



➤ De kunst van het vergelijken

Handvatten voor leren en verbeteren en
zorginkoopdialoog met de integrale vergelijking

Utrecht, 23-11-2021

SiRM. Strategies
in Regulated
Markets

Colofon

Titel van het rapport

De kunst van het vergelijken - Handvatten voor leren en verbeteren en zorginkoopdialoog met de integrale vergelijking

Auteurs

Jan-Peter Heida, Roderik Ponds en Tammo Zijlker

Contactpersoon: Roderik Ponds (roderik.ponds@sirm.nl)

Dit onderzoek is uitgevoerd in opdracht van de Nederlandse Zorgautoriteit.

Inhoud

Managementsamenvatting	1
1 Aanleiding, doel en aanpak	3
1.1 De NZa werkt in 2021 aan de doorontwikkeling van de integrale vergelijking op vier onderwerpen	3
1.2 SiRM onderzocht handvatten voor leren en verbeteren, zorginkoopdialoog en verminderen complexiteit	4
1.3 Het onderzoek is uitgevoerd in vier fases	4
2 De integrale vergelijking is gebaseerd op DEA, een techniek om relatieve efficiëntie ten opzichte van peers te meten	5
3 Zorgaanbieders zien mogelijkheden voor leren en verbeteren met IV onder drie voorwaarden	9
3.1 Groot deel van zorgaanbieders is neutraal tot zeer positief over leren en verbeteren met datagedreven vergelijkingen	9
3.2 Zorgaanbieders zien drie voorwaarden om te kunnen leren en verbeteren met de integrale vergelijking	11
4 Handvatten voor de zorginkoopdialoog hangen samen met rol bij tariefstelling en inhoudelijke doorontwikkeling	17
4.1 Formuleren van handvatten door partijen voor zorginkoopdialoog is lastig zolang het onduidelijk is hoe de relatie met tariefstelling is	17
4.2 Zorgaanbieders zijn verdeeld over de toepasbaarheid van de integrale vergelijking in de zorginkoopdialoog	18
4.3 Zorgkantoren wensen dezelfde inhoudelijke doorontwikkeling als aanbieders en hebben aantal voorwaarden voor toepassen iv bij zorginkoop	19
4.4 Bij de inzet van iv bij zorginkoop hechten partijen nog sterker aan inhoudelijke doorontwikkeling dan bij leren en verbeteren	20
5 Om concrete handvatten te bieden moeten efficiëntieverschillen worden verklaard	22
5.1 Uitgangspunt in dit hoofdstuk is dat DEA de verschillen in efficiëntie correct meet	23
5.2 Een tweede analysestap in de DEA-methode kan factoren die efficiëntieverschillen verklaren identificeren.	24
5.3 Goede visualisaties en heldere toelichting op de resultaten bieden concrete handvatten voor leren en verbeteren	29
5.4 Drie opties om het eerste contact tussen aanbieders voor leren en verbeteren te faciliteren	32

6	Aanpassing van de data en het model is nodig om de doorontwikkeling te realiseren	34
6.1	Door de gebruikte data te verbeteren kunnen efficiëntieverschillen mogelijk beter gemeten worden	35
6.2	Methodologische aanpassingen aan het model kunnen draagvlak voor toepassing en herkenbaarheid vergroten	38
6.3	De voorgestelde aanpassingen verhogen administratieve en uitvoeringslasten en vergen forse inspanningen van partijen	39
	Bijlage 1: Voorbeeld visualisaties	41
	Stap 1: Efficiëntiescore van de aanbieder en de vergelijkingsgroep	41
	Stap 2: Een overzicht van de verwachte impact van verklarende factoren voor de efficiëntiescore	42
	Stap 3: Score van de verklarende factoren in vergelijking met andere aanbieders	45

Managementsamenvatting

De NZa werkt sinds 2019 aan de ontwikkeling van de integrale vergelijking verpleeghuiszorg: een datagedreven vergelijking van zorgaanbieders die mogelijk onderdeel wordt van een nieuwe vorm van bekostiging. De gebruikte methode is Data Envelopment Analysis (DEA). DEA is een techniek om de relatieve efficiëntie van een aanbieder te meten op basis van een vergelijking met de meest efficiënte aanbieders die een vergelijkbare zorgproductie leveren (de peers). Na de in 2019 en 2020 uitgevoerde nulmeting heeft de NZa geadviseerd de integrale vergelijking op verschillende onderwerpen door te ontwikkelen met DEA als methode.

Onderzoek: handvatten voor leren en verbeteren, de zorginkoopdialoog en mogelijkheden voor verminderen ervaren complexiteit.

SiRM heeft in opdracht van de NZa tussen maart en november 2021 onderzoek gedaan naar één van de onderwerpen van de doorontwikkeling: het bieden van handvatten voor leren en verbeteren, voor de zorginkoopdialoog en voor mogelijkheden om de ervaren complexiteit te verminderen. Hiervoor zijn de wensen en voorwaarden van zorgaanbieders en zorgkantoren via interviews, een enquête en bijeenkomsten geïnventariseerd. Vervolgens is gekeken hoe deze wensen en randvoorwaarden gerealiseerd kunnen worden. De NZa stelde de gekozen methode (DEA) als uitgangspunt voor dit onderzoek. De vraag hoe de integrale vergelijking wordt toegepast voor regulering en bekostiging valt buiten het onderzoek.

Resultaten: inhoudelijk doorontwikkeling van de integrale vergelijking is op vier onderwerpen noodzakelijk voor leren en verbeteren.

Zorgaanbieders zien mogelijkheden voor leren en verbeteren met de integrale vergelijking mits aan drie voorwaarden wordt voldaan: inhoudelijke doorontwikkeling van de methode, faciliteren van het eerste contact voor leren en verbeteren, en duidelijkheid over hoe de integrale vergelijking zal worden gebruikt in de bekostiging. De gewenste inhoudelijke doorontwikkeling bestaat uit vier delen: inzicht in wie de peers zijn en waarom, de mogelijkheid om ook met andere aanbieders dan peers te vergelijken, spiegelinformatie over factoren die de berekende relatieve efficiëntiescore verklaren en een terugkoppeling die de ervaren complexiteit vermindert.

Aanbieders en zorgkantoren zien dezelfde inhoudelijke doorontwikkeling als randvoorwaarde voor het bieden van handvatten voor de zorginkoopdialoog. Het bleek voor partijen lastig om aanvullende wensen te formuleren zonder duidelijkheid over de rol van de integrale vergelijking bij tariefstelling. Hoe tarieven tot stand komen en de daarbij behorende onderhandelingsruimte vormen een belangrijk en integraal onderdeel van de zorginkoopdialoog. Gegeven de onzekerheid op dat gebied zijn zorgaanbieders verdeeld over de toepasbaarheid van de integrale vergelijking in de zorginkoopdialoog. Iets meer dan de helft van de geraadpleegde aanbieders (55%) beoordeelt dit neutraal tot positief, de rest als (zeer) beperkt.

De nadruk in dit onderzoek lag op de vraag onder welke omstandigheden en op welke manier de gewenste inhoudelijke doorontwikkeling voor leren en verbeteren, zorginkoopdialoog en reduceren ervaren complexiteit gerealiseerd kan worden.

Advies: om concrete handvatten te bieden moeten efficiëntieverschillen worden verklaard.

De integrale vergelijking kan handvatten bieden voor leren en verbeteren en de zorginkoopdialoog als deze niet alleen efficiëntieverschillen goed en correct meet maar deze ook zo goed mogelijk verklaart. Dat kan door een tweede analysestap toe te voegen aan de DEA-methode om factoren die verschillen verklaren te identificeren: second stage DEA. Hiervoor is per aanbieder informatie over potentieel verklarende factoren nodig. Door statistische analyses wordt vervolgens de invloed van deze factoren vastgesteld. Goede visualisaties van de resultaten hiervan via een online dashboard bieden concrete inzichten aan aanbieders en zorgkantoren over hoe efficiënt een aanbieder is, op welke factoren een aanbieder zich kan verbeteren en van welke aanbieder(s) geleerd kan worden. Deze inzichten kunnen het startpunt vormen voor leren en verbeteren tussen aanbieders onderling en voor de zorginkoopdialoog tussen zorgkantoren en aanbieders.

Overwegingen voor vervolg: realisatie vergt forse inspanning waarbij het op voorhand onduidelijk is of gewenste resultaat bereikt wordt.

Uitgangspunt bij second stage DEA is dat de efficiëntieverschillen tussen aanbieders goed worden gemeten. Op basis van de ervaringen met de nulmeting stelden veldpartijen dit tijdens dit onderzoek regelmatig ter discussie. Hoewel SiRM voor dit onderzoek geen eigen analyses heeft uitgevoerd, delen wij op basis van de beschikbare informatie over de nulmeting deze twijfels. Ook de NZa geeft in haar adviesrapport aan dat de data en het model uit de nulmeting verbeterd en uitgebreid moet worden. Door de data te verbeteren en uit te breiden kan in de toekomst mogelijk een betere meting van efficiëntieverschillen plaatsvinden. Het gaat dan om: betere kostendata met minder aannames, het meenemen van kapitaallasten, betere indeling van en weging binnen productgroepen en het duidelijke en goede definities van kwaliteitsverschillen als uitsluitingscriterium voor peers.

Het realiseren van second stage DEA en de genoemde verbeteringen vergt een forse inspanning van alle betrokken partijen waarbij het op voorhand onduidelijk is of het gewenste resultaat bereikt wordt. Of de voorgestelde verbeteringen en het uitvoeren van second stage DEA in de praktijk mogelijk zijn en ook de gewenste resultaten leveren kan pas worden vastgesteld met gegevens uit een eventuele vervolgmeting.

I Aanleiding, doel en aanpak

De NZa werkt in 2021 aan de doorontwikkeling van de integrale vergelijking verpleeghuiszorg: een datagedreven vergelijking van zorgaanbieders die mogelijk onderdeel wordt van een nieuwe vorm van bekostiging. Het door SiRM uitgevoerde onderzoek heeft betrekking op één van de vier onderwerpen van de doorontwikkeling: het bieden van handvatten voor leren en verbeteren, voor de zorginkoopdialoog en naar mogelijkheden om de ervaren complexiteit te verminderen. Tussen maart en november 2021 heeft SiRM het onderzoek uitgevoerd aan de hand van deskresearch, interviews, een enquête en werksessies.

1.1 De NZa werkt in 2021 aan de doorontwikkeling van de integrale vergelijking op vier onderwerpen

Sinds 2019 werkt de NZa aan de ontwikkeling van de integrale vergelijking verpleeghuiszorg: een datagedreven vergelijking van zorgaanbieders die mogelijk de basis wordt van een nieuwe vorm van bekostiging. Het achterliggende doel van de integrale vergelijking is drieledig:

- **Kwaliteit:** het leveren van zorg conform het kwaliteitskader bereikbaar maken voor alle zorgaanbieders.
- **Doelmatigheid:** een bekostiging die de sector als geheel ondersteunt en stimuleert om doelmatig te werken.
- **Leren en verbeteren:** inzicht in verschillen tussen zorgaanbieders om van elkaar te leren en te verbeteren.

De gekozen methode voor de integrale vergelijking is de Data Envelopment Analysis (DEA) waarmee in 2019 en 2020 een nulmeting is uitgevoerd. In hoofdstuk 2 worden deze methode en de uitgevoerde nulmeting kort toegelicht.

Op basis van de resultaten van de nulmeting, de duidingsperiode met zorgaanbieders en consultatie van de betrokken stakeholders ontwikkelt NZa in 2021 de integrale vergelijking door. Deze doorontwikkeling vindt plaats op vier onderwerpen:

- Het uitvoeren van aanvullende analyses (een regioanalyse en een impactanalyse).
- Het onderzoeken van verbeteringen op het gebied van datakwaliteit, administratieve lasten en de huidige modelkeuzes.
- Het beleidsmatig ondersteunen van partijen in het uitwerken van het vraagstuk kwaliteit en transparantie.

- Het onderzoeken van handvatten voor leren en verbeteren, de zorginkoopdialoog en het verminderen van de ervaren complexiteit.

Dit onderzoek gaat over de handvatten voor leren en verbeteren, zorginkoopdialoog en het verminderen van de ervaren complexiteit.

Op basis van de resultaten van de doorontwikkeling en de nadere uitwerking van de doorwerking van de integrale vergelijking op bekostiging, beslist de minister van VWS over het vervolg.

1.2 SiRM onderzocht handvatten voor leren en verbeteren, zorginkoopdialoog en verminderen complexiteit

Uit de (duiding van de) nulmeting werd geconcludeerd dat er onder andere nader onderzoek nodig was naar hoe de integrale vergelijking in de praktijk kan worden toegepast voor leren en verbeteren en voor de zorginkoopdialoog.¹ De integrale vergelijking wordt als complex ervaren door gebruikers en de veldpartijen en de resultaten blijken lastig te interpreteren. Dit belemmert toepassing in de praktijk.

Het onderzoek dat SiRM heeft uitgevoerd heeft drie doelen:

- 1 Opstellen van handvatten voor leren en verbeteren voor zorgaanbieders op basis van de uitkomsten van de integrale vergelijking.
- 2 Opstellen van handvatten voor de zorginkoopdialoog op basis van de uitkomsten van de integrale vergelijking voor zorgkantoren en zorgaanbieders.
- 3 Formuleren van mogelijkheden om de ervaren complexiteit van de IV te verminderen.

De NZa stelde de gekozen methode (DEA) als uitgangspunt voor dit onderzoek. De vraag hoe de integrale vergelijking wordt toegepast voor regulering en bekostiging valt buiten het onderzoek.

1.3 Het onderzoek is uitgevoerd in vier fases

Het onderzoek is in vier fases uitgevoerd tussen maart en november 2021. In de eerste fase zijn de wensen en ervaringen van aanbieders en zorgkantoren geïnventariseerd aan de hand van interviews, een enquête onder zorgaanbieders en werksessies. Aan het einde van fase 1 is er een brede themabijeenkomst georganiseerd waar de resultaten zijn gedeeld met de sector en aanvullende input is opgehaald. Bij deze bijeenkomst waren bijna 100 deelnemers bij betrokken. In de tweede fase zijn de mogelijkheden geïnventariseerd om deze wensen te realiseren. In de derde fase is een aantal opties in de vorm van voorbeeldwerkwijzen voor visualisaties in beeld gebracht. De resultaten zijn besproken in een werksessie. In de vierde fase is het eindrapport opgeleverd. Tijdens het onderzoek zijn er twee bijeenkomsten met een klankbordgroep georganiseerd om de resultaten te bespreken.

¹ Zie: NZa 2020: Advies Integrale vergelijking verpleeghuiszorg

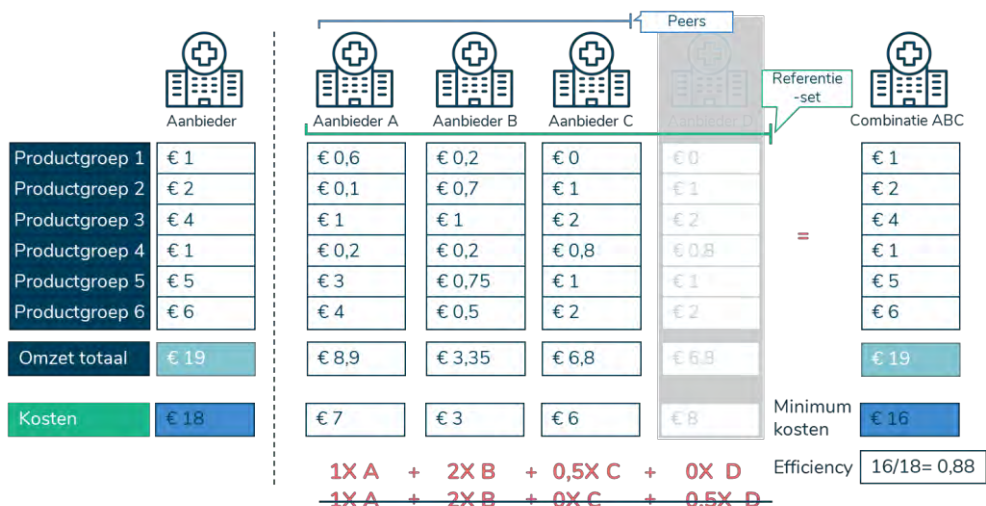
2 De integrale vergelijking is gebaseerd op DEA, een techniek om relatieve efficiëntie ten opzichte van peers te meten

De gebruikte methode in de integrale vergelijking is DEA. DEA is een techniek waarmee efficiëntieverschillen tussen aanbieders worden gemeten door modelmatig gegevens over de geleverde zorgproducten (de output) en de daarbij behorende kosten voor personeel en materieel (de inputs) te vergelijken. Per aanbieder wordt een efficiëntiescore berekend. Deze score drukt het verschil uit tussen de feitelijke kosten van een aanbieder en het minimale kostenniveau voor dezelfde zorgproductie (de kostengrens) dat op basis van een vergelijking met een combinatie van de meest efficiënte aanbieders (peers) mogelijk zou zijn.

Met de integrale vergelijking probeert de NZa kostenverschillen tussen verschillende aanbieders in kaart te brengen door aanbieders met elkaar te vergelijken. De methode die hiervoor gebruikt wordt, is Data Envelopment Analysis (DEA). DEA schat op basis van gegevens over de kosten en de productie van alle aanbieders een kostengrens (of productiegrens): de minimale kosten waarvoor een bepaalde zorgproductie geleverd zou moeten kunnen worden. Vervolgens wordt voor elke aanbieder een efficiëntiescore berekend: het verschil tussen de feitelijke kosten en de berekende minimale kosten op basis van de combinatie van meest efficiënte aanbieders die gezamenlijk een vergelijkbare productie hebben (de peers).

Een volledige beschrijving van de methode en het proces van dataverzameling in de nulmeting is te vinden in het advies integrale vergelijking verpleeghuiszorg². Basisbegrip van de berekening van de efficiëntiescore helpt om de inhoud van dit rapport goed te interpreteren. Daarom biedt dit hoofdstuk een beknopte toelichting van de DEA-methode zoals deze is toegepast door de NZa in de nulmeting. Iedere stap wordt toegelicht en geïllustreerd met een simpel rekenvoorbeeld waarin één aanbieder met vier andere aanbieders wordt vergeleken (Figuur 2.1). De andere hoofdstukken gaan in op de wensen en mogelijkheden om handvatten voor leren en verbeteren en de zorginkoopdialoog te bieden op basis van DEA en opties om de ervaren complexiteit te verminderen.

² NZa 2020: Advies Integrale vergelijking verpleeghuiszorg



Figuur 2.1 Illustratie van de DEA-methode die in de nulmeting van de integrale vergelijking is gehanteerd.

Voor het toepassen van de DEA-methode zoals in de nulmeting is gehanteerd zijn drie voorbereidende stappen noodzakelijk:

1. Per aanbieder wordt de productie in zes productgroepen bepaald:

Toelichting: De productie is de (gestandaardiseerde) omzet in zes productgroepen.

Tabel 2.1 toont een overzicht van deze zes productgroepen.

- Vier productgroepen zijn gerelateerd aan het kwaliteitskader verpleeghuiszorg. Voor deze vier productgroepen is de productie gelijk aan het aantal ZZP/VPT's vermenigvuldigd met de maximumprijs voor dit ZZP/VPT van de NZa.
- De overige twee productgroepen vallen niet onder het kwaliteitskader verpleeghuiszorg en zijn onderverdeeld in prestaties met en zonder verblijf. De productie is gedefinieerd als de omzet. Hier zijn geen standaard maximumprijzen voor beschikbaar.

Voorbeeld: Voor de aanbieder die wordt vergeleken met de andere aanbieders is de productie (de omzet) in de zes productgroepen weergegeven in Figuur 2.1: in totaal €19 miljoen.

Tabel 2.1 Overzicht van de categorieën van productie zoals deze tijdens de nulmeting zijn gedefinieerd.

Productgroep	Omschrijving
1	Gestandaardiseerde omzet voor ZZP / VPT 4
2	Gestandaardiseerde omzet voor ZZP / VPT 5-6 incl. behandeling
3	Gestandaardiseerde omzet voor ZZP / VPT 5-6 excl. behandeling
4	Gestandaardiseerde omzet voor ZZP / VPT 7-10
5	Omzet voor prestaties niet-kwaliteitskader met verblijf
6	Omzet voor prestaties niet-kwaliteitskader zonder verblijf

2. Per aanbieder wordt een onderliggend kostenniveau bepaald:

Toelichting: Voor de aanbieder worden de kosten die horen bij de productie bepaald. Deze kosten bestaan uit personeelslasten en materieel. Kapitaallasten zoals de kosten voor huisvesting laat de NZa buiten beschouwing. De aanbieders hadden ten tijde van de nulmeting (dataverzameling 2018) veelal nog te weinig personeel ten opzichte van het Kwaliteitskader Verpleeghuiszorg omdat zij nog in de ingroefase waren. Daarom zijn in de nulmeting personeelskosten bijgeschat. Gemiddeld zijn de personeelslasten in de bijschatting ongeveer 20% opgehoogd.

Voorbeeld: Voor de aanbieder die wordt vergeleken met de andere aanbieders zijn de kosten €18 miljoen (Figuur 2.1).

3. Er wordt een selectie gemaakt van aanbieders die model kunnen staan voor andere aanbieders:

Toelichting: Voordat aanbieders met elkaar vergeleken worden, wordt een selectie gemaakt van aanbieders die model zouden kunnen staan voor andere aanbieders. Deze groep aanbieders wordt de referentieset genoemd. Aanbieders zijn in de nulmeting onder twee voorwaarden opgenomen in de referentieset:

- De aangeleverde productie- en kostendata zijn van voldoende kwaliteit.
- Aanbieders maken kosten die representatief zijn om kwaliteit volgens het Kwaliteitskader Verpleeghuiszorg te kunnen leveren. Door het gebruik van de bijschatting voldeden alle zorgaanbieders tijdens de nulmeting aan deze tweede voorwaarde.

Voorbeeld: In het voorbeeld van Figuur 2.1 zijn alle aanbieders (A tot en met D) meegenomen in de referentieset. Deze kunnen onderling vergeleken worden.

De drie voorbereidende stappen maken het mogelijk de DEA-methode toe te passen. Per aanbieder wordt DEA daarna in vier stappen gedaan.

4. De aanbieder wordt (geautomatiseerd) vergeleken met alle mogelijke combinaties van andere zorgaanbieders die samen minimaal dezelfde productie leveren:

Toelichting: Volgens de DEA-methode wordt de aanbieder vergeleken met combinaties van andere zorgaanbieders. Met combinatie van andere aanbieders wordt het volgende bedoeld:

- Per aanbieder waarmee wordt vergeleken wordt één getal gekozen waarmee de productie in elke productgroep en de kosten worden vermenigvuldigd. Hiermee wordt de aanbieder in feite opgeschaald of juist afgeschaald met de aanname van gelijkblijvende verhouding tussen kosten en productie.
- De productie per productgroep en de kosten van de (geschaalde aanbieders) worden opgeteld tot één theoretische gecombineerde aanbieder. Dit is de combinatie die bedoeld wordt.

Van elke denkbare combinatie wordt alleen die combinatie van aanbieders geselecteerd die minimaal dezelfde productie in elke productgroep hebben als de aanbieder waarmee wordt vergeleken.

Voorbeeld: In het voorbeeld van Figuur 2.1 is een mogelijke combinatie getoond: 1x aanbieder A + 2 x aanbieder B + 0,5 x aanbieder C + 0 x aanbieder D. Wanneer de productie maal de coëfficiënt in elke productgroep wordt opgeteld is deze gelijk aan de productie van de aanbieder waarmee wordt vergeleken. Er zijn ook andere combinaties mogelijk die dezelfde productie leveren zoals bijvoorbeeld 1x aanbieder A + 2 x aanbieder B + 0 x aanbieder C + 0,5 x aanbieder D.

5. De combinatie van aanbieders die de productie tegen minimale kosten levert wordt geselecteerd als resultaat van het model. Hieruit volgen de peers.

Toelichting: Volgens de DEA methode wordt van alle denkbare combinaties die in stap vier zijn overwogen de combinatie gekozen die de minste kosten heeft (dit wordt de productie- of kostengrens genoemd). De aanbieders die in deze combinatie voorkomen zijn de peers voor de aanbieder die met andere aanbieders wordt vergeleken.

Voorbeeld: In het voorbeeld in Figuur 2.1 blijkt de combinatie van 1 x aanbieder A, 2 x aanbieder B en 0,5 x aanbieder C de combinatie die de productie tegen minimale kosten levert. Aanbieders A, B en C zijn daarom de peers. Aanbieder D levert dezelfde productie als aanbieder C, maar tegen hogere kosten. Deze wordt daarom niet in de combinatie meegenomen en de combinatie A,B en D is daarom ook doorgestreept.

6. Voor de aanbieder wordt een minimumkostenniveau bepaald:

Toelichting: De kosten die horen bij de combinatie uit de vorige stap vormen een ondergrens voor de kosten van de aanbieder die wordt onderzocht. Dit kostenniveau wordt in de nulmeting het minimumkostenniveau genoemd.

Voorbeeld: In Figuur 2.1 is het minimumkostenniveau gelijk aan €16 miljoen. Dit volgt uit de combinatie $1 \times €7 + 2 \times €3 + 0,5 \times €6$.

7. Voor de aanbieder wordt een efficiëntiescore bepaald:

Toelichting: De efficiëntiescore in de DEA-methode en de nulmeting is de verhouding tussen het minimumkostenniveau en de kosten van de aanbieder die vergeleken wordt met de andere. Een score van 0,9 wil dus zeggen: er is een combinatie van andere aanbieders die minimaal dezelfde productie kan leveren tegen 90% van de kosten. Op deze manier is de efficiëntiescore van DEA dus altijd lager dan 1.

Voorbeeld: In Figuur 2.1 is de efficiëntiescore gelijk aan 0,88. Dit getal volgt door de minimumkostenniveau van de aanbieder te delen door de werkelijke kosten van deze aanbieder (16/18)

3 Zorgaanbieders zien mogelijkheden voor leren en verbeteren met IV onder drie voorwaarden

Veel zorgaanbieders zien de potentie van datagedreven vergelijkingen en maken hier nu ook gebruik van om te leren en verbeteren. De integrale vergelijking kan in de ogen van aanbieders (ook) handvatten bieden voor leren en verbeteren van collega zorgaanbieders mits aan drie voorwaarden wordt voldaan: inhoudelijke doorontwikkeling van de methode, een partij die de dialoog voor het proces van leren en verbeteren op gang brengt, en duidelijkheid over of en hoe de integrale vergelijking zal worden gebruikt in de bekostiging.

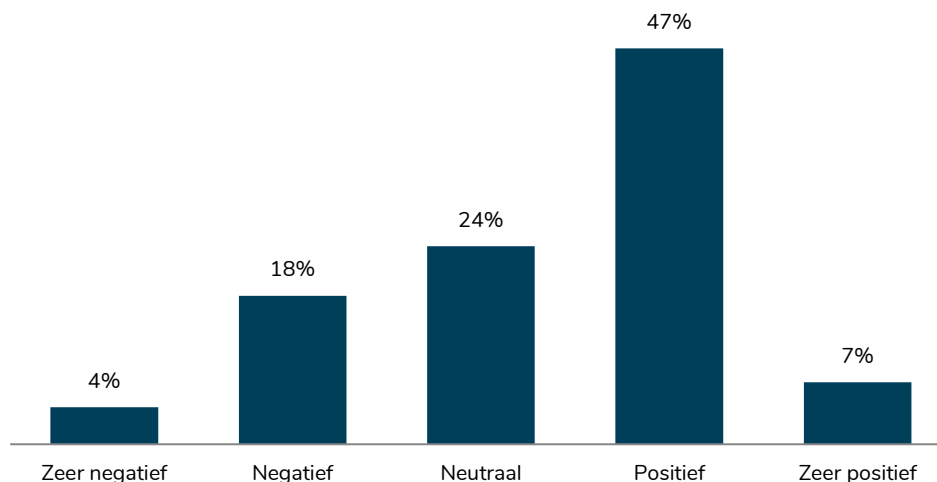
De gewenste inhoudelijke doorontwikkeling bestaat uit vier onderdelen: inzicht in wie de peers zijn en waarom, de mogelijkheid om ook met andere aanbieders dan peers te vergelijken, spiegelinformatie over factoren die de berekende relatieve efficiëntiescore verklaren en een terugkoppeling die de ervaren complexiteit vermindert.

3.1 Groot deel van zorgaanbieders is neutraal tot zeer positief over leren en verbeteren met datagedreven vergelijkingen

Zorgaanbieders zien mogelijkheden voor leren en verbeteren met behulp van onderlinge vergelijkingen op basis van data. Meer dan 75% van de ondervraagde zorgaanbieders in de verpleeghuissector is neutraal tot zeer positief over de potentie hiervan. Uit de enquête³ blijkt dat meer dan de helft van de zorgaanbieders (zeer) positief tegenover leren en verbeteren met behulp van datagedreven vergelijkingen staat (zie Figuur 3.1). Een kwart van de aanbieders staat hier neutraal tegenover.

³ De enquête is uitgevoerd in mei 2021. We hebben reacties van 69 zorgaanbieders ontvangen. De respondenten waren in de meeste gevallen bestuurders, controllers of zorgverkopers.

Hoe kijkt u aan tegen de potentie van leren en verbeteren op het gebied van bedrijfsvoering aan de hand van datagedreven vergelijkingen?



Figuur 3.1: Meer dan 75% van de zorgaanbieders staat neutraal tot positief tegenover de potentie van leren en verbeteren met behulp van datagedreven vergelijkingen.

Ongeveer 2/3 van de zorgaanbieders geeft aan op dit moment met datagedreven benchmarks te werken. De ActiZ benchmark (Kijk op data), de Berenschot benchmark, vergelijkingen van het zorgkantoor en iPVB zijn hierbij genoemde voorbeelden. De integrale vergelijking zou voor leren en verbeteren meerwaarde moeten bieden ten opzichte van deze bestaande benchmarks.

Zorgaanbieders geven aan dat (het resultaat van) een datagedreven vergelijking in hun ogen een startpunt voor leren en verbeteren is, maar op zichzelf niet voldoende is om te kunnen leren en verbeteren. Een datagedreven vergelijking kan aangeven op welke onderwerpen een aanbieder zich mogelijk kan verbeteren en van welke andere aanbieder(s) daarbij geleerd kan worden. Dit moet leiden tot een dialoog tussen aanbieders over die onderwerpen waarover verschillen zijn gevonden. Doel van deze dialoog is om de context achter de cijfers te begrijpen: inzicht in de gemaakte keuzes van de aanbieder is nodig om te leren (wat doet een andere aanbieder anders?) en vervolgens zelf te kunnen verbeteren (wat is er eventueel 'over te nemen' van de andere aanbieder?).

Zorgaanbieders (en ook zorgkantoren: zie hoofdstuk 4) willen leren en verbeteren over vraagstukken op de (relatief) korte termijn zoals het verbeteren van de efficiëntie bij bedrijfsvoering en over lange termijn vraagstukken in de ouderenzorg. Bij dit laatste gaat het vooral over de fundamentele vraag hoe er als maatschappij wordt omgegaan met de gevolgen van de dubbele vergrijzing voor de ouderenzorg. Mede op basis van de op 29 juni 2021 gehouden themabijeenkomst is het beeld dat in de ogen van aanbieders de rol voor datagedreven vergelijkingen bij dit type vragen wel kleiner is dan bij meer kortere termijn vraagstukken. Voor oplossingsrichtingen voor lange termijn vraagstukken ligt een meer kwalitatieve vergelijking en dialoog voor de hand.

3.2 Zorgaanbieders zien drie voorwaarden om te kunnen leren en verbeteren met de integrale vergelijking

Om te kunnen leren en verbeteren met behulp van de integrale vergelijking zien zorgaanbieders op basis van de nulmeting en de stand van zaken in mei en juni 2021 drie voorwaarden:

1. Inhoudelijke doorontwikkeling van de methode.
2. Het schetsen van het proces van leren en verbeteren.
3. Duidelijkheid over het gebruik van de integrale vergelijking voor bekostiging.

3.2.1 Inhoudelijke doorontwikkeling is noodzakelijk om concrete handvatten voor leren en verbeteren te bieden

Zorgaanbieders zien vier concrete onderwerpen voor inhoudelijke doorontwikkeling van de integrale vergelijking die noodzakelijk zijn voor het bieden van concrete handvatten voor leren en verbeteren.

Aanbieders willen weten wie de peers zijn en beter snappen hoe ze tot stand komen

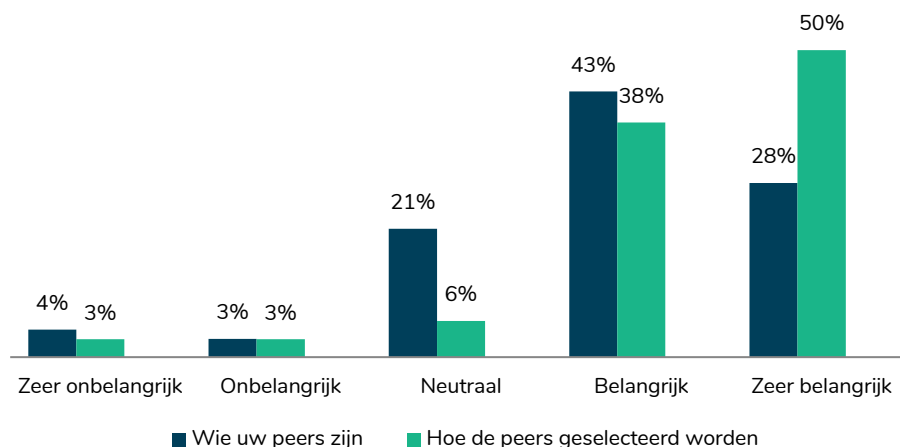
Om als zorgaanbieder te kunnen leren van peers is het belangrijk om te weten welke aanbieders dat zijn. De gebruikte methode in de integrale vergelijking (zie hoofdstuk 2) selecteert per aanbieder modelmatig peers: een combinatie van zorgaanbieders die gezamenlijk dezelfde zorgproductie tegen de minste kosten leveren. Zorgaanbieders kunnen leren van elkaar als inzichtelijk wordt wat deze andere aanbieders anders of beter doen en hoe dat kan worden toegepast in de eigen organisatie. Het voeren van een dialoog met deze peers is volgens veel aanbieders hierbij een belangrijke stap. Om dat te kunnen doen, moet bekend zijn wie de peers zijn. Ruim 70% van de aanbieders wil mede daarom weten wie de peers zijn (Figuur 3.2). Hiernaast geven vooral grote aanbieders aan dat vergelijken op het niveau van individuele verpleeghuizen of locaties (naast organisaties) grote meerwaarde biedt. Dit omdat er tussen locaties of huizen binnen een organisatie ook grote verschillen kunnen bestaan.

Inzicht in wie de peers zijn en hoe deze geselecteerd worden, is ook van belang voor de acceptatie en interpretatie van de uitkomsten. Bijna 90% van de aanbieders wil weten hoe de peers zijn geselecteerd (zie Figuur 3.2). Door te weten met wie een aanbieder wordt vergeleken en waarom kan een eigen oordeel van de vergelijkbaarheid van deze peers worden gemaakt.

Aanbieders hebben vaak een ander beeld van wat een vergelijkbare aanbieder is dan het begrip 'peers' in de gebruikte DEA-methode. Voor een aanbieder zijn andere aanbieders vergelijkbaar als ze een vergelijkbare omvang of zorgmix hebben en/of in dezelfde regio actief zijn. Het gaat daarbij om kenmerken van **individuele** aanbieders. Peers in het DEA-model zijn aanbieders die **gezamenlijk** een vergelijkbare zorgproductie hebben als de aanbieder in kwestie en daar gezamenlijk minder kosten voor maken. De individuele aanbieders die modelmatig als peer zijn aangewezen kunnen veel groter of kleiner zijn of een andere zorgmix hebben. Individuele peers worden daarom mogelijk niet als vergelijkbaar gezien door een aanbieder terwijl ze dat vanuit

DEA-perspectief wel zijn. Het expliciet maken van dit gebruik van het begrip vergelijkbaarheid en peers binnen DEA in de toelichting helpt om de (ervaren) complexiteit te verminderen.

Hoe belangrijk vindt u het van uw peers te weten:



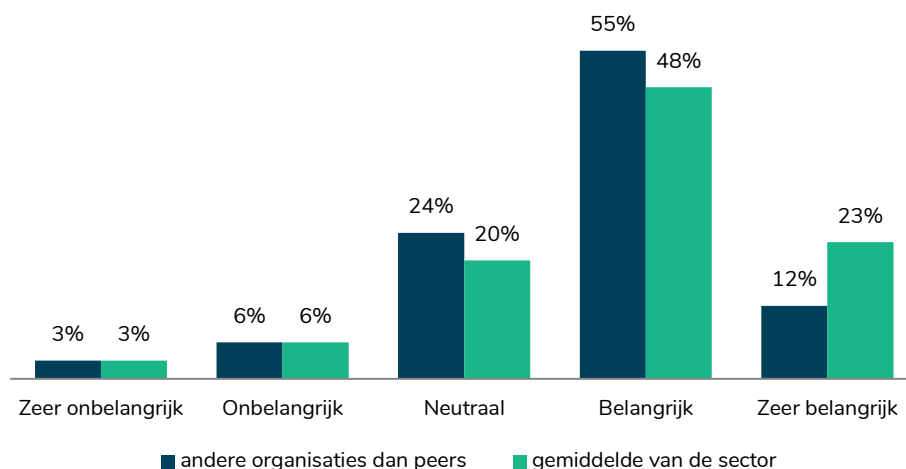
Figuur 3.2: Meer dan 70% van de aanbieders wil weten wie de peers zijn en ruim 90% wil weten hoe de peers zijn geselecteerd

Aanbieders willen in de terugkoppeling ook spiegelinformatie over andere (groepen) aanbieders dan peers

Voor leren en verbeteren vindt circa 70% van de aanbieders het van belang dat de integrale vergelijking ook mogelijkheden biedt om zichzelf te vergelijken met andere organisaties dan de peers en met het sectorgemiddelde (Figuur 3.3). De peers die worden geselecteerd via de methode van de integrale vergelijking volgen uit het DEA model. Aanbieders willen ook graag zelf (groepen van) aanbieders kunnen selecteren om zichzelf mee te vergelijken. Dat betreft vaak organisaties met een vergelijkbare omvang of uit dezelfde regio. De wensen zijn hierbij overigens niet uniform: sommige aanbieders zien vanuit concurrentieperspectief juist eerder een vergelijking met aanbieders uit andere regio's zitten. Daarmee lijkt de gedeelde wens de mogelijkheid om als aanbieder zelf andere vergelijkingsgroepen te kunnen selecteren op basis van eigen behoefte.

De mogelijkheid om zelf (groepen van) aanbieders te selecteren biedt ook voor peers de mogelijkheid om zichzelf te vergelijken als startpunt voor leren en verbeteren. De geselecteerde methode voor de integrale vergelijking biedt (vooral nog) voor peers alleen de terugkoppeling dat de aanbieder tot de meest efficiënte aanbieders behoort. In de nulmeting behoorden 34 van de 286 aanbieders tot de peers (circa 10%). Met een flexibele vergelijkingsgroep en aanvullende spiegelinformatie (zie volgende paragraaf) biedt de integrale vergelijking ook deze aanbieders mogelijkheden om te leren en verbeteren.

Hoe belangrijk vindt u het te kunnen vergelijken met:



Figuur 3.3: Rond de 70% van de aanbieders vindt het (zeer) belangrijk om zichzelf ook met andere organisaties dan de peers en het gemiddelde van de sector te vergelijken

Aanbieders hebben behoefte aan spiegelinformatie over factoren die de berekende relatieve efficiëntiescore verklaren

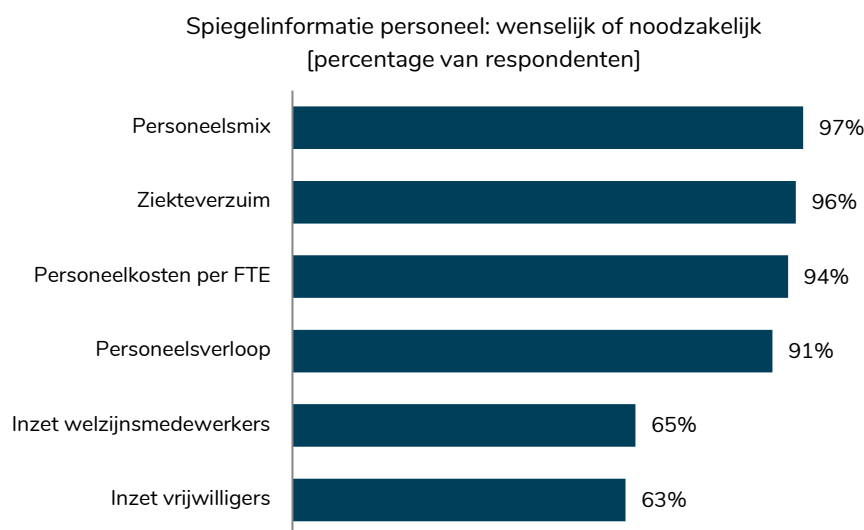
Om handvatten te bieden voor leren en verbeteren zou de integrale vergelijking naast verschillen in efficiëntie, ook inzicht in de factoren die hierop van invloed zijn, moeten bieden.

De nulmeting bood aanbieders inzicht in de relatieve efficiëntiescore (tussen 0-1) ten opzichte van hun (anonieme) peers en een daarmee samenhangend minimum en maximum kostenniveau. Om concrete handvatten te bieden voor leren en verbeteren is inzicht nodig in welke factoren deze (relatieve) score beïnvloeden en hoe aanbieders daarop scoren ten opzichte van peers en andere aanbieders. Door hierbij beïnvloedbare en niet-beïnvloedbare factoren te onderscheiden, kunnen er concrete onderwerpen waarop aanbieders kunnen leren en verbeteren worden geïdentificeerd. Aanbieders kunnen leren door gericht de dialoog aan te gaan met aanbieders die hierop beter 'scoren'.

Aanbieders die bijvoorbeeld een efficiëntiescore van 0,8 hebben, kunnen er via spiegelinformatie over achterliggende factoren achter komen dat dat deels komt door de relatief ongunstige score op ziekteverzuim en personeelsverloop. Als blijkt dat peers of andere aanbieders daar juist wel gunstig op scoren, kan men uitzoeken wat deze aanbieders op dit gebied anders doen (leren) en dat mogelijk ook gaan toepassen in de eigen organisatie (verbeteren).

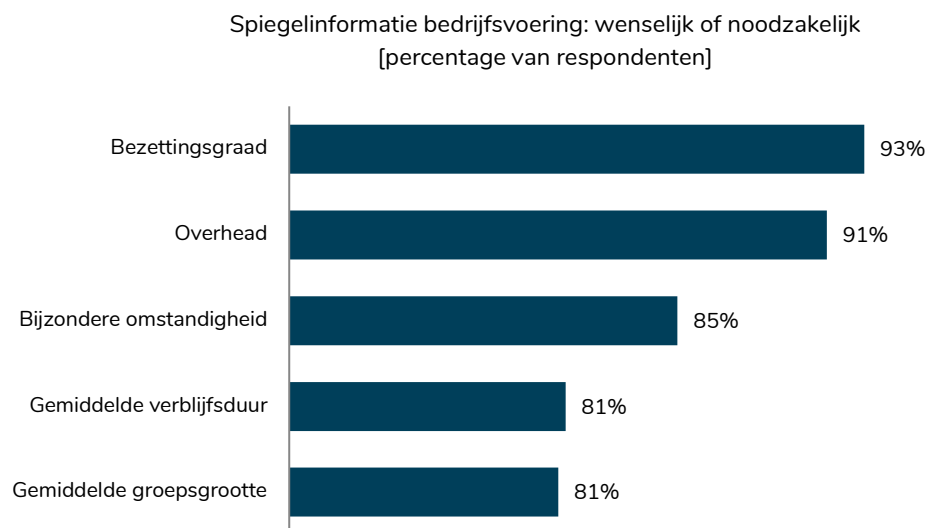
In de enquête hebben aanbieders onderwerpen op het gebied van personeel (Figuur 3.4) en bedrijfsvoering (Figuur 3.5) genoemd waarvoor ze spiegelinformatie als gewenst of noodzakelijk zien. Meer dan 90% van de aanbieders vindt spiegelinformatie over de personeelsmix, ziekteverzuim, personeelskosten en personeelsverloop wenselijk of noodzakelijk (Figuur 3.4). Daarnaast wordt spiegelinformatie over de inzet van welzijnsmedewerkers en vrijwilligers door meer dan 60% van de aanbieders als wenselijk of noodzakelijk gezien. Aanvullend hierop zijn onder andere onderwerpen als leeftijdsopbouw, verhouding inhuur-flex-vast personeel (en PNIL),

opleidingskosten, invulling van vacatures, tevredenheid personeel en de inzet FTE per client genoemd.



Figuur 3.4: Meer dan 90% van de aanbieders vindt spiegelinformatie over de personeelsmix, ziekteverzuim, personeelskosten en personeelsverloop wenselijk of noodzakelijk

Tussen de 80% en 90% van de aanbieders vindt spiegelinformatie over bezettingsgraad, overhead, bijzondere omstandigheden, verblijfsduur en groepsgrootte wenselijk of noodzakelijk (Figuur 3.5). Daarnaast benoemen verschillende aanbieders (verschillen in) het zorgaanbod, huisvesting, inzet op domotica en digitalisering en kwaliteit (inclusief kwaliteitsbeoordeling) als onderwerpen waarover zij spiegelinformatie willen hebben.



Figuur 3.5: Tussen de 80% en 90% van de aanbieders vindt spiegelinformatie over bezettingsgraad, overhead, bijzondere omstandigheden, verblijfsduur en groepsgrootte wenselijk of noodzakelijk

Aanbieders hebben behoefte aan een duidelijke terugkoppeling om de ervaren complexiteit van de integrale vergelijking te verminderen

Begrip van de onderliggende methode is voor aanbieders een belangrijke voorwaarde om de resultaten van de integrale vergelijking te kunnen gebruiken voor leren en verbeteren. Aanbieders vinden de stappen die in de toelichting bij de terugkoppeling zijn gezet tijdens de nulmeting positief, maar geven ook aan dat ze de integrale vergelijking nog steeds als (te) complex ervaren. Op dit moment is de DEA-methode nog onvoldoende helder voor veel aanbieders en wordt deze soms als een 'black-box' gezien. Dit staat de acceptatie van uitkomsten van de integrale vergelijking in de weg en daarmee ook het leren en verbeteren op basis van deze uitkomsten.

Aanbieders willen op hoofdlijnen de methode kunnen begrijpen en hebben niet zozeer de behoefte om de statistische achtergronden tot in detail te doorgronden. De ervaren complexiteit kan mogelijk verminderd worden door in de terugkoppeling ruimte te bieden aan:

- Concrete rekenvoorbeelden die voor elke aanbieder te herleiden zijn tot de door hen aangeleverde data: hoe leidt de door mij aangeleverde data tot een efficiëntiescore?
- Het uitdrukken van de resultaten van de integrale vergelijking in getallen of ratio's die herkenbaar en bruikbaar zijn in de praktijk
- Een duiding van de resultaten op basis van de factoren waarover het management beslist zoals groepsgrootte en personeelsinzet (zie alinea's hiervoor).

Zorgaanbieders zijn verdeeld over de meest gewenste vorm voor de terugkoppeling. Een deel geeft de voorkeur aan een (uitgebreide) infographic, een ander deel aan een online dashboard en weer een ander deel prefereert een rapportage met kentallen en toelichting. Omdat de achterliggende set aan gegevens en figuren hetzelfde is, is het goed mogelijk om elk van deze opties aan te bieden.

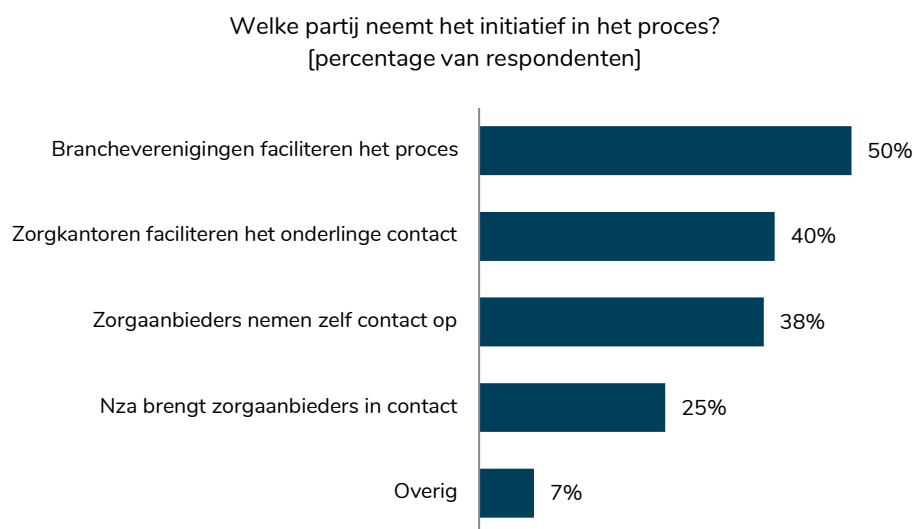
3.2.2 Voor leren en verbeteren is noodzakelijk het proces van leren en verbeteren met de IV te schetsen

Om te kunnen leren en verbeteren op basis van de integrale vergelijking wil een groot deel van de aanbieders ook dat het met elkaar in contact komen wordt gefaciliteerd. Aanbieders zien de data en resultaten van de integrale vergelijking als startpunt voor een inhoudelijk gesprek tussen organisaties. Veel aanbieders zien graag dat er naast het inzichtelijk maken van wie de peers zijn (zie hierboven) ook wordt geschetst hoe er met hen in contact kan worden gekomen en dat de organisatie van de dialoog voor leren en verbeteren wordt gefaciliteerd.

Brancheverenigingen (50%) en zorgkantoren (40%) worden daarbij relatief vaak genoemd als partijen die het initiatief voor deze eerste stap in het proces van leren en verbeteren op zich kunnen nemen (Figuur 3.6). Daarnaast geeft een deel van de aanbieders aan dat hier een rol voor zorgaanbieders zelf is weggelegd. Met name grotere zorgaanbieders kunnen hierbij volgens een deel van de sector een regiefunctie in de regio spelen. Daarbij is vaak de impliciete aanname dat voor leren en verbeteren het vooral gaat om leren en verbeteren met aanbieders in dezelfde regio. De peers in de integrale vergelijking kunnen echter juist door het hele land zitten. Omdat er tussen

aanbieders uit verschillende delen van het land vaak minder sterke informele netwerken bestaan, kunnen brancheverenigingen hier een rol spelen. Ook zorgkantoren zouden deze rol kunnen spelen maar dat vereist wel onderlinge afstemming: de peers kunnen immers over verschillende zorgkantorregio's verspreid zitten.

De bereidheid om kennis te delen is groot: op basis van de enquête blijkt dat 80% van de aanbieders zeker of waarschijnlijk bereid is om best practices te delen als zij peer zouden blijken te zijn. Wederkerigheid is daarbij wel een belangrijk aspect: 70% wil ook dat zij zelf kennis opdoen of nieuwe inzichten krijgen als zij zelf bepaalde kennis delen met andere aanbieders.



Figuur 3.6: Brancheverenigingen, zorgkantoren en zorgaanbieders zelf worden relatief vaak genoemd als partij om (de eerste stap) in het proces van leren en verbeteren te faciliteren

3.2.3 Voor acceptatie van het instrument voor leren en verbeteren moet duidelijk zijn hoe de IV gebruikt gaat worden voor bekostiging

Veel aanbieders geven aan dat voor de acceptatie van het instrument voor leren en verbeteren het duidelijk moet zijn hoe de integrale vergelijking wordt gebruikt in de bekostiging. Ten tijde van de inventarisatie onder aanbieders (voorjaar en zomer 2021) was het onduidelijk hoe en op welke termijn de integrale vergelijking van invloed gaat zijn op tarieven. Dit is een vraagstuk dat buiten de scope van dit onderzoek valt en parallel door de NZa nader wordt uitgewerkt. Aanbieders geven aan dit inzicht hierin ook vanuit het doel van leren en verbeteren van belang is, omdat bij sommigen het beeld bestaat dat de integrale vergelijking zal leiden tot kortingen op het tarief. Aanbieders waar de angst bestaat dat de integrale vergelijking tot kortingen zal leiden, zullen het instrument zonder duidelijkheid over de rol in de bekostiging ook niet accepteren voor leren en verbeteren.

4 Handvatten voor de zorginkoopdialoog hangen samen met rol bij tariefstelling en inhoudelijke doorontwikkeling

De integrale vergelijking kan volgens zorgkantoren en een deel van de zorgaanbieders handvatten bieden bij de zorginkoop als duidelijk wordt wat de rol bij de tariefstelling wordt en inhoudelijk doorontwikkeling plaatsvindt. Aanbieders zijn nu verdeeld over de toepasbaarheid van de integrale vergelijking in de zorginkoopdialoog. 55% beoordeelt dit (nu) als neutraal tot (zeer) goed en 39% als (zeer) beperkt. Om concrete handvatten te kunnen bieden voor de zorginkoopdialoog zien zorgkantoren op hoofdlijnen dezelfde inhoudelijke doorontwikkeling (zie hoofdstuk 3) voor zich als zorgaanbieders.

Door de integrale vergelijking eerst voor leren en verbeteren toe te passen en daarna voor de zorginkoopdialoog kan volgens sommige partijen het vertrouwen in het instrument worden opgebouwd.

4.1 Formuleren van handvatten door partijen voor zorginkoopdialoog is lastig zolang het onduidelijk is hoe de relatie met tariefstelling is

Zorgaanbieders en zorgkantoren geven aan dat het lastig is om concrete (wensen) voor handvatten voor de zorginkoopdialoog te formuleren als niet duidelijk is hoe de integrale vergelijking doorwerkt in tariefstelling. Ten tijde van de inventarisatie onder aanbieders en zorgkantoren (voorjaar en zomer 2021) was het onduidelijk of, hoe en op welke termijn de integrale vergelijking van invloed gaat zijn op de tariefstelling. De inhoud van de zorginkoopdialoog en de tariefstelling zijn met elkaar verbonden. Hoe tarieven tot stand komen en de daarbij behorende onderhandelingsruimte vormen een belangrijk en integraal onderdeel van de zorginkoopdialoog en daarmee met de mogelijke handvatten die de integrale vergelijking kan bieden. Zolang dat niet duidelijk is, vinden aanbieders en zorgkantoren het lastig aan te geven hoe de integrale vergelijking kan ondersteunen in de zorginkoopdialoog.

4.2 Zorgaanbieders zijn verdeeld over de toepasbaarheid van de integrale vergelijking in de zorginkoopdialoog

Zonder dat duidelijk is hoe de integrale vergelijking van invloed gaat zijn op de tariefstelling zijn zorgaanbieders verdeeld over de algemene toepasbaarheid van de integrale vergelijking in de zorginkoopdialoog. In de enquête beoordeelt 55% van de aanbieders de toepasbaarheid van de integrale vergelijking in de zorginkoop neutraal tot (zeer) goed en 39% van de aanbieders als (zeer) beperkt (Figuur 4.1).



Figuur 4.1: zorgaanbieders zijn verdeeld over de toepasbaarheid van de integrale vergelijking voor de zorginkoopdialoog

Aanbieders die de toepasbaarheid als neutraal tot (zeer) goed beoordelen zien de integrale vergelijking als startpunt voor een dialoog met het zorgkantoor op basis van een gelijkwaardige(re) uitgangspositie. De integrale vergelijking biedt vanuit hun perspectief handvatten voor de zorginkoop omdat zorgkantoren en aanbieders daarmee dezelfde kwantitatieve inzichten hebben. De informatie-asymmetrie die door sommige aanbieders wordt ervaren, wordt hierdoor verminderd. Een deel van de aanbieders geeft aan dat zorgkantoren nu over overkoepelende (spiegel-)informatie beschikken waarmee aanbieders onderling worden vergeleken die zij zelf niet hebben. Door een gedeeld beeld over onderwerpen waarop aanbieders afwijken van hun peers of andere aanbieders kan gericht gesproken worden over leren en verbeteren. Daarbij geven aanbieders als voorwaarde dat dit ook echt een dialoog is. Sommige aanbieders vrezen dat de integrale vergelijking leidt tot een inkoopproces waarbij er 'eenzijdig normen (inclusief risico op kortingen) worden opgelegd' op basis van de resultaten uit de integrale vergelijking. Door expliciet te maken hoe de integrale vergelijking in het inkoopproces wordt gebruikt kan dit beeld mogelijk worden weggenomen – onder de aanname dat dit anders gaat dan een deel van de aanbieders vreest.

Van de aanbieders die de toepasbaarheid als (zeer) beperkt beoordelen, ziet een deel het gebruik van één instrument (de integrale vergelijking) voor meerdere doelen als niet wenselijk. Leren en verbeteren vindt plaats in een dialoog waarin vertrouwen en openheid (over waarom een aanbieder bijvoorbeeld relatief minder goed scoort op een bepaald onderwerp) van belang zijn. Volgens een deel van de aanbieders gaat dat niet samen met een risico op tariefkortingen (als dat zou samenhangen met dezelfde minder goede score op een onderwerp). Als er bij het geven van bepaalde informatie negatieve financiële consequenties kunnen volgen, is er een prikkel voor strategisch gedrag (*gaming*) om de uitkomst van een indicator te beïnvloeden of alle aandacht te richten op het maximaliseren van bepaalde indicatoren (*teaching to the test*).⁴ Deze prikkel kan er toe leiden dat de voor leren en verbeteren noodzakelijke openheid over onderwerpen waarop een aanbieder juist minder goed scoort onder druk komt te staan. Daarnaast geeft een deel van de aanbieders aan dat ze het fundamenteel oneens zijn met gebruik van de integrale vergelijking (op basis van de nulmeting) voor een inkoopdialoog. Verschillen tussen aanbieders (op gebied van kwaliteit en cliëntmix) worden in hun ogen niet goed in beeld worden gebracht. Daarnaast wordt het als een risico ervaren dat er 'kaasschaafkortingen' volgen in plaats van een fundamentele discussie over de toekomst van de ouderenzorg.

Het oordeel over de toepasbaarheid bij de zorginkoopdialoog lijkt te worden beïnvloed door risico's en (angst-)beelden. Duidelijkheid over hoe de integrale vergelijking door zal werken in tarieven (zie paragraaf 4.1.) kan hier van invloed op zijn omdat dan expliciet is hoe reëel de ervaren risico's en angstbeelden zijn. Dit kan door een toelichting op hoe dit zal gaan gebeuren in combinatie met een rekenvoorbeeld om het basisidee van de doorwerking duidelijk te maken en in een later stadium mogelijk gevolgd door een simulatie om mogelijke verschillen met de bestaande tarieven duidelijk te maken per aanbieder. Deze duidelijkheid en inzichten kunnen vervolgens de beoordeling van de toepasbaarheid van de integrale vergelijking voor de zorginkoopdialoog veranderen.

4.3 Zorgkantoren wensen dezelfde inhoudelijke doorontwikkeling als aanbieders en hebben aantal voorwaarden voor toepassen iv bij zorginkoop

Leren en verbeteren is in de ogen van de geraadpleegde zorgkantoren een gedeelde wens en ambitie van aanbieders en zorgkantoren. Dit betreft zowel leren en verbeteren op de korte termijn als op de langere termijn op vraagstukken als arbeidsmarkt en innovatie die voor aanbieders en zorgkantoren onderdeel zijn van de zorginkoop. De integrale vergelijking moet vanuit dat perspectief voor zorgkantoren daarom ook bijdragen aan de brede maatschappelijk doelstelling om verpleeghuiszorg betaalbaar, doelmatig en toegankelijk te houden.

⁴ Zie bijvoorbeeld CPB 2014: Neveneffecten sturingsinstrumenten

Om vanuit dat doel concrete handvatten te bieden voor de zorginkoopdialoog zien zorgkantoren op hoofdlijnen dezelfde inhoudelijke doorontwikkeling voor zich als zorgaanbieders. Zorgkantoren zien de potentie voor een op data gebaseerde vergelijking van aanbieders als basis voor een zorginkoopdialoog en verschillende zorgkantoren maken daar nu (op basis van eigen data) ook al gebruik van. De in paragraaf 4.3.2. geformuleerde inhoudelijke wensen van de zorgaanbieders (o.a. inzicht in spiegelinformatie over factoren die samenhangen met verschillen in efficiëntie, flexibiliteit om ook te kunnen vergelijken met andere aanbieders peers) wordt gedeeld door zorgkantoren. Daarmee kan de integrale vergelijking via een gedeelde feitenbasis een goed startpunt voor een zorginkoopdialoog zijn met gerichte afspraken over leren en verbeteren als doel.

Zorgkantoren vinden dat er bij de introductie van de integrale vergelijking (waarbij peers door het hele land worden geselecteerd en niet alleen in de eigen regio) er wel voldoende ruimte binnen het eigen inkoopbeleid moet blijven om in te spelen op regio-specifieke vraagstukken. Vanuit deze voorwaarde leidt de invoering van de integrale vergelijking bij zorgkantoren tot de vraag hoe dat geborgd gaat worden. Specifiek aandachtspunt zijn aanbieders die actief zijn in verschillende zorgkantorregio's. De uitkomsten van de integrale vergelijking voor deze aanbieders zal dan worden gebruikt voor de zorginkoopdialoog van één aanbieder met meerdere zorgkantoren. De vraag is vervolgens of zorgkantoren de ruimte hebben om elk tot verschillende inkoopafspraken te komen op basis van dezelfde resultaten van de integrale vergelijking.

4.4 Bij de inzet van iv bij zorginkoop hechten partijen nog sterker aan inhoudelijke doorontwikkeling dan bij leren en verbeteren

Zorgkantoren en zorgaanbieders geven aan dat als de integrale vergelijking wordt toegepast bij de zorginkoopdialoog er (directe) financiële gevolgen aan de resultaten zijn gekoppeld. Dit maakt dat de vereisten aan de inhoud van het model en de kwaliteit van de data (nog) hoger zijn bij dan de toepassing voor leren en verbeteren. Verschillende aanbieders en zorgkantoren benoemen op basis van de (ervaringen met) de nulmeting een aantal inhoudelijke aandachtspunten bij de huidige methode:

- De gebruikte DEA methode focust op verpleeghuiszorg ZZP/VPT4-10 en daarmee alleen op een deel van de Wlz en een deel van ouderenzorg. Veel aanbieders bieden vaak ook andere zorg die vaak (deels) integraal met verpleeghuiszorg georganiseerd is in de organisatie.
- Kapitaallasten, vastgoed en (investeringsruimte voor) innovatie vallen nu buiten de integrale vergelijking.
- De (verschillen in) kwaliteit van de zorg wordt (nu) nog niet voldoende meegenomen in het DEA model.

Een meerderheid van de aanbieders en zorgkantoren is er door deze punten niet van overtuigd dat het in de nulmeting toegepaste DEA model de efficiëntieverschillen goed in beeld brengt en de

juiste peers identificeert. Bij de doorontwikkeling van het model en de dataverzameling achtten zij het van belang dat deze punten worden geadresseerd (zie ook hoofdstuk 6).

Om het vertrouwen in de resultaten te vergroten ziet een deel van de zorgkantoren en zorgaanbieders een volgtijdelijke ontwikkeling en invoering van de integrale vergelijking voor zich. De wens is daarbij om de integrale vergelijking eerst toe te passen voor leren en verbeteren voordat er directe financiële consequenties via tarieven aan worden verbonden. Hiermee kan vertrouwen in het instrument worden opgebouwd. Een breed draagvlak voor het instrument wordt van groot belang geacht en dat zou door een incrementele invoering (eventueel op basis van een door de sector gedragen roadmap) kunnen worden bereikt.

5 Om concrete handvatten te bieden moeten efficiëntieverschillen worden verklaard

Uitgangspunt voor leren en verbeteren is dat de DEA-methode verschillen in efficiëntie goed meet. Om concrete handvatten voor leren en verbeteren te bieden is het noodzakelijk om deze verschillen ook zo goed mogelijk te verklaren. Dat kan door een tweede analysestap toe te voegen aan DEA die verklarende factoren identificeert. Hiervoor moet eerst per aanbieder informatie over potentieel verklarende factoren worden verzameld. Via statistische analyses wordt vervolgens de invloed van deze factoren in kaart gebracht. Goede visualisaties van de resultaten hiervan bieden concrete inzichten aan aanbieders en zorgkantoren over de onderwerpen waar efficiëntieverbetering denkbaar is en van welke aanbieder(s) geleerd kan worden. Deze inzichten zijn het startpunt voor een dialoog gericht op leren en verbeteren tussen aanbieders en tussen zorgkantoren en aanbieders.

Dit hoofdstuk gaat in op de wensen en mogelijkheden voor het bieden van concrete handvatten op basis van DEA. Voor zowel leren en verbeteren, de zorginkoopdialoog als het verminderen van de (ervaren) complexiteit betekent dit dat de integrale vergelijking voor een aanbieder (peer of geen peer) antwoord moet geven op de volgende drie volgtijdelijke vragen:

1. Hoe efficiënt ben ik (en hoe is dat berekend)?
2. Op welke onderwerpen die mijn efficiëntie beïnvloeden kan ik verbeteren?
3. Van welke aanbieders kan ik leren op deze onderwerpen om dat te bereiken?

Dit hoofdstuk gaat in op het uitgangspunt waaronder DEA vraag 1 kan beantwoorden (paragraaf 5.1), hoe met het toevoegen van second stage DEA en goede visualisaties vragen 2 en 3 kunnen worden beantwoord (paragraaf 5.2 en 5.3) en wat opties zijn om het eerste contact voor leren en verbeteren te faciliteren (paragraaf 5.4).

Bij het gebruik van DEA voor leren en verbeteren is de kans aanwezig dat de focus vooral op onderwerpen komt te liggen die de (korte) termijn (statische) efficiëntie verbeteren. Veranderingen die op lange termijn tot efficiëntievoordelen leiden (dynamische efficiëntie) hebben vaak in de eerste jaren hogere kosten. Deze hogere kosten verlagen op korte termijn de efficiëntiescore. Afhankelijk van hoe de tariefstelling en zorginkoop plaats vindt, heeft dat voor een aanbieder mogelijk nadelige effecten op de tarieven. Leren en verbeteren op langere termijn zou dan ook op

een andere manier gestimuleerd moeten worden dan via een model dat ook toegepast wordt voor bekostiging.

In hoofdstukken drie en vier worden ook verschillende wensen en voorwaarden van aanbieders en zorgkantoren genoemd die betrekking hebben op de rol van de integrale vergelijking en DEA in bekostiging. Deze vraagstukken vallen buiten de scope van dit onderzoek en zijn onderdeel van de bredere doorontwikkeling van de integrale vergelijking.

5.1 Uitgangspunt in dit hoofdstuk is dat DEA de verschillen in efficiëntie correct meet

Een belangrijke voorwaarde voor het bieden van concrete handvatten voor leren en verbeteren en de zorginkoopdialoog is dat de DEA-methode de daadwerkelijke efficiëntieverschillen tussen aanbieders meet. Tijdens interviews en werksessies stelden veldpartijen op basis van hun ervaringen met de nulmeting dit regelmatig ter discussie. In hoofdstuk zes wordt een aantal opties geschetst die in een vervolgmeting de kwaliteit van de gemeten efficiëntieverschillen kan verbeteren. De NZa adresseert dit vraagstuk in de bredere doorontwikkeling van de integrale vergelijking.

Voor het verkennen van de mogelijkheden voor handvatten voor leren en verbeteren en de zorginkoopdialoog hanteren we het vooraf gestelde uitgangspunt dat de verschillen in efficiëntie goed worden gemeten. Voor de toepassing van de DEA-methode betekent dit dat de twee type variabelen die worden gebruikt voor elke aanbieder goed gemeten worden en vergelijkbaar zijn: de productie en het kostenniveau.

5.1.1 De productie wordt per aanbieder goed gemeten

In de nulmeting wordt de productie gemeten door zes categorieën van zorg te onderscheiden (op basis van ZZP's/VPT's, Tabel 2.1). Voor de betrouwbaarheid van de uitkomsten van DEA is het van belang dat aanbieders minimaal zorg leveren die voldoet aan de maatschappelijke ondergrens op kwaliteitsgebied en dat er geen of voor de uitkomst niet relevante verschillen tussen aanbieders bestaan in de productmix binnen elke categorie zorg. Aanbieders die zorg leveren die niet voldoet aan de minimale gewenste kwaliteit kunnen ten onrechte tot peer worden aangewezen omdat ze mogelijk het meest efficiënt zijn omdat ze deze lagere kwaliteit leveren. Voor een correcte meting van efficiëntieverschillen moet de kostengrens van DEA alleen gebaseerd zijn op aanbieders die zorg leveren die niet onder de maatschappelijke ondergrens van kwaliteit ligt. In hoofdstuk 6 gaan we in op hoe dat mogelijk beter te waarborgen is dan in de nulmeting. Verschillen in de productie binnen een zorgcategorie kunnen ontstaan als casemix van patiënten/cliënten tussen aanbieders binnen de gehanteerde productcategorieën verschilt.

5.1.2 Het kostenniveau wordt per aanbieder goed gemeten

Voor veel aanbieders wordt slechts een gedeelte van hun totale zorgproductie meegenomen in de integrale vergelijking. Dat betekent dat de totale kosten moeten worden verdeeld tussen een deel van de zorg dat wel en een deel van de zorg dat niet wordt meegenomen. Voor een goede

vergelijking tussen aanbieders is het noodzakelijk dat deze kostentoerekening op identieke wijze gebeurd door alle aanbieders. Zeker als de integrale vergelijking gebruikt gaat worden voor tariefstelling is het belangrijk dat de zorgaanbieders hun kosten op dezelfde manier toerekenen. Het gaat dan bijvoorbeeld over de toedeling van de overhead aan dat deel van de productie dat onder de integrale vergelijking valt; daarvoor kunnen diverse verdeelsleutels gebruikt worden (per m², per fte, per Euro omzet). Of over het toerekenen van personeel naar verschillende zorgproducten wat bij sommige zorgaanbieders met een urenregistratie wordt gedaan, en bij anderen forfaitair.

5.2 Een tweede analysestap in de DEA-methode kan factoren die efficiëntieverschillen verklaren identificeren.

Als het DEA-model de verschillen in efficiëntie goed meet, is de basis voor leren en verbeteren aanwezig, maar dat is niet voldoende. Uit de enquête en de reacties tijdens de duidingsperiode van de nulmeting blijkt dat de kernvraag die bestuurders met behulp van de IV willen beantwoorden is: *“Wat doen mijn peers anders dan mijn organisatie waardoor ze efficiënter zijn en wat kan ik daarvan leren?”*.

De DEA-methode biedt daar in de versie van de nulmeting nog geen antwoord op. Voor een aanbieder met een efficiëntiescore lager dan 1 wordt duidelijk dat er één of meerdere combinaties van aanbieders zijn die gezamenlijk vergelijkbare zorg efficiënter leveren. Idealiter wordt in de toekomst bekend welke aanbieders dat zijn en waarom ze efficiënter zijn, zodat ze met elkaar in gesprek kunnen gaan. Dit sluit ook aan bij de eisen die VWS stelt voor de integrale vergelijking⁵: *“Het is noodzakelijk dat zorgaanbieders en zorgkantoren weten waarop ze kunnen sturen en waarop de efficiëntiescore van de NZa is gebaseerd.”*

Om met DEA concrete handvatten te bieden voor leren en verbeteren, is het noodzakelijk om voor een aanbieder met een efficiëntiescore lager dan 1 ook de vraag te beantwoorden waarom andere aanbieders de zorg efficiënter leveren. Voor het verklaren van verschillen in efficiëntie die met DEA worden gemeten, bestaan technische mogelijkheden. Deze technieken worden samen ‘second stage DEA’⁶ genoemd. Deze tweede analysestap bestaat uit drie onderdelen die we hieronder nader toelichten:

1. Verzamelen van een longlist van potentieel verklarende factoren van efficiëntieverschillen.
2. Uitvoeren en terugkoppelen van verkennende analyses naar de relatie tussen individuele factoren en de efficiëntiescore.
3. Uitvoeren van regressieanalyses om tot een shortlist van verklarende factoren voor efficiëntieverschillen te komen.

⁵ Conform criterium 4.3 van de beleidsmatige toetsingscriteria bij beoordeling uitkomsten integrale vergelijking, opgenomen in de brief d.d. 14 november 2019 van de minister van VWS.

⁶ Zie bijvoorbeeld “Second stage DEA: Comparison of approaches for modelling the DEA score, A. Hoff, 2007, <https://doi.org/10.1016/j.ejor.2006.05.019>”

Het uitvoeren van deze onderdelen vergt onderzoek waarbij een goede betrokkenheid van het veld de inhoudelijke kwaliteit en het draagvlak vergroot. Dit kan worden vormgegeven door het onderzoek te laten begeleiden door een commissie die bestaat uit experts en vertegenwoordigers van het veld. Deze commissie neemt besluiten over de te nemen stappen en beoordeelt de resultaten van de analyses. Idealiter worden er vooraf ‘spelregels’ afgesproken over het proces: bijvoorbeeld dat (vertegenwoordigers van) alle partijen zich committeren aan de uitkomst en de beslissingen die de commissie maakt.

Startpunt is het verzamelen van een longlist van potentieel verklarende factoren voor efficiëntieverschillen tussen aanbieders

De analyse begint met het verzamelen van potentieel verklarende factoren voor efficiëntieverschillen. Deze factoren bieden spiegelinformatie voor de aanbieders en zorgkantoren: hoe scoort een aanbieder op factoren ten opzichte van de peers?

Een eerste longlist is opgesteld op basis van de input tijdens dit onderzoek

Uit deskresearch, interviews, de enquête onder aanbieders en de themabijeenkomst van 29 juni volgt onderstaand voorstel voor een eerste longlist van potentieel verklarende factoren (Tabel 5.1). Als wordt besloten om een second stage DEA uit te voeren ligt het voor de hand om de definitieve longlist samen met de sector vast te stellen.

Dit kan bijvoorbeeld door de volgende stappen te doorlopen:

- Ophalen mogelijke factoren: via een brainstorm met vertegenwoordigers van de sector, of een enquête zo veel mogelijk denkbare factoren opstellen.
- Kwalitatieve toets op uitlegbaarheid: op basis van literatuur en/of het raadplegen van experts uit de sector (mogelijk met de Delphi-methode) vaststellen hoe aannemelijk het is dat er een relatie is met efficiëntie. Dit zou ook in een bijeenkomst met de in de vorige paragraaf voorgestelde commissie kunnen worden vastgesteld.
- Inventarisatie beschikbaarheid data: samen met vertegenwoordigers van de sector nagaan of er data beschikbaar of te verzamelen is over deze factoren.

Factor	Toelichting en mogelijke relatie met kosten en efficiëntie ⁷
Personeel	
Personeelsmix	<ul style="list-style-type: none"> • Verdeling personeel over schalen en binnen schalen: hoe hoger gemiddelde schaal (en trede daarbinnen) hoe hoger de personeelskosten • Gemiddelde snelheid waarmee schalen worden doorlopen: hoe sneller hoe hoger de personeelskosten • Gemiddeld aantal dienstjaren: hoe hoger het aantal dienstjaren hoe meer ervaren (en mogelijk productiever) maar ook hogere personeelskosten • Aandeel externe inhuur (PNIL: personeel niet in loondienst): hoe hoger dit aandeel hoe hoger de personeelskosten

⁷ Ceteris paribus (twee verder identieke organisaties waarbij één een hogere score op die factor dan de ander)

Factor	Toelichting en mogelijke relatie met kosten en efficiëntie ⁷
Ziekteverzuim	<ul style="list-style-type: none"> Ziekteverzuimpercentage: hoger verzuim leidt tot hogere personeelskosten door inzet vervanging en/of extra ureninzet en mogelijk ook lagere productiviteit
Personeelskosten per FTE	<ul style="list-style-type: none"> De totale kosten (naast loon ook sociale lasten, onkostenvergoedingen etc.) per FTE: hoe hoger de kosten hoe lager de efficiëntie tenzij er een hogere productiviteit tegenover staat
Personeelsverloop	<ul style="list-style-type: none"> Uitstroompercentage (uitsplitsing eventueel naar oorzaken als pensionering etc.): hoe hoger de uitstroom hoe hoger de kosten voor aantrekken en inwerken nieuw personeel en hoe lager de productiviteit
Inzet welzijns-medewerkers	<ul style="list-style-type: none"> Inzet in FTE t.o.v. cliënten of totaal FTE: kan productiviteit zorgmedewerkers mogelijk vergroten door overname bepaalde taken en daardoor totale efficiëntie verhogen
Inzet vrijwilligers	<ul style="list-style-type: none"> Inzet in FTE t.o.v. cliënten of totaal FTE: kan productiviteit zorgmedewerkers mogelijk vergroten door overname bepaalde taken en daardoor totale efficiëntie verhogen
Bedrijfsvoering	
Bezettingsgraad	<ul style="list-style-type: none"> Aantallen cliënten t.o.v. capaciteit: hoe lager de bezettingsgraad hoe hoger de vaste kosten per client zijn en hoe lager de efficiëntie
Overhead	<ul style="list-style-type: none"> Uitgaven aan niet direct aan zorg gerelateerde activiteiten t.o.v. totale kosten of client: hogere overhead kan samenhangen met lagere efficiëntie
Bijzondere financiële omstandigheid	<ul style="list-style-type: none"> Tijdelijke financiële mee- of tegenvallers of eenmalige grote uitgaven: kunnen leiden tot een eenmalige lage of juist erg hoge (berekende) efficiëntie (idealiter al verrekend in kosten)
Gemiddelde verblijfsduur	<ul style="list-style-type: none"> Aantal dagen dat een cliënt gemiddeld in een instelling verblijft: een kortere verblijfsduur kan tot een lagere efficiëntie leiden vanwege hogere kosten voor opname en ontslagprocessen en kortdurende leegstand
Groepsgrootte	<ul style="list-style-type: none"> Gemiddeld aantal cliënten per groep: relatie met efficiëntie kan zowel positief als negatief zijn
Omgevingsfactoren	
Werkgebied/regio	<ul style="list-style-type: none"> Werkgebied of regio waar de aanbieder actief: specifieke regionale kenmerken⁸ kunnen leiden tot structureel hogere of lagere kosten
Stedelijkheid	<ul style="list-style-type: none"> Stedelijkheid (vaak gemeten via 5 klassen van oppervlakte-adressendichtheid) van werkgebied van (locaties) van aanbieder: stedelijkheid lijkt samen te hangen met hogere kosten
SES	<ul style="list-style-type: none"> Sociaal-economische-status van cliënten of van werkgebied: lagere SES-score lijkt samen te hangen met hogere zorgvraag gegeven een ZZP-indicatie en daarmee lagere gemeten efficiëntie

Tabel 5.1: voorstel voorlopige longlist van potentieel verklarende factoren voor efficiëntieverschillen op basis van interviews, sessies en themabijeenkomst 29 juni.

Bij elk van de factoren wordt bepaald of het voor de aanbieder beïnvloedbare of niet-beïnvloedbare factoren zijn

Voor de potentieel verklarende factoren moet worden bepaald of deze beïnvloedbaar of niet-beïnvloedbaar zijn door de zorgaanbieder. Voor leren en verbeteren en voor toepassen bij tarieven is het immers van belang of het binnen de macht van een aanbieder ligt om eventuele inefficiëntie

⁸ Idealiter wordt uit de voorgestelde second stage DEA duidelijk welke specifieke regionale kenmerken dat zijn

via het verbeteren van de onderliggende factoren daadwerkelijk te verminderen. Beïnvloedbare factoren staan centraal bij het leren en verbeteren: hier kan een instelling immers zelf op sturen en kan leren van andere aanbieders helpen. Niet-beïnvloedbare factoren zijn te gebruiken voor een correctie op de DEA-score (zie hoofdstuk 6) die als basis voor tarieven wordt gebruikt. Om deze redenen is onderscheid tussen beïnvloedbare of niet-beïnvloedbare factoren één van de toetsingscriteria die VWS hanteert bij de beoordeling van de uitkomsten van de integrale vergelijking⁹.

Voor een deel van de factoren is de beïnvloedbaarheid eenduidiger vast te stellen dan voor de rest. Zo kan een hoger personeelsverloop dan gemiddeld duiden op een minder goed HR-beleid (wat beïnvloedbaar is) of vooral samen hangen met een krappe regionale arbeidsmarkt (en daarmee niet echt beïnvloedbaar). Dit is mogelijk technisch vast te stellen door (ook) het relatieve personeelsverloop ten opzichte van aanbieders in de regio¹⁰ als indicator te gebruiken in het onderzoek.

Het valideren van een dergelijke analyse vindt idealiter plaats via de voorgestelde commissie bestaande uit vertegenwoordigers van het veld en experts. Deze commissie kan ook de onafhankelijke partij zijn die vaststelt of een factor wel of niet beïnvloedbaar is als dit via een analyse niet mogelijk is.

Er zijn drie situaties voor informatie over de potentieel verklarende factoren

Voor second stage DEA is meer informatie nodig dan ten tijde van de nulmeting voor de IV werd verzameld. Er zijn drie mogelijke situaties om deze informatie te verzamelen:

- De informatie is beschikbaar uit de uitvraag zoals bij de nulmeting gebruikt werd.
- De informatie kan worden afgeleid uit reeds beschikbare bronnen en in de toekomst mogelijk met één KIK-V¹¹ uitvraag voor de integrale vergelijking worden uitgevraagd.
- De informatie wordt nu nog niet (door alle aanbieders) verzameld en zal apart moet worden geregistreerd en verzameld.

Hierbij moet een afweging worden gemaakt of de meerwaarde van de informatie opweegt tegen de hogere administratieve lasten. Tijdens de themabijeenkomst van 29 juni 2021 bleek dat een groot deel van de daar aanwezige aanbieders het goed vindt om meer informatie aan te leveren als het efficiëntieverschillen verklaart, zeker als de benodigde gegevens al voor andere doeleinden worden verzameld.

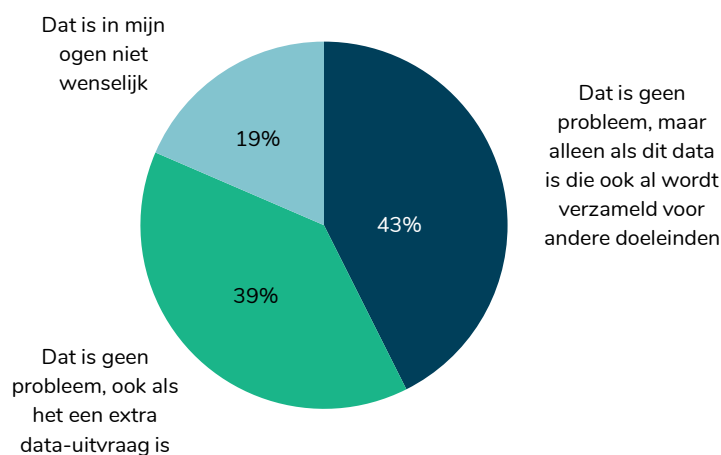
⁹ Conform criterium 4.2 van de beleidsmatige toetsingscriteria bij beoordeling uitkomsten integrale vergelijking, opgenomen in de brief d.d. 14 november 2019 van de minister van VWS.

¹⁰ Dit kan bijvoorbeeld door per aanbieder een unieke regio af te bakenen op basis van een aantal kilometers of reistijd, daarbinnen over alle andere aanbieders het gemiddelde personeelsverloop te bepalen en dat vervolgens af te trekken van het personeelsverloop van de aanbieder in kwestie., of door de arbeidsregio's voor zorg en welzijn van AZW/CBS te gebruiken.

¹¹ Binnen het programma Keteninformatie Kwaliteit Verpleeghuiszorg (KIK-V) maken ketenpartijen afspraken over het vaststellen en uitwisselen van (kwaliteits-)informatie zodat zorgaanbieders niet steeds dezelfde informatie hoeven aan te leveren aan verschillende partijen.

Naast (logischerwijs) een uniform gehanteerde definitie is het van belang dat de verzamelde gegevens betrekking hebben op deze periode en dezelfde rapportage-eenheid als de gegevens over kosten en productie.

Hoe kijkt u aan tegen het aanleveren van aanvullende data t.o.v. de nulmeting? [procent, n=54]



Figuur 5.1 Een ruime meerderheid van de aanwezige zorgaanbieders op de themabijeenkomst van 29 juni 2021 ziet geen probleem in het aanleveren van aanvullende data

5.2.1 Met verkennende analyses wordt relatie tussen potentieel verklarende factoren en efficiëntiescores in kaart gebracht en aan de gebruikers teruggekoppeld

Door verkennende analyses uit te voeren, wordt de relatie tussen potentieel verklarende factoren en verschillen in efficiëntiescores inzichtelijk. Het gaat dan bijvoorbeeld om analyses zoals het vergelijken van de (gewogen) gemiddelde efficiëntiescore tussen groepen aanbieders met verschillende efficiëntiescores. Met de resultaten van deze analyses kan meer context worden geboden aan de cijfers door mogelijke conclusies als: "De efficiëntie ligt in stedelijk gebied X% lager" en "Er zijn geen verschillen in efficiëntie tussen kleine en grote zorgaanbieders".

Dit type analyses verlaagt naar verwachting ook de ervaren complexiteit voor (een deel van de) gebruikers. Tijdens de nulmeting werd de efficiëntiescore als abstract ervaren, mede omdat de oorzaken van de verschillen niet zijn toegelicht. De resultaten van bovenstaande analyses bieden een eerste inzicht (op sectorniveau) hierin en zijn eenvoudig te interpreteren en te communiceren naar een brede doelgroep. Dat vermindert naar verwachting de ervaren complexiteit.

5.2.2 Met behulp van de regressietechnieken wordt een shortlist van verklarende factoren en hun gezamenlijk effect gemaakt

De verkennende analyse beschouwt elke factor afzonderlijk. Sommige factoren kunnen 'communicerende vaten' zijn: een hoge score op de ene factor hangt dan altijd samen met een hoge (of juist lage) score op een andere factor zonder dat duidelijk is welke belangrijker is. Een noodzakelijke vervolgstap is dan ook om de gezamenlijke invloed van de factoren in kaart te brengen en het relatieve belang van de verschillende factoren te bepalen. In second stage DEA

wordt met behulp van een regressietechniek, zoals Tobit-regressie¹² en/of machine learning technieken zoals LASSO-regressies¹³ geanalyseerd wat het gezamenlijk verband is tussen de potentieel verklarende factoren op de longlist en de gemeten efficiëntiescores van aanbieders. Deze techniek geeft een selectie van variabelen die het sterkst (en überhaupt) samenhangen met de efficiëntiescore als uitkomst. De coëfficiënt per variabele zegt iets over de richting en sterkte van de samenhang.

Uit deze analyse moet blijken welke factoren werkelijk samenhangen met de gemeten efficiëntiescores en in welke mate. De gevonden coëfficiënten worden weer omgerekend naar eenvoudigere maten die wat zeggen in de praktijk (zie 5.2.2) en besproken met vertegenwoordigers van de sector. Op deze manier ontstaat er een iteratief proces waarbij het resultaat een shortlist is van factoren waarvan de (statistische) relatie met de gemeten efficiëntiescores is vastgesteld. De interpretatie van de resultaten van dergelijke regressieanalyses kan complex zijn voor potentiële gebruikers. Deze complexiteit kan worden verminderd met heldere visualisaties en toelichting.

5.3 Goede visualisaties en heldere toelichting op de resultaten bieden concrete handvatten voor leren en verbeteren

Om de ervaren complexiteit te verminderen, is goede toelichting op en visualisatie van de integrale vergelijking essentieel. Zorgaanbieders en zorgkantoren zien graag een betere en transparantere toelichting op de berekening van de efficiëntiescore en de daarvoor geselecteerde peers. Daarnaast moeten de resultaten van de voorgestelde analyses in de vorige paragraaf ook op een heldere manier gevisualiseerd worden.

5.3.1 Transparantie over de peers en de berekening van de efficiëntiescore vergroot begrip en draagvlak voor de integrale vergelijking.

De ervaren complexiteit van de integrale vergelijking wordt verminderd door meer transparantie te geven over de berekening van de efficiëntiescore en over de peers.

VWS onderzoekt de juridische component van dit transparantievraagstuk in een parallel traject. Uitgangspunt in onderstaand advies is dat de identiteit van peers bekend mag worden gemaakt en dat aanbieders inzicht mogen hebben in elkaars gegevens.

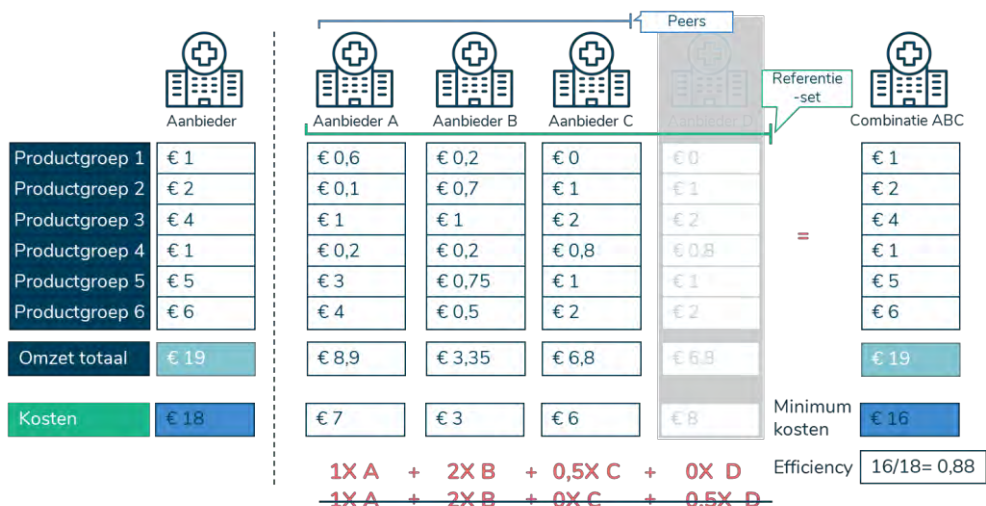
De terugkoppeling van de resultaten moet in ieder geval inzicht bieden in:

- De identiteit van de peers.
- De berekening van de efficiëntiescore.

De berekening van de efficiëntiescore kan grafisch worden weergegeven conform het voorbeeld in hoofdstuk 2. Dit voorbeeld wordt hier herhaald (Figuur 5.2).

¹² Tobit-regressie is een regressietechniek waarbij de uitkomsten van de afhankelijke variabelen begrensd zijn. Dit is voor DEA relevant waar de afhankelijke variabele, de efficiëntiescore, altijd tussen 0 en 1 ligt.

¹³ An Introduction to Statistical Learning, Gareth James et al (2013)



Figuur 5.2 Illustratie van de berekening van de efficiëntiescore in de nulmeting met behulp van de integrale vergelijking.

In de praktijk is het denkbaar om voor elke aanbieder inzichtelijk te maken:

- Hoe de aangeleverde data is 'vertaald' naar de kostenbasis en de productie voor het DEA model. Dit is in lijn met de eerste terugkoppeling van de nulmeting.
- Welke aanbieders peers zijn en (indien dit juridisch mogelijk is en er geen bezwaren zijn bij aanbieders) wat hun waarden zijn voor de kostenbasis en productie.
- Hoe de efficiëntiescore tot stand komt gegeven de peers en hun data: wat zijn de relatieve gewichten van elk van de peers die zijn gebruikt¹⁴ en hoe leiden deze tot de efficiëntiescore?

Op deze manier is voor elke (daarin geïnteresseerde) gebruiker eenvoudiger te volgen hoe de efficiëntiescore tot stand komt. Daarmee wordt DEA waarschijnlijk minder als een 'black box' ervaren.

5.3.2 Visualisatie van de verklarende factoren voor de efficiëntiescore geeft concrete handvatten voor leren en verbeteren.

De resultaten van de verklarende analyses uit paragraaf 5.2 zijn op verschillende wijzen te visualiseren. In deze paragraaf wordt beschreven hoe dit mogelijk kan en worden een aantal voorbeelden gegeven. In alle gevallen geldt dat het voor aanbieders en zorgkantoren wenselijk is om eigen vergelijkingsgroepen in te kunnen stellen. Dit maakt dat een online dashboard waarin aanbieders zelf keuzes maken voor vergelijkingsgroepen de meest voor de hand liggende terugkoppelingsvorm is. Binnen dit online dashboard kan dan de optie worden toegevoegd om de resultaten in de vorm van een standaardrapportage en/of infographic te exporteren. Daarmee wordt tegemoet gekomen aan de diversiteit aan wensen voor de vormen van terugkoppeling (zie hoofdstuk 3).

¹⁴ Het gaat hier om het bekend maken van de gewichten die worden gebruikt voor elke peer – het bepalen van de gewichten zelf vindt plaats met behulp van wiskundige optimalisatietechnieken.

Dynamische vergelijkingsgroepen

Voor leren en verbeteren is het gewenst om ook met andere aanbieders dan de peers te vergelijken. Dit betekent niet dat de efficiëntiescore opnieuw berekend wordt met andere peers, maar wel dat inzicht wordt gegeven hoe andere groepen dan de peers scoren op de verklarende factoren voor de efficiëntiescore.

In de integrale vergelijking staan de peers die volgen uit DEA centraal. De efficiëntiescore geeft immers een relatieve efficiëntie ten opzichte van deze peers. Peers zijn de aanbieders die gecombineerd dezelfde zorg leveren als een aanbieder maar tegen lagere kosten. Dit zijn niet perse ook de beste instellingen om van te leren op specifieke onderwerpen die leiden tot een hogere efficiëntie. Om te leren en verbeteren zal een aanbieder zich richten op de oorzaken van een lagere efficiëntiescore zoals naar voren komen uit de second stage DEA. Dit kan bijvoorbeeld een relatief hoog ziekteverzuim zijn. Aanbieders die een laag ziekteverzuim hebben zijn dan interessant om van te leren. Dit kunnen de peers zijn maar dat hoeft niet perse. Niet alle aanbieders met een laag ziekteverzuim zijn immers peer en niet alle peers hoeven een laag ziekteverzuim te hebben.

Peers kunnen zowel grote als kleine aanbieders zijn en door het hele land gevestigd zijn. Op het gebied van ziekteverzuim kan een aanbieder het liefst van aanbieders in de eigen regio – met dezelfde arbeidsmarktomstandigheden – leren. Met de mogelijkheid om voor de verklarende factoren voor de efficiëntiescore ook met andere vergelijkingsgroepen dan alleen de peers te vergelijken, wordt tegemoet gekomen aan de wens van de aanbieder in dit voorbeeld.

De volgende vergelijkingsgroepen lijken in ieder geval relevant:

- De peers die volgen uit de integrale vergelijking (met mogelijk onderscheid naar ‘eigen’ peers en alle peers).
- Aanbieders van vergelijkbare omvang.
- Door de instelling of zorgkantoor zelf te kiezen vergelijkingsgroep.
- Landelijk gemiddelde score op verklarende factoren.
- Regionaal gemiddelde score op verklarende factoren.

Voorbeelden voor visualisatie van de resultaten

In bijlage 1 zijn drie voorbeelden uitgewerkt voor de visualisatie van de verklarende factoren voor de efficiëntiescore. Bij elke visualisatie is de doelstelling om de resultaten te presenteren op een manier die voor de gebruiker snel te begrijpen is. De visualisaties worden vormgegeven binnen één online, interactief dashboard. In elk dashboard komen de drie genoemde stappen weer terug voor een aanbieder:

- **Stap 1: Hoe efficiënt ben ik en hoe is dat berekend?** De efficiëntiescore van de gebruiker, de peers en de totstandkoming van de efficiëntiescore conform Figuur 5.2.
- **Stap 2: Op welke onderwerpen die mijn efficiëntie beïnvloeden kan ik verbeteren?** De score op factoren die van invloed zijn op de efficiëntiescore. Dit biedt inzicht in de meest relevante onderwerpen waarop de aanbieder mogelijk kan leren en verbeteren.

- **Stap 3: Van welke aanbieders kan ik mogelijk leren op deze onderwerpen?** De scores van de gekozen vergelijkingsgroep op de factoren die van invloed zijn op de efficiëntie. Hiermee kunnen relevante aanbieders worden geselecteerd om te leren en verbeteren voor een specifiek onderwerp.

5.4 Drie opties om het eerste contact tussen aanbieders voor leren en verbeteren te faciliteren

Second stage DEA (sectie 5.2) en heldere visualisaties (sectie 5.3) geven aanbieders inhoudelijke aanknopingspunten voor leren en verbeteren: van wie en op welke onderwerpen kan ik leren en verbeteren? Voor het daadwerkelijke proces van leren en verbeteren moeten aanbieders vervolgens gericht met elkaar in gesprek gaan. In paragraaf 3.2.2 wordt toegelicht dat helderheid over de totstandkoming van dit contact nodig is. Uit de resultaten van de enquête onder zorgaanbieders lijken drie opties hiervoor denkbaar. Deze werkwijzen worden hieronder toegelicht.

Ongeacht de manier waarop aanbieders met elkaar in gesprek komen, kunnen de gesprekken of bijeenkomsten waarin kennis en ervaring worden uitgewisseld op verschillende manieren worden gevoerd. Het ligt voor de hand de aanbieders zelf te laten bepalen hoe ze dit willen doen. Als medewerkers van aanbieders elkaar al kennen, zal een bijeenkomst anders verlopen dan als dat niet het geval is. Ook kan het zijn dat de ene groep aanbieders behoefte heeft aan (externe) begeleiding en specifieke werkvormen terwijl de andere groep aanbieders zelf tot een vorm komt waarbij het gesprek wordt gevoerd.

Een aandachtspunt hierbij is nog wel dat de tijdshorizon voor het toepassen van DEA in tariefstelling en leren en verbeteren verschilt. Bij toepassing van DEA voor tariefstelling wordt naar verwachting elk jaar het model geschat waardoor de efficiëntiescore en de peers (mogelijk) ook elk jaar wijzigen. Leren en verbeteren is echter vaak een meerjarig proces. Het kost tijd om, nadat second stage DEA inzicht geeft in wat peers beter doen, vervolgens ook vast te stellen of en hoe van deze peers te leren is en het daarna toe te passen in de eigen organisatie om te verbeteren.

Optie 1: aanbieders nemen zelf contact op met elkaar

Aanbieders kunnen aan de hand van de resultaten van de integrale vergelijking zelf bepalen welke partijen relevant zijn om van te leren en verbeteren. Dit kunnen de peers zijn maar ook andere aanbieders die goed scoren op verklarende factoren waar een aanbieder zich op wil verbeteren. Het is daarvoor noodzakelijk contactgegevens van de aanbieders met elkaar uit te wisselen – bijvoorbeeld door een contactpersoon als eerste aanspreekpunt per aanbieder op te nemen in het online dashboard. In de nadere uitwerking moet bepaald worden wie toegang heeft tot deze contactgegevens en of contactgegevens pas worden uitgewisseld als beide organisaties daarmee akkoord zijn. Dit laatste is een optie als aanbieders het niet wenselijk vinden dat andere aanbieders apart van elkaar contact opnemen met dezelfde aanbieder die een peer is of het beste scoort op een bepaalde factor.

Optie 2: brancheverenigingen faciliteren het contact tussen de partijen

Brancheverenigingen zoals ActiZ kunnen het contact tussen aanbieders faciliteren en daarbij mogelijk ook (een deel van) het proces faciliteren. Hierbij kan mogelijk aansluiting worden gezocht bij reeds bestaande lerende netwerken. Ook kunnen brancheverenigingen partijen uit verschillende regio's waarschijnlijk laagdrempelig met elkaar in contact brengen en eventueel bijeenkomsten faciliteren. Een voordeel hiervan is dat het niet nodig is veel extra contactgegevens uit te wisselen binnen de integrale vergelijking.

Optie 3: zorgkantoren faciliteren het contact als onderdeel van de zorginkoopdialoog

Zorgkantoren kunnen als onderdeel van de zorginkoopdialoog ook stimuleren dat aanbieders van elkaar leren en verbeteren. Concreet kunnen zorgkantoren mogelijk het proces van leren en verbeteren faciliteren door relevante peers en aanbieders met elkaar in contact te brengen en bijeenkomsten te faciliteren. De mogelijkheden die zorgkantoren hiervoor hebben hangen deels samen met de keuzes die nog worden gemaakt op gebied van tariefstelling. Een open vraag is hierbij in hoeverre zorgkantoren dan alleen contact met peers in de eigen zorgkantorregio's kunnen faciliteren terwijl peers voor een aanbieder door het hele land heen gevestigd kunnen zijn.

6 Aanpassing van de data en het model is nodig om de doorontwikkeling te realiseren

Het realiseren van de omstandigheden voor de in hoofdstuk vijf geschetste doorontwikkeling vergt veranderingen in (de data voor) het DEA model en mogelijk ook in de toepassing ervan. Veranderingen in (de data voor) het model verbeteren naar verwachting de gemeten efficiëntieverschillen. Dit kan door het gebruik van betere kostendata met (veel) minder aannames, meenemen van kapitaallasten, andere indeling van en weging binnen de productgroepen en het gebruik van betere definities van kwaliteitsverschillen als uitsluitingscriterium voor peers. Deze wijzigingen verhogen wel de administratieve en uitvoeringslasten. Een eventuele vervolgmeting kan uitwijzen of dit in de praktijk mogelijk en haalbaar is.

In hoofdstuk vijf was het uitgangspunt voor second stage DEA dat DEA de efficiëntieverschillen tussen aanbieders goed meet. Tijdens dit onderzoek bleek dat een meerderheid van de geconsulteerde zorgaanbieders en zorgkantoren niet overtuigd is dat dit voor de nulmeting het geval is (zie o.a. paragraaf 4.4). Dit vormt een belemmering voor het draagvlak en daarmee voor leren en verbeteren en de zorginkoopdialoog. Hoewel wij voor dit onderzoek geen eigen analyses hebben gedaan op basis van de data van de nulmeting, delen wij de twijfels op basis van de beschikbare informatie of efficiëntieverschillen in de nulmeting goed genoeg gemeten zijn. Ook de NZa geeft in haar adviesrapport¹⁵ aan dat de data en het model in de nulmeting verbeterd en uitgebreid moeten worden.

Dit hoofdstuk schetst mogelijke aanpassingen aan het DEA-model en de gebruikte data om efficiëntieverschillen beter te meten. Een eventuele vervolgmeting inclusief een aangepaste data-uitvraag en second stage DEA kan bepalen of deze voorgestelde aanpassingen mogelijk zijn en ook betere resultaten opleveren.

Voor de doorontwikkeling zijn parallel aan dit onderzoek ook deeltrajecten die zich richten op de dataverzameling, het model en de vertaling naar tarieven. De keuze hoe om te gaan met de geschetste opties in dit hoofdstuk hangt daarom af van de uitkomst van alle trajecten in de

¹⁵ NZa 2020: Advies Integrale vergelijking verpleeghuiszorg – o.a. pagina 20 en 23

doorontwikkeling en op termijn ook van de resultaten van een eventuele vervolgmeting door de NZa.

6.1 Door de gebruikte data te verbeteren kunnen efficiëntieverschillen mogelijk beter gemeten worden

Een eventuele vervolgmeting kan de efficiëntieverschillen beter meten en verklaren door de onderliggende data te verbeteren. De gebruikte data kan ten opzichte van de nulmeting verbeterd worden door:

- meer actuele personeelsdata mee te nemen
- kapitaallasten mee te nemen
- de productgroepen beter in te delen
- kwaliteitsverschillen beter te definiëren, te meten en mee te nemen.

6.1.1 De gebruikte kostendata worden beter door actuelere data voor personeel te gebruiken

Door recentere kostendata te gebruiken zonder, of veel minder, aannames over personeelskosten, worden de kosten beter gemeten. Het vertrouwen in en draagvlak voor de resultaten wordt hierdoor naar verwachting groter.

In de nulmeting zijn noodgedwongen aannames gemaakt over de kosten die (moeten) worden gemaakt om te voldoen aan het kwaliteitskader. Gemiddeld zijn de personeelskosten per aanbieder met 20% verhoogd. Aanbieders waarvoor dit percentage (fors) hoger lag dan 20% vinden de gebruikte kostenniveaus vaak onrealistisch of arbitrair en daarmee de resultaten van het model ook. Aanbieders betwijfelen hiernaast of het benodigde personeel voor het behalen van het kwaliteitskader er op termijn zal zijn, waardoor het geschatte kostenniveau (en dus ook de norm van het kwaliteitskader) in de praktijk nooit kan worden bereikt. Door kostendata te verzamelen na de ingroeperiode van het kwaliteitskader, zijn deze aannames mogelijk niet meer nodig. Dit hangt in de praktijk af van de mate waarin alle aanbieders in staat zijn om de personeelsnormen van het kwaliteitskader te halen (waarvan de haalbaarheid wordt betwijfeld) of het eventueel loslaten of aanpassen van gestelde normen waardoor deze aannames niet meer hoeven te worden gemaakt.

6.1.2 De gebruikte kostendata worden beter door kapitaallasten mee te nemen

Het meenemen van kapitaallasten naast personeelslasten en kosten voor materieel, geeft een beter inzicht in efficiëntieverschillen. Weglaten van kapitaallasten vertekent mogelijk de gemeten efficiëntie. Investerings in vastgoed, IT of e-health kunnen namelijk leiden tot hogere productiviteit van het personeel (omdat looproutes in een gebouw korter zijn, er minder tijd aan administratie kwijt is, of er minder overheadkosten zijn, etc.). Gegeven een bepaalde omvang van de productie liggen de personeelslasten daardoor lager voor aanbieders die recent hebben geïnvesteerd maar hun kapitaallasten zijn wel hoger. Omdat het model alleen personeelslasten en materieel meeneemt, onderschat het model de kosten van deze aanbieder en daarmee de efficiëntie. Omgekeerd is het denkbaar dat de efficiëntie wordt onderschat voor instellingen met

relatief lage kapitaallasten. Per output zijn de personeelslasten mogelijk hoger dan gemiddeld terwijl dat wordt gecompenseerd door lagere kapitaallasten.

In de nulmeting zijn kapitaallasten (vastgoed en inventaris) niet meegenomen. Een belangrijke reden hiervoor is het gebrek aan goede gegevens. De NZa geeft in haar adviesrapport aan dat de geboekte kosten niet (altijd) representatief zijn voor de structurele kosten¹⁶. Het meenemen van kapitaallasten vergt het verzamelen van aanvullende gegevens die volgens nog nader te bepalen 'accounting rules' aangeleverd moeten worden. Om de administratieve lasten voor de nulmeting te beperken, is door de NZa besloten om dit niet te doen.

De NZa geeft in haar adviesrapport¹⁷ aan dat om de hierboven genoemde redenen op langere termijn de kapitaallasten wel zouden moeten worden meegenomen. Het ligt dan ook voor de hand om te onderzoeken of dit bij mogelijke vervolgmeting te realiseren is.

6.1.3 De indeling van en weging binnen productgroepen kan mogelijk beter

De indeling van en weging binnen productgroepen kan mogelijk beter. In de nulmeting wordt de geleverde zorg binnen het kwaliteitskader in zes productgroepen ingedeeld (op basis van ZZP's/VPT's, Tabel 2.1). De weging van de productie binnen productgroepen wordt gedaan op basis van de integrale tarieven voor ZZP's/VPT's.

Bij de indeling in zes productgroepen worden ZZP/VPT 7 tot en met 10 in een productgroep samengenomen. De vraag is of de variatie tussen de ZZP/VPT 7 tot en met 10 niet te groot is om dit als een productgroep te zien. Als dat zo is vermindert de kwaliteit van de gemeten efficiëntieverschillen. In deze groep valt nu zowel palliatieve zorg als intramurale intensieve zorg en begeleiding voor mensen met bijvoorbeeld een ernstige mate van dementie in combinatie met gedragsproblemen. Als de weging van de verschillende ZZP's binnen deze productgroep goed is kan de totale productie in de productgroep representatief zijn. Maar voor individuele aanbieders kunnen er peers worden gevonden die feitelijk heel andere ZZP's leveren dan de aanbieder zelf. Het aandeel van deze zorgtypen in de totale hoeveelheid geleverde zorg in Nederland is weliswaar klein waardoor de impact op de uitkomsten mogelijk beperkt is. In een mogelijke vervolgmeting kan empirisch getest worden of het wel of niet beter is om meer productgroepen te gebruiken.

Bij de weging binnen de productgroepen is het de vraag of in voldoende mate rekening wordt gehouden in het verschil tussen de leveringsvorm (intramuraal of VPT). Door het gebruik van de integrale kostprijs wegen de ZZP's in de output veel zwaarder dan de VPT's, terwijl bij de input alleen de zorgkosten worden meegenomen. De NZa heeft in de doorontwikkeling reeds onderzoek gedaan naar het verbeteren van deze weging door het vervangen van de integrale kostprijs voor alleen het zorgdeel van het ZZP/VPT.

Een mogelijk gevolg van het verhogen van het aantal productgroepen is dat er meer aanbieders als peers worden aangemerkt en de gemeten efficiëntieverschillen binnen de sector kleiner worden. Hiervoor moet een afweging worden gemaakt. De herkenbaarheid van de resultaten is voor

¹⁶ NZa 2020: Advies Integrale vergelijking verpleeghuiszorg – pagina 19

¹⁷ NZa 2020: Advies Integrale vergelijking verpleeghuiszorg – pagina 57

aanbieders en zorgkantoren mogelijk groter omdat de productcategorieën aansluiten bij de huidige ZZP/VPT structuur. Voor deze aanpassing aan het model is het niet nodig om extra data te verzamelen.

6.1.4 Kwaliteitsverschillen kunnen beter gedefinieerd, gemeten en meegenomen worden

Het beter definiëren en meenemen van kwaliteitsverschillen verkleint de kans dat de gemeten efficiëntieverschillen (deels) verschillen in kwaliteit reflecteren. Niet geobserveerde verschillen in kwaliteit die samenhangen met hogere of lagere kosten leiden tot fouten in de gemeten efficiëntieverschillen (zie ook een CPB memorandum¹⁸ uit 2005 waarin DEA-studies voor de Nederlandse ouderenzorg worden beoordeeld).

Uitgangspunt is dat instellingen alleen als peer worden aangewezen als ze zorg leveren die voldoet aan het kwaliteitskader. Dit is in de nulmeting niet gemeten maar de aanname is gemaakt dat elke instelling zorg conform het kwaliteitskader levert (of gaat leveren) en de hiervoor benodigde kosten zijn bijgeschat. De NZa heeft in haar advies¹⁹ eind 2020 geconcludeerd dat kwaliteit mogelijk op korte termijn - bij gebrek aan alternatieven – gedefinieerd kan worden als : de instelling heeft geen aanwijzing van de IGJ. De gevonden peers in de nulmeting bleken na inventarisatie door de NZa in de periode erna overigens geen dergelijke aanwijzing van de IGJ gehad te hebben.

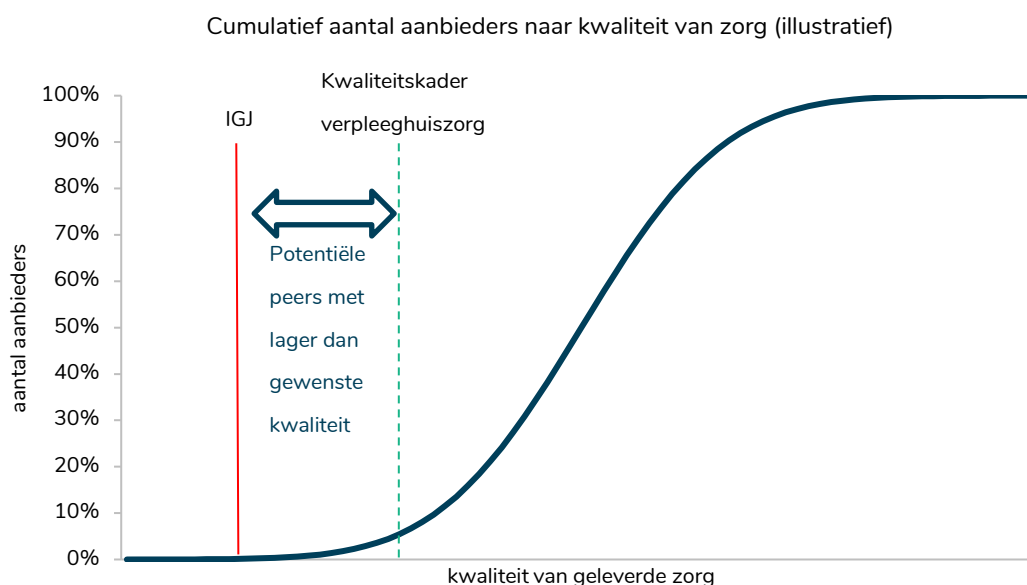
Het is ook met deze definitie in de praktijk echter mogelijk dat aanbieders die lagere kwaliteit leveren dan gewenst toch als peer worden aangewezen. De IGJ toetst de kwaliteit van instellingen namelijk risico-gedreven: op basis van meldingen of signalen dat de kwaliteit onvoldoende is. Als na onderzoek van de IGJ dat inderdaad zo blijkt te zijn, volgt een aanwijzing. Aanbieders waar de kwaliteit zwaar onder de maat is, worden via dit proces hoogstwaarschijnlijk geïdentificeerd en daarmee dus uitgesloten als peer. De IGJ meet echter niet structureel de kwaliteit van de zorg van aanbieders. Daarom is het denkbaar dat aanbieders waarover (nog) geen signalen bij de IGJ binnen zijn gekomen maar die wel lagere kwaliteit leveren dan het kwaliteitskader, toch als peer worden aangewezen (zie Figuur 6.1 ter illustratie).

Het gevolg van het gebruik van wel of geen aanwijzing van de IGJ als mogelijk toekomstig uitsluitingscriterium is dat de kostengrens mogelijk gebaseerd wordt op aanbieders die niet altijd voldoen aan het kwaliteitskader. Als lagere kwaliteit van zorg tot lagere kosten leidt, kan via lagere tarieven een risico op een ‘race to the bottom’ ontstaan op gebied van kwaliteit.

Kwaliteit dient volgens verschillende geïnterviewde aanbieders en zorgkantoren dan ook beter te worden gemeten en vooral beter te worden meegenomen in het model dan nu het geval is. Dit vereist wel dat kwaliteit(-verschillen) van de verpleeghuiszorg structureel worden gemeten. Dit is een taak voor het hele veld maar vindt nog beperkt plaats. Daarom is er op dit moment weinig inzicht in kwaliteitsverschillen op landelijk niveau.

¹⁸ CPB memorandum 121, 2005: Doelmatigheidswinsten in de AWBZ

¹⁹ NZa 2020: Advies Integrale vergelijking verpleeghuiszorg – pagina 30 en verder



Figuur 6.1: Door uit te gaan van wel of geen IGJ aanwijzing als kwaliteitscriterium, worden mogelijk peers geselecteerd die niet voldoen aan het kwaliteitskader verpleeghuiszorg - illustratief voorbeeld

Dat is voor de integrale vergelijking en het DEA model een nadeel omdat hierdoor mogelijk peers worden aangewezen die zorg leveren die niet voldoet aan het maatschappelijk gedefinieerd minimum. In algemene zin is dit een gemis omdat inzicht in kwaliteitsverschillen ook een grote maatschappelijke waarde heeft (om te kiezen voor cliënten, om te sturen vanuit overheid en financiers) en mogelijkheden biedt om ook op dat gebied van elkaar te leren en verbeteren. Dit laatste kan mogelijk ook via de integrale vergelijking plaatsvinden. Bijvoorbeeld door via scores op kwaliteitsindicatoren aanbieders ook zicht te geven op onderlinge kwaliteitsverschillen (naast het uitsluiten van aanbieders als potentiële peer als de kwaliteit van zorg onvoldoende is).

6.2 Methodologische aanpassingen aan het model kunnen draagvlak voor toepassing en herkenbaarheid vergroten

We zien naast verbeteringen aan de data ook mogelijke methodologische aanpassingen om het draagvlak voor toepassing te vergroten.

6.2.1 De resultaten van de voorgestelde 'second stage DEA' kunnen worden gebruikt om de efficiëntiescore te corrigeren voor niet-beïnvloedbare factoren

DEA berekent een efficiëntiescore die per aanbieder het verschil in efficiëntie t.o.v. de peers uitdrukt in een getal (zie hoofdstuk 2). Een deel van dat verschil kan komen door factoren waar een instelling geen invloed op heeft. Met second stage DEA kan de efficiëntiescore worden gecorrigeerd voor niet-beïnvloedbare factoren (zie hoofdstuk 5). Deze gecorrigeerde score zou vervolgens input voor toepassingen in tarieven kunnen zijn.

6.2.2 Gebruik van geclusterde referentiesets voor DEA op basis van herkenbare kenmerken vergroot herkenbaarheid peers maar verkleint gemeten efficiëntieverschillen

De herkenbaarheid van peers voor aanbieders wordt beter door kleinere referentiesets (de groep waarbinnen mogelijke peers worden gezocht) op basis van kenmerken als omvang en/of regio te gebruiken. Hierdoor zijn de peers van bijvoorbeeld grote aanbieders ook altijd grote aanbieders, komen de peers van een aanbieder uit Noord-Nederland ook altijd uit Noord-Nederland, of komen de peers van een aanbieder uit een G4 gemeente ook uit een G4 gemeente. Hierdoor zijn peers herkenbaarder voor aanbieders wat bijdraagt aan het begrip van en voor de methode.

De mogelijke consequentie hiervan is dat de gemeten efficiëntieverschillen kleiner en minder nauwkeurig worden. Het kleiner maken van de referentieset zorgt er voor dat de in DEA berekende minimale kostengrens op minder aanbieders wordt gebaseerd. De variatie tussen aanbieders is binnen een kleinere referentieset hoogstwaarschijnlijk kleiner, waardoor de gemeten efficiëntieverschillen naar verwachting ook kleiner en minder nauwkeurig worden wat ook doorwerkt bij toepassing in de tarieven.

De keuze om deze optie toe te passen is daarmee een afruil tussen grotere herkenbaarheid en begrip van de methode enerzijds en een lagere nauwkeurigheid en via de tarieven mogelijk ook een lagere impact op de doelmatigheid anderzijds. Een empirische test kan inzichtelijk maken hoe groot deze afruil is. Dit kan door voor een groep aanbieders de efficiëntiescores te berekenen op basis van een referentieset met en zonder bepaalde restricties en vervolgens te bepalen hoe groot de verschillen in de praktijk zijn. Het ligt wel voor de hand om eerst te bekijken of met de in hoofdstuk 5 en hierboven geschetste stappen de herkenbaarheid van peers en begrip van en voor de methode voldoende kan worden vergroot. Dat zijn namelijk stappen die geen impact hebben op de resultaten en de daaruit volgende tarieven.

6.3 De voorgestelde aanpassingen verhogen administratieve en uitvoeringslasten en vergen forse inspanningen van partijen

Het uitvoeren van de in paragraaf 6.1 en 6.2 genoemde aanpassingen aan de methodologie leidt naar verwachting tot extra administratieve lasten voor aanbieders en uitvoeringslasten voor de NZa.

Aanbieders zullen meer data moeten aanleveren dan voor de nulmeting. De impact hiervan op de administratieve lasten zal per aanbieder verschillen omdat dit mede afhankelijk is van de mate waarin aanbieders deze data zelf al verzamelen voor de interne bedrijfsvoering of verantwoording. Mogelijk worden op termijn alle gewenste data voor de integrale vergelijking via KIK-V aangeleverd. Als daarbij de gewenste dataset (grotendeels) gelijk blijft voor een langere periode is de toename van administratieve lasten voor aanbieders naar verwachting vooral tijdelijk.

De uitvoeringslasten voor de NZa nemen ook toe. Ten opzichte van het in de nulmeting gehanteerde DEA model moet immers het model worden uitgebreid met kapitaallasten, andere productgroepen en mogelijk aanvullende gegevens over kwaliteitsverschillen.

Het realiseren van de in hoofdstuk 5 en in paragraaf 6.2 en 6.3 beschreven stappen vergt een forse inspanning van alle betrokken partijen waarbij het op voorhand onduidelijk is of het gewenste resultaat bereikt wordt. Een mogelijke vervolgmeting kan uitwijzen of de beschreven verbeteringen en uitbreidingen in de praktijk mogelijk zijn.

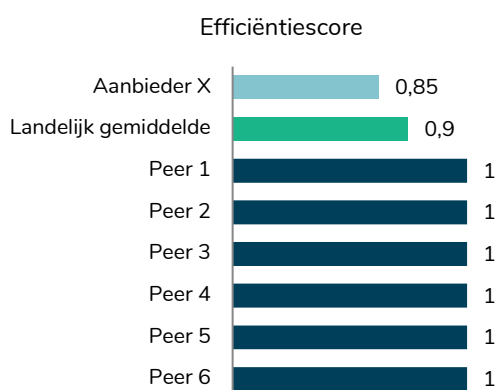
Bijlage 1: Voorbeeld visualisaties

Er zijn drie voorbeelden uitgewerkt voor de visualisaties voor de terugkoppeling van de resultaten (inclusief second stage DEA) van de integrale vergelijking. In elk van de voorbeelden komen dezelfde drie elementen terug die gezamenlijk de drie vragen die in het begin van hoofdstuk 5 staan beantwoorden:

- 1 De efficiëntiescore van de aanbieder en de vergelijkingsgroep. Deze visualisatie beantwoordt de vraag: Hoe efficiënt ben ik in vergelijking met andere instellingen (en hoe is dat berekend)? Deze resultaten volgen uit de DEA-methode zoals toegepast in de nulmeting.
- 2 De score op factoren die de efficiëntie beïnvloeden. Deze visualisatie beantwoordt de vraag: Op welke onderwerpen die de efficiëntie beïnvloeden kan mijn instelling verbeteren? Deze resultaten volgen uit second stage DEA. Second stage DEA vormt hiermee de brug van de DEA-methode naar een meer 'klassieke' benchmark.
- 3 De score op de factoren (waarop verbeterd kan worden) in vergelijking met andere aanbieders. Dit beantwoordt de vraag: Van welke aanbieders kan ik leren op de onderwerpen waar mijn organisatie zich verbeteren? Deze resultaten zijn vergelijkbaar met die van een 'klassieke' benchmark.

Stap 1: Efficiëntiescore van de aanbieder en de vergelijkingsgroep

Het eerste onderdeel van de visualisatie bestaat uit een figuur die de efficiëntiescore van de eigen instelling en de scores van de gekozen vergelijkingsgroep presenteert. Deze figuur is identiek in elk van de drie voorbeelden. Een staafgrafiek is het meest geschikt om deze verschillen te tonen (Figuur B1). De staafgrafiek toont de score van de aanbieder (hier aanbieder X met een score 0,85), de score van het landelijke gemiddelde (hier 0,9) en de peers (met naam - logischerwijs hebben deze allemaal score 1). Aanbieder X kan zelf andere vergelijkingsgroepen toevoegen.



Figuur B1: Voorbeeld illustratie van de eigen efficiëntiescore

Stap 2: Een overzicht van de verwachte impact van verklarende factoren voor de efficiëntiescore

Het tweede onderdeel van de visualisatie presenteert de resultaten van second stage DEA: op welke factoren die van invloed zijn op de efficiëntiescore scoort een aanbieder boven- of ondergemiddeld. Voor dit onderdeel zijn drie voorbeelden uitgewerkt.

Elk voorbeeld laat de scores zien op factoren die van invloed zijn op de efficiëntiescore en die beïnvloedbaar zijn. De drie voorbeelden verschillen in de gebruikte figuur en in de eenheid waarin de factoren wordt uitgedrukt.

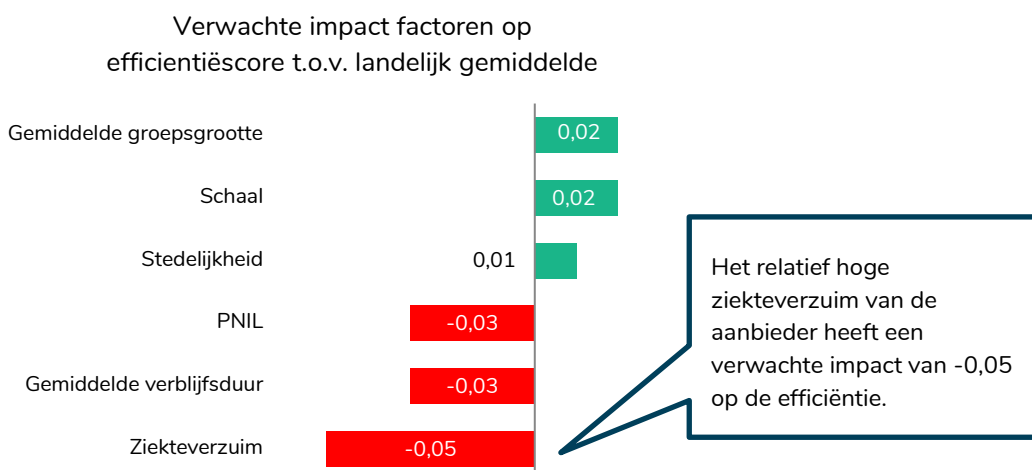
Voorbeeld illustratie 1: Staafgrafiek

De effecten van de verklarende factoren op de efficiëntiescore kunnen gevisualiseerd worden in een staafgrafiek (Figuur B2). De factoren zijn hier uitgedrukt in impact op de efficiëntiescore. Dit wordt berekend door:

- Het verschil tussen aanbieder en het landelijk gemiddeld op een bepaald onderwerp te berekenen. Bijvoorbeeld het ziekteverzuim van een aanbieder is 8% en landelijk is dit 3% - het verschil is 5%-punt.
- Dit verschil wordt vermenigvuldigd met de coëfficiënt uit second stage DEA die de relatie van die factor met efficiëntie score vast heeft gesteld. Uit de second stage DEA analyse kan bijvoorbeeld blijken dat 1%-punt samenhangt met een efficiëntiescore die 0,01 punt lager is.
- De uitkomst van de berekening laat zien hoe groot het effect is van het hoger of lager scores op verklarende factoren op de efficiëntiescore. Voor de aanbieder in kwestie leidt de uitkomst van het ziekteverzuim tot een score van -0,05 ten opzichte van het landelijk gemiddelde (onderste balk in Figuur B2).

Figuur B2 laat zo voor iedere verklarende factor het verschil zien in de efficiëntiescore tussen de huidige score en het landelijk gemiddelde. Onderwerpen waarop een aanbieder beter scoort zijn positief, onderwerpen waarop een aanbieder minder scoort zijn negatief.

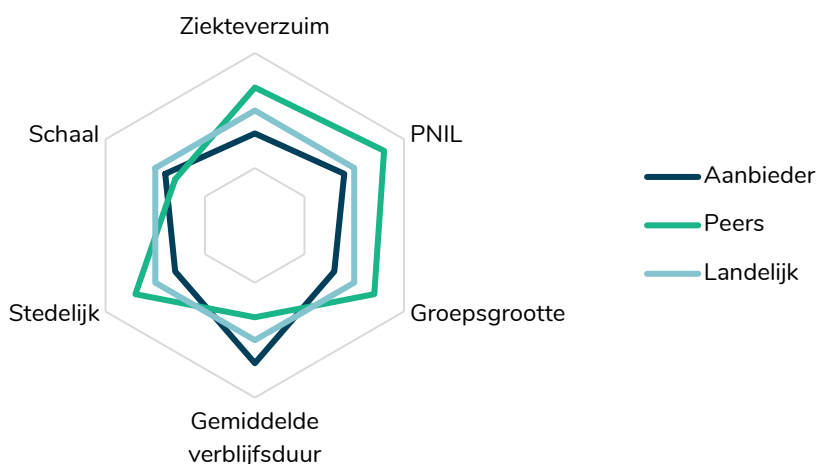
Het voordeel van deze optie is dat direct duidelijk wordt op welk onderwerp leren en verbeteren de grootste impact heeft op de efficiëntiescore omdat er rekening wordt gehouden met zowel het verschil in de factor zelf als de impact die dat heeft op de efficiëntiescore. Een (ander) voordeel van deze werkwijze is dat de figuur (staafgrafiek) herkenbaar is en eenvoudig te begrijpen is. Een mogelijk nadeel kan zijn dat de eenheid (verschillen in efficiëntiescore t.o.v. landelijk gemiddelde) niet direct herkenbaar is en wat uitleg vergt.



Figuur B2 Door middel van een staafgrafiek kan het effect van de verklarende factoren op de efficiëntiescore in één overzicht worden weergegeven.

Voorbeeld illustratie 2: Spin

In deze werkwijze worden de verklarende factoren voor de DEA score in één overzicht weergegeven (Figuur B3). Per verklarende factor kan worden getoond of de aanbieder relatief goed of minder goed scoort ten opzichte van zijn peers en/of het landelijk gemiddelde. Deze scores kunnen bijvoorbeeld gestandaardiseerd worden met Z-scores. Hogere waarden verder naar buiten duiden op een betere, hogere efficiëntiescore. Lagere waarden verder naar binnen wijzen op een slechtere, lagere efficiëntiescore.



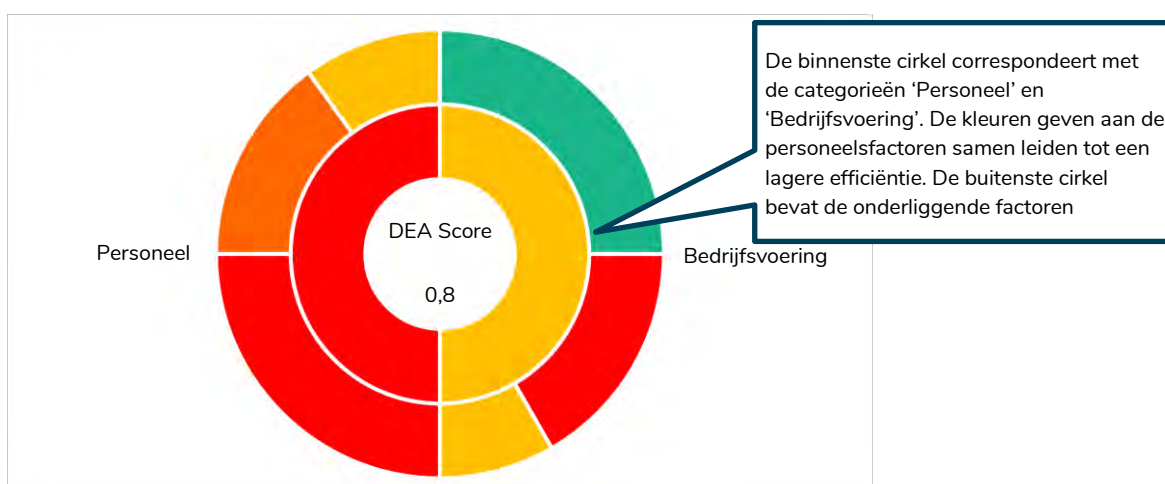
Figuur B3 Door middel van een spingrafiek kunnen de scores van verschillend vergelijkingsgroepen tegelijkertijd worden weergegeven.

Een voordeel van deze werkwijze is dat de verklarende factoren in goed in één overzicht kunnen worden weergegeven en dat er verschillende vergelijkingsgroepen tegelijk kunnen worden gevisualiseerd. Een belangrijk nadeel is dat met deze optie het verwachte effect op de efficiëntiescore niet wordt weergegeven. Het verbeteren op een factor met een klein verschil ten opzichte van anderen maar die zwaar bijdraagt aan de efficiëntiescore kan meer impact hebben op

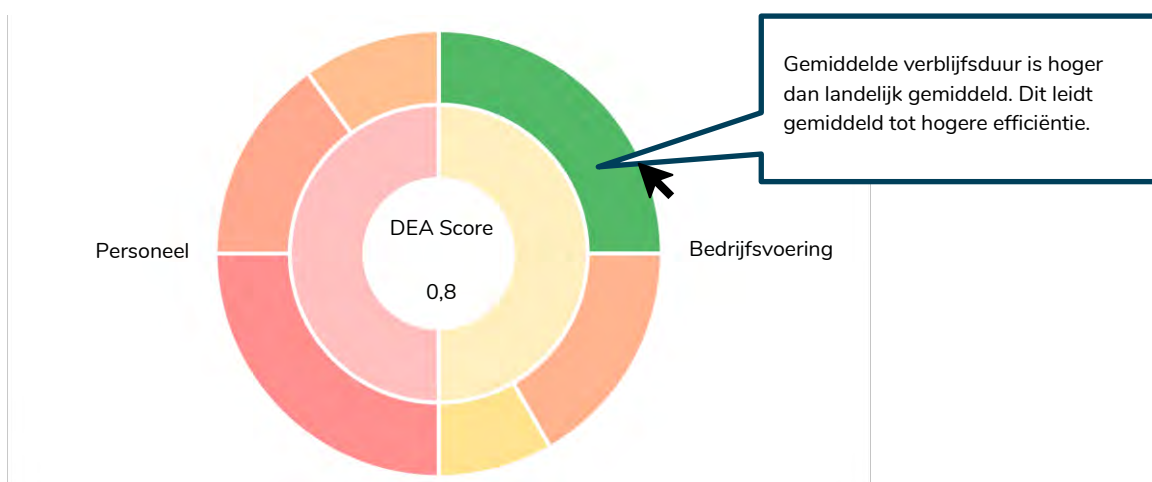
dan verbeteren op een factor waarop je sterk afwijkt maar waar de bijdrage aan de efficiëntiescore beperkt is. Dat wordt in deze figuur niet direct duidelijk.

Voorbeeld illustratie 3: Sunburst

In dit voorbeeld wordt de gebruikers een interactief dashboard getoond zoals in Figuur B4 en B5. In deze werkwijze wordt de gebruiker in één oogopslag overzicht gegeven van zijn DEA score en de factoren die dit beïnvloeden door middel van een Sunburst grafiek. Elk vak hoort bij een verklarende factor. De kleur geeft aan of die de score positief of negatief beïnvloedt. Door op de vakken te klikken kan de gebruiker steeds gedetailleerdere informatie over dat onderdeel zoeken. SiRM heeft een dergelijke tool toegepast voor de benchmark "[Kwaliteitsindicatoren ziekenhuizen](#)". Een voordeel van deze werkwijze is dat deze het meest interactief is, de gebruiker kan zelf doorklikken naar voor hem relevante inzichten. Een nadeel is dat het voor gebruikers mogelijk wat complexer is om in één keer te snappen.



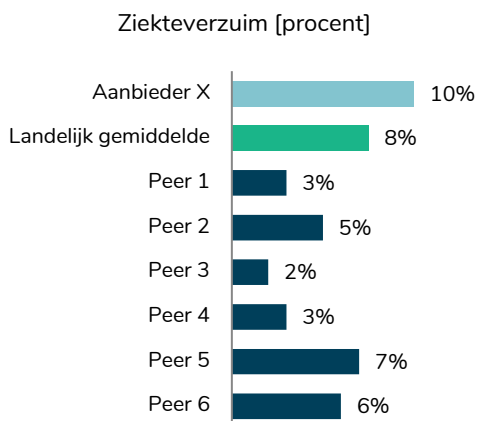
Figuur B4 Door middel van een sunburst grafiek wordt het effect van de verklarende factoren op de efficiëntiescore door middel van kleuren gevisualiseerd. Ieder vak in de buitenste ring correspondeert met een verklarende factor. De vakken aan de linkerkant van de grafiek horen bij personeelsfactoren, de vakken aan de rechterkant horen bij factoren die raken aan de bedrijfsvoering.



Figuur B5 Door in de sunburst grafiek op de gekleurde vakken te klikken krijgt de gebruiker een toelichting te zien bij de verklarende factor en het effect op zijn efficiëntiescore.

Stap 3: Score van de verklarende factoren in vergelijking met andere aanbieders

Het derde onderdeel van de visualisatie geeft inzicht in hoe andere aanbieders scoren op de door de gebruiker geselecteerde verklarende factor. Door te kijken hoe je als aanbieder scoort ten opzichte van andere op een onderwerp kunnen relevante partijen om mee te leren en verbeteren worden geselecteerd. Dit derde onderdeel is vergelijkbaar met een 'reguliere' benchmark.



Figuur B6 Voorbeeld illustratie van de benchmark van de verklarende factor ziekteverzuim.

Een concreet voorbeeld is (weer) ziekteverzuim. Als een aanbieder in het tweede onderdeel ziet dat hij hoog scoort op ziekteverzuim en dit zijn efficiëntiescore negatief beïnvloed kan hij in dit onderdeel hoe andere aanbieders scoren hierop. Via deze vergelijking wordt duidelijk van welke aanbieders hij mogelijk kan leren. Een staafgrafiek is ook voor visualisatie dit onderdeel zeer geschikt (Figuur B6) waarin een aanbieder zelf kan kiezen met welke vergelijkingsgroep en type aanbieders hij zich wil vergelijken. In dit voorbeeld is dat het landelijk gemiddelde en de peers maar dit kan bijvoorbeeld ook rij met aanbieders in de regio zijn.