

Efficiëntie, concurrentie en globale budgetten in de zorg

Rudy Douven en Jan Boone

Wanneer de zorguitgaven harder stijgen dan het inkomen dan wordt het steeds moeilijker om de zorg collectief te financieren. De stijgende zorguitgaven kunnen dan andere collectieve uitgaven, zoals onderwijs, defensie etc. uithollen. We bekijken in dit artikel of het gebruik van budgetten, in een systeem van gereguleerde concurrentie, een oplossing is om de zorguitgaven te beteugelen zonder dat het ten koste gaat van de welvaart. We laten zien dat onder bepaalde voorwaarden budgetten de welvaart kunnen verhogen. Dit is het geval wanneer budgetten de allocatieve efficiëntie verhogen zonder dat dit ten koste gaat van de productieve efficiëntie.

1 Inleiding

Een belangrijke uitdaging in het zorgstelsel van gereguleerde concurrentie is de verbetering van de efficiëntie in de zorg. Het Taskforce rapport (Ministerie van VWS, 2012) geeft aan dat de invoering van het nieuwe zorgstelsel heeft geleid tot een toename van de doelmatigheid en een hogere productiviteit in delen van de zorg. De hogere productiviteit heeft echter ook een nadeel omdat het heeft geleid tot een sterke groei van het zorgvolume en oplopende zorgkosten. Het Taskforce rapport laat zien dat dit hoge zorgvolume niet altijd wenselijk is. Zo worden onder druk van het aanbod de indicaties opgerekt en richt het aanbod zich steeds meer op een lichtere zorgvraag, waardoor zorg geleverd wordt aan een steeds grotere groep mensen. De meest recente cijfers laten weliswaar een lage groei van de zorguitgaven voor 2012 en begin 2013 zien, maar een belangrijke uitdaging voor de toekomst is om de structurele groei van de zorguitgaven terug te brengen naar een houdbaar groeipad.

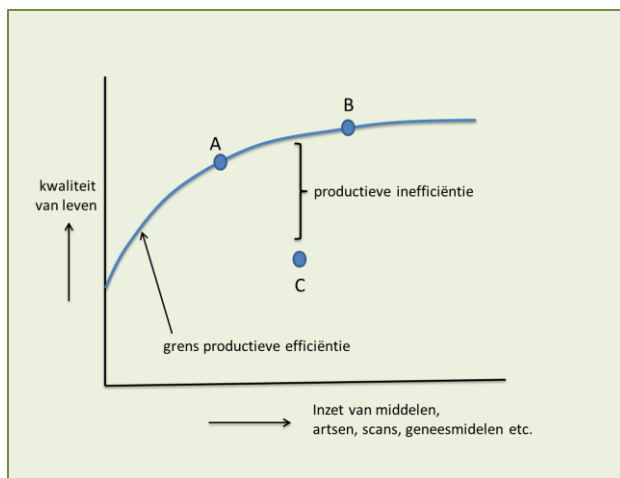
In dit artikel behandelen we de aanbodkant van de zorg en gaan we na hoe de doelmatigheid in de zorg verbeterd kan worden. Hierbij maken we een onderscheid tussen productieve en allocatieve efficiëntie. We beargumenteren dat het huidig stelsel van gereguleerde concurrentie vooral gericht is op het verbeteren van de productieve efficiëntie, maar dat er meer prikkels nodig zijn voor een verbetering van de allocatieve efficiëntie in de zorg. We geven aan dat overbehandeling in de zorg een probleem is, omdat zorgaanbieders de neiging hebben meer te letten op de medische effectiviteit van behandelingen dan op de kosten. Toekomstig beleid

dient meer gebruik te maken van het kosteneffectiviteitscriterium. Bij dit criterium worden de baten van een behandeling afgezet tegen de kosten. Dit kan de allocatieve efficiëntie in de zorg verbeteren. Globale budgetten kunnen helpen bij het bereiken van meer kosteneffectieve zorg. Een belangrijke voorwaarde is wel dat artsen behandelingen voldoende moeten kunnen prioriteren; dat wil zeggen dat ze bij hun keuze tussen wel of niet behandelen van een patiënt voldoende onderscheid moeten kunnen maken tussen welke behandelingen wel en niet kosteneffectief zijn. Daarnaast laten we zien dat een globaal budget effectiever is wanneer het wordt toegepast op een gegeven patiëntenpopulatie, en er bij de bepaling van het budget rekening wordt gehouden met het risicoprofiel van de populatie en de geleverde kwaliteit van de zorg.

Bij de beschrijving van de theorie van allocatieve en productieve efficiëntie maken we gebruik van de zorgliteratuur uit de VS (zoals Baicker en Chandra 2011; Chandra, Cutler en Song 2011; Chandra, Jena en Skinner 2011; Chandra en Skinner 2012) en het Centraal Planbureau (Douven 2013). Het doel van het artikel is om deze theorieën uit te leggen en vervolgens toe te passen op de Nederlandse zorgmarkt en te bezien of er belangrijke beleidslessen zijn te trekken.

2 Productieve en allocatieve efficiëntie

In de zorg bestaan vele vormen van efficiëntie. Een specialist kan hartoperaties technisch gezien perfect uitvoeren. Dat wil bijvoorbeeld zeggen dat de hartoperatie van hoge kwaliteit is, relatief niet duur is en naar de tevredenheid van de patiënt verloopt. Dit hoeft echter niet te betekenen dat de keuze om de patiënt te behandelen met een hartoperatie optimaal is. Misschien was het goedkoper en effectiever geweest om met een goedkoop geneesmiddel hetzelfde resultaat te bereiken. Misschien was het zelfs beter geweest om de patiënt helemaal niet te behandelen en zou de totale welvaart hoger zijn geweest wanneer we hetzelfde bedrag aan een betere verkeersveiligheid of hogere dijken hadden besteed. Om de verschillende vormen van efficiëntie beter te kunnen duiden, onderscheiden we twee soorten efficiëntie.

Figuur1 Productieve en allocatieve efficiëntie

Productieve efficiëntie. Gegeven dat er een behandeling nodig is, probeer je bij productieve efficiëntie met de beschikbare middelen zoveel mogelijk gezondheidswinst te bereiken. Dit kan efficiëntie in kosten betreffen. Bijvoorbeeld, wanneer er voor een ziekte twee verschillende behandelingen mogelijk zijn die hetzelfde resultaat opleveren, dan is het efficiënter om de goedkoopste behandeling uit te voeren. Maar het kan ook een efficiënt gebruik van input of productiefactoren betreffen. Bij gelijke input is het efficiënter om die behandeling te kiezen waarvan de baten, in termen van gezondheidswinst, het grootst zijn. In Figuur 1 geven we dit weer. Op de horizontale as staan de middelen om de gezondheid te verbeteren, zoals specialisten, verzorgers, ziekenhuizen, geneesmiddelen, scans en type behandeling. Hoe meer we van deze middelen gebruiken, hoe hoger de zorguitgaven zijn. Op de verticale as staan de baten van behandelen; de gezondheidswinst die bereikt wordt door het inzetten van gezondheidszorg. De grens van productieve efficiëntie geeft weer welke gezondheidswinst valt te behalen bij een optimale inzet van deze middelen. Hierdoor krijgt de grens van productieve efficiëntie een concave vorm. In Figuur 1 noemen we het productieniveau in punt C productief inefficiënt. Immers, met dezelfde middelen kan meer gezondheidswinst worden bereikt wanneer die middelen efficiënter worden ingezet.

Allocatieve efficiëntie. Een behandeling is allocatief efficiënt wanneer de behandeling echt nodig is, of wanneer de baten van de behandeling groter zijn dan de baten van dezelfde uitgaven elders in de economie. Dit komt erop neer dat geld in de economie daar moet worden uitgegeven waar het nut het grootst is. Wanneer één extra euro in onderwijs of defensie meer oplevert dan één extra euro in de zorg, dan is het efficiënter om deze euro aan onderwijs uit te geven. Dit uitgangspunt wijkt af van een meer bredere definitie van allocatieve efficiëntie waarbij het gaat

om zowel de verdeling van middelen binnen de zorgsector als tussen de zorgsector en andere sectoren in de economie. In ons betoog valt een efficiënte verdeling van middelen binnen de zorgsector onder productieve efficiëntie. Wanneer zorg een gewoon consumptiegoed zou zijn, zouden consumenten zelf de afweging maken tussen het consumeren van zorg en het nut van andere goederen in de economie. De zorg wordt voor het overgrote deel echter collectief gefinancierd; de overheid bepaalt in belangrijke mate welke basiszorg voor iedereen belangrijk is. Wanneer we de collectieve zorguitgaven willen beperken, dan is dit eigenlijk alleen nodig wanneer de marginale baten van hogere collectieve zorguitgaven lager zijn dan de baten van marginale collectieve uitgaven in andere sectoren van de economie. Wanneer dit niet het geval is, mag de zorg in principe verder groeien. Uiteindelijk is dit een keuze van de overheid, en hierin zit ook opgesloten welke mate van solidariteit we als maatschappij in de zorg wenselijk achten. Wanneer overheden convenanten of budgetten in de zorg aankondigen om de uitgaven te beperken, dan kan dat gezien worden als een signaal dat de zorg allocatief inefficiënt is.

In Figuur 1 worden twee verschillende keuzes weergegeven op de grens van productieve efficiëntie: A en B. De keuze geeft aan hoeveel we maatschappelijk bereid zijn te betalen aan zorg in verhouding tot andere goederen in de economie. In punt A kiezen we voor minder zorg en lagere belastingen (ofwel meer andere bestedingen); in punt B kiezen we juist voor meer zorg. Bij punt A kan men bijvoorbeeld denken aan een kleiner basispakket, hogere verplichte eigen risico's, of een minder snelle diffusie van nieuwe technologie dan in punt B. Bij die keuze is het maatschappelijk gezien belangrijk hoe we extra gezondheidswinst of kwaliteit van leven waarderen. Wanneer we maatschappelijk gezien punt A als optimaal beschouwen, dan is punt B allocatief inefficiënt. Immers, de extra benodigde middelen en de euro's die we aan gezondheidszorg uitgeven om van A naar B te komen, wegen niet op tegen de extra gezondheidswinst en het verlies van andere goederen in de economie.

3 Kosteneffectiviteit en kwaliteit van de zorg

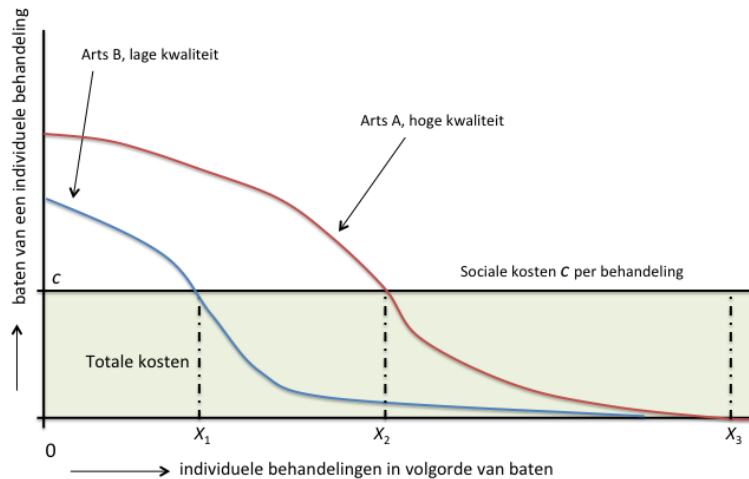
Iedereen wil een goede kwaliteit zorg en veel mensen vinden dat niet goed genoeg en willen zelfs de beste kwaliteit van zorg. Om consumenten deze keuzes te kunnen geven is kwaliteitsinformatie nodig om een beter inzicht te krijgen in het handelen van de arts. Toch is inzicht in de kwaliteit van de geleverde zorg nog maar beperkt aanwezig.

Wanneer het gaat over de kwaliteit van voetballers dan weten veel mensen wat goede en minder goede voetballers zijn. Goede voetballers kunnen immers wekelijks worden geobserveerd door miljoenen mensen. Bij artsen zijn de kwaliteitsverschillen veel lastiger te achterhalen, omdat de resultaten van hun activiteiten veel moeilijker zijn te observeren en te meten. Maar ook bij artsen zouden de kwaliteitsverschillen wel eens groot kunnen zijn. Artsen die een hogere kwaliteit leveren, kunnen meer patiënten behandelen. Zo laten Chandra en Staiger

(2007) zien dat specialisatie in het behandelen van hartaanvallen leidt tot een hogere kwaliteit per behandeling en dat als gevolg ook meer patiënten konden worden behandeld. De artsen trokken dus niet alleen meer patiënten aan omdat ze een betere kwaliteit leverden, maar ook omdat hun toegevoegde waarde per behandeling hoger is. Dit geldt vooral daar waar er sprake is van specialisatie in de zorg. Hoewel het in Nederland nog niet voorkomt, zou men bij het bekostigen van artsen rekening willen houden met het effect dat een betere kwaliteit zorg leidt tot een groter behandelvolume.

Een hypothetisch voorbeeld geven we in Figuur 2, waarbij we een situatie veronderstellen waarbij er één arts is die een groep patiënten behandelt. We beschouwen twee verschillende typen artsen. Arts A heeft langere scholing en training ondergaan en is beter op de hoogte van de nieuwste medische ontwikkelingen in zijn vakgebied dan arts B. We veronderstellen dat beide artsen productief efficiënt behandelen maar hun opleiding zorgt ervoor dat hun productieve efficiëntiegrens, zoals aangegeven in Figuur 1, verschilt. Dit laten we in Figuur 2 zien door voor beide artsen de behandelingen te rangschikken in volgorde van baten, waarbij de behandelingen met de hoogste baten links zijn gelegen en de laagste baten rechts. In de Figuur 2 is de kwaliteit van arts A hoger dan van arts B, omdat de baten van een behandeling voor arts A voor iedere patiënt groter zijn dan voor arts B. Wanneer we voor het gemak de sociale kosten van een behandeling gelijk aan c veronderstellen, dan stellen economen dat de behandelingen door arts A kosteneffectiever zijn dan door arts B.

Wanneer artsen bekend zijn met hun eigen kwaliteit van behandelen en daarnaast geschoold en getraind zijn om alleen dan te behandelen wanneer een behandeling kosteneffectief is, krijgen we vanzelf het gewenste optimum (aangenomen dat een arts voldoende betaald wordt voor zijn verrichtingen). Dit optimum van kosteneffectiviteit geven we in Figuur 1 weer. Arts B voert in het optimum X_1 behandelingen uit en arts A X_2 behandelingen.

Figuur 2 Kwaliteit en kosteneffectiviteit

Dit geschetste optimum van kosteneffectiviteit is in de praktijk moeilijk te bereiken omdat de zorg wordt gekenmerkt door een groot aantal marktimperfecties. Een eerste imperfectie is dat artsen de neiging hebben om patiënten te helpen, zelfs wanneer dat hoegenaamd geen gezondheidswinst oplevert. Chandra en Skinner (2012) stellen bijvoorbeeld dat het merendeel van de artsen opgeleid en getraind, en dus intrinsiek gemotiveerd zijn, om mensen beter te maken en dat ze alles in het werk zullen stellen om de gezondheid van patiënten te bevorderen, ook als de baten in de verste verte niet opwegen tegen de kosten.

In economische termen betekent dit dat ze meer geneigd zijn om medische effectiviteit als criterium te nemen voor het wel of niet behandelen van patiënten. Dit is een belangrijk verschil met het kosteneffectiviteitscriterium omdat ze bij het medische effectiviteitscriterium niet of nauwelijks op de kosten van een behandeling letten. Wanneer een arts geneigd is om alleen op het medische effectiviteitscriterium te letten dan resulteert dat in een toename van het aantal behandelingen. In Figuur 2 wordt dit weergegeven. Arts A heeft dan de neiging om in plaats van X_2 , X_3 behandelingen uit te voeren en arts B in plaats van X_1 , X_3 behandelingen. De extra behandelingen zijn dus wel effectief, maar niet kosteneffectief.

Naast het feit dat artsen de neiging hebben exclusief op de medische effectiviteit van een behandeling te letten, spelen nog een aantal andere imperfecties in de zorg een rol die artsen stimuleren om meer in richting van X_3 te behandelen. Zo kan de arts bij het bepalen van een behandeling beïnvloed worden door de wens van de patiënt. Patiënten hebben de premie immers al betaald en wensen (of eisen) dan een behandeling zolang deze positief bijdraagt aan hun gezondheid. Deze wens van de patiënt zal groter zijn bij een volledige verzekering, of wanneer er geen, of lage, eigen betalingen bestaan. Dit mechanisme kan nog versterkt worden wanneer artsen per behandeling betaald krijgen. Artsen kunnen dan hun inkomen verhogen door meer te behandelen. Bij specialisten die betaald

worden per behandeling, speelt dit een grotere rol dan bij specialisten die in loondienst werken (Douven et al. 2012).

4 Globale budgetten in de zorg

Wanneer we ervan uitgaan dat artsen voornamelijk letten op de medische effectiviteit van een behandeling dan is voor een betalende partij, zoals de overheid of een verzekeraar, de uitdaging om een betalingssysteem te bedenken waarbij artsen aangespoord worden om minder te behandelen. In dit artikel beschouwen we een globaal budget voor (individuele) zorgaanbieders. Bij een globaal budget krijgen zorgverleners een vast budget, waarbij ze als tegenprestatie een minimum aantal behandelingen moeten uitvoeren. Een minimum aantal behandelingen veronderstellen we hier om te voorkomen dat een arts een globaal budget ontvangt, zonder daarvoor een tegenprestatie te leveren. In de praktijk zijn ook andere mechanismen denkbaar. Zo bestaan er in de VS momenteel zogenoemde ‘Accountable Care Organizations’ (ACO) op de publieke verzekeringsmarkt en ‘Alternative Quality Contracts’ (AQC) op de private verzekeringsmarkt. De ACO en AQC staat voor een groep zorgaanbieders die verantwoordelijk is voor een vooraf gedefinieerde verzekerdenpopulatie. Vaak spreekt de verzekeraar met de ACO of AQC een globaal budget af. Voor de bepaling van dit budget vindt risicoverevening plaats, dat wil zeggen dat het budget wordt afgestemd op de risicokenmerken van de populatie. Achteraf kan het globale budget worden aangepast op basis van de geleverde kwaliteit. Bij een AQC wordt dit gebaseerd op zestig kwaliteitsindicatoren (Song et al. 2012). Bij een globaal budget wordt dus het delen en dragen van risico’s omtrent uitgaven en medische effectiviteit van behandeling anders vorm gegeven. Een belangrijke voorwaarde voor een budget is wel dat zorgaanbieders voldoende inzicht hebben in de medische effectiviteit en uitgaven. Een andere belangrijke voorwaarde is dat het introduceren van een budget de concurrentie in de zorg niet mag dwarsbomen (Boone en Douven 2012). Hier gaan we in dit artikel nader op in.

Wanneer de betalende partij volledige informatie heeft over alle aspecten van de behandeling, zoals de kwaliteit, productiekosten en de kosteneffectiviteit van behandelingen, dan is de bepaling van een globaal budget voor een arts in Figuur 2 eenvoudig en kunnen we rekening houden met de kwaliteitsverschillen tussen artsen. Arts A krijgt een hoger globaal budget dan arts B, waarbij arts A meer behandelingen moet uitvoeren dan arts B (X_2 tegenover X_1).

In de praktijk is er een aantal aandachtspunten bij het operationaliseren van deze oplossing. Een eerste vraag is of artsen voldoende kennis hebben om behandelingen te kunnen prioriteren op basis van het kosteneffectiviteitscriterium. Immers, wanneer arts A minimaal X_2 behandelingen dient uit te voeren, dan zijn dat idealiter alle behandelingen gelegen in het interval $[0, X_2]$. We gaan hier dieper op in de eerste tekstbox.

Een tweede vraag is of artsen zo makkelijk te bewegen zijn om alleen kosteneffectieve behandelingen uit te voeren. Artsen zullen dit moeilijk vinden, zeker wanneer ze geschoold zijn om te behandelen op basis van het medische effectiviteitscriterium. Om wat meer inzicht te krijgen, maken we hier twee veronderstellingen over de nutsfunctie van de arts: (i) artsen ondervinden disnut voor elke patiënt die ze niet mogen behandelen waar wel een effectieve ingreep voorhanden is en (ii) dit disnut is hoger naarmate de effectiviteit van de behandeling hoger is. Deze veronderstellingen zorgen ervoor dat een arts behandelingen prioriteert op basis van medische effectiviteit. Verder levert het afwijzen van een bepaalde patiënt met gegeven X een hoger disnut op voor arts A dan B in Figuur 2, omdat de behandelingen van arts A effectiever zijn.

Gegeven deze nutsstructuur van de arts, is het niet optimaal voor de overheid om kosteneffectieve zorg af te dwingen. De arts moet dan gecompenseerd worden voor het disnut over de effectieve behandelingen die ze niet mag uitvoeren. Wanneer dit disnut van de arts groot is, kan het optimaal zijn om de arts toe te staan meer te behandelen dan X_1 en X_2 voor respectievelijk arts B en arts A. Dus afhankelijk van de nutsstructuur van de arts, kan het optimaal zijn om enige overbehandeling toe te staan, maar niet tot aan het punt X_3 . In termen van Figuur 1 betekent dit dat het allocatief efficiënt kan zijn om zorg toe te laten die niet kosteneffectief is.

Op de lange termijn ligt het meer voor de hand om artsen meer bewust te maken van het kosteneffectiviteitscriterium, en opleidingen hierop aan te passen. Op deze manier wordt de nutsfunctie van nieuwe artsen aangepast. In Figuur 1 betekent dit dat op de lange termijn het allocatief inefficiënte punt B wordt verschoven naar het meer allocatief efficiënte punt A.

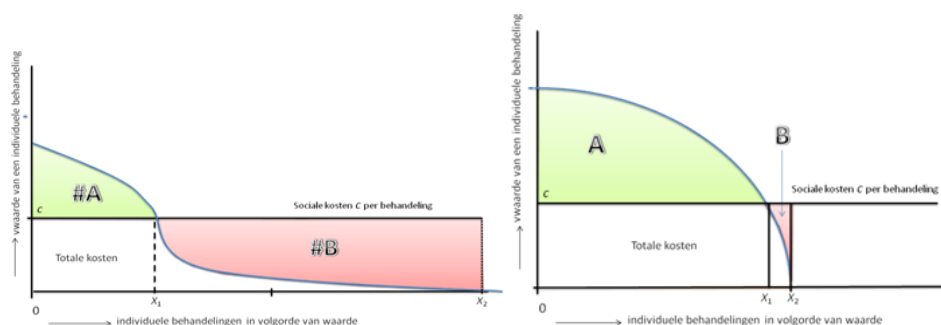
Kunnen artsen behandelingen prioriteren?

Een globaal budget werkt welvaartsverhogend wanneer artsen behandelingen kunnen prioriteren in volgorde van kosteneffectiviteit. Het idee hierachter is eenvoudig. Wanneer er overbehandeling is, kunnen we met minder behandelingen toe. Door het aantal behandelingen te begrenzen, bijvoorbeeld door het aantal behandelingen iets lager te stellen dan vorig jaar, wordt een arts gedwongen om te gaan kiezen: welke behandelingen blijf ik wel doen en welke behandelingen ga ik niet meer doen?

De arts dwingen om keuzes te maken is niet eenvoudig. Zo weet een arts niet welke behandelingen in de toekomst nog gaan komen. Een arts kan ook besluiten om niet te kiezen en alle behandelingen gewoon uit te voeren. Wanneer het aantal behandelingen op is, ontstaan er wachtlijsten. In Figuur 3 laten we twee soorten behandelingen zien. In het linkerpaneel van Figuur 3 geldt dat $\#A < \#B$. Dat wil zeggen, wanneer alle behandelingen $[0, X_2]$ worden uitgevoerd, is het resultaat niet-kosteneffectief. Een globaal budget verhoogt dan altijd de welvaart, ook wanneer artsen behandelingen niet goed kunnen prioriteren. Wachtlijsten zullen toenemen, maar dit is voor een groter deel efficiënt dan inefficiënt (Boone en Douven 2012). Bij de behandeling in het rechterpaneel geldt dat $\#A > \#B$. Wanneer artsen niet kunnen prioriteren, leidt inperking van het aantal behandelingen altijd tot welvaartsverlies. Mogelijke voorbeelden van behandelingen behorend bij linkerpaneel zijn behandelingen van rugpijn, of het plaatsen van stents in bloedvaten en voor het rechterpaneel anti-retrovirale therapie bij HIV- en AIDS-patiënten (Chandra en Skinner 2012).

Empirische evidentie over prioriteren is schaars. Baicker et al. (2006) vinden aanwijzingen hiervoor bij verloskunde. Een andere aanwijzing is Van de Vijzel et al. (2011). Zij laten zien dat het afschaffen van de budgettering in 2001 vooral heeft geleid tot een toename van behandelingen zonder duidelijke diagnose, hetgeen impliceert dat de kosteneffectiviteit ervan onduidelijk was. In de praktijk zal een arts niet precies tussen behandelingen kunnen discrimineren. De conclusie is dat een globaal budget meer zin heeft bij behandelingen waarbij de arts de kosteneffectiviteit goed kan inschatten, of wanneer de gemiddelde behandeling niet kosteneffectief is zoals in het linkerpaneel van Figuur 3.

Figuur 3 Twee soorten behandelingen



Een derde aspect is dat de betalende partij vaak gebrek aan informatie over kwaliteit heeft. Wanneer de kwaliteit bekend is, zou de betalende partij daar rekening mee willen houden en de hogere-kwaliteitsarts A meer behandelingen laten uitvoeren dan arts B. Wanneer deze kwaliteitsinformatie ontbreekt, bestaat er asymmetrische informatie tussen de arts en de betalende partij. Arts B met lage kwaliteit kan dan gebruik proberen te maken van de asymmetrische informatie en claimen dat zij de behandelingen die arts A kan uitvoeren ook zelf zou kunnen uitvoeren (averechtse selectie).

De vraag is nu hoe de betalende partij het globale budget gaat vormgeven. Stel, de betalende partij is bekend met het verschil in de kwaliteit van artsen maar weet niet precies welke van de twee artsen ze nu voor zich heeft. De vraag is dan hoe de betalende partij een globaal budget moet opstellen met als tegenprestatie dat de arts een minimum aantal behandelingen uitvoert. Het contract moet in ieder geval zo worden opgesteld dat het aantrekkelijk is voor beide type artsen om het contract uit te voeren (individual rationality constraint). De betalende partij kan dan twee contracten opstellen: één globaal budget met een minimaal aantal behandelingen X_a , bedoeld voor arts A, en één globaal budget met minimaal behandelingen X_b , bedoeld voor arts B. Wanneer de betalende partij arts B voor zich heeft dan bestaat het risico dat deze arts ook voor het contract behorende bij arts A kiest en X_a behandelingen wil uitvoeren (incentive compatibility constraint). Wanneer we ervan uitgaan dat artsen graag willen behandelen, bestaat er als het ware een neiging tot overbehandeling, omdat de minder kwaliteit arts de hoge kwaliteit arts wil na-apen. In Boone en Douven (2013) maken we gebruik van de techniek ‘mechanism design’ en laten we zien dat onder bepaalde voorwaarden de betalende

partij dit probleem van na-apen kan proberen te voorkomen door een optimaal contract te introduceren.

Het optimale contract heeft twee opvallende eigenschappen. Ten eerste wordt het contract bedoeld voor lage-kwaliteitsarts B zo opgesteld dat het voor hoge-kwaliteitsarts A onaantrekkelijk wordt om B na te apen. Dat wil zeggen arts B hoeft minder te behandelen (lage X_b) en krijgt daarvoor een kleine vergoeding. Ten tweede wordt het contract bedoeld voor de hoge-kwaliteitsarts A zo opgesteld dat dit contract voor de lage-kwaliteitsarts B niet haalbaar is. Het contract eist daarom een bepaalde mate van overbehandelen van de hoge kwaliteitsarts A (hoge X_a). Asymmetrische informatie (vergeleken met complete informatie) kan dus leiden tot hogere kosten en meer overbehandeling. Het gedrag van artsen dat we hier schetsen is in feite het spiegelbeeld van het standaardvoorbeeld uit de arbeidseconomie, waarbij goede werknemers zich kunnen voordoen als minder goede werknemers (zie volgende tekstbox).

Niet contracteerbaarheid van kwaliteit leidt tot een lagere productie

Laffont en Tirole (1993) laten zien dat werkgevers soms slimme prikkels moeten gebruiken om te voorkomen dat werknemers gaan onderpresteren. Dit probleem kan optreden wanneer de werkgever wel de productie van een werknemer kan observeren, maar niet de kwaliteit van die productie kan contracteren. Beschouw het volgende voorbeeld. Als we veronderstellen dat een goed geschoolde werknemer meer productie kan leveren dan een minder goed geschoolde werknemer, dan zal het lastig zijn voor de minder goed geschoolde werknemer om dezelfde productie te leveren als de goed geschoolde werknemer. Andersom ligt dat echter anders. De goed geschoolde werknemer heeft een prikkel om zich te gaan gedragen als een minder goed geschoolde werknemer en kan kiezen voor minder productie en meer vrije tijd. De werkgever moet dit gedrag van de hoog geschoolde werknemer zien te vermijden. Macho-Stadler en Perez-Castrillo (2001) laten zien dat dit probleem van averechtse selectie in sommige gevallen kan worden opgelost door de werkgever informatie-rente te laten betalen. Een optie voor de werkgever is om de minder goed geschoolde werknemer minder te laten produceren en een lager loon uit te betalen. De winst voor de werkgever bestaat er dan uit dat de goed geschoolde werknemer zich niet gaat gedragen als minder goed geschoolde werknemer. Een andere optie is om de goede werknemer extra te betalen om te voorkomen dat deze zich gaat gedragen als een minder goed geschoolde werknemer.

In Boone en Douven (2013) laten we zien dat de situatie verandert wanneer beide artsen op de markt zijn en concurreren om één selectief contract. Bij een gebrek aan kwaliteitsinformatie zal de betalende partij dan altijd een contract opstellen dat bedoeld is voor de lagekwaliteitsaanbieder. De reden is eenvoudig. Wanneer de betalende partij een contract afsluit, bedoeld voor de lagekwaliteitsaanbieder, dan is hij er zeker van dit (goedkopere) contract wordt uitgevoerd. Wanneer kwaliteitsinformatie ontbreekt, lukt het namelijk niet om een contract op te stellen voor de hoge-kwaliteitsaanbieder. Stel dat de betalende partij wel probeert met de hoge-kwaliteitsaanbieder te contracteren, dan bestaat er altijd het risico dat de lagekwaliteitsaanbieder (vermomd als hoge-kwaliteitsaanbieder) het contract gaat uitvoeren, waardoor per saldo de welvaart lager uitvalt. Om dit te voorkomen moet de betalende partij de lagekwaliteitsaanbieder een budget geven, terwijl ze weinig

patiënten hoeft te behandelen. Maar zo'n contract is ook aantrekkelijk voor een hogekwaliteitsarts. Een dergelijk contract kan niet incentive compatible worden aangeboden. Dit is een belangrijke bevinding. Het betekent dat verzekeraars, wanneer ze moeten kiezen tussen twee soorten behandelingen waarvan de kwaliteit onbekend is, alleen voor de hogerekwaliteitsbehandeling zullen kiezen wanneer de aanbieder dat ook echt kan aantonen. In andere gevallen zullen ze eerder kiezen voor de (goedkopere) lagerekwaliteitsbehandelingen.

We concluderen dat allocatieve efficiëntie in de zorg niet eenvoudig is te bereiken. Ten eerste hebben artsen de neiging om alleen te letten op de medische effectiviteit van zorg, waardoor er veel behandelingen plaatsvinden en waarvan de baten lager zijn dan de kosten. Ten tweede wordt dit effect versterkt, omdat de informatie van de overheid of de verzekeraar gebrekkig is. Zo kan de kwaliteit van zorg moeilijk worden geobserveerd en gecontracteerd. Dit kan het voor artsen aantrekkelijker maken om meer te behandelen, zeker wanneer artsen vooral letten op het effectiviteitscriterium en het daarnaast ook financieel aantrekkelijk is. Een andere bevinding is dat lagerekwaliteitsartsen een grote marktmacht hebben wanneer kwaliteitsinformatie ontbreekt. Ze kunnen zich dan voordoen als hogekwaliteitsarts en zullen gecontracteerd worden voor een hoger budget dan wanneer kwaliteitsinformatie wel beschikbaar is.

5 Globale budgetten en concurrentie

In 2006 is er een stelsel van gereguleerde concurrentie ingevoerd. Het idee van concurrentie is dat het zowel de productieve als allocatieve efficiëntie in de zorg bevordert. Er zijn aanwijzingen dat de invoering van gereguleerde concurrentie de productieve efficiëntie heeft bevordert. Een vaak genoemd voorbeeld is het actieve zorginkoopbeleid van zorgverzekeraars op de geneesmiddelenmarkt waardoor er vaker goedkopere generieke middelen werden voorgeschreven (Douven en Meijer 2008). Ook een stijging van de arbeidsproductiviteit van 1,8% per jaar in de curatieve zorg gedurende 2001-2010 duidt op het eerste gezicht op efficiëntiewinst in de zorg (CBS 2012). Bijvoorbeeld, de arbeidsproductiviteit is toegenomen door de opkomst van ZBC's die eenvoudige operaties, zoals van knieën, heupen of staaroperaties, efficiënter wisten te organiseren. Zoals in de inleiding vermeld, lijkt het hier echter voornamelijk om productieve efficiëntie te gaan. Zoals het rapport van de Taskforce laat zien, duidt de hoge volumegroei in de zorg ook op overbehandeling: Het aantal behandelingen waarvan de baten mogelijk lager zijn dan de kosten is de afgelopen jaren sterk gegroeid. In Figuur 2 betekent dit dat er meer behandelingen in de richting van X_3 plaatsvinden.

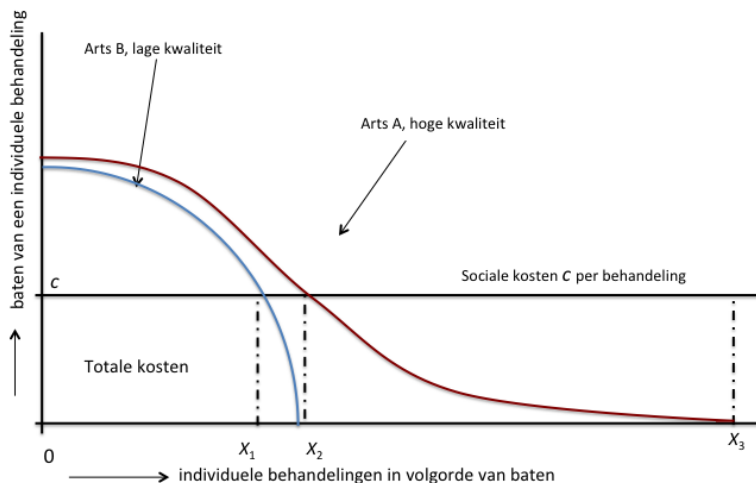
De vraag die we hier willen beantwoorden is in hoeverre concurrentie in de zorg de allocatieve efficiëntie bevordert. Om deze vraag te beantwoorden gaan we in Boone en Douven (2014) na in hoeverre concurrentie geschikt is om overbehandeling in de zorg terug te dringen. Om de effecten van concurrentie scherp te krijgen vergelijken we het gebruik van globale budgetten in twee extreme

situaties. Een situatie waarbij consumenten zelf zorgaanbieders mogen bezoeken met een situatie waarbij het zorgaanbod voor consumenten beperkt is. Bij dit laatste kan men denken aan postcodezorg. Bij postcodezorg mag men alleen zorgaanbieders in de eigen regio bezoeken. Voorbeelden hiervan zijn de ziekenhuiszorg in Portugal (Barros 2011) of de huisartszorg in Nederland, waarbij huisartsen veraf gelegen patiënten mogen weigeren.

Een belangrijk onderscheid dat we maken is dat er twee groepen consumenten zijn: een groep die bewust kiest voor kwaliteit, ook al moet de consument daar ver voor reizen, en een groep consumenten die indifferent is voor kwaliteit, en een voorkeur heeft voor de dichtstbijzijnde arts. Uit empirisch onderzoek blijkt immers dat veel consumenten kiezen voor het dichtstbijzijnde zorgaanbieder (Varkevisser et al. 2012). Dit kan enerzijds betekenen dat consumenten een bewuste afweging maken tussen kwaliteit en reisafstand, maar ook dat veel consumenten zich niet verdiepen in mogelijke kwaliteitsverschillen tussen artsen en erop vertrouwen dat de kwaliteit van de arts goed is.

De eerste observatie die we in Boone en Douven (2014) maken is dat wanneer aanbieders verschillen in kwaliteit er altijd een bepaalde vorm van inefficiëntie optreedt als consumenten zelf hun aanbieder mogen kiezen. Dit komt omdat preferenties van consumenten niet in lijn liggen met wat optimaal is vanuit het oogpunt van kosteneffectiviteit. Dit is eenvoudig in te zien met behulp van Figuur 2. Stel dat artsen zelf rekening houden met de kosteneffectiviteit van een behandeling, dus arts A voert alleen behandelingen uit tot X_2 , en arts B tot X_1 (dit is een andere nutsfunctie dan die we hierboven gebruikt hebben, daar werd het nut bepaald door effectiviteit). Op het eerste gezicht lijkt deze oplossing allocatief efficiënt. Maar een patiënt die op basis van reisafstand arts B prefereert boven A, maar die niet door B behandeld wordt, zal naar A toe reizen om te zien of zij hem wil behandelen. Als de overheid het aantal behandelingen contracteert en rekening houdt met de reisafstand van de consument, zal er arbitrage plaatsvinden tussen de artsen. Arts B zal dan behandelingen uitvoeren die niet kosteneffectief zijn en arts A zal dan niet alle kosteneffectieve behandelingen kunnen uitvoeren. Dus ook bij volledige informatie vinden er in het allocatief efficiënte optimum behandelingen plaats die niet kosteneffectief zijn!

Bovenstaande redenatie gaat uit van de ideale situatie dat artsen zelf rekening houden met de kosteneffectiviteit. Dit is echter niet het geval. Ook de overheid gebruikt het kosteneffectiviteit criterium nog nauwelijks voor de afbakening van het basispakket. Het College van Zorgverzekeringen (CVZ) beheert in Nederland het basispakket. Het CVZ heeft in het verleden wel eens behandelingen niet toegelaten tot het basispakket, maar dat was meer op basis van effectiviteit dan op basis van kosteneffectiviteit. Er is hier dus nog een wereld te winnen.

Figuur 4 Hogere kwaliteit hoeft niet altijd welvaartsverhogend te zijn

Protontherapie goed of slecht voor de allocatieve efficiëntie?

De meest patiënten die een kankertherapie ondergaan krijgen fotontherapie (radiotherapie) om de groei van kwaadaardige tumoren te stoppen. Op 29 juli jl. besloot de minister om ook protontherapie toe te laten als behandeling tegen kanker. Voor protontherapie zijn grotere investeringen nodig dan voor fotontherapie en de aanschaf van een protonenapparaat loopt al snel tegen 100 miljoen euro. Vier umc's in Nederland kregen een licentie om zo'n apparaat aan te schaffen. Een dergelijke investering is alleen dan zinvol wanneer er voldoende baten met protontherapie gegenereerd worden. In de VS is er grote discussie ontstaan over de (kosten)effectiviteit van protontherapie. Zo uitten gezondheidseconomen Baicker en Chandra (2011) bedenkingen tegen de invoering van protontherapie, omdat deze behandelingen niet kosteneffectief zijn. Ook de minister van VWS heeft deze bedenkingen, zo stelt ze in de Staatscourant van 1 augustus 2013 dat er onderzoek nodig is naar de "exacte vaststelling en/of verhoging van de meerwaarde van protontherapie gekoppeld aan studies naar kosteneffectiviteit...". Hier zien we een voorbeeld waarbij er kwaliteitsverschillen kunnen optreden. Artsen van universitair medische centra mogen behandelingen verrichten die artsen in algemene ziekenhuizen niet kunnen doen. Onze analyse laat zien dat wanneer blijkt dat protontherapie weliswaar iets effectiever is dan fotontherapie, maar minder kosteneffectief is dan fotontherapie, dan bestaat het risico dat veel patiënten voor een protonbehandeling zullen kiezen. Het gevolg is dat de kosten oplopen en de welvaart daalt. De overheid houdt weliswaar een restrictie aan op het aantal patiënten dat met protontherapie mag worden behandeld, maar een mogelijke aanschaf van de apparaten zal hoe dan ook terugverdiend moeten worden door de vier umc's. Dus de kosten lopen vooruit op de baten. Internationale concurrentie lijkt een belangrijke rol gespeeld te hebben bij de aanschaf van de apparaten, zo was protontherapie wel al beschikbaar in het buitenland maar nog niet in Nederland. Wanneer Nederland op wetenschappelijk gebied mee wil doen aan de nieuwste medische ontwikkelingen en wil concurreren met het buitenland, kan een dergelijke investering wenselijk zijn en op lange termijn de Nederlandse welvaart kunnen verhogen. Bij de toelating van protontherapie lijkt voorzichtigheid geboden. Of het de welvaart in Nederland verhoogt, en of we hier te maken hebben met een 'race to the top' of 'race to the bottom' zal de toekomst moeten uitwijzen

6 Patiëntenmobiliteit: een indicatie voor kwaliteit

Concurrentie kan de welvaart verhogen, omdat het kwaliteit meer zichtbaar maakt. Wanneer de overheid of verzekeraar onvoldoende informatie hebben over de kwaliteit van de aanbieder, kunnen ze wel proberen deze informatie op een indirecte manier te verkrijgen. Ze kunnen proberen uit keuzegedrag van patiënten informatie te halen. Bijvoorbeeld, wanneer patiënten kiezen op kwaliteit dan geeft dit informatie. Goedekwaliteitsaanbieders zullen dan meer patiënten trekken dan minder-goedekwaliteitsaanbieders. De betalende partij moet dus niet alleen rekening houden met het feit dat hogekwaliteitsaanbieders meer patiënten kunnen behandelen, ze zullen ook geografisch gezien meer mensen aantrekken. Een optimaal globaal budget houdt hier rekening mee.

Merk op dat er hier een belangrijk verschil is tussen de VS en Nederland. De eerder genoemde ACO- en AQC-organisaties uit de VS hebben een globaal budget voor een vooraf vastgestelde populatie. Mobiliteit van patiënten speelt dus alleen een rol binnen het netwerk van aanbieders. De kwaliteit van de zorgaanbieder wordt achteraf vastgesteld op basis van zestig vooraf vastgelegde kwaliteitsindicatoren (Song et al. 2012). Deze globale budgetten verstoren concurrentie tussen zorgaanbieders dus niet. In Nederland ligt dat anders. Zo houdt het huidige macrobeheersingsinstrument (MBI) in de ziekenhuissector geen rekening met kwaliteitsverschillen tussen aanbieders. Dit kan averechts werken. Bijvoorbeeld wanneer bij twee zorgaanbieders van verschillende kwaliteit de lagerekwaliteitszorgaanbieder te veel behandelt, waardoor het macrobudget wordt overschreden, wordt de hogerekwaliteitszorgaanbieder bij het toepassen van het MBI hiervoor gestraft (Boone en Douven 2012). Een ander probleem is dat er in Nederland in vergelijking met de VS nog weinig selectieve contractering plaatsvindt, waardoor het lastiger is om budgetten vast te stellen op een lager niveau dan op macroniveau. Bij selectieve contracten zijn er immers goed gedefinieerde groepen verzekerden te onderscheiden, waardoor het makkelijker is om een globaal budget te construeren. Het construeren van budgetten is ook makkelijker bij aanbieders waarvoor de populatie vastligt, zoals huisartsen, dan waarbij de patiëntenpopulatie kan fluctueren, zoals ziekenhuizen.

Bij ACO-organisaties in de VS wordt de populatie waarvoor ze verantwoordelijk zijn, bepaald door de overheid. Ze zijn verantwoordelijk voor de zorg van deze patiënten en daardoor kan ook het budget dat de ACO krijgt afgestemd worden op de patiëntenpopulatie. Opmerkelijk is echter dat patiënten ook zorgaanbieders buiten de ACO-organisatie mogen bezoeken en dat die kosten voor rekening van de ACO komen. Het idee hierachter is dat zorgaanbieders binnen een ACO dan sterker geprikkeld worden om goede zorg te leveren en om patiënten zoveel mogelijk binnen hun netwerk te houden.

7 Conclusie

In dit artikel geven we aan dat budgetten de allocatieve efficiëntie in de zorg kunnen verbeteren. Hiervoor zijn een aantal voorwaarden nodig. Een eerste voorwaarde is dat aanbieders voldoende inzicht hebben in de kosten en medische effectiviteit van de zorg. Aanbieders kunnen dan zelf een inschatting maken welke behandelingen wel en niet kosteneffectief zijn. Een andere voorwaarde is dat het invoeren van budgetten niet ten koste mag gaan van de productieve efficiëntie in de zorg. Dit noopt tot gerichte budgetten die de concurrentie in de zorg niet (te veel) verstoren. Dit zijn budgetten die afgestemd zijn op de patiëntenpopulatie en waarbij rekening wordt gehouden met kwaliteitsverschillen tussen aanbieders. Hoe budgetten er precies moeten uitzien en ingebed moeten worden in een stelsel van gereguleerde concurrentie, verdient nader onderzoek. Er zijn nog veel open vragen, maar het is wel duidelijk dat dit, ook internationaal gezien, een van de belangrijke toekomstige uitdagingen zal zijn op zorggebied.

Auteurs

Rudy Douven (e-mail: R.Douven@cpb.nl) is wetenschappelijk medewerker bij het Centraal Planbureau en geaffilieerd aan de Erasmus Universiteit Rotterdam. Rudy heeft aan het artikel gewerkt tijdens zijn Harkness fellowship aan Harvard Medical School, dat wordt gesponsord door het ministerie van VWS en het Commonwealth Fund uit New York. Jan Boone is professor aan de Universiteit aan Tilburg en verbonden aan het Centraal Planbureau. Jan bedankt NWO voor financiële ondersteuning (Vici beurs 453.07.00). De auteurs danken collega's van het ministerie van VWS, het CPB, en twee anonieme referenten voor commentaar op een eerdere versie van dit artikel.

Literatuur

- Baicker, K., K.S. Buckles en A. Chandra, 2006, Geographic variation in the appropriate use of cesarean delivery, *Health Affairs*, vol. 25(5): w335-w367.
- Baicker, K., en A. Chandra, 2011, Aspirin, Angioplasty, and Proton Beam Therapy: the economics of smarter health care, prepared for Jackson Hole Economic Policy Symposium.
- Barros, P.P., 2011, Portugal, Health System Review, European Observatory on Health Care Systems and Policies European Observatory on Health Care Systems and Policies, WHO.
- Boone, J., en R. Douven, 2012, Budgetten en marktwerking in de zorg: waarom en hoe?, in: E. Schut en M. Varkevisser (eds), *KVS preadviezen 2012*, Een economisch gezonde gezondheidszorg, pp. 45-66.
- Boone, J., en R. Douven, 2014, Provider competition and over-utilization in health care, Working Paper, forthcoming.

- CBS, 2012, *Gezondheid en zorg in cijfers 2012*, Den Haag: Centraal Bureau voor de Statistiek.
- Chandra, A., D. Cutler en Z. Song, 2011, Who ordered that? The economics of Treatment Choices in Medical Care, in: M.V. Pauly, T.G. McGuire en P.P. Barros (eds), *Handbook of Health Economics vol. II*, pp. 397-432, Amsterdam: Elsevier.
- Chandra A., A.B. Jena en J.S. Skinner, 2011, The pragmatist's guide to comparative effectiveness research, *Journal of Economic Perspectives*, vol. 25(2): 27-46.
- Chandra, A., en J. Skinner, 2012. Technology Growth and Expenditure Growth in Health Care, *Journal of Economic Literature, American Economic Association*, vol. 50(3): 645-80.
- Chandra, A., en D. Staiger, 2007, Productivity spillovers in healthcare: Evidence from the treatment of heart attacks, *Journal of Political Economy*, vol. 115(1): 103-40.
- Douven, R., 2013, Het grote grijze gebied in de zorg, in: C. van Ewijk, A. van der Horst en Paul Besseling (eds), *Toekomst voor de zorg*, Centraal Planbureau, Den Haag.
- Douven, R., en A. Meijer, 2008, Prijsvorming van generieke geneesmiddelen; forse prijsdaling in het nieuwe zorgstelsel, CPB Document 175, Centraal Planbureau, Den Haag.
- Douven, R., R. Mocking en I. Mosca, 2012, The effect of physician fees and density differences on regional variation in hospital treatments, CPB Discussion Paper 208, CPB, Den Haag.
- Laffont, J., en J. Tirole, 1993, *A theory of incentives in procurement and regulation*, MIT press.
- Ministerie van VWS, 2012, Naar beter betaalbare zorg, Rapport Taskforce Beheersing Zorguitgaven, juni 2012.
- Maco-Stadler, I., en J.D. Perez-Castrillo, 2001, *An Introduction to the Economics of Information: Incentives and Contracts*, Oxford University Press.
- Song, Z., D.G. Safran, B.E. Landon, M.B. Landrum, Y. He, R.E. Mechanic, M. Day en M.E. Chernew, 2012, The 'Alternative Quality Contract,' Based on a Global Budget, Lowered Medical Spending And Improved Quality, *Health Affairs*, vol. 31(8):1885-94.
- Varkevisser, M., S. A. van der Geest en E. Schut, 2012, Do patients choose hospitals with high quality ratings? Empirical evidence from the market for angioplasty in the Netherlands, *Journal of Health Economics*, vol. 31(2): 371-78.
- Vijsel, A.R. van de, P.M. Engelfriet en G.P. Westert, 2011, Rendering hospital budgets volume based and open ended to reduce waiting lists: does it work?, *Health Policy*, vol. 100(1): 60-70.