt het doen van sectorbrede uitspraken onmogelijk. De studies beperken zich tot uitspraken over doelmatigheid binnen clusters van instellingen met soortgelijke zorgzwaarte.
3. *Onvergelijkbaarheid door verschillen in regionale omstandigheden*  
Tenslotte bestaat geen goed beeld van de invloed van regionale factoren als arbeidsmarktomstandigheden op de doelmatigheid. Scores van instellingen in verschillende zorgkantorregio's zijn dus niet zomaar met elkaar te vergelijken.
4. *Steekproef levert geen representatief en betrouwbaar beeld op*  
Voor verpleeg- en verzorgingshuizen staat verder een kleine, niet representatieve steekproef betrouwbare uitspraken op sectorniveau in de weg.

### 6.2 Kort overzicht sterke en zwakke punten benchmark studies

In onderstaande tabel zetten we de sterke en zwakke punten van de vijf benchmark studies kort naast elkaar. Zoals hierboven gesteld, is voor alle studies de conclusie eensluidend: sectorbrede uitspraken over te behalen doelmatigheidswinsten zijn niet mogelijk. Aan de hand van de vier criteria in de tabel geven we aan welke studie het best scoort: dat is de studie van PWC (2002) naar de thuiszorg.

**Tabel 6.1 Sterke en zwakke punten van doelmatigheidsmeting in de vijf benchmark studies**

	Arcares (2004)	PWC (2002)	Prismant (2005)	SCP (1996)	Kooreman (1994)
Kwaliteit van de zorg	(-) Veel kwaliteitsaspecten buiten doelmatigheidsmeting; afzonderlijke uitgebreide meting kwaliteit uit cliëntenperspectief; maar onduidelijk verband met doelmatigheidsscore	(-) Idem Arcares	(- -) Buiten beschouwing	(-) Idem Arcares, maar kwaliteit (beperkt) gemeten uit instellingsperspectief; verband doelmatigheid en kwaliteit negatief	(-) Idem SCP, maar verband doelmatigheid en kwaliteit niet statistisch significant
Zorgzwaarte	(-) Doelmatigheidsscores alleen vergelijkbaar binnen clusters met soortgelijke zorgzwaarte	(-) Idem Arcares	(+/-) Onduidelijk of correctie geslaagd	(+) Indruk dat correctie hiervoor is geslaagd	(+) Idem SCP
Regionale factoren	(- -) Buiten beschouwing	(+/-) Onderzocht, maar grootte van effect onduidelijk	(- -) Buiten beschouwing	(-) Buiten beschouwing	(+/-) Alleen 'grote-stad' effect
Steekproef	(- -) Klein, niet representatief	(++) Bijna gehele populatie, representatief	(+) Idem PWC	(+) Idem PWC	(+) Idem PWC
Conclusie	Geen sectorbrede uitspraak over doelmatigheid mogelijk	Idem	Idem	Idem	Idem

Noot: Verklaring van gebruikte symbolen: (++) zeer goed; (+) goed; (+/-) onduidelijk; (-) slecht; (- -) zeer slecht.

### 6.3 Mogelijke verbeteringen

Vergeleken met de eerdere benchmark studies Kooreman (1994) en SCP (1996) winnen Arcares (2004) en PWC (2002) aan betrouwbaarheid van de gegevens en aan informatie over kwaliteit van de zorg. De thuiszorgstudie van PWC (2002) scoort ook goed wat de grootte van de steekproef betreft. Prismant (2005) behoort tot de 'buitencategorie' omdat dit onderzoek kwaliteit buiten beschouwing laat en ook geen rekening houdt met verschillende regionale omstandigheden. Welke verbeteringen zouden in de toekomst tot een betere doelmatigheidsmeting leiden?

#### (1) Meer productdimensies in doelmatigheidsmeting

Voor een betere doelmatigheidsmeting is het noodzakelijk om meer productdimensies mee te nemen. Bij verpleeghuizen kan het bijvoorbeeld gaan om het aantal cliënten met doorligwonden. Er bestaat wel een afweging tussen meer productdimensies in de analyse en voldoende onderscheidend vermogen. Hoe meer productsoorten in de DEA analyse, hoe meer 'unieke' instellingen er zijn en dus hoe meer instellingen 100% efficiënt zijn. In de thuiszorg zijn de grenzen snel bereikt: daar bestaan immers maar ongeveer 100 instellingen. Bij verpleeg-

en verzorgingshuizen liggen deze grenzen ruimer. Een andere afweging bij het opnemen van nieuwe productdimensies is dat alle productsoorten even zwaar meetellen in de DEA analyse. De vraag is of we 'productie' (bijvoorbeeld verpleegdagen) even zwaar moeten wegen als 'kwaliteit' (bijvoorbeeld bejegening).

## **(2) Betrouwbare gegevens voor een groter aantal verpleeg- en verzorgingshuizen**

Belangrijkste prioriteit voor verpleeg- en verzorgingshuizen is het verbeteren van de gegevensvoorziening. Arcares (2004) heeft weliswaar een brede set van betrouwbare gegevens verzameld, maar de steekproef is veel te klein om rekening te kunnen houden met verschillen in zorgzwaarte, kwaliteit en regionale omstandigheden. Wellicht dat gegevens van de Inspectie Gezondheidszorg over kwaliteit van de zorg ook zijn te gebruiken bij de analyse van doelmatigheid.

Met een grote steekproef lijken verschillen in zorgzwaarte goed op te vangen met een voldoende aantal product- en kostensoorten in de DEA, zo laten Kooreman (1994) en SCP (1996) zien. Dit heeft de voorkeur boven het onderscheiden van clusters van instellingen met soortgelijke zorgzwaarte. Allereerst zijn clustergrenzen altijd enigszins willekeurig. En als een grote mate van homogeniteit bestaat binnen een cluster, dan is er het probleem dat sectorbrede uitspraken onmogelijk zijn.

## **(3) Houd rekening met verschillende omgevingsfactoren**

PWC (2002) laat zien dat factoren buiten de invloed van het management de doelmatigheid beïnvloeden, denk bijvoorbeeld aan regionale arbeidsmarktomsomstandigheden. Om hiervoor te corrigeren kunnen omgevingsfactoren bijvoorbeeld als verklarende variabele in een regressieanalyse van de doelmatigheidsscores worden opgenomen.

## **(4) Gebruik innovaties in methoden**

De ontwikkeling van methoden van doelmatigheidsmeting kent een snelle voortgang. Zo kan DEA nu ook beter omgaan met meetfouten: hiervoor is een 'stochastische DEA' ontwikkeld. Wanneer betere gegevens beschikbaar komen, dan kan met deze nieuwe methoden doelmatigheid nog beter worden gemeten.

### **Tot slot**

Ook als de doelmatigheidsmeting op al deze manieren wordt verbeterd, blijven de resultaten met onzekerheid omgeven. Correctie voor niet-waargenomen factoren die de *gemeten* doelmatigheid buiten het management om beïnvloeden is beperkt. Meetfouten blijven altijd spelen. De thuiszorg kent maar 100 instellingen, waardoor correctie voor factoren als externe factoren beperkt blijft. Verschillende methoden van doelmatigheidsmeting met ieder hun voor-

en nadelen kunnen een ander beeld geven. Uit analyses blijken de verschillen in doelmatigheidsscores in grote mate een ‘zwart gat’: recente kwalitatieve studies naar het effect van bedrijfsvoering op de score geven geen duidelijk beeld (Arcare, 2004, PWC, 2002); kwantitatieve studies laten verreweg de meeste variantie in doelmatigheid onverklaard (SCP, 1996, Kooreman, 1994). Kooreman (1994a, p. 346) merkt daarom op dat een score vooral een ‘signaalwaarde’ heeft:

*A low score is primarily an indication of the presence of special circumstances. Whether this is tantamount to “true” inefficiency is a question that has to be addressed next, not necessarily by means of the same data and statistical methods.*

## Referenties

Arcares (ATOS, Stichting Cliënt en Kwaliteit, Customers Choice, Economic Programs, Van Loveren en partners, PWC Accountants, Prismant), 2004, *Benchmark verpleeg en verzorgingshuizen 2003. Prestaties van aanbieders gemeten*, onderzoek in opdracht van Arcares and Ministerie van VWS, Den Haag.

Blank, J.L.T., 1998, Benchmarken of de kunst van het vergelijken, *ESB*, 83 (4154), p. 432-434.

Blank, J.L.T. and E. Eggink, 1996, *Een empirisch onderzoek naar de productiestructuur van verpleeghuizen in Nederland*, SCP, Den Haag.

Dell, M. and L.J.R. Vandermeulen, 2005, *Arbeidsproductiviteit in de zorg*, Prismant, Utrecht.

Eggink, E. and J.L.T. Blank, 2001, *Verstandig verzorgd. Een empirisch onderzoek naar de efficiëntie van de intramurale zorg voor verstandelijk gehandicapten*, SCP, Den Haag.

Inspectie Gezondheidszorg, 2004, *Verpleeghuizen garanderen minimale zorg niet*, Den Haag.

Kooreman, P., 1994, Nursing home care in The Netherlands: a nonparametric efficiency analysis, *Journal of Health Economics*, 13 (3), pp. 301-316.

Kooreman, P. 1994a, Data envelopment analysis and parametric frontier estimation: complementary tools, *Journal of Health Economics*, 13 (3), pp. 345-346.

Prismant, 2005, *Arbeidsproductiviteit in de zorg*, Utrecht.

PWC Consulting, 2002, *Benchmarkonderzoek 2000 verscherpt inzicht in prestaties en bedrijfsvoering thuiszorginstellingen*, PWC, Almere/Utrecht.

Weech-Maldonado, Robert, Vince Mor en Adetokunbo Oluwole, 2004, Nursing home costs and quality of care: is there a tradeoff?, *Expert Review of Pharmacoeconomics and Outcomes Research*, 4 (1), pp. 99-110.