

## Regionale effecten in de risicoverevening





Amsterdam, mei 2018  
In opdracht van het ministerie van VWS

## Regionale effecten in de risicoverevening

WOR 910

Marloes Lammers  
Sandra Vriend  
Caren Tempelman  
Roderik Ponds (Atlas voor Gemeenten)



seo economisch onderzoek

“De wetenschap dat het goed is”

*SEO Economisch Onderzoek doet onafhankelijk toegepast onderzoek in opdracht van overheid en bedrijfsleven. Ons onderzoek helpt onze opdrachtgevers bij het nemen van beslissingen. SEO Economisch Onderzoek is gelieerd aan de Universiteit van Amsterdam. Dat geeft ons zicht op de nieuwste wetenschappelijke methoden. We hebben geen winst oogmerk en investeren continu in het intellectueel kapitaal van de medewerkers via promotietrajecten, het uitbrengen van wetenschappelijke publicaties, kennisnetwerken en congresbezoek.*

SEO-rapport nr. 2018-45

ISBN 978-90-6733-922-3

**Informatie & Disclaimer**

SEO Economisch Onderzoek heeft op de verkregen informatie en data geen onderzoek uitgevoerd dat het karakter draagt van een accountantscontrole of due diligence. SEO is niet verantwoordelijk voor fouten of omissies in de verkregen informatie en data.

**Copyright © 2018 SEO Amsterdam.** Alle rechten voorbehouden. Het is geoorloofd gegevens uit dit rapport te gebruiken in artikelen, onderzoeken en collegesyllabi, mits daarbij de bron duidelijk en nauwkeurig wordt vermeld. Gegevens uit dit rapport mogen niet voor commerciële doeleinden gebruikt worden zonder voorafgaande toestemming van de auteur(s). Toestemming kan worden verkregen via [secretariaat@seo.nl](mailto:secretariaat@seo.nl)

## Samenvatting

*Het huidige regiocriterium in het vereveningsmodel voor de somatische zorg is samengesteld op basis van een meerstappenplan dat veel tijd kost tijdens de jaarlijkse update van het vereveningsmodel. Een model waarin de regiokenmerken direct worden meegenomen in de schatting van het risicovereveningsmodel (en vervolgens tien regioclusters gevormd worden) is eenvoudiger, transparanter en heeft een vergelijkbare verevenende werking.*

Het ministerie van VWS wil alternatieven voor het huidige regiocriterium in het somatische risicovereveningsmodel verkennen. Het huidige regiocriterium is samengesteld op basis van een meerstappenplan dat veel tijd kost tijdens de jaarlijkse update van het vereveningsmodel. Bovendien is de aanvullende verevenende werking van de regioclusters beperkt. Om de verevening te verbeteren en de schattingsmethode te vereenvoudigen rekent SEO Economisch Onderzoek in dit onderzoek een aantal modellen door waarin op verschillende manieren rekening gehouden wordt met regiokenmerken in de risicoverevening. Tabel S. 1 vat de geschatte modellen samen.

**Tabel S. 1** Geschatte modellen

Naam	Omschrijving	Kenmerken
M00	OT 2018 exclusief regiokenmerk	Geen regiokenmerken
<b>M0</b>	Regiomodel OT 2018	Normkosten somatische zorg, niet-westerse migratieachtergrond, regionaal klantenpotentieel, luchtvervuiling (PM10), huurwoningen, OAD
<b>M0*</b>	Regiomodel met aangepaste set regiokenmerken	Constance, niet-westerse migratieachtergrond, westerse migratieachtergrond, ziekteverzuim, SES laag inkomen
<b>M1</b>	Loslaten indeling 10 gelijke groepen	Zelfde als M0
<b>M1*</b>	Loslaten indeling 10 gelijke groepen	Zelfde als M0*
<b>M2</b>	Clustering van kenmerken, aan de hand van meerkosten	Zelfde als M0, excl. normkosten
<b>M2*</b>	Clustering van kenmerken, aan de hand van meerkosten	Zelfde als M0*, excl. constante
<b>M3</b>	Regiovariabelen rechtstreeks opnemen	Niet-westerse migratieachtergrond, regionaal klantenpotentieel
<b>M3*</b>	Regiovariabelen rechtstreeks opnemen	Niet-westerse migratieachtergrond, westerse migratieachtergrond, SES laag inkomen
<b>M4</b>	Principale-componentenanalyse	Twee principale componenten geconstrueerd uit niet-westerse migratieachtergrond, regionaal klantenpotentieel, luchtvervuiling (PM10), huurwoningen, OAD
<b>M4*</b>	Principale-componentenanalyse	Twee principale componenten geconstrueerd uit: percentages Marokkanen, Turken, Surinamers, Antillianen, overig niet-westers, MOE-landers, overig westers en zonder migratieachtergrond, percentage speciaal onderwijs, overlast, percentage corporatiewoning, percentages goede, redelijke en slechte woningen, bomen/bos, natuurlijk terrein, geluidsoverlast, benzeen, koolmonoxide, ammoniak, stikstofdioxide, ozon, zwaveldioxide, fijnstof PM2,5, zeer lage, lage, midden en hoge SES en ziekteverzuim
<b>M5</b>	Rechtstreeks opnemen en clusteren	Zelfde als M3
<b>M5*</b>	Rechtstreeks opnemen en clusteren	Zelfde als M3*

Bron: SEO Economisch Onderzoek

Binnen elke technische variant (te herkennen aan het oplopende cijfer) is telkens eerst een model geschat waarin de regiokenmerken zijn opgenomen die op dit moment worden gebruikt in het risicovereveningsmodel voor de somatische zorg. Dit zijn de kenmerken niet-westerse migratieachtergrond, regionaal klantenpotentieel, luchtvervuiling (PM10), percentage huurwoningen en OAD (stedelijkheid). Binnen elke technische variant is vervolgens ook een model geschat waarin een nieuwe set regiokenmerken is opgenomen. Dit zijn de varianten met een ster in Tabel S. 1. Deze nieuwe regiokenmerken zijn geselecteerd op basis van een literatuuronderzoek naar regiokenmerken die een bewezen causaal verband hebben met gezondheid. Het literatuuronderzoek is uitgevoerd door Atlas voor Gemeenten. Alleen indicatoren die niet (sterk) samenhangen met het ZVW zorgaanbod zijn meegenomen.

### Toetsingskader

Alle varianten zijn getoetst aan de hand van de toetsingscriteria zoals bekend in de risicoverevening: verevenende werking, doelmatigheid, beheersbare complexiteit en validiteit en meetbaarheid.

De verevenende werking op individuniveau en zorgverzekeraarsniveau is voor alle modellen ongeveer gelijk. Er zijn wel verschillen op regionaal niveau. Sommige modellen resulteren in grote negatieve resultaten voor de G4, of juist voor de randen van het land (periferie) en scoren daarom slechter op verevenende werking.

Overall presteert het model M5\* het best. Dit model bestaat uit twee stappen:

1. Schatten van risicovereveningsmodel somatische zorg inclusief drie regiokenmerken op pc4-niveau: percentage verzekerdenjaren met een niet-westerse migratieachtergrond, percentage verzekerden met een westerse migratieachtergrond en percentage lage inkomens (lage SES). Ieder kenmerk is ingedeeld in vier klassen op basis van zoveel mogelijk homogene meerkosten in deze klassen;
2. Op basis van de geschatte coëfficiënten op de regiokenmerken in stap 1 de bijdrage van deze regiokenmerken aan de voorspelde zorgkosten bepalen. Op basis van deze voorspelde zorgkosten de verzekerden indelen in tien zoveel mogelijk gelijke regioclusters.

Groot voordeel van een model op individuniveau is dat het risicovereveningsmodel in twee keer geschat wordt, inclusief regiokenmerken. Dit kost minder onderzoektijd dan het huidige achtstapenplan. Een ander voordeel is dat de regio-effecten in de eerste stap transparant zijn: bijvoorbeeld een hoger aandeel personen met een niet-westerse migratieachtergrond leidt tot hogere zorgkosten. Bovendien resulteert het model in één regiokenmerk: de indeling in tien regioclusters. Introductie van extra (regio)kenmerken maakt de verzekerdenraming bewerkelijker en is daardoor ongewenst.

**Tabel S. 2**    **Overzicht afweging toetsingscriteria**

	M0	M0*	M1	M1*	M2	M2*	M3	M3*	M4	M4*	M5	M5*
Verevenende werking	++	++	++	++	+	++	+	++	+	+	+	++
Doelmatigheid	-	+	-	+	-	-	-	+	-	+	-	+
Beheersbare complexiteit	--	--	--	--	-	-	-/+	-/+	-	-	+	+
Validiteit en meetbaarheid	-	++	-	++	-	++	-	++	-	++	-	++

Bron: SEO Economisch Onderzoek

# Inhoud

<b>Samenvatting</b> .....	<b>i</b>
<b>1 Inleiding</b> .....	<b>1</b>
1.1 Onderzoeksdoel.....	1
1.2 Onderzoeksaanpak.....	1
<b>2 Huidige techniek voor samenstellen van het regiocriterium</b> .....	<b>3</b>
<b>3 Omgevingskenmerken en gezondheid</b> .....	<b>7</b>
3.1 Omgevingskenmerken als proxy voor gegevens op individuniveau.....	8
3.2 Omgevingskenmerken die gezondheidsverschillen veroorzaken.....	9
3.3 Nieuw te testen regiokenmerken.....	11
3.4 Validiteit huidige regiokenmerken.....	13
<b>4 Technische varianten regiomodel</b> .....	<b>17</b>
4.1 Uitgangsmoedellen en model met aangepaste set regiokenmerken (M00, M0 en M0*).....	17
4.2 Regiomodel waarbij indeling in 10 gelijke groepen is losgelaten (M1 en M1*).....	18
4.3 Model op basis van clustering van kenmerken, aan de hand van meerkosten (M2 en M2*).....	21
4.4 Model waarin regiovariabelen rechtstreeks zijn opgenomen (M3 en M3*).....	23
4.5 Model op basis van clustering van kenmerken, zonder gebruik van meerkosten (M4 en M4*).....	24
4.6 Model waarin regiovariabelen rechtstreeks zijn opgenomen, vervolgens indeling in regioclusters (M5 en M5*).....	26
4.7 Alternatieve technieken die hier niet verder verkend worden.....	26
<b>5 Meerkosten en modelschattingen</b> .....	<b>29</b>
5.1 Meerkosten en verzekerdenjaren.....	29
5.2 Normbedragen.....	34
5.3 Maatstaven voor verevenende werking.....	36
<b>6 Toetsingskader</b> .....	<b>41</b>
<b>7 Conclusie</b> .....	<b>45</b>
<b>Literatuur</b> .....	<b>47</b>
<b>Bijlage A Kaartjes met residuen</b> .....	<b>49</b>

Bijlage B	Regiomodel met nieuwe kenmerken (model M0*-regio) .....	55
Bijlage C	Meerkosten geselecteerde regiokenmerken bij indeling in 10 gelijke groepen .....	59
Bijlage D	Normbedragen .....	61
Bijlage E	Verevenende werking varianten M0 en M3 .....	71
Bijlage F	Data en databewerking.....	73
Bijlage G	Reproductie regiomodel M0.....	77



# 1 Inleiding

*Wat zijn de mogelijkheden voor verbetering van het regiocriterium in het somatische risicovereveningsmodel? Dit onderzoek presenteert alternatieven die mogelijk de verevenende werking verbeteren en/of minder onderzoekstijd vergen dan de huidige methodiek.*

Het ministerie van VWS wil alternatieven voor het huidige regiocriterium in het somatische risicovereveningsmodel verkennen. Enerzijds gaat het om het verbeteren van de verevenende werking via het regiocriterium, zonder daarbij te compenseren voor zorgaanbodfactoren. Anderzijds gaat het om een eenvoudigere methode voor het samenstellen van het regiocriterium. Momenteel wordt het regiocriterium namelijk samengesteld in een meerstappenplan, dat veel onderzoekstijd vergt tijdens de jaarlijkse update van het vereveningsmodel. Het ministerie van VWS vraagt zich daarom af of een eenvoudigere aanpak mogelijk is, die daarnaast ook de verevening verbetert.

## 1.1 Onderzoeksdoel

Doel van dit onderzoek is nagaan of een betere modellering van de regionale effecten in de verevening mogelijk is. Het ministerie van VWS stelt de volgende onderzoeksdoelen:

1. Onderzoeken of – binnen de bestaande methodiek – het loslaten van de 10 gelijke groepen de verevening verbetert. Het gaat hierbij om de indeling (eventueel anders dan in groepen van gelijke aantallen) en om het aantal groepen (eventueel meer of minder dan 10);
2. Bespreking van minimaal drie alternatieve methoden om een regiocriterium samen te stellen. Hierbij wordt uiteengezet wat de voor- en nadelen zijn van de methoden;
3. Doorrekenen en beoordelen van de gepresenteerde modellen.

Randvoorwaarde bij het verkennen van een eenvoudigere aanpak is dat de verevenende werking via de morbiditeitscriteria niet minder wordt en dat bij eventueel gebruik van regiovariabelen geen variabelen worden gebruikt die het zorgaanbod betreffen (bijvoorbeeld het aantal ziekenhuisbedden in de regio).

Het regiocriterium is ook opgenomen in het GGZ-model. Dit onderzoek richt zich op de somatische zorg; mocht blijken dat er een zinvolle alternatieve modellering mogelijk is, dan zal deze in een later stadium wellicht ook kunnen worden toegepast in het GGZ-model.

## 1.2 Onderzoeksaanpak

Dit onderzoek gaat eerst na hoe binnen het huidige regiomodel de verevenende werking via het regiocriterium verbeterd kan worden. Hoofdstuk 2 presenteert de huidige schattingsmethode voor samenstellen van het regiocriterium. Hoofdstuk 3 geeft een literatuuroverzicht van regiofactoren die theoretisch samenhangen met gezondheid en mogelijk de verevenende werking van het regiocriterium kunnen verbeteren. Alleen regiofactoren die niet samenhangen met ZVW-aanbod worden besproken. Deze set van regiofactoren is vervolgens – tezamen met de huidige factoren – getoetst in het huidige regiomodel. Dit resulteert in een voorstel voor een nieuwe set regiofactoren.

Het onderzoek presenteert vervolgens een aantal alternatieve methoden om het regiocriterium samen te stellen en de belangrijkste voor- en nadelen van deze methodieken (Hoofdstuk 4). Deze alternatieve methoden zijn zowel geschat op de huidige set regiofactoren (voor de verdeling van ZVW-gelden in 2018) als op de nieuwe set regiofactoren. De schattingsresultaten en bijbehorende vereveningsmaatstaven van de verschillende modellen staan in Hoofdstuk 5. De modellen zijn geschat op het OT2018-bestand met gegevens uit 2015. Hoofdstuk 6 beoordeelt de modellen aan de hand van de criteria uit het toetsingskader en Hoofdstuk 7 concludeert.

## 2 Huidige techniek voor samenstellen van het regiocriterium

*Het huidige regiocriterium wordt samengesteld door eerst het risicovereveningsmodel te schatten exclusief regiofactoren, en vervolgens de residuen (werkelijke kosten minus voorspelde kosten) van deze regressie te relateren aan regiokenmerken. Vervolgens wordt Nederland ingedeeld in 10 regio's bestaande uit postcodegebieden met een vergelijkbare waarde van het voorspelde residu.*

Het regiocriterium bestaat uit 10 regio's (dummyvariabelen) die elk een (vrijwel) gelijk aantal verzekerdenjaren omvatten. De regio's zijn samengesteld uit gebieden verdeeld over heel Nederland. De indeling in de 10 regio's wordt jaarlijks geactualiseerd, dit gebeurt met behulp van een meerstappenplan:

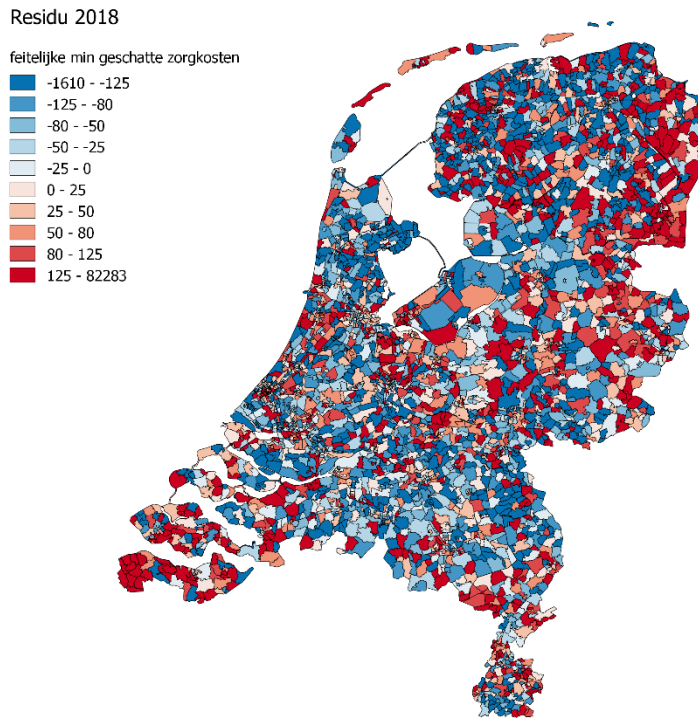
1. Schatten van het somatische vereveningsmodel *exclusief* regiocriterium;
2. Bepalen van het residu (werkelijke kosten minus voorspelde kosten) op viercijferig postcode-niveau;
3. De residuen uit stap 2 verklaren uit zogenaamde regiokenmerken en de gewogen gemiddelde normkosten uit de eerste regressie.<sup>1</sup> Alleen grote postcodegebieden met meer dan 500 verzekerdenjaren worden meegenomen in de schatting;
4. Op basis van de regressie in stap 3 een voorspeld residu op viercijferig postcodeniveau berekenen voor de grote postcodegebieden;
5. De grote postcodegebieden indelen in 10 groepen met een (vrijwel) gelijk aantal verzekerdenjaren op basis van hun voorspelde residu;
6. Op basis van de regressie in stap 3 een voorspeld residu op gemeenteniveau berekenen voor *alle* postcodegebieden;
7. Op gemeenteniveau indelen in 10 groepen met een (vrijwel) gelijk aantal verzekerdenjaren op basis van hun voorspelde residu;
8. De grote postcodegebieden indelen op basis van de indeling op postcodeniveau (stap 5) en de kleine postcodegebieden indelen op basis van de indeling op gemeenteniveau (stap 7).

De postcodegebieden die behoren tot één regio zijn verspreid over heel Nederland. Figuur 2.1 toont het residu op viercijferig postcodeniveau van het somatische model *exclusief* regiocriterium. Dit is de afhankelijke variabele in het regiomodel. Er zijn geen grote aangesloten gebieden in Nederland aan te wijzen als gebieden waarbij sterke over- of ondercompensatie plaatsvindt. Wel is zichtbaar dat in Oost-Groningen en Zeeuws-Vlaanderen veel postcodegebieden een grote ondercompensatie hebben. Ook is bijvoorbeeld in Utrecht een opvallend verschil zichtbaar tussen wijken: ondercompensatie in wijken als Overvecht en overcompensatie in het centrum en Oost. Mogelijk zijn de verschillen in de residuen slechts voor een klein deel te verklaren uit verschillen in kenmerken op regionaal niveau, waardoor geen duidelijke relatie tussen regio en het residu zichtbaar is. De ruimtelijke verdeling van residuen lijkt stabiel wanneer ook de residuen van het somatische model 2016 en 2017 worden bekeken (voor de kaartjes zie Bijlage A).

---

<sup>1</sup> De regiokenmerken zijn op viercijferig postcodeniveau. Missende waarden binnen de kenmerken worden aangevuld met het gewogen gemiddelde (naar verzekerdenjaren) op gemeenteniveau.

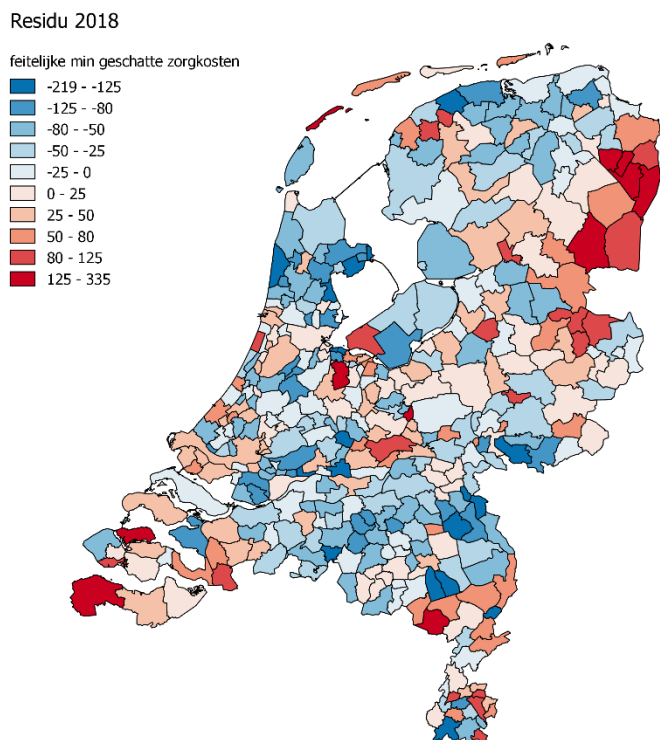
**Figuur 2.1** Geen sterk regionaal patroon in overcompensatie of ondercompensatie van residuen somatisch model 2018 *exclusief* regiocriterium



Bron: OT2018-bestand, bewerking SEO Economisch Onderzoek en Atlas voor Gemeenten.

Postcodegebieden met een klein aantal verzekerdenjaren zijn gevoelig voor uitschieters in het residu, Figuur 2.2 laat daarom de residuen van het somatische model *exclusief* regiocriterium op gemeenteniveau zien. Ondercompensatie lijkt dan deels geconcentreerd te zijn in krimpgebieden als Oost-Groningen, Oost-Drenthe, Noordoost-Overijssel en Zeeuws-Vlaanderen. Daarmee lijkt er ruimte te zijn voor een verbetering van de verevenende werking middels het regiocriterium.

**Figuur 2.2** Op gemeenteniveau ondercompensatie in krimpgebieden van residuen somatisch model 2018 *exclusief* regiocriterium



Bron: OT2018-bestand, bewerking SEO Economisch Onderzoek en Atlas voor Gemeenten.



### 3 Omgevingskenmerken en gezondheid

*Van verschillende regiokenmerken is in eerder onderzoek aangetoond dat zij een effect hebben op gezondheid. Deze kenmerken zijn: luchtvervuiling, geluidsbelasting, kwaliteit van de woning, groenvoorzieningen, sociaaleconomische status in de wijk en werkomstandigheden. Er zijn ook individuele kenmerken van personen die van invloed zijn op hun gezondheid, maar die niet op individuniveau zijn opgenomen in het risicovereveningsmodel. Het gaat hier voornamelijk om de migratieachtergrond van personen.*

Verschillen in gezondheid (en daaraan gerelateerd zorgkosten) tussen mensen kunnen het gevolg zijn van een groot aantal factoren. Voor een belangrijk deel zijn dat factoren op individueel niveau. Het gaat dan bijvoorbeeld om leeftijd, genetische aanleg, of leefstijlfactoren als roken, gezond eten en lichaamsbeweging (Harpers en Eysink, 2011). De individuele morbiditeitscriteria in combinatie met leeftijd, sociaaleconomische status (SES) en aard van inkomen (AvI) meten dit direct of indirect in het somatische vereveningsmodel. Daarnaast kunnen omgevingskenmerken een rol spelen. Het gaat hierbij om factoren die spelen in de (directe) omgeving van een individu en van invloed zijn op de gezondheid. Het bestaan van omgevingsfactoren zorgt er – in theorie – voor dat twee identieke personen verschillen in gezondheid en daarmee zorgkosten omdat ze in een andere wijk of buurt wonen. Het bestaan van deze omgevingsfactoren kan daarmee leiden tot verschillen in zorgkosten tussen verzekeraars, ook nadat verevend is voor verschillen in individuele kenmerken van verzekerden. Idealiter wordt het regiocriterium in het somatische vereveningsmodel dan ook gebaseerd op relevante omgevingsfactoren.

Dit hoofdstuk geeft een overzicht van regiofactoren die volgens de (wetenschappelijke) literatuur een effect kunnen hebben op gezondheid. Dit overzicht vormt de basis voor een selectie van regiokenmerken waarvan op theoretische gronden een relatie met gezondheid te verwachten is en waarvoor data beschikbaar is voor Nederland op viercijferig postcodeniveau. De geselecteerde kenmerken worden vervolgens getest in het regiomodel en varianten daarop.

Alle indicatoren voldoen aan de volgende randvoorwaarden:

#### 1. Validiteit

Een kenmerk is alleen valide als er een theoretische grond is om te verwachten dat het kenmerk van invloed is op gezondheidsverschillen, waardoor het verschillen in zorgkosten kan voorspellen. Voor regiokenmerken zijn er twee varianten: (a) het kenmerk corrigeert voor gezondheidsverschillen op individuniveau, maar kan op individuniveau niet worden opgenomen (b) het kenmerk corrigeert voor gezondheidsverschillen op regioniveau, nadat al gecorrigeerd is voor individuele gezondheidsverschillen op basis van DKG's, FKG's en andere individuele kenmerken in het vereveningsmodel;

#### 2. Geen samenhang met zorgaanbod

Uitgangspunt is dat compensatie voor verschillen in ZVW zorgaanbod ongewenst is (ongewenste aanbodvariatie).

### 3.1 Omgevingskenmerken als proxy voor gegevens op individuniveau

Op individueel niveau is het effect van sociaaleconomische factoren via de SES- en AvI-scores meegenomen. In combinatie met morbiditeitskenmerken, leeftijd en huishoudenssamenstelling is daarmee voor een belangrijk deel gecorrigeerd voor de factoren op individueel niveau die verschillen in gezondheid (en zorgkosten) teweegbrengen. De vraag is of er daarnaast nog additionele factoren op persoonsniveau zijn die ervoor zorgen dat personen structureel meer of minder zorgkosten hebben.

Eén van die factoren is mogelijk het hebben van een specifieke migratieachtergrond. Mensen met een niet-westerse migratieachtergrond hebben gemiddeld hogere zorgkosten dan mensen zonder migratieachtergrond, zo blijkt uit studies van onder andere het CBS (CBS, 2016). Voor een deel is dat te verklaren door de gemiddeld lagere sociaaleconomische status – waarmee via de SES al rekening is gehouden. Maar niet alle verschillen zijn daardoor te verklaren. Ook fysiologische verschillen spelen een rol: na correctie voor allerlei andere factoren die ook van invloed zijn, blijken mensen met een niet-westerse migratieachtergrond bijvoorbeeld vaker diabetes te hebben (CBS, 2014). Deze verschillen worden via de morbiditeitskenmerken meegenomen. Het SCP geeft aan dat ook culturele verschillen in voedings- en bewegingspatronen een rol kunnen spelen (SCP, 2009). Kinderen met een niet-westerse achtergrond bewegen gemiddeld minder vaak en – meisjes in het bijzonder – zijn minder vaak lid van een sportclub. Ook komt roken significant vaker voor onder sommige groepen met een bepaalde migratieachtergrond (CBS, 2016). Dit maakt dat het meenemen van het aandeel mensen met een bepaalde migratieachtergrond als aanvullende indicator voor de hand ligt. Daarbij kan zowel een onderscheid worden gemaakt naar westerse en niet-westerse migratieachtergrond als een aantal deelgroepen (zoals Marokkaans, Turks, Surinaams en Antilliaans).

Een andere mogelijke factor is het bestaan van specifieke gezondheidsproblemen bij kinderen in een huishouden. Via de morbiditeitskenmerken en aard van inkomen - arbeidsongeschiktheid wordt voornamelijk rekening gehouden met langdurige ziektes of handicaps bij volwassenen. De drempelwaarden voor opname in een FKG zijn bijvoorbeeld meestal gebaseerd op referentie-innamen van medicijngebruik voor volwassenen.<sup>2</sup> Ook hebben kinderen per definitie geen arbeidsongeschiktheidsuitkering. Maar huishoudens met kinderen met een bepaalde lichamelijke of zintuiglijke beperking of die langdurig ziek zijn, hebben logischerwijs wel hogere zorgkosten. Een indirecte manier om dat te meten is via het aandeel kinderen dat naar het speciaal onderwijs gaat.

---

<sup>2</sup> Bij de FKGs ‘schildklieraandoeningen’, ‘psychose, alzheimer en verslaving’, ‘depressie’, ‘astma’ en ‘epilepsie’ wordt wel rekening gehouden met een andere referentie-inname bij kinderen.



## 3.2 Omgevingskenmerken die gezondheidsverschillen veroorzaken

### 3.2.1 Luchtvervuiling en geluid

Luchtvervuiling en geluidsoverlast vormen twee belangrijke omgevingsfactoren die een negatief effect (kunnen) hebben op de gezondheid. Luchtvervuiling en geluidsoverlast worden samen met factoren die samenhangen met de kwaliteit van het binnenmilieu (tabaksrook en Radon) door het RIVM milieufactoren genoemd. Gezamenlijk zouden deze milieufactoren bijna 6 procent van de totale ziektelast veroorzaken (Knol en Staatse, 2005; volksgezondheidszorg.info). Daarbinnen vormt met name fijnstof een belangrijke factor, maar ook (verkeers-)geluid en andere vormen van luchtvervuiling zijn hierbij van belang (Hänninen en Knol, 2011). Het ligt dan ook voor de hand om indicatoren te onderzoeken die samenhangen met ruimtelijke verschillen in omvang van luchtvervuiling en geluidsoverlast.

Het RIVM maakt een onderscheid tussen de effecten van fijnstof, ozon en stikstofdioxide. Fijnstof is de verzamelnaam voor zwevende deeltjes in de lucht van kleine grootte. Verschillende studies laten zien dat er een negatief effect uitgaat van blootstelling aan fijnstof op de gezondheid van mensen (Fischer et al., 2015; Maas et al., 2015). Hoewel vrijwel overal in Nederland aan de normen voor maximale fijnstofblootstelling wordt voldaan, is het aannemelijk dat op plekken met meer fijnstof er grotere negatieve gezondheidseffecten zijn dan op plekken met weinig fijnstof. Dit komt omdat ook een klein beetje fijnstof gezondheidsproblemen kan veroorzaken. Er kan een onderscheid worden gemaakt tussen relatief grove (ook wel PM10 - met een diameter tussen 2,5 en 10  $\mu\text{m}$ ) en kleinere fijnstofdeeltjes (PM2,5 - met een diameter kleiner dan tussen 2,5 $\mu\text{m}$ ) waarbij de verwachte negatieve effecten van de kleinere fijnstofdeeltjes groter is omdat deze niet in de bovenste luchtwegen worden tegengehouden (volksgezondheidszorg.info). De huidige maatstaf 'luchtvervuiling' in het risicovereveningsmodel betreft de grove deeltjes (PM10). Dit onderzoek test zowel PM2,5 als PM10 als mogelijk relevante regiokenmerken.

Naast fijnstof zijn er specifieke gassen waar een negatief effect van uitgaat op de gezondheid en waar (mede daardoor) in Nederland op lokaal niveau de concentratie van wordt gemeten. Het gaat dan onder andere om stikstofdioxide (NO<sub>2</sub>). Stikstofdioxide is vooral het resultaat van verbrandingsprocessen en komt onder andere via de uitstoot van verkeer in de lucht. Plekken met een hoge verkeersintensiteit kennen dan ook hogere concentraties stikstofdioxide. Daarnaast komt via de industrie en energiecentrales stikstofdioxide in de lucht. Een andere stof waar negatieve effecten van uitgaan op de gezondheid is ozon ('smog'). Ozon is niet direct gerelateerd aan bijvoorbeeld verkeer of industrie maar ontstaat uit de reactie van andere stoffen met zonlicht en komt daarom in Nederland vooral in de zomer voor.

Stoffen die ook een negatief effect hebben maar die in Nederland in veel minder sterke mate de lucht verontreinigen zijn benzeen, zwaveldioxide en koolmonoxide. Voor benzeen geldt dat de concentratie in Nederland fors is afgenomen (door verschillende maatregelen) maar dat er rond de chemische industrie in de regio Rijnmond nog wel een relatief sterke concentratie is (CBS, PBL, RIVM en WUR, 2014). Ook voor zwaveldioxide en koolmonoxide geldt dat de concentratie in Nederland fors is afgenomen en in tegenstelling tot vroeger niet meer als een (groot) probleem wordt gezien als het gaat om luchtverontreiniging. Zwaveldioxide komt in wat hogere concentraties

voor op plekken met raffinaderijen en zeescheepvaart zoals Rotterdam, IJmuiden en Vlissingen (CBS, PBL, RIVM en WUR, 2014). De hierboven benoemde soorten luchtvervuiling zijn over het algemeen gerelateerd aan verkeer en vervuilende industrie. Een andere vorm van luchtvervuiling is ammoniak dat vooral gerelateerd is aan (intensieve) veeteelt. In een recent onderzoek van het RIVM bleek dat er een negatieve relatie was tussen de concentratie van ammoniak en de longfunctie van mensen (Maassen et al., 2016).

Van al van deze genoemde vormen van luchtvervuiling zijn gegevens op viercijferig postcodegebied beschikbaar. In veel gevallen zijn deze gegevens niet gebaseerd op daadwerkelijke metingen in elk viercijferig postcodegebied in Nederland maar op een combinatie van metingen en modelmatige schattingen (volksgezondheidszorg.info).

Naast luchtvervuiling kan blootstelling aan geluid ook tot gezondheidsklachten leiden. Geluidsoverlast kan leiden tot slaapttekorten, stress en gebrek aan concentratievermogen wat via een hogere bloeddruk en effecten op hormonen kan leiden tot somatische klachten zoals hart- en vaatziekten (Kempen en Houthuijs, 2008)). Belangrijke bronnen van geluid zijn lucht-, weg- en spoorverkeer. De mate van geluidsbelasting is ingeschat door het RIVM en heeft betrekking op de cumulatieve geluidbelasting van weg-, rail- en vliegverkeer, windturbines en industrie.<sup>3</sup> Er worden verschillende klassen onderscheiden op basis van de gemiddelde geluidsbelasting over een etmaal waarbij geluidsoverlast in de nacht zwaarder meeweegt dan overdag.

### 3.2.2 Groen, overlast, woning en werk

Naast luchtvervuiling en geluidsoverlast kunnen ook kenmerken van de directe woonomgeving van invloed zijn op de gezondheid van mensen en de daaruit voortvloeiende zorgkosten. Er zijn verschillende studies die concluderen dat er een positieve relatie bestaat tussen het aanbod van groen (parken, plantsoen, bossen of andere natuur) en de (ervaren) gezondheid van mensen (Maas et al., 2009; Maas et al., 2006; de Jong et al., 2012). Deze relatie blijkt er ook te zijn na correctie voor individuele kenmerken. Dit is van belang omdat de kans groot is dat mensen met een hoger inkomen én gezonder leven én zich vaker een woning kunnen veroorloven in de buurt van een park of bos. De aanwezigheid van groen in de directe woonomgeving zou meer mogelijkheden bieden voor beweging, ontspanning en sociale contacten wat direct of indirect een positief effect zou hebben op de gezondheid (Staatsen et al., 2016).

De directe woonomgeving kan ook een nadelig effect hebben op de gezondheid als er sprake is van leefbaarheidsproblemen en in het bijzonder overlast en criminaliteit (Ruijsbroek et al., 2016). Dat kan leiden tot stress, slaapttekort en uiteindelijk tot onder andere meer hart- en vaatziekten (Groenewegen et al., 2018).

Naast woonomgeving kan ook de woning zelf van invloed zijn op de gezondheid. Woningen die te maken hebben met een hoge luchtvochtigheid (en mogelijk schimmels) en/of een slechtere kwaliteit van de lucht (door slechte ventilatie, de aanwezigheid van asbest of de aanwezigheid van tabaksrook) kunnen leiden tot negatieve gezondheidseffecten. Een deel van deze factoren (zoals tabaksrook) hangt weer samen met kenmerken van het individu of huishouden. Voor andere factoren

---

<sup>3</sup> Zie [https://geluid.rivm.nl/geluid/geluidbel\\_achtergrond.html](https://geluid.rivm.nl/geluid/geluidbel_achtergrond.html)

geldt dat er een relatie is met bouwperiode. De periode waarin een woning is gebouwd, bepaalt voor een belangrijk deel de kwaliteit van de materialen en mate van ventilatie enzovoorts. Zo blijkt bijvoorbeeld dat zichtbare schimmels zich nauwelijks meer voordoen in (koop-)woningen gebouwd na 1980 maar juist regelmatig aanwezig zijn in vooroorlogse woningen (Dongen en Vos, 2007). Ook kunnen er verschillen zijn in de *gemiddelde* kwaliteit van een koopwoning ten opzichte van een huurwoning bij een woningbouwcorporatie (die immers vooral bedoeld zijn voor mensen met een laag inkomen).

Ten slotte kan ook de werkomgeving een rol spelen (Eysink et al., 2012). In sommige sectoren wordt bijvoorbeeld met gevaarlijke stoffen gewerkt of is de kans op ongelukken op de werkvloer groot en in andere sectoren is de werkdruk relatief hoog of zitten mensen het grootste deel van de dag stil achter een computer (Hooftman et al., 2017). Dit maakt dat twee dezelfde personen die werkzaam zijn in verschillende sectoren toch een ander gezondheidsrisico hebben en daarmee mogelijk andere zorgkosten. Naast kenmerken van de woonomgeving zou daarom idealiter ook een indicator voor de werkomgeving worden meegenomen.

### 3.3 Nieuw te testen regiokenmerken

Op basis van bovenstaand literatuuroverzicht zijn indicatoren op het niveau van viercijferige postcodes verzameld. Tabel 3.1 geeft een overzicht van de indicatoren die een proxy vormen voor kenmerken op individueel niveau waarvoor geen gegevens zijn: migratieachtergrond en speciaal onderwijs. Bij speciaal onderwijs gaat het om het aantal kinderen (afgerond op tientallen vanwege privacyredenen) dat is ingeschreven op het speciaal basis- of voortgezet onderwijs als percentage van alle 4- tot en met 19-jarigen

**Tabel 3.1** Migratieachtergrond en speciaal onderwijs dienen als proxy voor (ontbrekende) kenmerken op individueel niveau

Onderwerp	Kenmerk	Bron
Migratieachtergrond - breed	% mensen met niet-westerse achtergrond	CBS
	% mensen met westerse achtergrond	CBS
Migratieachtergrond – deelgroepen	% mensen met Marokkaanse achtergrond	CBS
	% mensen met Turkse achtergrond	CBS
	% mensen met Surinaamse achtergrond	CBS
	% mensen met Antilliaanse achtergrond	CBS
	% mensen met overige niet-westerse achtergrond	CBS
	% mensen met een Midden- en Oost-Europese (MOE) achtergrond	CBS
	% mensen met overige westerse achtergrond	CBS
Speciaal onderwijs	% kinderen (4 tot en met 19 jaar) speciaal onderwijs	CBS

Bron: SEO Economisch Onderzoek en Atlas voor Gemeenten

Tabel 3.2 geeft een overzicht van de indicatoren die de omgevingsfactoren meten waarvan op basis van het literatuuroverzicht een positieve dan wel negatieve bijdrage aan de gezondheid en zorgkosten verondersteld kan worden. Bij luchtvervuiling gaat het in alle gevallen om een indicatie van de concentratie van een bepaalde stof (of stoffen) op de locatie gemeten in  $\mu$ gram per m<sup>3</sup>. Bij geluidsbelasting is er veelal sprake van variatie binnen een viercijferig postcodegebied (bijvoorbeeld omdat een deel direct aan een drukke weg woont en een ander deel niet). In plaats van een gemiddelde

klasse voor het hele gebied te berekenen is er daarom voor gekozen om het aandeel mensen dat te maken heeft met weinig, gemiddeld of veel geluidsbelasting per viercijferig postcodegebied te berekenen.

Groen in de buurt is berekend als het aandeel van een bepaald type groen in de totale oppervlakte van een gebied. Voor verschillen in leefbaarheid is gebruikgemaakt van een door Atlas voor Gemeenten in eerder onderzoek ontwikkelde index voor overlast en onveiligheid (Marlet en Woerkens, 2007a en 2007b). Dit is een gewogen index van 10 verschillende overlastindicatoren variërend van drugsoverlast tot woninginbraken. Deze index is ook opgenomen in het bijstandsverdeelmodel voor de verdeling van bijstandsgelden over gemeenten (Tempelman et al., 2017). In dit model wordt op individuniveau de kans op bijstand verklaard. Daarnaast is het kenmerk SES dat nu al op individuniveau in het risicovereveningsmodel aanwezig is, geaggregeerd naar viercijferig postcode-niveau als aanvullende indicator van leefbaarheid/overlast.

Er zijn geen data beschikbaar die een beeld geven van het aandeel mensen naar woonplek dat werkt in sectoren met een bepaald gezondheidsrisico. Daarom is een inschatting gemaakt op basis van de verschillen in het ziekteverzuimpercentage per sector, het aantal banen per sector per gemeente (werkplek) en de woon-werkpendel. Het resultaat is het geschatte gemiddelde ziekteverzuimpercentage onder werknemers naar woonplek van de werknemers. Deze indicator is vanwege de databeschikbaarheid van pendel berekend op het niveau van gemeenten.

**Tabel 3.2**    **Overzicht indicatoren voor omgevingsfactoren**

Onderwerp	Indicator	Bron
Luchtvervuiling	concentratie benzeen (C6H6)	RIVM
	concentratie koolmonoxide (CO)	RIVM
	concentratie ammoniak (NH3)	RIVM
	concentratie stikstofdioxide (NO2)	RIVM
	concentratie ozon (O3)	RIVM
	concentratie fijnstof PM10	RIVM
	concentratie fijnstof PM2,5	RIVM
	concentratie zwaveldioxide (SO2)	RIVM
Geluidsbelasting	% inwoners in pc4-gebied in specifieke geluidsklassen – drie hoofdgroepen laag, midden en hoog <sup>4</sup>	RIVM
Woningvoorraad	% woningen naar bouwperiode, onderscheid naar goede woningen (vanaf 2001), slechte woningen (1945-1970) en midden goede woningen (overige bouwjaren)	CBS
	% woningen dat in bezit is van woningbouwcorporaties	CBS
Groen in de buurt	% bomen, grasland, plantsoenen en parken in totale oppervlakte	Kadaster en CBS
	% bos in totale oppervlakte	CBS
	% (overig) droog natuurlijk terrein (heide, duinen etc.)	CBS
Leefbaarheid	Index overlast en onveiligheid	Atlas voor gemeenten
	SES op pc4-niveau	OT 2018
Werkomstandigheden	% werknemers in ziekteverzuim naar woonplek	Inschatting op basis van CBS en LISA

Bron: SEO Economisch Onderzoek en Atlas voor Gemeenten

<sup>4</sup> De originele indeling in 10 klassen is hierbij geaggregeerd tot drie hoofdklassen

### 3.4 Validiteit huidige regiokenmerken

Tabel 3.3 toont de huidige regiokenmerken in het regiomodel voor de somatische zorg (eerste kolom). Een aantal kenmerken komt terug in het literatuuronderzoek in de voorgaande paragrafen: niet-westerse migratieachtergrond, luchtvervuiling en huurwoningen/corporatiewoningen. Het percentage corporatiewoningen is een mogelijk alternatief voor het percentage huurwoningen. Huurwoningen worden ook aangeboden in de private sector en bevatten dus ook duurdere huurwoningen waarbij achterstallig onderhoud vermoedelijk minder aan de orde is. Anderzijds bieden woningbouwcorporaties ook koopwoningen die mogelijk van betere kwaliteit zijn dan de gemiddelde huurwoning.

Drie van de huidige regiokenmerken worden in de (wetenschappelijke) literatuur niet genoemd als regiokenmerken die een verschil in gezondheid tussen de inwoners kan verklaren. Dit zijn de normkosten somatische zorg, regionaal klantenpotentieel en OAD/stedelijkheid. De factoren zijn dus mogelijk niet valide en/of hangen samen met het zorgaanbod.

**Tabel 3.3 Relatie huidige regiokenmerken en nieuw te testen regiokenmerken**

Huidige regiokenmerken (OT 2018)	Voorstel nieuw regiokenmerk
Normkosten somatische zorg	X
Niet-westerse migratieachtergrond (%)	Niet-westerse migratieachtergrond (%) Westerse migratieachtergrond (%) Deelgroepen migratieachtergrond (%)
Regionaal klantenpotentieel (relatief)	X
Luchtvervuiling (PM10)	Verschillende maatstaven
Huurwoningen (%)	Corporatiewoningen (%) Woningen naar bouwperiode (in 3 groepen)
OAD1	X
OAD2	X
OAD3	X
OAD4	X
OAD5	X
	Speciaal onderwijs (%) Geluidsbelasting (in 3 groepen) Groen in de buurt (3 maatstaven) Overlast index SES op pc4-niveau (in 4 groepen) Ziekteverzuim werknemers (%)

Bron: SEO Economisch Onderzoek

#### *Normkosten somatische zorg*

De normkosten somatische zorg zijn opgenomen in het regiomodel als correctie voor de interactie van de regiokenmerken met de individuele kenmerken in het risicovereveningsmodel (WOR 821 – onverklaarde regionale variatie). WOR 034 – regio in detail – biedt meer achtergrondinformatie. In dit rapport wordt een probleem geconstateerd dat optreedt bij het schatten van een model in twee stappen: zowel de kenmerken op individuniveau als de kenmerken op regioniveau worden verkeerd geschat, indien zij met elkaar samenhangen (wat aannemelijk is).

Dit kan worden opgelost door ook de relevante regiokenmerken direct mee te schatten in het model op individuniveau. Dan worden zowel de kenmerken op individuniveau als de kenmerken op regioniveau zuiver geschat. Destijds was een model op individuniveau geen optie, omdat er dan te veel vereveningskenmerken zouden worden toegevoegd aan het model. Het toevoegen van meerdere vereveningskenmerken bemoeilijkt de uitvoering, bijvoorbeeld bij de verzekerdenraming door het ZIN.

Door de normkosten somatische zorg níet op te nemen, wordt verondersteld dat er geen relatie is tussen de individuele kenmerken en de regiofactoren. Dat is niet aannemelijk. Door de normkosten somatische zorg wél op te nemen, wordt verondersteld dat de (complexe) relatie tussen individuele kenmerken en de regiofactoren in één parameter te vangen is. Ook dat is niet aannemelijk. Bijvoorbeeld: stel dat het percentage groen in de wijk van invloed is op de zorgkosten. En in 'groene' wijken wonen veel mensen met hoge inkomens/hoge SES. Dan is er een sterke relatie tussen een regiokenmerk en SES. Een sterke relatie met de FKG's ligt echter niet voor de hand. Het toevoegen van één parameter is dus niet voldoende en bemoeilijkt de interpretatie van de (resterende) correlatie tussen de residuen en de regiokenmerken. In dit onderzoek wordt daarom ook de mogelijkheid verkend om de normkosten weg te laten uit het regiomodel.

Binnen het huidige regiomodel is het ook mogelijk om alle individuele kenmerken uit het risico-vereveningsmodel – geaggregeerd naar pc4-niveau op te nemen in het regiomodel. Dit doet meer recht aan de complexe relatie tussen individuele kenmerken en regionale kenmerken. In dat geval resulteert echter nog steeds een onzuivere schatting van de kenmerken op individuniveau. Bovendien zal het toevoegen van een groot aantal variabelen aan het regiomodel het schatten van de regioclusters niet eenvoudiger of transparanter maken.

Een technische factor die ontbreekt in het huidige regiomodel is de constante. Door een constante toe te voegen is er automatisch een referentiegroep in de regressie, waardoor minder kenmerken hoeven te worden toegevoegd. Bijvoorbeeld: bij de OAD-indeling kan dan één OAD-categorie worden weggelaten. De coëfficiënten op de overige OAD-categorieën kunnen dan worden geïnterpreteerd als afwijking van de referentiecategorie. Op die manier kan ook de significantie van OAD beter worden beoordeeld: op dit moment zit de constante immers ook 'verscholen' in de OAD-categorieën. En wanneer OAD (of een andere categorale variabele) niet meer in het regiomodel zit, zullen andere variabelen de weggelaten constante gaan oppikken en daardoor niet meer zuiver geschat worden.

#### *Regionaal klantenpotentieel en OAD*

Regionaal klantenpotentieel en OAD (stedelijkheid) hangen sterk met elkaar samen: een grote stad heeft doorgaans ook een hoger regionaal klantenpotentieel. Er is in de wetenschappelijke literatuur geen relatie gevonden tussen stedelijkheid en gezondheid. Wél zijn er relaties tussen factoren die in steden vaker voorkomen (fijnstof, geluidsbelasting, minder groen) en gezondheid. Bij correctie voor deze directe oorzaken van gezondheid zou het dus niet meer nodig moeten zijn om voor stedelijkheid te corrigeren. Stedelijkheid/omgevingsadressendichtheid zelf is daarmee niet valide als kenmerk in het regiomodel. Als er in steden sprake is van meer mensen met hogere zorgkosten doordat zij ongezonder zijn, dan zou dat moet blijken uit de individuele kenmerken waarmee in

het model reeds rekening wordt gehouden of de regionale kenmerken die dit gezondheidsverschil veroorzaken.<sup>5</sup>

Bovendien is er een risico dat regionaal klantenpotentieel en omgevingsadressendichtheid samenhangen met zorgaanbodfactoren. Het zorgaanbod in grote steden is hoger. Het risico bestaat dus dat via deze kenmerken alsnog aanbod wordt meegenomen in het regiomodel. Dit is inderdaad het geval: WOR 821 (pp. 16) vermeldt dat de hogere OAD-klassen een deel van de verklaringskracht van de aanbodfactoren (afstand tot huisarts en afstand tot ziekenhuis) opvangen.

---

<sup>5</sup> Om deze reden is het regionaal klantenpotentieel uiteindelijk ook niet opgenomen in het verdeelmodel voor de bijstandsbudgetten, zie Tempelman, Vriend, Kroon, Marlet en Woerkens (2017): Verfijning bijstandsverdeelmodel 2018.





## 4 Technische varianten regiomodel

*Er zijn verschillende alternatieve manieren om in het somatische vereveningsmodel voor regiokenmerken te compenseren. Sommige methoden resulteren in een indeling in regioclusters, andere nemen kenmerken op individuniveau mee.*

Het huidige regiocriterium is samengesteld op basis van een meerstappenplan dat veel tijd kost tijdens de jaarlijkse update van het vereveningsmodel. Bovendien is de aanvullende verevenende werking van de regioclusters beperkt. Om de verevening te verbeteren en de schattingsmethode te vereenvoudigen presenteert dit hoofdstuk enkele alternatieve manieren om rekening te houden met regiokenmerken in de risicoverevening.

Samenvattend gaat het om de volgende modellen:

1. Uitgangsmodel somatische zorg exclusief het huidige regiocriterium (M00);
2. Uitgangsmodel somatische zorg inclusief het huidige regiocriterium (M0);
3. Regiomodel met aangepaste set regiokenmerken (M0\*);
4. Regiomodel waarbij indeling in 10 gelijke groepen is losgelaten (M1 en M1\*);
5. Model op basis van clustering van kenmerken, aan de hand van meerkosten (M2 en M2\*);
6. Model waarin regiovariabelen rechtstreeks zijn opgenomen (M3 en M3\*);
7. Model op basis van clustering van kenmerken, zonder gebruik van meerkosten (M4 en M4\*);
8. Model waarin regiovariabelen rechtstreeks zijn opgenomen, vervolgens indeling in regioclusters (M5 en M5\*).

Dit hoofdstuk beschrijft deze varianten. De varianten zonder ster gaan telkens uit van de huidige regiokenmerken 2018. De varianten met ster gaan uit van de nieuwe regiokenmerken zoals geïdentificeerd in de literatuurstudie van Hoofdstuk 3 (zie voor een overzicht Tabel 3.1 en Tabel 3.2). Hoofdstuk 5 beoordeelt alle modellen op hun verevenende werking.

### 4.1 Uitgangsmodellen en model met aangepaste set regiokenmerken (M00, M0 en M0\*)

Model M00 is het uitgangsmodel voor de somatische zorg exclusief regiocriterium. Het model dient als referentie voor de verschillende modellen met regiokenmerken. Model M0 is het uitgangsmodel voor de somatische zorg met het regiomodel dat wordt geschat volgens het meerstappenplan uit Hoofdstuk 2. Het bevat de huidige regiokenmerken: percentage personen met een niet-westerse migratieachtergrond, percentage huurwoningen, luchtvervuiling, OAD, regionaal klantenpotentieel en de normkosten somatische zorg.

Model M0\* hanteert hetzelfde meerstappenplan, maar hanteert andere regiokenmerken, namelijk percentage personen met een westerse migratieachtergrond, percentage personen met een niet-

westerse migratieachtergrond, lage SES<sup>6</sup> en het percentage ziekteverzuim. De normkosten somatische zorg zijn niet opgenomen als technische factor, een constante is wel opgenomen als technische factor.

Deze kenmerken zijn als volgt geselecteerd (zie voor een uitgebreide beschrijving van de procedure inclusief schattingsresultaten Bijlage B). In een eerste stap nemen we alle huidige regiokenmerken uit de OT2018 én alle kenmerken die zijn geïdentificeerd in de literatuurstudie in Hoofdstuk 3 op in het regiomodel. Vervolgens verwijderen we het meest insignificante regiokenmerk of een kenmerk met een onlogische relatie. Dit proces herhalen we tot alleen significante variabelen met logische tekens overblijven.

#### *Voor- en nadelen van de nieuwe regiokenmerken*

De  $R^2$  van het regiomodel met de nieuwe regiokenmerken is 5,2 procent, wat lager is dan het huidige regiomodel (6,7 procent). De lage  $R^2$  kan diverse oorzaken hebben. Wellicht zijn kenmerken op regioniveau beperkt van belang in het verklaren van de variantie in zorgkosten die overblijft na schatten van het risicovereveningsmodel exclusief regiokenmerk. Aan de andere kant is het mogelijk dat er regionale gezondheidseffecten zijn die hier niet onderzocht zijn en die wel een (grote) invloed hebben op zorgkosten. Welk kenmerk dit zou moeten zijn is onduidelijk. Immers alle regiofactoren die volgens eerder onderzoek een effect hebben op gezondheid zijn meegenomen.

Voordeel van model M0\* ten opzichte van model M0 is dat controversiële factoren (OAD, regionaal klantenpotentieel, normkosten) niet zijn opgenomen, en dat bovendien alle regiokenmerken het verwachte en logische teken hebben. Dit in tegenstelling tot bijvoorbeeld luchtvervuiling dat in de OT2018 (en dus ook in model M0) een onverwacht negatieve relatie met zorgkosten heeft.

Bijlage E toont de verevenende werking van nog een andere variant op model M0: model M0\*\*. Dit model bevat een groot aantal regiokenmerken, namelijk alle regiokenmerken met een significante relatie met de residuen, ook de kenmerken met een onlogische relatie. De hoofdstuk behandelt dit model verder niet.

## 4.2 Regiomodel waarbij indeling in 10 gelijke groepen is losgelaten (M1 en M1\*)

Dit model borduurt voort op de bestaande methodiek, maar laat de eis los dat er 10 groepen worden gevormd van gelijke grootte. In vergelijking met model M0 is er alleen een aanpassing in stap 5 en 7 van de bepaling van het regiocriterium (de voorspelde waarden van residuen verdelen in 10 groepen met een (vrijwel) gelijk aantal verzekerdenjaren), zie Hoofdstuk 2. De overige stappen blijven hetzelfde.

---

<sup>6</sup> Lage SES is een betere maatstaf voor lage sociaal economische status dan zeer lage SES. Personen met een zeer laag inkomen zijn een zeer hybride groep en bevatten bijvoorbeeld ondernemers met een slecht jaar, studenten etc. (zie ook Vriend et al., 2017).

We schatten dit model met alle regiokenmerken uit OT2018 (inclusief normkosten als technische factor). Dit is model M1. We schatten dit model ook met de aangepaste set regiokenmerken (exclusief de normkosten als technische factor maar inclusief constante) op basis van de literatuurstudie (M1\*).

Om clusters te vormen die het meest op elkaar lijken in termen van het voorspelde residu, maken we gebruik van Wards methode – zie Box 4.1. Deze methode wordt ook gehanteerd bij de indeling in DKG's. De methode van Ward maakt geen keuze voor het aantal clusters. Er is gekozen voor een indeling in 11 clusters, zie Box 4.1.

### Box 4.1 Methode van Ward

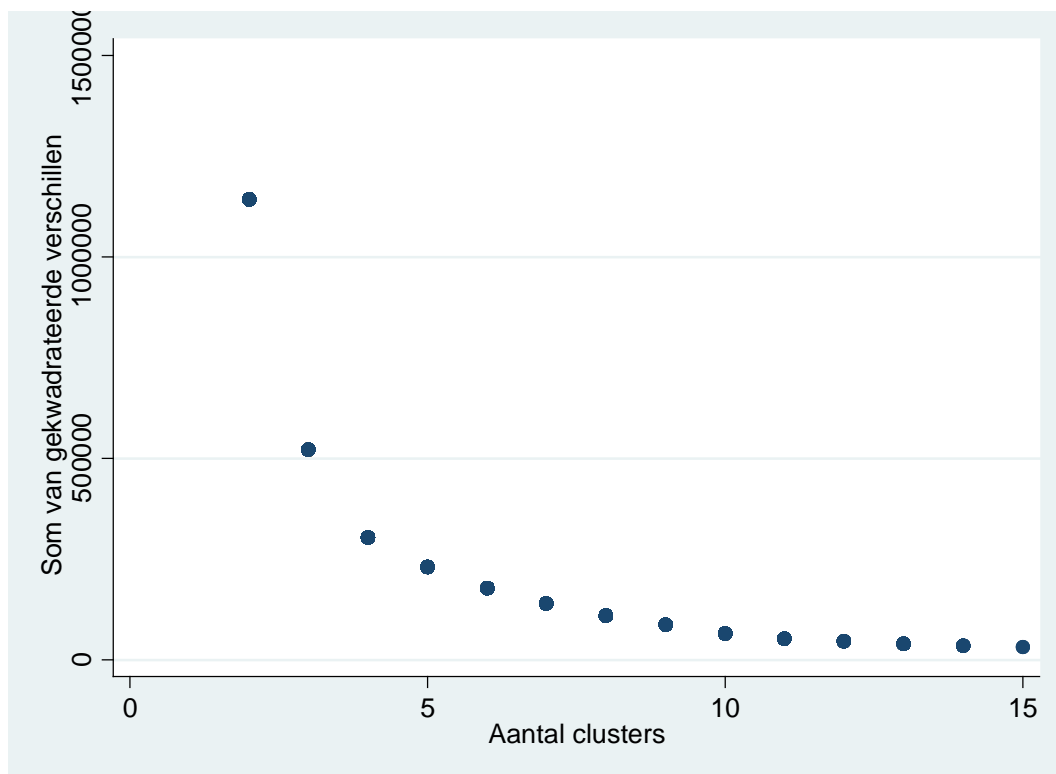
De methode van Ward kiest de clusters op zo'n manier dat de verschillen binnen een cluster zo klein mogelijk zijn, en de verschillen tussen clusters zo groot mogelijk. De grootte van de clusters kan verschillen. Bijvoorbeeld: indien er een kleine groep postcodes is met hoge (voorspelde) meerkosten, een grote groep met voorspelde meerkosten ongeveer gelijk aan 0 en een kleine groep met lage (voorspelde) meerkosten, dan zal de methode resulteren in twee kleine groepen en één grote groep.

In vergelijking met een methode die uitgaat van tien gelijke groepen kan Wards methode de compensatie van sterk afwijkende postcodes verbeteren. Dit is een voordeel van deze methode ten opzichte van de huidige methode. Nadeel van deze methode in vergelijking met de alternatieve methoden die hieronder staan beschreven, is dat niet wordt afgeweken van de bestaande – bewerkelijke – methodiek in het regiomodel.

De methode van Ward maakt geen keuze voor het aantal clusters. Deze keuze wordt doorgaans bepaald door te bekijken hoe de homogeniteit binnen de clusters toeneemt naarmate er meer clusters gevormd worden. Figuur 4.1 laat zien hoe dit gebeurt voor model M1, het regiomodel met de huidige regiokenmerken 2018. Het vormen van steeds meer clusters verkleint de som van de gekwadrateerde verschillen tussen het voorspelde residu van het individu en het gemiddelde voorspelde residu van zijn cluster. Met andere woorden: de homogeniteit binnen een cluster neemt toe. Na het vormen van het elfde cluster vindt geen grote winst in homogeniteit meer plaats. We gaan daarom voor model M1 uit van 11 clusters.

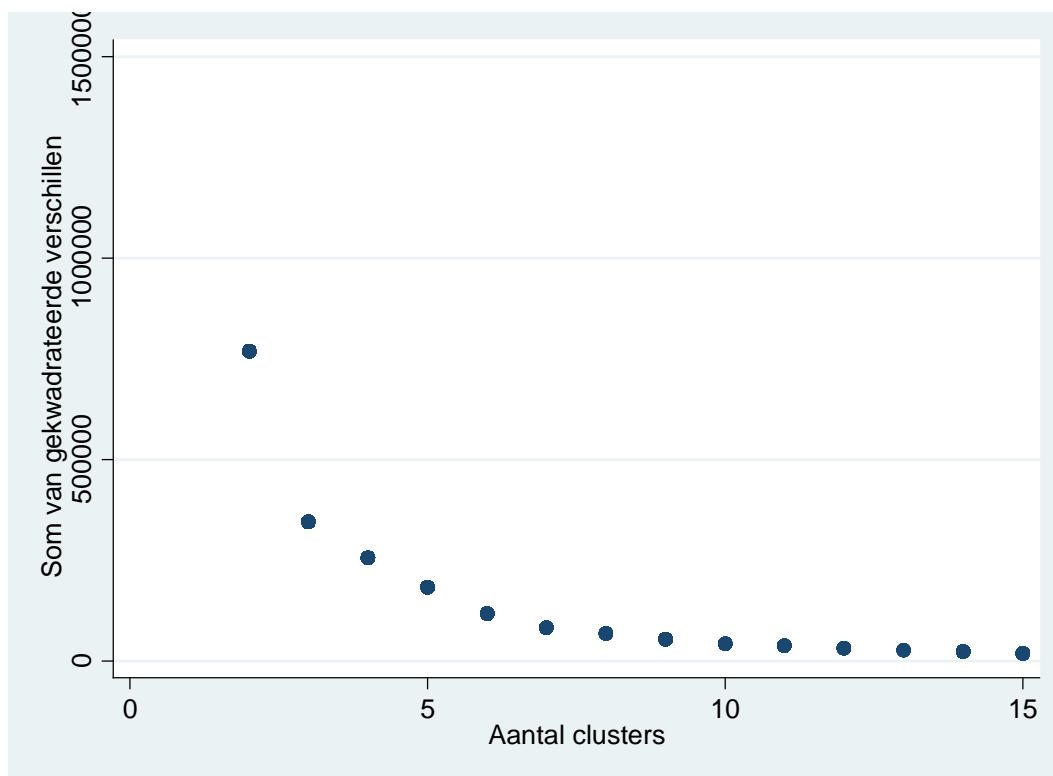
Voor de vergelijkbaarheid van de modellen met de huidige en nieuwe regiocriteria gaat model M1\* ook uit van 11 clusters. De som van de gekwadrateerde verschillen voor dit model vertoont een vergelijkbaar verloop, zie Figuur 4.2. Wel is de som al bij een klein aantal clusters een stuk kleiner, wat duidt op een betere verdeling in regioclusters wanneer gebruik wordt gemaakt van de nieuwe regiocriteria.

Figuur 4.1 Na vormen van het 11<sup>e</sup> cluster verbetert de afstand tussen de clusters nog minimaal (model M1)



Bron: OT2018-bestand, bewerking SEO Economisch Onderzoek

**Figuur 4.2** Na vormen van het 11<sup>e</sup> cluster verbetert de afstand tussen de clusters nog minimaal (model M1\*)



Bron: OT2018-bestand, bewerking SEO Economisch Onderzoek

#### *Voor- en nadelen van loslaten indeling in 10 gelijke groepen*

Mogelijk voordeel van Wards methode is dat de compensatie van sterk afwijkende postcodegebieden verbeterd kan worden en de impact van het regiokenmerk daardoor kan toenemen. Mogelijk risico is wel dat dit model minder stabiel is, omdat de grootte van de groepen niet vastligt. Niet alleen de indeling in de 10 groepen kan dan verschuiven, maar ook een verandering van de grootte van de groepen van jaar op jaar kan ervoor zorgen dat een postcodegebied bij een ander cluster wordt ingedeeld.

### 4.3 Model op basis van clustering van kenmerken, aan de hand van meerkosten (M2 en M2\*)

Een alternatief voor het huidige regiomodel is om een regio-indeling te baseren op een clustering van meerkosten, waarbij voorafgaand aan de clustering de postcodegebieden zijn ingedeeld in groepen op basis van een combinatie van regiokenmerken.

Model M2 gaat uit van de regiokenmerken uit de OT2018 (exclusief technische factor): percentage personen met een niet-westerse migratieachtergrond, regionaal klantenpotentieel, luchtvervuiling,

percentage huurwoningen en OAD. Alleen OAD is een categorische variabele, alle andere variabelen zijn op een continue schaal. We delen daarom eerst alle kenmerken in categorieën in.<sup>7</sup> Iedere variabele bestaat dan uit drie categorieën met in iedere categorie een ongeveer gelijk aantal verzekerdenjaren; dus een laag, gemiddeld, hoog percentage huurwoningen, et cetera.<sup>8</sup> Op die manier ontstaan  $3 * 3 * 3 * 3 * 5 = 405$  potentiële groepen met een unieke combinatie van de kenmerken uit het regiomodel. Hiervan zijn 262 groepen gevuld.

Postcodegebieden met minimaal 500 verzekerdenjaren worden ingedeeld op basis van hun eigen kenmerken (bijvoorbeeld laag percentage niet-westerse migratieachtergrond, laag regionaal klantenpotentieel, lage luchtvervuiling, lage huurwoningen, OAD5). Postcodegebieden met minder dan 500 verzekerdenjaren worden ingedeeld op basis van de gewogen gemiddelde kenmerken in de gemeente waartoe zij behoren (weging met aantal verzekerdenjaren). Dit bevordert de stabiliteit van de clusterindeling. Wanneer kleine postcodegebieden op basis van hun eigen kenmerken in een groep worden ingedeeld, bestaat immers het risico dat één klein postcodegebied een eigen groep gaat vormen met zeer afwijkende waarden van de kenmerken (bijvoorbeeld hoog percentage met een migratieachtergrond in een plattelandsgemeente). Met de huidige schattingsmethoden zijn 7 van de 262 groepen alsnog 'kleine' groepen, met minder dan 500 verzekerdenjaren. Het gaat in totaal om ruim 1.000 verzekerdenjaren.

Per groep worden vervolgens de gewogen gemiddelde meerkosten ofwel het residu berekend. Vervolgens worden deze groepen ingedeeld in regioclusters met de methode van Ward. Voor de vergelijkbaarheid gaat model M2 ook uit van 11 clusters. De 262 groepen worden dus ingedeeld in 11 clusters, op basis van hun gewogen gemiddelde meerkosten.

Dezelfde methodiek als voor model M2 is eveneens toegepast op de aangepaste set regiokenmerken op basis van de literatuurstudie. Dit is model M2\*. Het betreft dan de kenmerken die resulteren uit het aangepaste regiomodel M0\*: westerse migratieachtergrond, niet-westerse migratieachtergrond, ziekteverzuim en lage SES. In dit geval resulteren dus  $3 * 3 * 3 * 3 = 81$  potentiële groepen.

*Voor- en nadelen model op basis van clustering van kenmerken, aan de hand van meerkosten*

Belangrijkste voordeel van deze methode is dat geen tweede regressie op de residuen meer hoeft te worden uitgevoerd: stap 3 in het huidige meerstappenplan voor de indeling in regioclusters vindt dan niet meer plaats. Dit maakt de bepaling van het regiocriterium eenvoudiger en minder bewerkelijk. De winst is echter beperkt: net als in de huidige methodiek dient eerst het vereveningsmodel exclusief regiocriterium te worden geschat, en in plaats van een tweede regressie dient er een indeling in klassen plaats te vinden. Voor model M2 geldt dat net als in model M0 het nog altijd mogelijk is dat er een relatie is tussen de regiovariabelen (stedelijkheid) en aanbodvariabelen (afstand tot het dichtstbijzijnde ziekenhuis) waardoor aanbodfactoren indirect wel een rol spelen.

<sup>7</sup> Wanneer we dit niet doen vormt ieder postcodegebied zijn eigen groep en reduceert deze methode tot alternatief 1: Model op basis van clustering van meerkosten.

<sup>8</sup> Een uitzondering is personen met een niet-westerse migratieachtergrond: het percentage personen met een niet-westerse migratieachtergrond is zeer scheef verdeeld over Nederland. Er is daarom voor gekozen om de grenzen voor niet-westerse migratieachtergrond op 10 procent (gemiddelde van Nederland) en 20 procent te zetten.

## 4.4 Model waarin regiovariabelen rechtstreeks zijn opgenomen (M3 en M3\*)

Een ander alternatief voor het huidige regiomodel is om regiovariabelen direct op te nemen in het risicovereveningsmodel. Deze variant is verkend in WOR 821 met de variabelen achterstandswijk (als alternatief voor het percentage personen met een niet-westerse migratieachtergrond) en stedelijkheid/OAD. Dit model had op individuniveau een vergelijkbare verevenende werking als het huidige regiomodel, maar resulteerde op regionaal niveau in een daling van het resultaat van de G4 van € 3 naar - € 10.

We gaan uit van de regiokenmerken uit de OT2018 (exclusief technische factoren): percentage personen met een niet-westerse migratieachtergrond, regionaal klantenpotentieel, luchtvervuiling, percentage huurwoningen en OAD. In WOR 821 is de selectie van regiokenmerken gebaseerd op de absolute verhoging in de  $R^2$  van het regiomodel dat de residuen van het somatische risicovereveningsmodel verklaart. Eerst is OAD toegevoegd, vervolgens het percentage personen met een niet-westerse migratieachtergrond en tot slot de overige regiokenmerken (pp21 WOR 821). Het kenmerk dat als eerst wordt toegevoegd heeft de grootste kans om de  $R^2$  te verhogen, eventuele samenhang met andere kenmerken die ook de zorgkosten verhogen wordt immers opgepakt door dat ene kenmerk. Het is daarom te verwachten dat de twee variabelen die als eerste worden toegevoegd, de  $R^2$  het meest verhogen. Om die reden hanteren we een andere manier om de belangrijkste regiokenmerken voor het risicovereveningsmodel te selecteren. In een eerste stap nemen we een brede set van regiokenmerken direct op in het risicovereveningsmodel. Vervolgens verwijderen we het meest insignificante regiokenmerk of een kenmerk met een onlogische relatie. Dit proces herhalen we tot alleen significante variabelen met logische tekens overblijven. In dit ‘stappenplan’ is ieder kenmerk ingedeeld in drie klassen met een gelijk aantal verzekerdenjaren.

In het uiteindelijke model is ieder kenmerk ingedeeld in vier klassen. Dit is gedaan omdat de modellen met drie klassen resulteerden in relatief grotere afwijkingen op regionaal niveau (zie Bijlage E). Dit is volgens enkele leden van de begeleidingscommissie ongewenst. In een eerdere conceptversie van dit rapport is gerekend met een indeling in tien klassen met een gelijk aantal verzekerdenjaren. Dit resulteerde echter in onlogische normbedragen, die niet altijd constant opliepen of juist aflieden maar soms ook alterneerde. Ook was de verevenende werking op regioniveau van de modellen met indeling in tien klassen minder goed dan die van de modellen met indeling in vier klassen (zie Bijlage E). Dit eindrapport presenteert daarom in de hoofdtekst de verevenende werking van de modellen met indeling in vier klassen.

Voor model M3 is het bovenstaande proces gevolgd met als uitgangspunt alle regiokenmerken in de OT2018. Uiteindelijk resulteert een model met het percentage personen met een niet-westerse migratieachtergrond en regionaal klantenpotentieel. Als dit model geschat wordt met de aangepaste set regiokenmerken op basis van de literatuurstudie, dan resulteren: niet-westerse migratieachtergrond, westerse migratieachtergrond en lage SES in de wijk. Dit is model M3\*.

Een enkel lid van de begeleidingscommissie heeft voorkeur voor een model inclusief OAD. Daarom is getoetst of toevoegen van het kenmerk OAD in model M3\* wél een significante en logische relatie heeft met de meerkosten. Dit blijkt niet het geval: bij toevoegen van OAD aan M3\* is er geen logische relatie tussen stedelijkheid en meerkosten: er resulteert een alternerend patroon:

- 15 voor OAD 1, 13 voor OAD2, -7 voor OAD3, 6 voor OAD4 en 0 voor OAD5. OAD is daarom ook in tweede instantie niet toegevoegd aan model M3\*.<sup>9</sup>

*Voor- en nadelen van model waarin regiokenmerken rechtstreeks zijn opgenomen*

Groot voordeel van een model op individuniveau is dat het risicovereveningsmodel in één keer geschat wordt, inclusief regiokenmerken. Dit kost veel minder onderzoektijd. Bovendien kunnen de effecten zowel voor grote als voor kleine postcodegebieden in één keer geschat worden. Omdat postcodes met meer verzekerdenjaren een groter gewicht krijgen in het bepalen van de coëfficiënten, hebben de kleine postcodegebieden een minder grote invloed op de gemeten effecten. Een ander voordeel is dat de regio-effecten transparant zijn: bijvoorbeeld luchtvervuiling leidt tot hogere zorgkosten. Tot slot wordt zowel het effect van regiokenmerken als het effect van de kenmerken in het risicovereveningsmodel op individuniveau alleen zuiver geschat wanneer alle factoren direct op individuniveau zijn opgenomen in het model. Nadeel is dat het regiocriterium bestaat uit meerdere kenmerken, wat de uitvoering bemoeilijkt wat betreft de verzekerdenraming en de ophoging naar de verzekerdenraming bij de berekening van de normbedragen.

## 4.5 Model op basis van clustering van kenmerken, zonder gebruik van meerkosten (M4 en M4\*)

Een ander alternatief voor het huidige regiomodel is om als regiokenmerk een beperkt aantal samengestelde kenmerken op te nemen in het vereveningsmodel. Dit alternatief neemt dus ook regiokenmerken direct op, maar reduceert het aantal op te nemen regiokenmerken tot een beperkt aantal samengestelde kenmerken.

De samengestelde kenmerken volgen uit een zogenaamde principale-componentenanalyse – zie Box 4.2.

### Box 4.2 Principale-componentenanalyse

Principale-componentenanalyse is een methode om het aantal gecorreleerde kenmerken te reduceren, terwijl het grootste deel van de informatie uit deze kenmerken behouden blijft. De analyse resulteert in een (kleiner) aantal zogenaamde principale componenten die niet gecorreleerd zijn. De eerste component bevat zoveel mogelijk van de variatie in de oorspronkelijke kenmerken. Elk van de volgende componenten omvat zoveel mogelijk van de resterende variatie.

De componenten zijn lineaire combinaties van de achterliggende kenmerken. In welke mate de achterliggende kenmerken meetellen in een component hangt af van het gewicht dat zij in dat component hebben. Het gewicht hangt af van de mate waarin een achterliggend kenmerk bijdraagt aan een bepaald gedeelte van de variatie in de achterliggende kenmerken.

Omdat de principale componenten lineaire combinaties van de achterliggende kenmerken zijn, zijn de waarden die de principale componenten hebben niet eenduidig te interpreteren. De waarden worden immers bepaald door zowel de schaal waarop de achtergrondkenmerken zijn gemeten en de gewichten die de verschillende achtergrondkenmerken krijgen.

Door principale componenten te construeren kan het aantal regiokenmerken dat in het risicovereveningsmodel wordt opgenomen, worden gereduceerd. In plaats van alle relevante achtergrondkenmerken mee te nemen, kunnen de meest belangrijke principale componenten worden opgenomen. Zo kan het aantal kenmerken in het model beperkt worden gehouden, terwijl wel rekening wordt gehouden met een groot deel van de variatie in de regiokenmerken.

<sup>9</sup> Qua verevenende werking op regioniveau resulteert in model M3\* + OAD ook een sterker negatief resultaat voor de G4 en achterstandswijken, ten opzichte van model M3\* (zonder OAD). Zie Bijlage E



De principale-componentenanalyse is allereerst toegepast op de regiokenmerken uit de OT2018 (exclusief technische factoren): het percentage met een niet-westerse migratieachtergrond, regionaal klantenpotentieel, luchtvervuiling, percentage huurwoningen en OAD. Deze variabelen zijn in min of meerdere mate aan elkaar gerelateerd. Bijvoorbeeld grootstedelijke gebieden (OAD 1) hebben doorgaans een hoog regionaal klantenpotentieel (in vergelijking met omliggende gebieden wonen er veel mensen), meer mensen met een niet-westerse migratieachtergrond, meer huurwoningen en meer luchtvervuiling. Deze regiokenmerken worden samengevat in twee principale componenten, waarbij rekening wordt gehouden met het aantal verzekerdenjaren in een bepaald postcodegebied. We selecteren de twee principale componenten die het belangrijkste deel van de variantie in kenmerken verklaren, delen deze in tien gelijke klassen in en nemen deze klassen met bijbehorende restricties (de normkosten tellen over alle klassen van de component op tot nul) direct op in het somatische vereveningsmodel.

Postcodes krijgen een gewicht dat evenredig is aan het aantal verzekerdenjaren. Omdat postcodes met meer verzekerdenjaren een groter gewicht krijgen in het bepalen van factoren, zijn de kleine postcodegebieden niet van grote invloed op de factoren. We nemen daarom zowel grote als kleine postcodegebieden mee in de analyse.

Model M4 construeert twee principale componenten uitgaande van alle regiokenmerken uit de OT2018 (exclusief normkosten somatische zorg). De eerste component bevat 62 procent van de variatie in de regiokenmerken, de tweede component voegt daar nog eens 17 procent aan toe, zodat de twee factoren samen bijna 80 procent van de variatie in de regiokenmerken verklaren. Model M4 neemt deze twee componenten, ingedeeld in tien klassen met een gelijk aantal verzekerdenjaren, direct op.

Model M4\* past dezelfde procedure toe op de aangepaste set regiokenmerken op basis van de literatuurstudie. Alle regiokenmerken die ook in de opbouw naar model M0\* zijn getoetst, zijn hierin meegenomen. Er zijn geen regiokenmerken vooraf verwijderd op basis van de uitkomsten uit de voorgaande modellen. De eerste twee componenten die hieruit volgen verklaren tezamen 44 procent van de variatie in de regiokenmerken. Met name migratieachtergrond, overlast, corporatiewoning, zeer lage SES, benzeen en stikstofdioxide hebben een hoog gewicht in de eerste component. In de tweede component zijn speciaal onderwijs, corporatiewoning, woningen van slechte kwaliteit en (zeer) lage SES van belang. Model M4\* neemt deze twee componenten, ingedeeld in tien klassen met een gelijk aantal verzekerdenjaren, direct op.

#### *Voor- en nadelen principale-componentenanalyse*

Voordeel van de principale-componentenanalyse is dat een groot aantal regiovariabelen kan worden meegenomen als verklaring van de variatie in zorgkosten op regioniveau. De procedure reduceert immers een groot aantal variabelen tot de twee belangrijkste componenten. Ook al wordt dus een groot aantal regiokenmerken meegewogen, het aantal variabelen in het risicovereveningsmodel verandert in beperkte mate. Bovendien kan het risicovereveningsmodel in één keer geschat worden, aangezien de componenten uitsluitend gebaseerd zijn op de regiokenmerken.

Nadeel van de principale-componentenanalyse is dat de samengestelde factor niet is gebaseerd op minimaliseren van de variantie van de *meerkosten*, maar op minimaliseren van de variantie tussen *regiokenmerken*. De regiokenmerken die volgen uit de literatuurstudie en zijn meegenomen in model

M4\* hebben echter allemaal theoretisch een relatie met de zorgkosten, zie Hoofdstuk 3. Ten opzichte van M3/M3\* is een nadeel dat de componenten minder transparant zijn: ze bestaan uit meerdere achterliggende kenmerken waardoor de geschatte relatie met zorgkosten niet theoretisch te beoordelen is en het niet direct duidelijk is welke regiokenmerken in welke mate van belang zijn. Bovendien hebben de waarden van de principale componenten geen logische interpretatie aangezien het een samenstelling is van tal van (gestandaardiseerde) achterliggende regiokenmerken.

## 4.6 Model waarin regiovariabelen rechtstreeks zijn opgenomen, vervolgens indeling in regioclusters (M5 en M5\*)

Een laatste alternatief is om in een eerste stap de regiovariabelen direct op te nemen in het risicovereveningsmodel, en in een tweede stap de postcodegebieden te clusteren in 10 groepen op basis van de zorgkosten zoals voorspeld door de regiovariabelen uit stap 1. Dit alternatief is tijdens het onderzoek door ESHPM ingebracht.

Model M5 is in de eerste stap precies hetzelfde als model M3: de kenmerken percentage personen met een niet-westerse migratieachtergrond en regionaal klantenpotentieel zijn meegenomen, ieder in vier klassen. Model M5\* is in de eerste stap precies hetzelfde als model M3\*: de kenmerken niet-westerse migratieachtergrond, westerse migratieachtergrond en lage SES op pc-4 niveau zijn meegenomen, ieder in vier klassen.

### *Voor- en nadelen van model waarin regiokenmerken rechtstreeks zijn opgenomen en vervolgens geclusterd*

Dit model heeft alle voordelen van het model waarin regiokenmerken rechtstreeks zijn opgenomen (M3/M3\*). Extra voordeel is echter dat het regiocriterium uit één kenmerk bestaat, wat in vergelijking met model M3/M3\* de uitvoering makkelijker maakt wat betreft de verzekerdensraming en de ophoging naar de verzekerdensraming bij de berekening van de normbedragen. Bovendien geeft dit de mogelijkheid om in de toekomst in de eerste stap nog extra regiokenmerken toe te voegen. Extra nadeel in vergelijking met model M3/M3\* is dat er in deze variant nog steeds twee stappen nodig zijn om het vereveningsmodel te kunnen schatten.

## 4.7 Alternatieve technieken die hier niet verder verkend worden

Er zijn nog andere manieren mogelijk om regiokenmerken op te nemen in het risicovereveningsmodel. We noemen er twee en de redenen waarom zij niet verder zijn verkend.

### **Model op basis van directe clustering van meerkosten**

Het is mogelijk om de een regio-indeling direct te baseren op een clustering van meerkosten. In dit geval wordt Wards methode dus niet gebaseerd op het *voorspelde* residu<sup>10</sup> – zoals in de huidige methode waarbij de indeling in 10 groepen van gelijke grootte is losgelaten – maar direct op het residu zelf. De somatische meerkosten zijn immers gelijk aan het residu van het somatische risicovereveningsmodel (werkelijke kosten – normbedrag).

<sup>10</sup> Van het risicovereveningsmodel exclusief regiocriterium, gemiddeld op postcode4 niveau.

Reden waarom deze methode niet verder is verkend, is dat regionale meerkosten die worden gedreven door aanbodfactoren ook (volledig) gehonoreerd worden. Daarnaast wordt hiermee doelmatig beleid van de zorgverzekeraars niet beloond. Als een zorgverzekeraar regionaal actief is en bewerkstelligt dat de zorgkosten dalen door doelmatig beleid, wordt deze doelmatigheidswinst vervolgens via deze methode verevend. Bovendien ontbreekt een theoretische fundering bij een directe indeling in regioclusters op basis van de meerkosten.

### **Hiërarchisch model waarin regiovariabelen rechtstreeks zijn opgenomen**

Het is mogelijk om de regiovariabelen direct op te nemen in het risicovereveningsmodel, waarbij het somatische risicovereveningsmodel geen OLS meer is, maar een hiërarchisch model ofwel multiniveaumodel. In aanvulling op de OLS neemt een multiniveaumodel ook niet-geobserveerde (random) effecten op verschillende niveaus op.

Redenen waarom deze methode niet verder is verkend zijn:

- De random effecten zorgen voor een langere schattingsstijd van het model. Voor het bijstandsverdeelmodel 2015 en 2016 is een multiniveaumodel gehanteerd op een steekproef. Op integrale gegevens, voor alle individuen in Nederland, kon zo'n model met niet-geobserveerde kenmerken op drie niveaus door zijn complexiteit soms helemaal niet geschat worden;
- Een multiniveaumodel is alleen een zuivere schatter wanneer de random effecten op alle niveaus (multivariaat) normaal verdeeld zijn met constante variantie en gemiddelde nul (Leeuw en Meijer, 2007). Aan deze aanname is vaak niet voldaan.

Beide problemen met hiërarchische modellen (complexiteit en onzuivere schatter door stringente aannames) maken dit geen goede kandidaat voor het risicovereveningsmodel. Daarom wordt ook in het bijstandsverdeelmodel vanaf 2017 geen hiërarchisch model meer geschat (Tempelman et al., 2016).



## 5 Meerkosten en modelschattingen

*De verschillende varianten voor het regiomodel hebben weinig gevolgen voor de verevenende werking op individuniveau. Ook de normbedragen voor de andere factoren veranderen nauwelijks. Op regionaal niveau zijn wel verschillen zichtbaar in verevenende werking.*

Dit hoofdstuk presenteert de meerkosten, modelschattingen en vereveningscriteria voor de verschillende modellen. Tabel 5.1 vat deze modellen samen en geeft weer welke regionale kenmerken in het model zitten.

**Tabel 5.1 Samenvatting geschatte modellen**

Naam	Omschrijving	Kenmerken
M00	OT 2018 exclusief regiokenmerk	Geen regiokenmerken
M0	Regiomodel OT 2018	Normkosten somatische zorg, niet-westerse migratieachtergrond, regionaal klantenpotentieel, luchtvervuiling (PM10), huurwoningen, OAD
M0*	Regiomodel met aangepaste set regiokenmerken	Constante, niet-westerse migratieachtergrond, westerse migratieachtergrond, ziekteverzuim, SES laag inkomen
M1	Loslaten indeling 10 gelijke groepen	Zelfde als M0
M1*	Loslaten indeling 10 gelijke groepen	Zelfde als M0*
M2	Clustering van kenmerken, aan de hand van meerkosten	Zelfde als M0, excl. normkosten
M2*	Clustering van kenmerken, aan de hand van meerkosten	Zelfde als M0*, excl. constante
M3	Regiovariabelen rechtstreeks opnemen	Niet-westerse migratieachtergrond, regionaal klantenpotentieel
M3*	Regiovariabelen rechtstreeks opnemen	Niet-westerse migratieachtergrond, westerse migratieachtergrond, SES laag inkomen
M4	Principale-componentenanalyse	Twee principale componenten geconstrueerd uit niet-westerse migratieachtergrond, regionaal klantenpotentieel, luchtvervuiling (PM10), huurwoningen, OAD
M4*	Principale-componentenanalyse	Twee principale componenten geconstrueerd uit: percentages Marokkanen, Turken, Surinamers, Antillianen, overig niet-westers, MOE-landers, overig westers en zonder migratieachtergrond, percentage speciaal onderwijs, overlast, percentage corporatiewoning, percentages goede, redelijke en slechte woningen, bomen/bos, natuurlijk terrein, geluidsoverlast, benzeen, koolmonoxide, ammoniak, stikstofdioxide, ozon, zwaveldioxide, fijnstof PM2,5, zeer lage, lage, midden en hoge SES en ziekteverzuim
M5	Rechtstreeks opnemen en clusteren	Zelfde als M3
M5*	Rechtstreeks opnemen en clusteren	Zelfde als M3*

Bron: SEO Economisch Onderzoek.

### 5.1 Meerkosten en verzekerdenjaren

Tabel 5.2 laat zien dat de meerkosten tussen regioclusters soms sterk verschillen, al naar gelang de methode om de regioclusters samen te stellen. Opvallend is dat de meerkosten van de regioclusters in model M2 en M2\* een aanzienlijk grotere bandbreedte hebben dan bij de huidige methode (M0),

M0\*), de huidige methode bij loslaten van 10 gelijke groepen (M1, M1\*), en het rechtstreeks opnemen en daarna clusteren (M5, M5\*).

Dit komt omdat bij model M2 (M2\*) geen regressiemodel is gebruikt, dus niet de voorspelde residuen maar de ‘kale’ residuen op pc4-niveau worden direct gebruikt om clusters te vormen. Hierdoor vangt het model meer variatie op. De  $R^2$  van het huidige regiomodel is slechts 6,5 procent, dus een groot deel van de variatie in regionale meerkosten is onverklaard. Deze hele variatie wordt wel gebruikt in model M2. De vraag is of dit wenselijk is: de ‘extra’ variatie die wordt meegenomen wordt immers niet veroorzaakt door theoretische factoren. De modellen M2 en M2\* zijn naar verwachting ook minder stabiel. Bijvoorbeeld als het risicovereveningsmodel wijzigt, wijzigen de meerkosten op regionaal niveau mee, en deze wijzigingen worden één op één opgepakt door model M2 (M2\*) omdat direct gebruikgemaakt wordt van de gemiddelde meerkosten per klasse.

De standaarddeviatie op regioclusterniveau van modellen M2 en M2\* (berekend met de ‘kale’ residuen) is logischerwijs ook hoger dan in de modellen die gebruikmaken van de voorspelde residuen (waarbij de standaarddeviatie die is berekend op basis van de voorspelde residuen is gebruikt). Het regiomodel slaat immers enige variatie ‘plat’: de voorspelde residuen zullen nooit zulke uitschieters kennen als de werkelijke residuen. De laagste standaarddeviatie wordt bereikt door model M1\*, het model met de nieuwe kenmerken en loslaten van de 10 gelijke groepen.

**Tabel 5.2** Bandbreedte van de meerkosten per regiocluster fors hoger bij modellen M2, M2\*

Cluster	M0	M0*	M1	M1*	M2	M2*	M5	M5*
Regiocluster 1	57	58	76	69	574	252	45	46
Regiocluster 2	38	26	63	47	264	175	22	41
Regiocluster 3	11	18	52	24	161	96	6	11
Regiocluster 4	8	8	22	28	108	52	4	4
Regiocluster 5	7	-4	9	10	66	23	4	3
Regiocluster 6	-7	-17	4	0	37	7	-13	1
Regiocluster 7	-17	-24	-13	-20	0	-9	-17	-13
Regiocluster 8	-34	-16	-38	-16	-31	-26	-29	-19
Regiocluster 9	-34	-22	-25	-29	-79	-50		-32
Regiocluster 10	-28	-26	-34	-30	-183	-97		-30
Regiocluster 11			-162	-13	-422	-167		
Std. dev. Clusterniveau (in duizenden)	50.390	45.491	48.109	43.313	74.584	85.819		48.479

Bron: OT2018-bestand, berekeningen SEO Economisch Onderzoek. Clusters gesorteerd in oplopende volgorde van meerkosten.

Het verschil tussen de methodes is ook zichtbaar in het aantal verzekerdenjaren per regiocluster (zie Tabel 5.3). Model M0 en M0\* zijn gebaseerd op 10 (ongeveer) even grote groepen. Model M1 en M1\* laten dit principe los, maar zijn nog wel gebaseerd op een regiomodel. De voorspelde residuen verschillen niet sterk tussen pc4-gebieden, waardoor geen (zeer) kleine groepen ontstaan.<sup>11</sup> Dit geldt niet voor modellen M2 en M2\*: omdat de (werkelijke) residuen sterk verschillen

<sup>11</sup> Een uitzondering is regiocluster 11 in model M1. Dit is een regiocluster met zeer weinig verzekerdenjaren bestaande uit pc4-gebieden met zeer hoge meerkosten én zeer lage meerkosten. Gemiddeld zijn de meerkosten negatief, maar de standaarddeviatie in deze groep is ongeveer even groot als het gemiddelde zelf.

tussen de klassen die de basis vormen voor de clusterindeling, kan een aantal zeer kleine clusters ontstaan. Dit bevestigt de verwachte instabiliteit van deze methode: de kleine klassen worden gedreven door een beperkt aantal postcodegebieden met een afwijkende combinatie van regiokenmerken en gemiddeld zeer hoge (of juist lage) meerkosten. Deze gemiddeld zeer hoge of lage meerkosten kunnen juist ontstaan doordat het om zeer kleine groepen gaat: bij kleine groepen worden de gemiddelde meerkosten sterk gedreven door een enkele uitschieter.

Merk ook op dat bij model M5/M5\* de resulterende regioclusters niet van gelijke grootte zijn. Dit komt omdat in de eerste stap de regiokenmerken zijn ingedeeld in klassen. Bijvoorbeeld in model M5 gaat het om twee kenmerken, ieder in vier klassen. Dit zorgt ervoor dat uit de eerste stap  $4 * 4 = 16$  waarden resulteren voor de voorspelde regionale zorgkosten op basis van de geschatte regio-coëfficiënten. Eén van deze waarden omvat zelfs 20 procent van de verzekerden, enkele andere waarden omvatten tot 15 procent van de verzekerden. Hierdoor zijn niet 10 maar 8 clusters zijn opgenomen in dit model. Bij model M5\* gaat het om 3 kenmerken waardoor het aantal verzekerden per regiocluster minder sterk van elkaar verschilt.

Door in de eerste stap de regiokenmerken in te delen in klassen wordt het stramien van het vereveningsmodel gevolgd en wordt rekening gehouden met een mogelijk niet-lineaire relatie tussen regiokenmerken en zorgkosten. Indien 10 gelijke groepen toch gewenst zijn, is een alternatief om de regiokenmerken in de eerste stap als continue variabele op te nemen. In dat geval dient in de eerste stap wel een constante meegenomen te worden om de regio-coëfficiënten zuiver te schatten. Zo wordt voorkomen dat de weggelaten constante zich gaat ‘verdelen’ over de continue variabelen die zijn opgenomen in het model.

**Tabel 5.3** M2 en M2\* identificeren een aantal groepen met een klein aantal verzekerdenjaren

Cluster	M0	M0*	M1	M1*	M2	M2*	M5	M5*
Regiocluster 1	1.649.516	1.652.680	182.648	548.272	1.403	935	1.523.443	1.573.345
Regiocluster 2	1.667.966	1.656.315	635.017	1.465.471	3.355	1.632	1.481.593	1.850.382
Regiocluster 3	1.663.555	1.664.218	1.471.303	1.064.648	19.596	173.841	2.607.616	1.474.516
Regiocluster 4	1.660.450	1.674.526	1.760.059	1.049.323	261.306	1.469.269	3.401.296	1.293.095
Regiocluster 5	1.700.974	1.676.883	1.918.512	1.568.717	1.140.944	3.629.397	685.790	954.173
Regiocluster 6	1.670.007	1.672.303	2.866.652	2.511.110	2.367.281	2.856.200	2.440.322	1.946.635
Regiocluster 7	1.672.199	1.677.422	2.889.640	3.202.218	8.108.999	2.728.774	2.424.135	1.797.717
Regiocluster 8	1.677.600	1.669.710	2.055.046	2.653.418	3.938.360	4.854.838	2.107.021	2.309.191
Regiocluster 9	1.653.088	1.662.750	1.971.836	1.232.097	768.922	920.139		1.924.507
Regiocluster 10	1.655.861	1.664.411	918.522	1.037.078	55.896	35.044		1.547.655
Regiocluster 11			1.982	338.865	5.155	1.147		

Bron: OT2018-bestand, berekeningen SEO Economisch Onderzoek. Clusters gesorteerd in oplopende volgorde van meerkosten.

In model M3 en M3\* zijn de regionale kenmerken direct op individuniveau opgenomen. Tabel 5.4 laat het aantal verzekerdenjaren en de meerkosten per klasse zien. De grenswaarden voor de indeling in vier klassen zijn gekozen op basis van de meerkosten bij een indeling van het kenmerk in 10 gelijke klassen, zie Bijlage C. Op basis van deze meerkosten zijn één groep met negatieve meerkosten geselecteerd, één groep met meerkosten rond nul en twee groepen met positieve meerkosten.

De meerkosten voor een gebied kunnen hoger worden dan de meerkosten voor regiocluster 10 in M0: een regio die in M3\* zowel veel personen met een niet-westerse als westerse migratieachtergrond heeft en een hoog aandeel mensen in een regio met SES laag inkomen heeft nu in totaal € 97 aan meerkosten. Regio's die op al deze factoren laag scoren hebben gemiddelde meerkosten van € -45. Door het grote aantal verzekerdenjaren per klasse zal dit resultaat naar verwachting stabiel zijn.

**Tabel 5.4 Meerkosten per regiokenmerk vertonen oplopend patroon**

Kenmerk	M3		M3*	
	Verzekerden jaren	Meerkosten	Verzekerden jaren	Meerkosten
Regionaal klantenpotentieel:				
- 1 (minder dan 20)	2.164.659	-28		
- 2 (20 tot 40)	2.611.579	-19		
- 3 (40 tot 80)	3.201.940	-8		
- 4 (80 of meer)	8.693.040	15		
Niet-westerse migratieachtergrond:				
- 1 (minder dan 10%)	10.057.576	-11	10.057.576	-11
- 2 (10% tot 20%)	3.550.573	2	3.550.573	2
- 3 (20% tot 30%)	1.520.034	22	1.520.034	22
- 4 (30% of meer)	1.543.034	43	1.543.034	43
Westerse migratieachtergrond:				
- 1 (minder dan 8%)			6.582.359	-13
- 2 (8% tot 12%)			6.567.089	6
- 3 (12% tot 16%)			2.063.253	12
- 4 (16% of meer)			1.458.517	15
SES laag inkomen:				
- 1 (minder dan 18%)			5.579.519	-21
- 2 (18% tot 22%)			5.976.166	-3
- 3 (22% tot 26%)			3.937.876	22
- 4 (26% of meer)			1.177.656	39

Bron: OT2018-bestand, berekeningen SEO Economisch Onderzoek.

Tot slot zijn in model M4 en M4\* principale componenten op individuniveau opgenomen die zijn geconstrueerd uit de regionale kenmerken. Tabel 5.5 laat het aantal verzekerdenjaren en de meerkosten per klasse van elk van de componenten zien. De meerkosten lopen in principe op met de omvang van de component, alhoewel de klassenindeling van de principale componenten geen logische interpretatie heeft. De meerkosten zijn in omvang vergelijkbaar met M3 en M3\*.

**Tabel 5.5 Meerkosten per principale component vertonen oplopend patroon**

Principale componenten	M4		M4*	
	verzekerdenjaren	meerkosten	verzekerdenjaren	meerkosten
Principale component 1 o.b.v. oude regiokenmerken:				
- 1 (minder dan -1,99)	1.671.407	-25		
- 2 (tussen -1,99 en -1,48)	1.664.082	-4		
- 3 (tussen -1,48 en -1,14)	1.678.921	-15		



Principale componenten	M4		M4*	
	verzekerd-jaren	meerkosten	verzekerd-jaren	meerkosten
- 4 (tussen -1,14 en -0,72)	1.654.413	-17		
- 5 (tussen -0,72 en -0,32)	1.673.240	-5		
- 6 (tussen -0,32 en 0,20)	1.671.177	6		
- 7 (tussen 0,20 en 0,72)	1.659.352	7		
- 8 (tussen 0,72 en 1,44)	1.667.064	17		
- 9 (tussen 1,44 en 2,47)	1.667.592	-3		
- 10 (meer dan 2,47)	1.663.968	38		
Principale component 2 o.b.v. oude regiokenmerken:				
- 1 (minder dan -1,39)	1.683.182	-4		
- 2 (tussen -1,39 en -0,71)	1.656.804	23		
- 3 (tussen -0,71 en -0,35)	1.665.419	28		
- 4 (tussen -0,35 en -0,12)	1.665.240	9		
- 5 (tussen -0,12 en 0,09)	1.665.169	8		
- 6 (tussen 0,09 en 0,32)	1.675.194	0		
- 7 (tussen 0,32 en 0,55)	1.666.338	-9		
- 8 (tussen 0,55 en 0,78)	1.663.311	-18		
- 9 (tussen 0,78 en 1,07)	1.664.484	-17		
- 10 (meer dan 1,07)	1.666.075	-20		
Principale component 1 o.b.v. nieuwe regiokenmerken:				
- 1 (minder dan -3,58)			1.667.416	-13
- 2 (tussen -3,58 en -2,65)			1.670.810	-14
- 3 (tussen -2,65 en -1,94)			1.670.606	-7
- 4 (tussen -1,94 en -1,26)			1.661.911	-11
- 5 (tussen -1,26 en -0,53)			1.666.156	-5
- 6 (tussen -0,53 en 0,20)			1.677.115	1
- 7 (tussen 0,20 en 1,18)			1.660.041	0
- 8 (tussen 1,18 en 2,62)			1.665.198	9
- 9 (tussen 2,62 en 4,43)			1.673.700	8
- 10 (meer dan 4,43)			1.658.267	33
Principale component 2 o.b.v. nieuwe regiokenmerken:				
- 1 (minder dan -3,47)			1.669.739	-24
- 2 (tussen -3,47 en -2,49)			1.664.608	-13
- 3 (tussen -2,49 en -1,70)			1.667.100	-19
- 4 (tussen -1,70 en -1,06)			1.667.313	-16
- 5 (tussen -1,06 en -0,39)			1.666.892	-1
- 6 (tussen -0,39 en 0,34)			1.667.992	-10
- 7 (tussen 0,34 en 1,23)			1.668.547	13
- 8 (tussen 1,23 en 2,51)			1.675.617	4
- 9 (tussen 2,51 en 4,18)			1.663.544	22
- 10 (meer dan 4,18)			1.659.865	44

Bron: OT2018-bestand, berekeningen SEO Economisch Onderzoek.

## 5.2 Normbedragen

Tabel 5.6 toont de geschatte normbedragen voor de regioclusters. Deze volgen logischerwijs de berekende meerkosten uit Tabel 5.2. Omdat het somatische vereveningsmodel een lineair model is (ROLS) zijn na opname van de regioclusters de gemiddelde meerkosten per cluster (nagenoeg) gelijk aan nul. De overige normbedragen schuiven hooguit enkele euro's per verzekerdenjaar, zie Bijlage D voor een volledig overzicht van de normbedragen van alle modellen. De grootste verschuivingen zijn te zien bij de bijstandsgerechtigden. Bijstandsgerechtigden komen immers relatief vaak voor in grote steden en wonen vaak in huurwoningen. Ook hebben personen met een (niet-westerse) migratieachtergrond een hogere kans op bijstand en hebben bijstandsgerechtigden een laag inkomen (lage SES).

**Tabel 5.6 Normbedragen regiocriterium volgt meerkosten per regiocluster**

Vereveningscriterium	M0	M0*	M1	M1*	M2	M2*	M5	M5*
Regiocluster 1	59	63	79	74	573	259	47	50
Regiocluster 2	39	29	66	51	265	175	23	43
Regiocluster 3	11	20	53	27	161	98	7	13
Regiocluster 4	8	8	23	30	109	55	4	4
Regiocluster 5	7	-4	9	11	67	25	3	1
Regiocluster 6	-7	-17	4	0	38	8	-13	2
Regiocluster 7	-18	-25	-14	-21	0	-10	-18	-14
Regiocluster 8	-35	-18	-39	-18	-32	-27	-30	-20
Regiocluster 9	-34	-25	-26	-32	-79	-51		-32
Regiocluster 10	-30	-29	-35	-34	-184	-96		-34
Regiocluster 11			-157	-16	-421	-169		

Bron: OT2018-bestand, berekeningen SEO Economisch Onderzoek.

Ook voor modellen M3 en M3\* geldt dat de normbedragen de gemiddelde meerkosten per kenmerk volgen, zie Tabel 5.7. De normbedragen zijn beperkt: in regio's met veel personen met een niet-westerse migratieachtergrond is het normbedrag in M3\* € 31 en in regio's met weinig personen met een niet-westerse migratieachtergrond is het normbedrag € -6. Voor de overige individuele kenmerken geldt dat de normbedragen vrijwel gelijk blijven, zie Bijlage D.

**Tabel 5.7 Relatief lage normbedragen per regiokenmerk**

Kenmerk	M3	M3*
Regionaal klantenpotentieel:		
- 1 (minder dan 20)	-23	
- 2 (20 tot 40)	-13	
- 3 (40 tot 80)	-5	
- 4 (80 of meer)	12	
Niet-westerse migratieachtergrond:		
- 1 (minder dan 10%)	-7	-6
- 2 (10% tot 20%)	-3	-2
- 3 (20% tot 30%)	15	14
- 4 (30% of meer)	35	31
Westerse migratieachtergrond:		

Kenmerk	M3	M3*
- 1 (minder dan 8%)		-9
- 2 (8% tot 12%)		4
- 3 (12% tot 16%)		8
- 4 (16% of meer)		13
SES laag inkomen:		
- 1 (minder dan 18%)		-20
- 2 (18% tot 22%)		-2
- 3 (22% tot 26%)		19
- 4 (26% of meer)		42

Bron: OT2018-bestand, berekeningen SEO Economisch Onderzoek.

Tot slot geldt ook voor modellen M4 en M4\* dat de normbedragen de gemiddelde meerkosten per kenmerk volgen, zie Tabel 5.8. De normbedragen zijn ook hier beperkt in omvang en variëren van € -32 tot 36 in model M4 en van € -32 tot 47 in model M4\*. Voor de overige individuele kenmerken geldt dat de normbedragen vrijwel gelijk blijven, zie Bijlage D.

**Tabel 5.8** Relatief lage normbedragen per principale componenten in model M4 en M4\*

Principale componenten	M4	M4*
Principale component 1 o.b.v. oude regiokenmerken:		
- 1 (minder dan -1,99)	-32	
- 2 (tussen -1,99 en -1,48)	-4	
- 3 (tussen -1,48 en -1,14)	-15	
- 4 (tussen -1,14 en -0,72)	-14	
- 5 (tussen -0,72 en -0,32)	-2	
- 6 (tussen -0,32 en 0,20)	5	
- 7 (tussen 0,20 en 0,72)	7	
- 8 (tussen 0,72 en 1,44)	18	
- 9 (tussen 1,44 en 2,47)	0	
- 10 (meer dan 2,47)	36	
Principale component 2 o.b.v. oude regiokenmerken:		
- 1 (minder dan -1,39)	8	
- 2 (tussen -1,39 en -0,71)	24	
- 3 (tussen -0,71 en -0,35)	24	
- 4 (tussen -0,35 en -0,12)	6	
- 5 (tussen -0,12 en 0,09)	1	
- 6 (tussen 0,09 en 0,32)	-5	
- 7 (tussen 0,32 en 0,55)	-12	
- 8 (tussen 0,55 en 0,78)	-20	
- 9 (tussen 0,78 en 1,07)	-14	
- 10 (meer dan 1,07)	-13	
Principale component 1 o.b.v. nieuwe regiokenmerken:		
- 1 (minder dan -3,58)		-29
- 2 (tussen -3,58 en -2,65)		-17
- 3 (tussen -2,65 en -1,94)		-7
- 4 (tussen -1,94 en -1,26)		-7
- 5 (tussen -1,26 en -0,53)		1
- 6 (tussen -0,53 en 0,20)		6

Principale componenten	M4	M4*
- 7 (tussen 0,20 en 1,18)		7
- 8 (tussen 1,18 en 2,62)		9
- 9 (tussen 2,62 en 4,43)		10
- 10 (meer dan 4,43)		27
Principale component 2 o.b.v. nieuwe regiokenmerken:		
- 1 (minder dan -3,47)		-32
- 2 (tussen -3,47 en -2,49)		-17
- 3 (tussen -2,49 en -1,70)		-20
- 4 (tussen -1,70 en -1,06)		-15
- 5 (tussen -1,06 en -0,39)		1
- 6 (tussen -0,39 en 0,34)		-10
- 7 (tussen 0,34 en 1,23)		16
- 8 (tussen 1,23 en 2,51)		6
- 9 (tussen 2,51 en 4,18)		25
- 10 (meer dan 4,18)		47

Bron: OT2018-bestand, berekeningen SEO Economisch Onderzoek.

Tabel 5.9 vat de verschuivingen in de normbedragen per model en per vereveningskenmerk samen. Voor de modellen mét regiokenmerken geldt dat alleen de (gewogen) gemiddelde absolute verschuiving in de normbedragen van de kenmerken leeftijd/geslacht, SES en/of AVI méér bedraagt dan een euro. Alleen in de modellen op individuniveau (M3/M3\*, M4/M4\*) wordt direct rekening gehouden met de samenhang tussen individuele kenmerken en kenmerken op regioniveau, waardoor ook de individuele kenmerken zuiver geschat worden. De verschuiving die dit teweegbrengt in de kenmerken op individuniveau is echter beperkt.

**Tabel 5.9** Kleine verschuivingen in normbedragen (GGAV t.o.v. M0, per vereveningskenmerk)

Verevenings-criterium	M00	M0	M0*	M1	M1*	M2	M2*	M3	M3*	M4	M4*	M5	M5*
Leeftijd/ geslacht	0,7	Nvt	1,5	0,1	1,6	1,0	1,1	1,8	1,3	1,2	1,1	1,7	1,5
FKG	0,3	Nvt	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1
pDKG	0,0	Nvt	0,0	0,0	0,0	0,0	0,0	0,1	0,0	0,1	0,0	0,1	0,0
sDKG	0,0	Nvt	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
HKG	1,0	Nvt	1,0	0,0	1,0	0,0	0,0	0,0	0,0	0,0	1,0	0,0	0,0
AVI	1,5	Nvt	0,4	0,1	0,3	0,3	0,3	1,2	0,3	1,0	0,3	1,1	0,4
SES	3,4	Nvt	1,9	0,0	1,9	1,0	1,1	1,3	1,2	0,8	1,6	1,3	1,2
PPA	0,3	Nvt	0,3	0,0	0,3	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
MHK	0,1	Nvt	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
FDG	0,0	Nvt	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
VGG	0,0	Nvt	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

Bron: OT2018-bestand, berekeningen SEO Economisch Onderzoek. Bedragen afgerond op tientallen eurocenten.

## 5.3 Maatstaven voor verevenende werking

Alle modellen worden beoordeeld op hun verevenende werking aan de hand van de bekende kwantitatieve normen (hoofdstuk WOR 871):

Tabel 5.10 Voor elk model worden 7 vereveningsmaatstaven beoordeeld

Niveau	Maatstaf	Omschrijving
Individu	R <sup>2</sup>	% verklaarde variantie van kosten
	CPM	% van de absolute verschillen in kosten tussen individuen dat door het model wordt verklaard
Subgroep	GGAA	gewogen gemiddelde absolute verschil tussen werkelijke kosten en normatieve kosten, gemiddeld per subgroep. weging met aantal verzekerdenjaren per subgroep
	Resultaat	gewogen gemiddelde financiële resultaat (normkosten – werkelijke kosten) per verzekerdenjaar voor (a) 15% met laagste kosten in t-3, (b) 15% met hoogste kosten in t-1 en (c) VGG(t-1)>0
Verzekeraar	GGAA	gewogen gemiddelde absolute verschil tussen werkelijke kosten en normatieve kosten, gemiddeld per verzekerdenjaar
	Bandbreedte	verschil tussen meest positieve en meest negatieve financiële resultaat, gemiddeld per verzekerdenjaar
	GGARV	gewogen gemiddelde absolute resultaatverschuiving, tussen 2 modellen, gemiddeld per verzekerdenjaar.

Bron: WOR 871 en SEO Economisch Onderzoek.

Voor beoordeling van de regiokenmerken is ook de prestatie op regioniveau van belang. De regiokenmerken hebben immers vooral een invloed op regioniveau. Aanvullend zijn daarom ook op postcodeniveau en op gemeenteniveau het resultaat, de bandbreedte van het resultaat en de GGARV berekend. Het resultaat is daarnaast ook berekend voor de G4, de G21 en de rest van NL, voor achterstandswijken en voor de Randstad versus aanpalende periferie.<sup>12</sup>

Een goed referentiepunt om de relatieve prestaties van de modellen op regioniveau te kunnen beoordelen ontbreekt. Immers: een gemiddeld resultaat dat dicht bij nul ligt is niet per definitie gewenst. Afwijkingen van nul kunnen bijvoorbeeld ook ontstaan door verschillen in zorgaanbod (bijvoorbeeld veel aanbod in de G4) en regionaal inkoopbeleid van verzekeraars. De behandelprijzen tussen ziekenhuizen verschillen immers enorm (Douven et al., 2018). Onverklaarbaar grote uitschieters (van meer dan € 10 of minder dan - € 10) lijken echter ongewenst.

Met deze beslisregel in gedachten kunnen we stellen dat de verevenende werking van de modellen mét regiokenmerken de voorkeur heeft boven de verevenende werking van het model M00 zonder regiokenmerken. De afwijkingen voor elk van de regionale indelingen zijn immers groter in absolute waarden voor model M00, zie Tabel 5.11.

Tabel 5.11 laat ook zien dat bij vrijwel alle modellen de varianten met de nieuwe kenmerken (\*) minder gunstig uitpakken voor de grote gemeenten, en gunstiger voor de rest van Nederland en voor achterstandswijken. Dat deze modellen gunstiger uitpakken voor achterstandswijken is waarschijnlijk te danken aan het kenmerk 'lage SES' dat als nieuw kenmerk is opgenomen. Grote steden komen iets minder gunstig uit, waarschijnlijk omdat OAD en regionaal klantenpotentieel niet meer in deze modellen zitten. Model M2 wijkt wederom af en laat zowel bij de G4 als bij achterstandswijken grote negatieve resultaten zien. Over het algemeen is te zien dat de wijze waarop rekening gehouden wordt met regio op individuniveau niet tot grote verschillen in de verevenende werking leidt.

<sup>12</sup> Het gaat hier om de Randstad, de intermediaire zone om de Randstad (o.a. Gelderland en Brabant) en de rest van Nederland

Tabel 5.11 Substantiële verschillen tussen de modellen op regionaal niveau

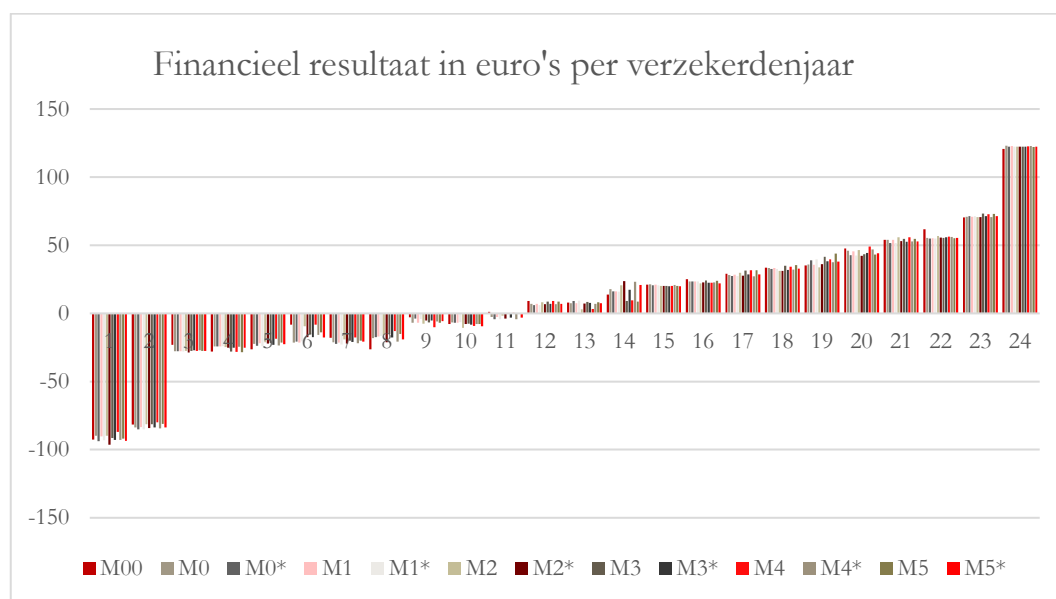
Maatstaf	M00	M0	M0*	M1	M1*	M2	M2*	M3	M3*	M4	M4*	M5	M5*
<b>Individu</b>													
R <sup>2</sup> x 100%	32,1	32,1	32,1	32,1	32,1	32,1	32,1	32,1	32,1	32,1	32,1	32,1	32,1
CPM x 100%	31,9	31,9	31,9	31,9	31,9	31,8	31,9	31,9	31,9	31,9	31,9	31,9	31,9
<b>Subgroep</b>													
GGAA (€/verjaar)	808	808	808	808	808	808	808	808	808	808	808	808	808
Resultaat 15% laagste kosten t-3 (€/verjaar)	106	104	105	104	105	104	104	104	105	104	105	104	105
Resultaat 15% hoogste kosten t-3 (€/verjaar)	-134	-134	-134	-134	-134	-134	-134	-134	-134	-134	-134	-134	-134
Resultaat VGG>0 (€/verjaar)	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>Resultaat regionaal niveau (€/verjaar)</i>													
G4	-25,3	-1,9	-3	0,1	-3,3	-11	-7,4	3,9	-3,8	0,7	-11,3	2,8	-3,3
G21	-13,3	2,5	-0,4	3,2	-0,5	-1,2	-2,4	6,9	-1,6	3,1	-4,5	5,9	-0,7
Rest van NL	5,7	-1,1	0,2	-1,4	0,2	0,5	1	-3	0,7	-1,3	1,9	-2,5	0,3
Achterstandswijken	-34,8	-4,6	3,3	-3,3	4,1	-19,9	-7,1	-5,4	-1,7	-9,3	-2,6	-5,5	-3,1
Randstad	-4,3	-3,6	-4,4	-2,9	-4,3	-1,3	-3,2	1,7	-2,7	3	-4,4	2,5	-2,2
Intermediaire zone	10,6	6,4	8,5	6,2	8,7	3,8	6,2	7,3	7,6	3	7	6,7	5,9
Periferie	-10,4	-4,1	-6,4	-5,2	-6,9	-4,3	-4,5	-16,9	-8,3	-11,8	-3,5	-17,6	-6,2
<i>Resultaat PC4-niveau (€/verjaar)</i>													
Gemiddeld	0	0	0	0	0	0	0	0	0	0	0	0	0
Bandbreedte	93314	93216	93217	93305	93297	93237	93236	93279	93286	93253	93334	93282	93300
GGARV	24		18,8	6,7	19,3	25,7	22,3	17,3	19,7	19,7	18,9	17,5	19,9
<i>Resultaat gemeenteniveau (€/verjaar)</i>													
Gemiddeld	0	0	0	0	0	0	0	0	0	0	0	0	0
Bandbreedte	554	567	597	570	604	549	603	554	590	547	604	554	607
GGARV (€/verjaar)	16,6	Nvt	9	2,5	9,1	14,6	12,3	9,9	10,1	12,4	10	10,5	10,5
<b>Risicodragers</b>													
GGAA (€/verjaar)	24	23	23	23	23	22	23	23	23	22	24	23	23
Bandbreedte	213	213	216	213	216	212	219	214	215	210	216	214	216
GGARV (€/verjaar)	3,1		1,6	0,3	1,7	1,8	1,2	1,8	1	3,7	1,2	2	0,9

Bron: OT2018-bestand, berekeningen SEO Economisch Onderzoek.

De bandbreedte tussen verzekeraars neemt iets toe voor de modellen met de nieuwe kenmerken (\*), de GGARV op verzekeraarsniveau ten opzichte van model M0 is klein: tussen de € 1 en € 2 voor de meeste modellen. Alleen in model M4 ligt de GGARV met € 3,70 iets hoger.

Figuur 5.1 bevestigt dat de resultaatverschuivingen op niveau van de verzekeraar niet heel groot zijn, maar wel betekenisvol. De grootste resultaat verschuiving van één verzekeraar is € 13 per verzekerdenjaar (voor model M4, ten opzichte van model M0).

Figuur 5.1 Kleine resultaatverschuivingen



Bron: OT2018-bestand, berekeningen SEO Economisch Onderzoek. Risicodragers zijn gesorteerd naar opklimmend resultaat voor M0.





## 6 Toetsingskader

*Alle modellen scoren (ongeveer) even goed op verevenende werking. Model M5\* presteert op alle criteria het best.*

Onderstaande tabel vat de werking van de modellen op de toetsingscriteria samen. Onder doelmatigheid wordt in deze context ook verstaan: geen samenhang met zorgaanbod.

**Tabel 6.1** Overzicht afweging toetsingscriteria: model M5\* scoort het best

	M0	M0*	M1	M1*	M2	M2*	M3	M3*	M4	M4*	M5	M5*
Verevenende werking	++	++	++	++	+	++	+	++	+	+	+	++
Doelmatigheid	-	+	-	+	-	-	-	+	-	+	-	+
Beheersbare complexiteit	--	--	--	--	-	-	-/+	-/+	-	-	+	+
Validiteit en meetbaarheid	-	++	-	++	-	++	-	++	-	++	-	++

Bron: WOR 871, waardering toetsingscriteria per model SEO Economisch Onderzoek. Onder beheersbare complexiteit vallen 'eenvoud en transparantie', 'stabiliteit en homogeniteit' en 'beschikbaarheid en betrouwbaarheid'

### Verevenende werking

De verevenende werking is op individuniveau en op verzekeraarsniveau voor alle modellen (ongeveer) gelijk.

Op regionaal niveau zijn er wel verschillen: grote negatieve uitschieters zijn te zien bij model M2 voor de achterstandswijken (clustering van kenmerken, oude regiovariabelen) en M4\* voor de G4 (factoranalyse). Deze modellen krijgen daarom een minder goede score op verevenende werking. Model M3, M4 en M5 tonen juist een grote ondercompensatie van de periferie. Ook deze modellen krijgen daarom een minder goede score op verevenende werking. De bandbreedtes op pc4-niveau en op gemeenteniveau tonen geen systematische verschillen tussen de modellen.

Verder valt op dat de GGARV – zowel op regionaal niveau als op verzekeraarsniveau – voor de modellen met de oude kenmerken (zonder ster) sterk uiteenloopt, terwijl deze bij de modellen met de nieuwe kenmerken (met ster) niet erg afhankelijk is van de gebruikte techniek. Dit doet vermoeden dat de nieuwe kenmerken meer stabiel zijn (zie verder onder 'beheersbare complexiteit').

### Doelmatigheid:

Doelmatigheid hangt in dit geval samen met validiteit en verevenende werking: enerzijds moet het kenmerk gebaseerd zijn op gezondheidsverschillen (validiteit), anderzijds moet het ondoelmatig gedrag voorkomen (door selecteren van gebieden met grote overcompensatie of mijden van gebieden met grote ondercompensatie).

Onder doelmatigheid wordt in deze context ook verstaan: geen samenhang met zorgaanbod. OAD en regionaal klantenpotentieel hangen samen met zorgaanbod, waardoor modellen met deze ken-

merken slechter scoren. Ook model M2/M2\* honoreren mogelijk een groot deel van de aanbodfactoren door het direct clusteren van residuen uit het vereveningsmodel, zonder tussenliggend verklaringsmodel. Om die reden scoren ook deze modellen slechter op doelmatigheid.

### **Beheersbare complexiteit**

Hieronder vallen ‘eenvoud en transparantie’, ‘stabielheid en homogeniteit’ en ‘beschikbaarheid en betrouwbaarheid’.

Eenvoud en transparantie: één van de primaire doelen van het onderzoek was te komen tot een methodiek die eenvoudiger is dan het huidige regiomodel. Behalve ingewikkeld is het huidige regiomodel ook niet transparant: met een regressie op de residuen worden gebieden ingedeeld in clusters, waarbij gebieden die in één cluster terechtkomen (dus) een verschillende combinatie van kenmerken krijgen. Het huidige model, en ook de andere modellen gebaseerd op het clusteren van regio's aan de hand van residuen (M0, M0\*, M1, M1\*, M2, M2\*), zijn dus niet eenvoudig en transparant. Ook de factoranalyse (M4, M4\*) is niet eenvoudig en transparant: hoewel schatten van een regiomodel wordt voorkomen, zijn de factoren samengesteld uit een gewogen gemiddelde van kenmerken. Het model op individuniveau (M3, M3\*) daarentegen is ook niet eenvoudig omdat deze drie kenmerken kent, namelijk westerse migratieachtergrond, niet-westerse migratieachtergrond en lage SES op pc4-niveau, terwijl de overige modellen maar één kenmerk kennen (modellen die resulteren in een regio-indeling), of twee kenmerken (factoranalyse). Hoe meer kenmerken, hoe ingewikkelder de verzekeerderaming en de ophoging naar de verzekeerderaming bij de berekening van de normbedragen.

De tweestapmodellen met een schatting op individuniveau en vervolgens samenstellen van tien regioclusters (M5, M5\*) scoren het best op eenvoud: dit model werkt met één kenmerk, en is bovendien in twee transparante stappen te schatten. Daarnaast behoudt dit model de flexibiliteit om in de toekomst nog kenmerken toe te voegen. Dit is bij model M3/M3\* lastig, omdat ieder extra kenmerk dat wordt opgenomen in het vereveningsmodel de verzekeerderaming en ophoging naar de verzekeerderaming ingewikkelder maakt.

Stabielheid en homogeniteit: *de kenmerken zelf* zijn naar verwachting stabiel: een (groot) pc4-gebied met een hoog percentage personen met een migratieachtergrond heeft volgend jaar niet ineens een laag percentage personen met een migratieachtergrond. Ook de G4 blijft altijd de G4 (stedelijkheid, regionaal klantenpotentieel) en wijken met een relatief hoog percentage lage inkomens trekken niet ineens veel tweeverdieners.

De stabielheid van *het geschatte effect van kenmerken* op zorgkosten is bekend voor de kenmerken die al verschillende jaren worden gebruikt: voor luchtvervuiling geldt dat deze in de OT2018 een negatief teken had, terwijl het in onderzoek van het jaar daarvoor een positief teken had. Ook is er geen duidelijke en stabiele relatie tussen stedelijkheid/regionaal klantenpotentieel in de regiomodellen die de afgelopen jaren zijn geschat. Alleen het kenmerk niet-westerse migratieachtergrond heeft een stabiele en logische relatie in het regiomodel. Naar verwachting geldt dit ook voor westerse migratieachtergrond. Het kenmerk 'lage SES op pc4-niveau' en het kenmerk 'ziekteverzuim' zijn nieuw.<sup>13</sup> De relatie met de meerkosten op basis van 2015-data is logisch. Voor het kenmerk 'lage

<sup>13</sup> Ziekteverzuim is alleen opgenomen in de modellen die gebruikmaken van de nieuwe regiokenmerken en resulteren in een clusterindeling (M0\*/M1\*/M2\*).

SES' geldt bovendien dat deze zowel in het huidige regiomodel (en varianten daarop), alsook in het model op individuniveau overblijft als logische en significante variabele. Dit geeft vertrouwen in de stabiliteit van dit kenmerk. Per saldo lijken de modellen die naast niet-westerse migratieachtergrond ook westerse migratieachtergrond en lage SES op pc4-niveau meenemen stabielier dan de modellen met de 'huidige' regiokenmerken.

Beschikbaarheid en betrouwbaarheid: alle geteste regiokenmerken (zowel de 'huidige' als de 'nieuwe') zijn beschikbaar en betrouwbaar. De bron van de gegevens is meestal CBS, en soms RIVM. Het kenmerk ziekteverzuim is een constructie van Atlas voor Gemeenten, op basis van CBS- en LISA-gegevens. Daarmee is dit kenmerk minder goed beschikbaar. Bij gebruik van dit kenmerk zal jaarlijks aan Atlas voor Gemeenten een update van het kenmerk gevraagd moeten worden. Dit kenmerk wordt alleen gebruikt in modellen M0\*, M1\* en M2\*. Deze modellen scoren daarom iets minder goed op beschikbaarheid.

**Validiteit en meetbaarheid:**

Alle kenmerken zijn meetbaar (gebruikte gegevens zijn volledig en juist beschikbaar, indeling in klassen is niet arbitrair maar objectief, op basis van meerkosten per klasse). De 'oude' regiokenmerken zijn niet allemaal valide (OAD/regionaal klantenpotentieel/normkosten). De 'nieuwe' regiokenmerken zijn wel allemaal valide, want het resultaat van een literatuuronderzoek. De modellen die zijn gebaseerd op de 'nieuwe' regiokenmerken (met ster) krijgen daarom een betere score op validiteit dan de modellen die gebaseerd zijn op de 'oude' regiokenmerken (zonder ster).



## 7 Conclusie

*Een model dat de regiokenmerken rechtstreeks opneemt en vervolgens tien regioclusters samenstelt heeft veel voordelen. Grootste voordeel is dat het model in twee stappen geschat kan worden. Dit kost minder onderzoekstijd dan het huidige achtstappenplan. Ander voordeel is dat de schatting steeds resulteert in één regiokenmerke in het somatische model. Hierdoor wordt de verzeikerdensraming niet (verder) bemoeilijkt.*

Dit onderzoek verkent alternatieven voor het huidige regiocriterium in het somatische risicovereveningsmodel. Enerzijds gaat het om het verbeteren van de verevenende werking via het regiocriterium, zonder daarbij te compenseren voor zorgaanbodfactoren. Anderzijds gaat het om een eenvoudigere methode voor het samenstellen van het regiocriterium. Momenteel wordt het regiocriterium namelijk samengesteld in een meerstappenplan, dat veel onderzoekstijd vergt tijdens de jaarlijkse update van het vereveningsmodel.

Op basis van de (wetenschappelijke) literatuur is in dit onderzoek een overzicht gemaakt van regionale factoren die een effect kunnen hebben op gezondheid. Het gaat hier enerzijds om kenmerken die niet op individuniveau in het vereveningsmodel kunnen worden opgenomen (zoals migratieachtergrond) en anderzijds om regionale kenmerken die van invloed zijn op gezondheid (bijvoorbeeld fijnstof, geluidsbelasting, overlast, groen, enzovoorts). Alleen indicatoren die niet samenhangen met ZVW zorgaanbod komen in aanmerking. Drie van de regiokenmerken in het huidige regiomodel worden in de (wetenschappelijke) literatuur niet genoemd als regiokenmerken die een verschil in gezondheid tussen de inwoners kunnen verklaren. Dit zijn de normkosten somatische zorg, regionaal klantenpotentieel en OAD/stedelijkheid. Deze factoren zijn dus mogelijk niet valide en/of hangen samen met zorgaanbod. In dit onderzoek is daarom steeds een modelvariant met de huidige regiokenmerken geschat alsmede een variant die uitgaat van de nieuwe regiokenmerken zoals geïdentificeerd in de literatuurstudie.

De volgende alternatieve manieren om rekening te houden met regiokenmerken in de risicoverevening zijn onderzocht:

1. Uitgangsmodel somatische zorg exclusief regiocriterium;
2. Uitgangsmodel somatische zorg inclusief het huidige regiocriterium;
3. Regiomodel met aangepaste set regiokenmerken;
4. Regiomodel waarbij indeling in 10 gelijke groepen is losgelaten;
5. Model op basis van clustering van kenmerken, aan de hand van meerkosten;
6. Model waarin regiokenmerken rechtstreeks zijn opgenomen;
7. Model op basis van clustering van kenmerken, zonder gebruik van meerkosten, die rechtstreeks zijn opgenomen;
8. Model waarin regiovariabelen rechtstreeks zijn opgenomen, vervolgens indeling in regioclusters.

Voor al deze varianten zijn de voor- en nadelen in kaart gebracht en is gekeken naar de verevenende werking op allerlei niveaus (individu-, subgroep-, regio- en verzekeraarsniveau). Op basis hiervan is het voorkeursmodel van de onderzoekers het model dat de regiokenmerken rechtstreeks opneemt en op basis van de geschatte coëfficiënten in deze eerste stap tien regioclusters samenstelt.

De regiokenmerken die in de eerste stap zijn opgenomen zijn: niet-westerse migratieachtergrond, westerse migratieachtergrond en SES laag inkomen in de regio.

Bijlage A laat kaartjes van Nederland zien met de residuen op pc4-niveau en gemeenteniveau, zowel voor het model exclusief regiocriterium, voor het uitgangsmodel M0 en voor het voorkeursmodel M5\*. Deze kaartjes laten wederom zien dat de modellen elkaar qua verevenende werking op regio-niveau niet veel ontlopen. Op basis van het toetsingscriterium 'verevenende werking' kan dus nauwelijks onderscheid worden gemaakt tussen het uitgangsmodel M0 en het voorkeursmodel M5\*.

Op alle andere toetsingscriteria scoort model M5\* duidelijk beter dan model M0. Groot voordeel van dit model is dat het in twee stappen geschat kan worden. In vergelijking met het huidige achtstappenplan kost dit minder onderzoekstijd. Een ander voordeel is dat de regio-effecten in de eerste stap transparant zijn: bijvoorbeeld een hoger aandeel personen met een niet-westerse migratieachtergrond leidt tot hogere zorgkosten. De andere onderzochte varianten hebben als nadeel dat zij nog altijd een meerstappenplan vereisen (variant 3 tot en met 5) of dat zij minder transparant zijn (bijvoorbeeld variant 7). In vergelijking met het model waarin regiokenmerken rechtstreeks zijn opgenomen (variant 6) resulteert model M5\* slechts in één regiokenmerk, in plaats van drie regiokenmerken. Hoe meer regiokenmerken, hoe moeilijker de verzekerdensraming en ophoging naar de verzekerdensraming bij de jaarlijkse schatting van de normbedragen.

## Literatuur

- CBS (2014). Gezondheid en zorg in cijfers (Den Haag: Centraal Bureau voor de Statistiek)
- CBS, PBL, RIVM, WUR (2014). Benzeen in lucht, 1995-2013 (indicator 0457, versie 09 , 9 oktober 2014) – Compendium voor de Leefomgeving - [www.clo.nl](http://www.clo.nl)
- CBS (2016). Jaarrapport Integratie 2016 (Den Haag: Centraal Bureau voor de Statistiek)
- CBS (2016). Jaarrapport Integratie 2016 (Den Haag: Centraal Bureau voor de Statistiek)
- Dongen J. van, Vos H. (2007). Gezondheidsaspecten van woningen in Nederland. (Delft; TNO)
- Douven R., Burger M., Schut E. (2018). Grote prijsverschillen ziekenhuiszorg, ondanks concurrentie. Economisch Statistische Berichten, 29 maart 2018.
- Eysink P. et al. (2012). Ziektelast van ongunstige arbeidsomstandigheden in Nederland (Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu RIVM)
- Fischer P. et al. (2015): Air Pollution and Mortality in Seven Million Adults: The Dutch Environmental Longitudinal Study (DUELS) In: Environmental Health Perspectives 123(7): pp. 697-70
- Groenewegen P. et al. (2018). Neighbourhood social and physical environment and general practitioner assessed morbidity. In: Health & Place 49, pp. 68–84
- Hänninen O., Knol A. (2011). European perspectives on environmental burdens of disease. Estimates for nine stressors in six European countries (Helsinki: National Institute for Health and Welfare (THL))
- Harbers M., Eysink P.E.D. (2011). Wat zijn determinanten van gezondheid? In: Volksgezondheid Toekomst Verkenning, Nationaal Kompas Volksgezondheid. Bilthoven: RIVM
- Hooftman W.E. (2017). Nationale Enquête Arbeidsomstandigheden 2016: Methodologie en globale resultaten (Leiden / Heerlen: TNO / CBS)
- Jong de et al. (2012). Perceived green qualities were associated with neighborhood satisfaction, physical activity, and general health: results from a cross-sectional study in suburban and rural Scania, southern Sweden. In: Health Place 18, pp: 1374–1380.
- Kempen E. van, Houthuijs D. (2008). Omvang van de effecten op gezondheid en welbevinden in de Nederlandse bevolking door geluid van weg- en railverkeer (Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu -RIVM)
- Knol A., Staatse B. (2005). Trends in the environmental burdens of disease in the Netherlands 1980-2000 (Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu)

- Leeuw J., Meijer E. de (2007). *Handbook of Multilevel Analysis*. Springer Press.
- Maas J. et al. (2006). Green space, urbanity, and health: how strong is the relation? In: *Journal of Epidemiology and Community Health* 60: pp. 587–592
- Maas R. et al. (2015). *Luchtkwaliteit en gezondheidswinst* (Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu – RIVM)
- Maas J. et al. (2009). Morbidity is related to a green living environment. In: *Journal of Epidemiology and Community Health* 63: pp. 967–973
- Maassen K. et al. (2016). *Veehouderij en gezondheid omwonenden* (Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu – RIVM)
- Marlet G.A., Woerkens C.M. van (2007). Weg uit de wijk, in *Economisch statistische berichten*, 4502, 26-1-2007
- Marlet G.A., Woerkens C.M. van (2007): *Op weg naar Early Warning. Omvang, oorzaak en ontwikkeling van problemen in de wijk* (Atlas voor Gemeenten, Utrecht)
- Ruijsbroek A. et al. (2016): The interplay between neighbourhood characteristics: the health impact of changes in social cohesion, disorder and unsafety feelings. In: *Health and Place* 39(1): pp. 1-8
- SCP (2009). *Jaarrapport Integratie 2009* (Den Haag: Sociaal en Cultureel Planbureau)
- Staatsen B. et al. (2016). *Gezonde leefomgeving, gezonde mensen* (Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu – RIVM)
- Tempelman C., Lammers M., Vriend S., Smits T. (2016): *Verdeelmodel inkomensdeel Participatiewet. SEO-rapport*
- Tempelman C., Vriend S., Kroon L., Marlet G., Woerkens C. van (2017). *Verfijning bijstandsverdeelmodel 2018. SEO-rapport*
- Vriend S., Knoef M., Lammers M., Weel B. ter (2017). *Inkomensmobiliteit in Nederland. SEO rapport*
- WOR 034 (2005). *Regio in detail. APE rapport*.
- WOR 821 (2016) *Onverklaarde (regionale) variatie. SiRM rapport*. Visser J., Mazzola G., Drunen P. van, Hoendervanger J., Stam P.
- WOR 871 (2017). *Toetsingskader 2016. Ministerie van VWS, cluster risicoverevening*.
- WOR 875 (2017). *Overall Toets 2018. iBMG-projectteam risicoverevening*.

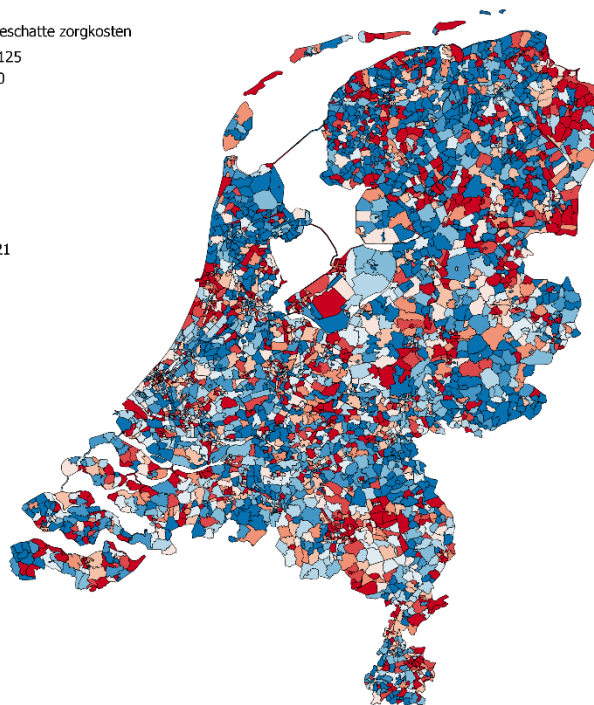
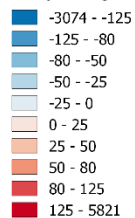


## Bijlage A Kaartjes met residuen

Figuur A. 1 Residuen somatische zorg 2016 *exclusief* regiocriterium, pc4-niveau

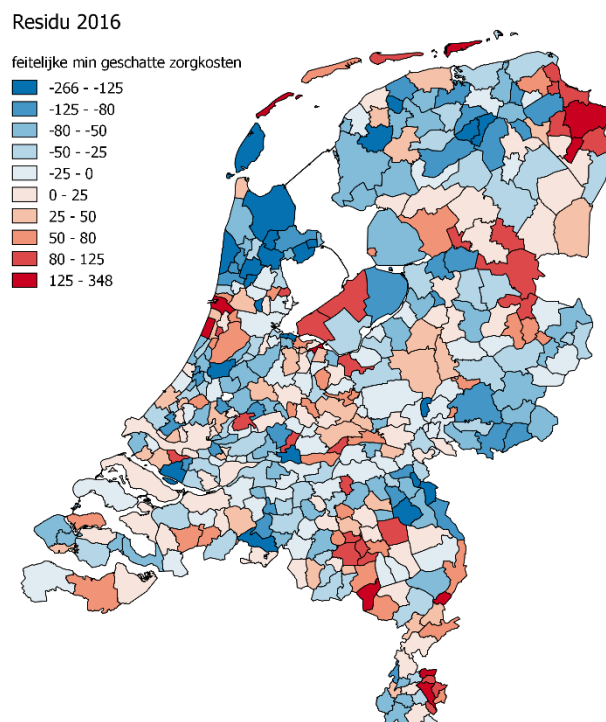
Residu 2016

feitelijke min geschatte zorgkosten



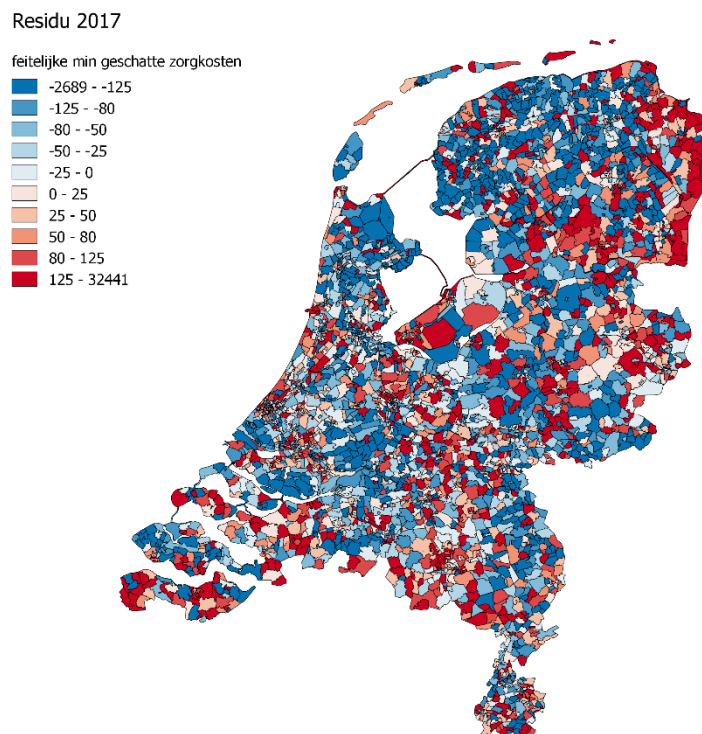
Bron: OT2016-bestand, bewerking SEO Economisch Onderzoek en Atlas voor Gemeenten.

**Figuur A. 2 Residuen somatische zorg 2016 *exclusief* regiocriterium, gemeenteniveau**



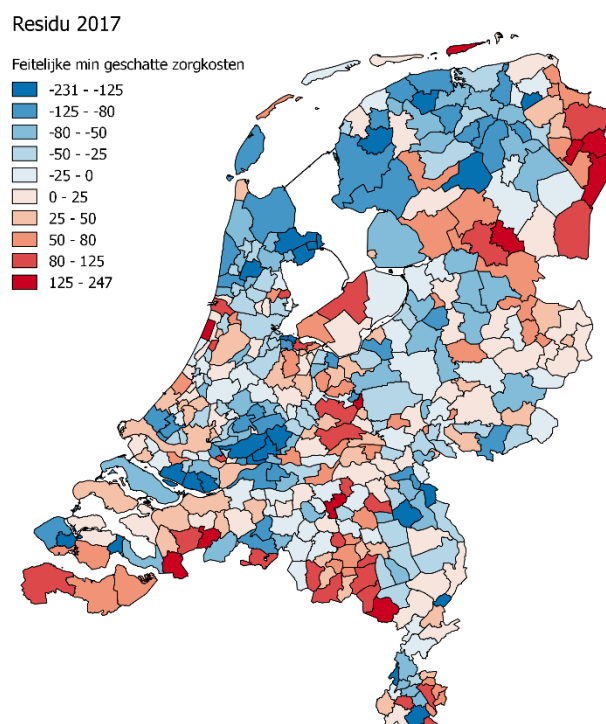
Bron: OT2016-bestand, bewerking SEO Economisch Onderzoek en Atlas voor Gemeenten.

**Figuur A. 3 Residuen somatische zorg 2017 *exclusief* regiocriterium, pc4-niveau**



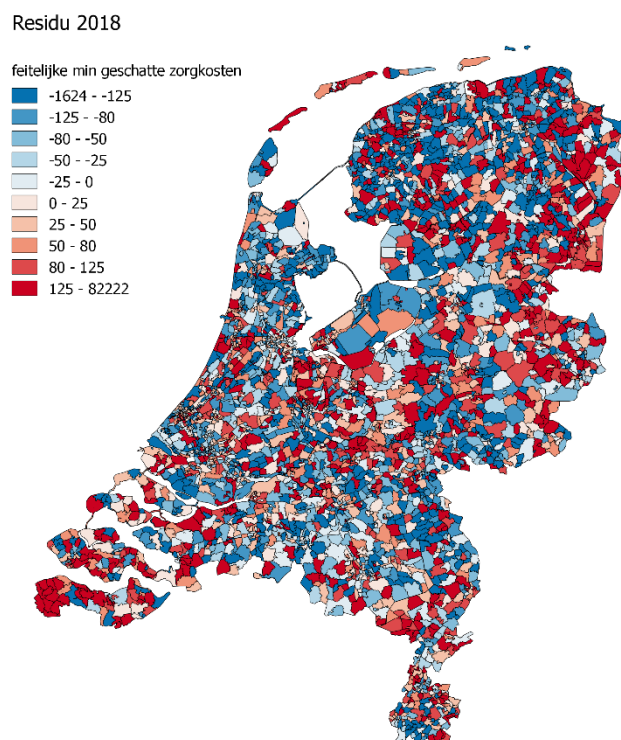
Bron: OT2017-bestand, bewerking SEO Economisch Onderzoek en Atlas voor Gemeenten.

**Figuur A. 4** Residuen somatische zorg 2017 *exclusief* regiocriterium, gemeenteniveau



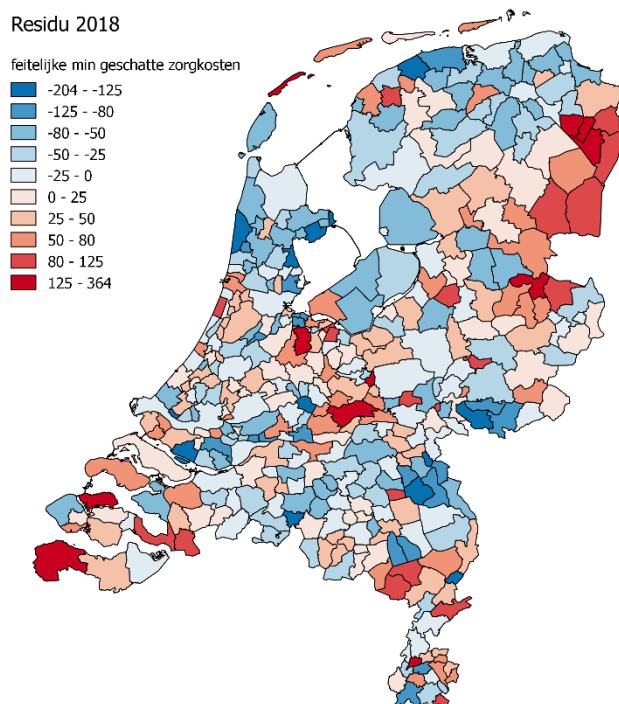
Bron: OT2017-bestand, bewerking SEO Economisch Onderzoek en Atlas voor Gemeenten.

**Figuur A. 5** Residuen somatische zorg uitgangsmode M0, pc4-niveau



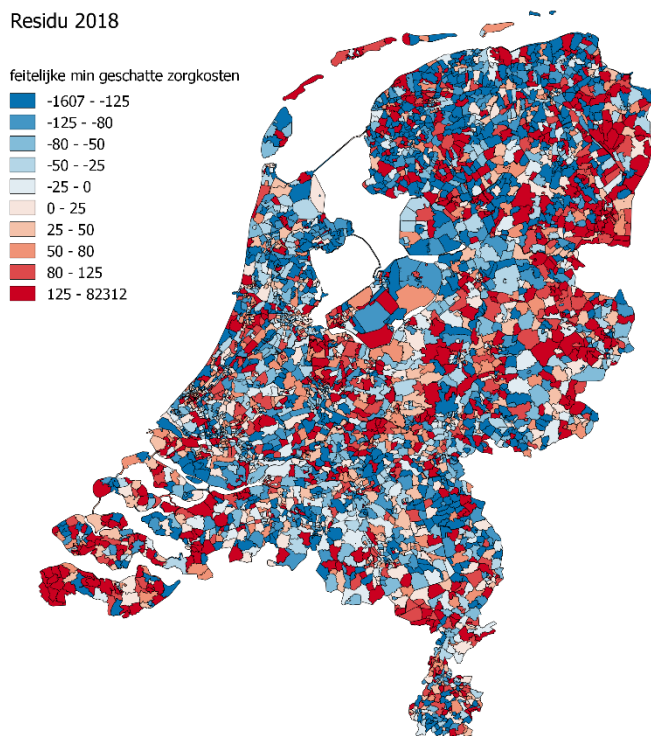
Bron: OT2018-bestand, bewerking SEO Economisch Onderzoek en Atlas voor Gemeenten.

Figuur A. 6 Residuen somatische zorg uitgangsmodel M0, gemeenteniveau



Bron: OT2018-bestand, bewerking SEO Economisch Onderzoek en Atlas voor Gemeenten.

Figuur A. 7 Residuen somatische zorg voorkeursmodel M5\*, pc4-niveau

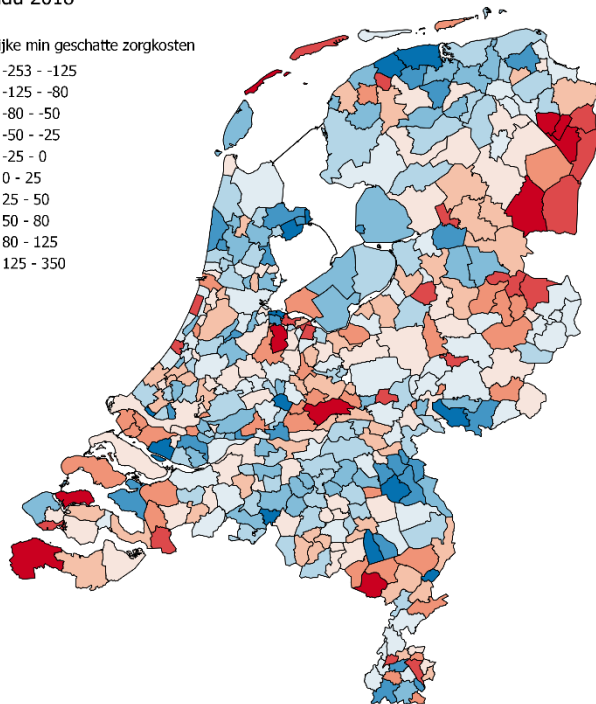
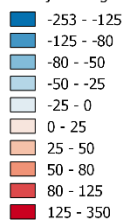


Bron: OT2018-bestand, bewerking SEO Economisch Onderzoek en Atlas voor Gemeenten.

**Figuur A. 8** Residuen somatische zorg voorkeursmodel M5\*, gemeenteniveau

Residu 2018

feitelijke min geschatte zorgkosten



Bron: OT2018-bestand, bewerking SEO Economisch Onderzoek en Atlas voor Gemeenten.



## Bijlage B Regiomodel met nieuwe kenmerken (model M0\*-regio)

Om tot een nieuwe set regiokenmerken te komen die in het regiomodel een significant effect hebben én bovendien een logisch teken hebben, is het volgende stappenplan gehanteerd:

1. Start met het huidige regiomodel (M0), neem alle kenmerken uit Tabel 3.3 – zie eerste kolom Tabel B.1.;
2. Vervang fijnstof PM10 door fijnstof PM2,5 en kijk of de R<sup>2</sup> van het regiomodel verbetert:
  - Ja, dus vervang fijnstof PM10 door fijnstof PM2,5;
3. Vervang percentage huurwoningen door percentage corporatiewoningen, en kijk of de R<sup>2</sup> van het regiomodel verbetert:
  - Nee, dus vervang percentage huurwoningen niet door percentage corporatiewoningen;
4. Schat nu het regiomodel met alle oude regiokenmerken (waarbij fijnstof PM10 is vervangen door fijnstof PM2,5) en alle nieuwe regiokenmerken;
5. Bekijk of de controversiële kenmerken OAD en regionaal klantenpotentieel nog significant zijn:
  - OAD niet, dus OAD eruit. Regionaal klantenpotentieel wel, dus blijft erin;
6. Verwijder vervolgens één voor één de minst significante variabele en herschat het model, totdat alle kenmerken significant zijn op 10 procent niveau – zie tweede kolom Tabel B.1.;<sup>14</sup>
7. Verwijder vervolgens één voor één de kenmerken met een onlogisch teken en herschat het model, totdat alle tekens logisch zijn. Begin met het meest insignificante kenmerk;<sup>15</sup>
8. Verwijder vervolgens weer één voor één de minst significante variabele en herschat het model, totdat alle kenmerken significant zijn op 10 procent niveau – zie derde kolom Tabel B.1.

De kenmerken die in dit gereduceerde model overblijven zijn: lage SES op pc4-niveau en het percentage Surinamers + Antillianen (gezamenlijk).

Als we beginnen met het huidige regiomodel met alleen de huidige regiofactoren en één voor één de kenmerken met onlogische tekens en niet-significante kenmerken verwijderen, dan resulteert alleen het regiokenmerk percentage personen met een niet-westerse migratieachtergrond. De kenmerken OAD, percentage huurwoningen en luchtvervuiling hebben immers allemaal onlogische tekens. Ten eerste is er geen duidelijke relatie tussen OAD en de meerkosten: de laagste meerkosten gelden in sterk stedelijk gebied (OAD1), iets minder stedelijk gebied (OAD2) heeft bijna de hoogste meerkosten en bij OAD3 zijn de meerkosten weer bijna net zo hoog als in OAD1. Luchtvervuiling leidt volgens het regiomodel tot dalende meerkosten en percentage huurwoningen ook. Beide zijn

<sup>14</sup> De 'technische factor', in dit geval de normkosten, wordt nooit verwijderd.

<sup>15</sup> Onlogische tekens waren: de verschillende maten voor luchtvervuiling hebben een negatief effect op de meerkosten en zijn daarom verwijderd, het percentage bos heeft een positief effect op de meerkosten en is daarom verwijderd, de index voor overlast heeft een negatief effect op de meerkosten en is daarom verwijderd, het percentage ziekteverzuim heeft een negatief effect op de meerkosten en is daarom verwijderd, het percentage personen met een westerse migratieachtergrond heeft een negatief effect op de meerkosten en is daarom verwijderd, en het percentage SES zeer laag inkomen heeft een positief effect op de meerkosten en is daarom verwijderd. Overigens werden veel van deze kenmerken ook insignificant na verwijderen van andere kenmerken.

in tegenspraak met de theoretisch te verwachten relatie op basis van het literatuuroverzicht.<sup>16</sup> Na verwijderen van deze onlogische kenmerken is het regionaal klantenpotentieel niet langer significant en wordt daarom ook verwijderd.

Als we hetzelfde stappenplan hanteren maar de normkosten als ‘technische factor’ vervangen door een constante als ‘technische factor’, blijven meer regiokenmerken overeind. De resterende regiokenmerken zijn in dat geval: percentage personen met een westerse migratieachtergrond, percentage personen met een niet-westerse migratieachtergrond, lage SES en het percentage ziekteverzuim – zie vierde kolom Tabel B.1. Het model in kolom 4, geschat volgens de huidige systematiek, is model M0\*-regio. De regioclusters die uit dit regiomodel resulteren worden opgenomen in het risicovereveningsmodel op individuniveau. Het resulterende model is model M0\* zoals beschreven en getoetst in de hoofdtekst.

Een enkel lid van de begeleidingscommissie heeft voorkeur voor een model inclusief OAD. Daarom is getoetst of het kenmerk OAD in model M0\*-regio wél significant en logisch is. OAD is immers vroeg in het stappenplan al verwijderd uit het model, maar zou in het gereduceerde model M0\*-regio wel weer van toegevoegde waarde kunnen zijn. Dit blijkt niet het geval: ook bij toevoegen van OAD aan M0\*-regio is er geen logische relatie tussen stedelijkheid en meerkosten: de laagste meerkosten gelden in sterk stedelijk gebied (OAD1), iets minder stedelijk gebied (OAD2) heeft de hoogste meerkosten en bij OAD3 zijn de meerkosten weer bijna net zo hoog als in OAD1. OAD is daarom ook in tweede instantie niet toegevoegd aan model M0\*-regio.

---

WOR 821 geeft als verklaring voor het negatieve teken dat huurders vaak relatief jong en daardoor gezond zijn. In het model op individuniveau is echter al gecontroleerd voor leeftijd. Leeftijd op individuniveau kan dus geen verklaring vormen voor het negatieve teken op de regiofactor ‘percentage huurwoningen’. Het is bovendien niet aannemelijk dat de leeftijdssamenstelling op regioniveau een additioneel effect heeft bovenop leeftijd op individuniveau. Dan zou eenzelfde persoon in een ‘jongere’ buurt (met meer huurwoningen) lagere zorgkosten moeten hebben omdat zijn burens relatief jong zijn.



Tabel B. 1 (Niet-)Westerse migratieachtergrond, ziekteverzuim en lage SES blijven over als relevante regiokenmerken

	Huidige kenmerken	Nieuwe kenmerken		
	M0 regio	Significant	Significant en logische tekens	M0* regio
Normkosten	0,04***	0,04***	-0,01***	
Constante				-302,7***
OAD1	-72,2**			
OAD2	-47,4			
OAD3	-64,0*			
OAD4	-47,5			
OAD5	-45,4			
Huurwoningen (%)	-0,4*			
Fijnstof PM10	-3,3**			
Regionaal klantenpotentieel	0,2***	0,1**		
Niet-westerse migratieachtergrond (%)	1,7***			1,1***
Marokkanen (%)		2,2***		
Surinamers (%)		4,5***		
Antillianen (%)		8,2**		
Surinamers + Antillianen (%)			3,4***	
Westerse migratieachtergrond (%)		2,3***		1,4***
Ziekteverzuim (%)		0,1		0,6***
SES zeer laag inkomen (%)		-2,1***		
SES laag inkomen (%)		3,0***	1,2***	3,2***
SES midden inkomen (%)				
Index overlast		0,8*		
Natuurlijk terrein (%)		0,8**		
Koolmonoxide		0,1*		
Stikstofdioxide (no2)		-5,6***		
Ozon		-3,4***		
<b># postcode4 gebieden</b>	<b>3.190</b>	<b>3.190</b>	<b>3.190</b>	<b>3.190</b>
<b>R2</b>	<b>6,5</b>	<b>9,5</b>	<b>2,7</b>	<b>5,2</b>

Bron: OT2018-bestand, regiokenmerken 2018, regiokenmerken Atlas voor Gemeenten. Bewerking SEO Economisch Onderzoek. \* = p<0,10 \*\* = p<0,05 \*\*\* = p<0,01.



## Bijlage C Meerkosten geselecteerde regio-kenmerken bij indeling in 10 gelijke groepen

Tabel C. 1 Bij indeling in 10 gelijke groepen lopen meerkosten niet altijd op

Kenmerk	M3		M3*	
	Verzekerden jaren	Meerkosten	Verzekerden jaren	Meerkosten
Regionaal klantenpotentieel:				
- 1 (minder dan 16)	1.832.755	-32		
- 2 (tussen 16 en 29)	1.739.170	-18		
- 3 (tussen 29 en 41)	1.451.435	-17		
- 4 (tussen 41 en 65)	1.684.410	-5		
- 5 (tussen 65 en 87)	1.632.374	-6		
- 6 (tussen 87 en 113)	1.718.452	16		
- 7 (tussen 113 en 142)	1.700.597	23		
- 8 (tussen 142 en 171)	1.622.883	-1		
- 9 (tussen 171 en 209)	2.116.787	26		
- 10 (209 of meer)	1.172.354	9		
Niet westerse migratieachtergrond:				
- 1 (minder dan 1,8%)	1.667.945	-8	1.668.498	-8
- 2 (tussen 1,8% en 2,7%)	1.667.593	-12	1.667.178	-11
- 3 (tussen 2,7% en 3,7%)	1.675.661	-11	1.666.372	-13
- 4 (tussen 3,8% en 5,3%)	1.662.439	-25	1.668.069	-20
- 5 (tussen 5,3% en 7,3%)	1.671.709	1	1.670.567	-4
- 6 (tussen 7,3% en 9,6%)	1.666.312	-12	1.669.373	-7
- 7 (tussen 9,6% en 12,9%)	1.669.698	3	1.664.258	-1
- 8 (tussen 12,9% en 18,4%)	1.655.809	1	1.663.816	4
- 9 (tussen 18,4% en 28,5%)	1.672.532	19	1.669.544	18
- 10 (28,5% of meer)	1.661.521	43	1.663.542	43
Westerse migratieachtergrond:				
- 1 (minder dan 4,3%)			1.671.627	-7
- 2 (tussen 4,3% en 5,8%)			1.662.879	-13
- 3 (tussen 5,8% en 6,9%)			1.671.527	-13
- 4 (tussen 6,9% en 8,1%)			1.671.321	-22
- 5 (tussen 8,1% en 9,0%)			1.663.338	0
- 6 (tussen 9,0% en 9,9%)			1.666.948	12
- 7 (tussen 9,9% en 10,9%)			1.677.010	10
- 8 (tussen 10,9% en 12,2%)			1.661.750	7
- 9 (tussen 12,2% en 15,2%)			1.660.809	11
- 10 (meer dan 15,2%)			1.664.008	14
SES laag inkomen:				
- 1 (minder dan 14%)			1.671.507	-17
- 2 (tussen 14% en 15,8%)			1.662.751	-16
- 3 (tussen 15,8% en 17,6%)			1.668.581	-24

Kenmerk	M3		M3*	
	Verzekerden jaren	Meerkosten	Verzekerden jaren	Meerkosten
- 4 (tussen 17,6% en 18,8%)			1.667.016	-25
- 5 (tussen 18,8% en 19,8%)			1.669.221	-1
- 6 (tussen 19,8% en 20,9%)			1.664.301	2
- 7 (tussen 20,9% en 22,1%)			1.668.013	1
- 8 (tussen 22,1% en 23,6%)			1.674.889	20
- 9 (tussen 23,6% en 25,3%)			1.662.204	28
- 10 (meer dan 25,3%)			1.662.734	34

Bron: OT2018-bestand, berekeningen SEO Economisch Onderzoek.

# Bijlage D Normbedragen

Tabel D. 1 Normbedragen somatische zorg modelvarianten

Verevenings-criterium	M00	M0	M0*	M1	M1*	M2	M2*	M3	M3*	M4	M4*	M5	M5*
M, 0, t-1	9.395	9.393	9.391	9.392	9.391	9.392	9.392	9.391	9.391	9.392	9.392	9.391	9.391
M, 0, t	2.752	2.751	2.749	2.751	2.749	2.750	2.749	2.749	2.749	2.750	2.750	2.749	2.749
M, 1-4	1.896	1.895	1.895	1.895	1.895	1.894	1.895	1.894	1.894	1.895	1.895	1.894	1.894
M, 5-9	1.746	1.746	1.746	1.746	1.746	1.745	1.746	1.745	1.746	1.746	1.746	1.745	1.746
M, 10-14	1.771	1.772	1.772	1.772	1.772	1.770	1.772	1.772	1.772	1.772	1.772	1.772	1.772
M, 15-17	1.841	1.841	1.842	1.841	1.842	1.840	1.841	1.842	1.842	1.842	1.842	1.842	1.842
M, 18-24	1.652	1.652	1.650	1.652	1.649	1.652	1.650	1.649	1.650	1.650	1.650	1.650	1.650
M, 25-29	1.664	1.663	1.659	1.663	1.659	1.663	1.661	1.660	1.660	1.660	1.660	1.660	1.660
M, 30-34	1.668	1.667	1.663	1.666	1.663	1.667	1.664	1.664	1.664	1.664	1.664	1.664	1.664
M, 35-39	1.743	1.742	1.740	1.742	1.740	1.742	1.741	1.740	1.740	1.741	1.741	1.740	1.740
M, 40-44	1.829	1.829	1.828	1.829	1.828	1.828	1.828	1.828	1.828	1.829	1.828	1.828	1.828
M, 45-49	1.934	1.934	1.934	1.935	1.934	1.934	1.934	1.934	1.934	1.935	1.934	1.934	1.934
M, 50-54	2.089	2.090	2.090	2.090	2.090	2.089	2.089	2.090	2.090	2.090	2.090	2.090	2.090
M, 55-59	2.349	2.350	2.350	2.350	2.350	2.349	2.350	2.351	2.350	2.350	2.350	2.351	2.350
M, 60-64	2.625	2.626	2.627	2.626	2.627	2.626	2.627	2.627	2.627	2.627	2.627	2.627	2.627
M, 65-69	3.130	3.130	3.130	3.130	3.130	3.131	3.130	3.132	3.130	3.131	3.130	3.132	3.130
M, 70-74	3.501	3.501	3.502	3.501	3.503	3.502	3.502	3.504	3.502	3.503	3.502	3.504	3.503
M, 75-79	3.860	3.859	3.862	3.860	3.862	3.862	3.862	3.863	3.862	3.862	3.861	3.863	3.862
M, 80-84	4.140	4.139	4.144	4.139	4.144	4.142	4.144	4.144	4.144	4.142	4.143	4.144	4.144
M, 85-89	4.684	4.681	4.688	4.682	4.688	4.685	4.687	4.687	4.687	4.684	4.686	4.687	4.688
M, 90+	5.130	5.126	5.134	5.126	5.134	5.130	5.133	5.132	5.133	5.129	5.132	5.132	5.133
V, 0, t-1	8.273	8.271	8.270	8.271	8.270	8.271	8.270	8.270	8.270	8.271	8.270	8.270	8.270
V, 0, t	2.414	2.412	2.411	2.412	2.411	2.412	2.411	2.411	2.410	2.411	2.411	2.411	2.410
V, 1-4	1.693	1.693	1.692	1.693	1.692	1.691	1.692	1.692	1.692	1.692	1.692	1.692	1.692
V, 5-9	1.711	1.711	1.711	1.711	1.711	1.710	1.711	1.711	1.711	1.711	1.711	1.711	1.711
V, 10-14	1.796	1.797	1.798	1.797	1.798	1.796	1.797	1.797	1.798	1.798	1.798	1.797	1.798
V, 15-17	1.945	1.946	1.947	1.946	1.947	1.945	1.946	1.946	1.947	1.947	1.946	1.946	1.947
V, 18-24	1.903	1.903	1.900	1.903	1.900	1.904	1.902	1.900	1.901	1.901	1.901	1.900	1.901
V, 25-29	2.378	2.377	2.373	2.377	2.373	2.377	2.375	2.374	2.374	2.374	2.374	2.374	2.373
V, 30-34	2.504	2.503	2.500	2.502	2.500	2.502	2.501	2.499	2.500	2.500	2.501	2.500	2.500
V, 35-39	2.177	2.176	2.175	2.176	2.175	2.175	2.175	2.174	2.175	2.175	2.176	2.174	2.175
V, 40-44	1.958	1.959	1.959	1.959	1.959	1.958	1.958	1.958	1.959	1.958	1.959	1.958	1.959
V, 45-49	2.025	2.026	2.027	2.026	2.027	2.025	2.026	2.026	2.027	2.026	2.027	2.026	2.027
V, 50-54	2.111	2.112	2.112	2.112	2.112	2.111	2.112	2.112	2.112	2.112	2.112	2.112	2.112
V, 55-59	2.213	2.214	2.214	2.214	2.214	2.214	2.214	2.215	2.214	2.214	2.214	2.215	2.214
V, 60-64	2.361	2.362	2.364	2.362	2.364	2.362	2.363	2.364	2.363	2.363	2.363	2.364	2.364
V, 65-69	2.716	2.716	2.717	2.716	2.717	2.717	2.717	2.719	2.717	2.718	2.717	2.719	2.717
V, 70-74	2.965	2.965	2.967	2.965	2.968	2.967	2.967	2.968	2.967	2.967	2.967	2.968	2.968
V, 75-79	3.271	3.270	3.274	3.270	3.274	3.273	3.274	3.275	3.274	3.273	3.273	3.275	3.274
V, 80-84	3.541	3.540	3.545	3.540	3.545	3.543	3.545	3.545	3.545	3.543	3.544	3.545	3.545
V, 85-89	4.079	4.077	4.083	4.077	4.083	4.081	4.083	4.082	4.083	4.080	4.082	4.082	4.083

Verevenings-criterium	M00	M0	M0*	M1	M1*	M2	M2*	M3	M3*	M4	M4*	M5	M5*
V, 90+	4.627	4.623	4.631	4.623	4.631	4.628	4.631	4.629	4.630	4.626	4.629	4.629	4.631
Geen FKG	-256	-256	-256	-256	-256	-256	-256	-256	-256	-256	-256	-256	-256
Glaucoom	162	162	162	161	162	162	162	161	162	161	162	161	162
Schildklieraan- doening	71	71	71	71	71	71	71	71	71	71	71	71	71
Psychose, Alzheimer en verslaving	196	195	195	195	195	196	195	194	195	194	195	194	195
Depressie	204	206	206	206	206	205	205	205	205	205	205	205	205
Chronische pijn excl. opioïden	906	905	904	905	904	905	905	906	905	906	904	906	905
Neuropathisch pijn complex	1.836	1.836	1.836	1.836	1.836	1.836	1.836	1.836	1.836	1.836	1.835	1.836	1.836
Hoog choleste- rol	90	89	89	89	89	90	89	89	89	89	89	89	89
Diabetes type II zonder hyper- tensie	507	503	502	503	502	504	503	503	503	503	503	503	503
COPD / Zware astma	1.526	1.525	1.524	1.525	1.524	1.525	1.525	1.525	1.525	1.526	1.524	1.525	1.524
Astma	408	408	408	408	407	407	407	408	408	408	407	408	407
Diabetes type II met hypertensie	852	849	847	848	847	849	848	849	848	849	848	849	848
Epilepsie	663	664	664	664	664	664	664	664	664	663	663	664	664
Ziekte van Crohn / Colitis Ulcerosa	532	534	534	534	534	534	534	534	534	534	534	534	534
Hartaandoenin- gen	1.598	1.598	1.597	1.598	1.597	1.598	1.597	1.598	1.597	1.598	1.597	1.598	1.597
Auto-immuun- ziekten (add- ons)	6.490	6.491	6.491	6.491	6.491	6.491	6.491	6.491	6.491	6.491	6.491	6.491	6.491
Reuma	683	684	684	684	684	684	683	684	684	684	684	684	684
Parkinson	1.842	1.843	1.844	1.843	1.844	1.842	1.843	1.843	1.843	1.842	1.843	1.843	1.843
Diabetes type I	1.817	1.814	1.813	1.814	1.813	1.815	1.814	1.815	1.814	1.815	1.814	1.815	1.814
Transplantaties	1.394	1.394	1.394	1.394	1.394	1.394	1.394	1.394	1.394	1.394	1.394	1.394	1.394
Cystic fibrosis / Pancreasenzym- en	2.772	2.775	2.775	2.775	2.775	2.774	2.774	2.774	2.775	2.773	2.775	2.774	2.775
Aandoeningen hersenen / rug- genmerg: MS	3.417	3.420	3.420	3.420	3.420	3.419	3.420	3.419	3.419	3.420	3.419	3.419	3.420
Aandoeningen hersenen / rug- genmerg: Overig	2.586	2.586	2.587	2.586	2.587	2.586	2.587	2.586	2.587	2.585	2.587	2.586	2.587
Kanker	2.187	2.188	2.187	2.188	2.187	2.188	2.187	2.187	2.187	2.188	2.188	2.187	2.187
Hormoongevoe- lige tumoren	815	816	816	816	816	816	816	816	816	815	816	816	816
HIV / AIDS	4.895	4.892	4.888	4.892	4.888	4.892	4.889	4.884	4.888	4.884	4.891	4.885	4.887
Nieraandoenin- gen	7.793	7.794	7.793	7.794	7.793	7.794	7.793	7.793	7.793	7.794	7.793	7.793	7.793
Psoriasis	440	439	439	439	438	440	439	439	439	439	439	439	439

Verevenings-criterium	M00	M0	M0*	M1	M1*	M2	M2*	M3	M3*	M4	M4*	M5	M5*
Pulmonale arteriële hypertensie	17.884	17.885	17.885	17.884	17.885	17.885	17.883	17.884	17.884	17.884	17.885	17.884	17.884
Kanker (add-ons)	13.034	13.035	13.036	13.035	13.036	13.036	13.035	13.035	13.035	13.035	13.035	13.035	13.036
Groeistoomnissen (add-ons)	3.952	3.954	3.953	3.954	3.953	3.954	3.954	3.953	3.953	3.953	3.953	3.954	3.953
Extreem Hoge Kosten cluster 1	108326	108321	108323	108322	108322	108325	108324	108321	108323	108324	108324	108321	108323
Extreem Hoge Kosten cluster 2	197632	197636	197635	197637	197635	197639	197634	197638	197635	197637	197635	197638	197634
Extreem Hoge Kosten cluster 3	420447	420442	420452	420443	420452	420446	420449	420450	420452	420447	420448	420449	420451
Geen pDKG	-205	-205	-205	-205	-205	-205	-205	-205	-205	-205	-205	-205	-205
pDKG 1	608	608	608	608	608	608	608	608	608	608	608	608	608
pDKG 2	1.070	1.070	1.071	1.070	1.071	1.070	1.070	1.071	1.070	1.071	1.070	1.071	1.070
pDKG 3	1.122	1.122	1.122	1.122	1.122	1.122	1.122	1.123	1.122	1.122	1.122	1.123	1.122
pDKG 4	1.395	1.394	1.394	1.395	1.394	1.395	1.394	1.395	1.394	1.395	1.394	1.395	1.394
pDKG 5	1.849	1.848	1.848	1.848	1.848	1.849	1.848	1.848	1.848	1.849	1.848	1.848	1.848
pDKG 6	2.088	2.089	2.089	2.089	2.089	2.089	2.089	2.088	2.089	2.088	2.089	2.088	2.089
pDKG 7	4.162	4.162	4.162	4.162	4.162	4.162	4.162	4.162	4.162	4.162	4.162	4.162	4.162
pDKG 8	4.301	4.300	4.300	4.300	4.300	4.301	4.301	4.300	4.300	4.300	4.301	4.300	4.301
pDKG 9	4.986	4.986	4.987	4.986	4.987	4.986	4.986	4.986	4.986	4.985	4.986	4.986	4.986
pDKG 10	5.826	5.827	5.827	5.827	5.827	5.827	5.827	5.826	5.827	5.826	5.827	5.826	5.827
pDKG 11	9.617	9.618	9.619	9.618	9.618	9.618	9.619	9.618	9.618	9.617	9.618	9.618	9.618
pDKG 12	12.628	12.628	12.627	12.628	12.627	12.628	12.628	12.627	12.628	12.628	12.628	12.628	12.628
pDKG 13	10.982	10.983	10.983	10.983	10.983	10.984	10.983	10.983	10.983	10.984	10.983	10.983	10.983
pDKG 14	66.125	66.127	66.127	66.127	66.127	66.127	66.127	66.126	66.127	66.126	66.127	66.126	66.127
pDKG 15	51.548	51.544	51.544	51.543	51.544	51.544	51.544	51.545	51.544	51.544	51.545	51.545	51.544
Geen sDKG	-100	-100	-100	-100	-100	-100	-100	-100	-100	-100	-100	-100	-100
sDKG 1	830	829	829	829	829	829	829	829	829	829	830	829	829
sDKG 2	1.977	1.977	1.977	1.977	1.977	1.976	1.977	1.976	1.977	1.976	1.977	1.977	1.977
sDKG 3	4.845	4.844	4.844	4.844	4.844	4.844	4.844	4.844	4.844	4.844	4.845	4.844	4.844
sDKG 4	7.445	7.445	7.445	7.445	7.445	7.445	7.445	7.445	7.445	7.445	7.445	7.445	7.445
sDKG 5	11.664	11.663	11.663	11.663	11.664	11.663	11.663	11.663	11.663	11.663	11.664	11.663	11.663
sDKG 6	16.345	16.344	16.344	16.344	16.344	16.344	16.344	16.344	16.344	16.344	16.344	16.344	16.344
sDKG 7	47.938	47.938	47.939	47.938	47.939	47.936	47.940	47.938	47.940	47.938	47.939	47.939	47.940
Geen HKG	-45	-46	-45	-46	-45	-46	-46	-46	-46	-46	-45	-46	-46
CPAP-apparaat	123	122	122	122	121	123	122	122	122	123	122	122	122
Therapeutische elastische kousen	550	551	550	550	550	551	550	550	550	551	550	550	550
Voorzieningen voor stomapatiënten	1.343	1.344	1.344	1.344	1.344	1.344	1.344	1.344	1.344	1.344	1.344	1.344	1.344
Vernevelaar met toebehoren	2.071	2.068	2.069	2.068	2.069	2.071	2.070	2.070	2.070	2.071	2.070	2.071	2.070
Middelen voor urine-opvang	1.889	1.889	1.890	1.889	1.890	1.889	1.890	1.890	1.890	1.889	1.889	1.890	1.890

Verevenings- criterium	M00	M0	M0*	M1	M1*	M2	M2*	M3	M3*	M4	M4*	M5	M5*
Injectiespuiten met toebehoren (excl. diabetes)	2.078	2.078	2.078	2.078	2.078	2.080	2.079	2.079	2.079	2.079	2.078	2.079	2.079
Zuurstofapparaten met toebehoren	4.004	4.003	4.003	4.003	4.003	4.003	4.003	4.004	4.003	4.004	4.004	4.004	4.003
Voedingshulpmiddelen (excl. zuigelingen)	5.824	5.824	5.824	5.824	5.824	5.823	5.824	5.824	5.823	5.824	5.823	5.824	5.824
Slijmuitzuigapparatuur	19.945	19.943	19.944	19.943	19.944	19.944	19.942	19.944	19.943	19.941	19.944	19.944	19.943
Draagbare in-fuuspompen	9.121	9.123	9.122	9.123	9.122	9.121	9.124	9.122	9.122	9.122	9.123	9.122	9.123
AVI referentie 0-17 en 65+	0	0	0	0	0	0	0	0	0	0	0	0	0
IVA, 18-34	1.112	1.109	1.111	1.109	1.112	1.111	1.112	1.115	1.112	1.115	1.110	1.116	1.112
IVA, 35-44	1.646	1.641	1.641	1.641	1.641	1.645	1.643	1.645	1.642	1.645	1.642	1.646	1.642
IVA, 45-54	1.235	1.230	1.229	1.230	1.229	1.232	1.231	1.233	1.230	1.233	1.230	1.233	1.230
IVA, 55-64	867	864	863	864	863	866	864	866	864	866	864	866	864
Arbeidsongeschikten excl. IVA, 18-34	257	253	254	253	255	254	255	258	255	257	254	258	255
Arbeidsongeschikten excl. IVA, 35-44	446	442	442	443	442	444	443	445	443	445	442	445	443
Arbeidsongeschikten excl. IVA, 45-54	493	490	489	490	489	491	490	491	489	491	489	491	489
Arbeidsongeschikten excl. IVA, 55-64	425	423	422	423	422	424	423	424	423	424	423	424	422
Bijstandsgerechtigden, 18-34	241	230	231	230	231	233	233	236	233	236	231	236	233
Bijstandsgerechtigden, 35-44	260	248	247	248	247	253	250	253	250	253	248	253	249
Bijstandsgerechtigden, 45-54	338	327	325	327	325	331	328	330	328	330	327	330	327
Bijstandsgerechtigden, 55-64	379	368	366	367	366	371	369	370	368	371	368	370	368
Studenten, 18-34	-170	-166	-169	-166	-169	-166	-168	-173	-168	-171	-169	-172	-169
Zelfstandigen, 18-34	-93	-89	-90	-89	-89	-90	-90	-92	-91	-92	-89	-92	-91
Zelfstandigen, 35-44	-139	-135	-135	-135	-135	-137	-136	-137	-136	-137	-135	-137	-136
Zelfstandigen, 45-54	-171	-166	-164	-165	-164	-167	-166	-167	-166	-168	-165	-167	-166
Zelfstandigen, 55-64	-235	-229	-228	-229	-228	-231	-229	-231	-229	-231	-228	-231	-229
Hoogopgeleiden, 18-34	-13	-10	-11	-9	-10	-10	-11	-14	-11	-13	-10	-13	-11
Hoogopgeleiden, 35-44	-81	-78	-78	-78	-78	-79	-78	-81	-78	-81	-77	-81	-78
Referentiegroep, 18-34	27	25	26	25	26	25	26	28	26	27	26	28	26



<b>Verevenings- criterium</b>	M00	M0	M0*	M1	M1*	M2	M2*	M3	M3*	M4	M4*	M5	M5*
Referentiegroep, 35-44	-17	-17	-17	-17	-17	-17	-17	-16	-17	-16	-17	-16	-17
Referentiegroep, 45-54	-57	-56	-56	-56	-56	-56	-56	-56	-56	-56	-56	-56	-56
Referentiegroep, 55-64	-88	-87	-87	-87	-87	-87	-87	-87	-87	-87	-87	-87	-87
Regiocluster 1		59	63	79	74	573	259					47	50
Regiocluster 2		39	29	66	51	265	175					23	43
Regiocluster 3		11	20	53	27	161	98					7	13
Regiocluster 4		8	8	23	30	109	55					4	4
Regiocluster 5		7	-4	9	11	67	25					3	1
Regiocluster 6		-7	-17	4	0	38	8					-13	2
Regiocluster 7		-18	-25	-14	-21	0	-10					-18	-14
Regiocluster 8		-35	-18	-39	-18	-32	-27					-30	-20
Regiocluster 9		-34	-25	-26	-32	-79	-51						-32
Regiocluster 10		-30	-29	-35	-34	-184	-96						-34
Regiocluster 11				-157	-16	-421	-169						
SES Zeer laag, 18-	86	76	73	75	73	80	76	78	75	78	74	78	75
SES Zeer laag, 18-64	28	23	20	23	20	25	22	23	21	23	21	23	21
SES Zeer laag, 65+	490	487	484	487	484	488	485	488	485	489	485	488	485
SES Laag, 18-	32	29	27	29	27	30	27	31	27	30	27	31	27
SES Laag, 18- 64	22	19	17	19	17	20	18	21	18	20	18	21	18
SES Laag, 65+	-2	-3	-5	-3	-5	-3	-5	-2	-5	-2	-5	-2	-5
SES Midden, 18-	-21	-19	-19	-19	-19	-20	-20	-19	-19	-19	-20	-19	-19
SES Midden, 18-64	10	11	11	11	11	11	11	12	11	11	11	12	11
SES Midden, 65+	-144	-144	-144	-144	-144	-144	-144	-144	-144	-144	-144	-144	-144
SES Hoog, 18-	-58	-51	-48	-51	-48	-53	-49	-54	-49	-54	-48	-54	-49
SES Hoog, 18- 64	-44	-40	-37	-40	-37	-41	-38	-42	-38	-41	-37	-42	-38
SES Hoog, 65+	-268	-264	-261	-264	-261	-265	-262	-266	-262	-266	-261	-266	-262
Referentie 0-17 jarigen	0	0	0	0	0	0	0	0	0	0	0	0	0
Blijvend ihh. 18- 64	-405	-397	-389	-397	-389	-400	-391	-396	-389	-397	-393	-396	-388
Blijvend ihh. 65- 79	-2.663	-2.664	-2.658	-2.664	-2.658	-2.662	-2.658	-2.663	-2.657	-2.664	-2.660	-2.663	-2.657
Blijvend ihh. 80+	-4.430	-4.431	-4.427	-4.431	-4.427	-4.429	-4.428	-4.429	-4.427	-4.430	-4.428	-4.429	-4.427
Instromend ihh. 18-64	8.867	8.870	8.876	8.870	8.877	8.868	8.874	8.871	8.875	8.870	8.874	8.870	8.876
Instromend ihh. 65-79	10.110	10.109	10.114	10.109	10.114	10.111	10.113	10.110	10.113	10.108	10.112	10.110	10.113
Instromend ihh. 80+	4.010	4.011	4.015	4.012	4.015	4.012	4.014	4.012	4.014	4.011	4.014	4.012	4.014

Verevenings- criterium	M00	M0	M0*	M1	M1*	M2	M2*	M3	M3*	M4	M4*	M5	M5*
Eenpersoons hh. 18-64	11	10	9	10	9	11	10	8	9	8	10	8	9
Eenpersoons hh. 65-79	70	68	69	68	69	69	69	68	69	67	69	67	69
Eenpersoons hh. 80+	605	605	605	605	605	605	605	605	605	604	605	605	605
Overig hh. 18-64	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3
Overig hh. 65-79	-31	-30	-31	-30	-31	-30	-31	-30	-30	-30	-31	-30	-30
Overig hh. 80+	152	153	151	153	151	152	151	152	151	153	152	152	151
Geen MHK	-542	-542	-542	-542	-542	-542	-542	-542	-542	-542	-542	-542	-542
Ten minste 1 jaar kosten in de top 30%	113	113	113	113	113	113	113	113	113	113	113	113	113
2 jaar kosten in de top 10%	2.332	2.331	2.331	2.331	2.331	2.331	2.331	2.331	2.331	2.331	2.331	2.331	2.331
3 jaar kosten in de top 15%	2.158	2.157	2.157	2.157	2.157	2.157	2.157	2.157	2.157	2.158	2.157	2.157	2.157
3 jaar kosten in de top 10%	3.471	3.470	3.471	3.470	3.471	3.470	3.471	3.471	3.471	3.471	3.471	3.471	3.471
3 jaar kosten in de top 7%	5.246	5.244	5.245	5.244	5.245	5.244	5.245	5.245	5.245	5.245	5.245	5.245	5.245
3 jaar kosten in de top 4%	8.404	8.403	8.404	8.403	8.404	8.403	8.403	8.403	8.403	8.403	8.404	8.403	8.403
3 jaar kosten in de top 1,5%	16.747	16.746	16.747	16.746	16.747	16.746	16.747	16.746	16.747	16.746	16.747	16.746	16.747
3 jaar kosten in de top 0,5%	43.527	43.526	43.527	43.526	43.527	43.526	43.527	43.527	43.527	43.528	43.527	43.527	43.527
Geen FDG	-17	-17	-17	-17	-17	-17	-17	-17	-17	-17	-17	-17	-17
FDG 1	526	528	528	528	528	528	528	528	528	527	528	527	528
FDG 2	1.374	1.376	1.375	1.376	1.375	1.376	1.376	1.375	1.375	1.376	1.376	1.375	1.376
FDG 3	1.407	1.409	1.409	1.409	1.409	1.409	1.409	1.409	1.409	1.409	1.410	1.409	1.409
FDG 4	8.602	8.606	8.605	8.606	8.605	8.607	8.604	8.603	8.604	8.603	8.605	8.603	8.603
Geen VVG	-171	-171	-171	-171	-171	-171	-171	-171	-171	-171	-171	-171	-171
V&V-kosten t-1 in top 2,5%	1.888	1.887	1.888	1.887	1.888	1.888	1.888	1.889	1.888	1.888	1.887	1.889	1.888
V&V-kosten t-1 in top 2%	3.347	3.347	3.347	3.347	3.347	3.347	3.347	3.348	3.347	3.348	3.347	3.348	3.347
V&V-kosten t-1 in top 1,5%	6.397	6.395	6.395	6.395	6.395	6.396	6.396	6.397	6.396	6.397	6.395	6.397	6.396
V&V-kosten t-1 in top 1%	9.982	9.980	9.980	9.980	9.980	9.982	9.980	9.982	9.980	9.982	9.980	9.982	9.980
V&V-kosten t-1 in top 0,5%	11.310	11.307	11.308	11.307	11.308	11.310	11.308	11.310	11.308	11.310	11.308	11.310	11.308
V&V-kosten t-1 in top 0,25% en 18-	52.909	52.905	52.905	52.905	52.905	52.906	52.905	52.907	52.905	52.906	52.905	52.907	52.905
V&V-kosten t-1 in top 0,25% en 18+	28.533	28.529	28.530	28.529	28.530	28.530	28.530	28.532	28.531	28.531	28.530	28.532	28.530
Regionaal klan- tenpotentieel 1								-23					
Regionaal klan- tenpotentieel 2								-13					

Verevenings- criterium	M00	M0	M0*	M1	M1*	M2	M2*	M3	M3*	M4	M4*	M5	M5*
Regionaal klan- tenpotentieel 3								-5					
Regionaal klan- tenpotentieel 4								12					
Niet-westerse migratie 1								-7	-6				
Niet-westerse migratie 2								-3	-2				
Niet-westerse migratie 3								15	14				
Niet-westerse migratie 4								35	31				
Westerse migratie 1									-9				
Westerse migratie 2									4				
Westerse migratie 3									8				
Westerse migratie 4									13				
SES laag inko- men 1									-20				
SES laag inko- men 2									-2				
SES laag inko- men 3									19				
SES laag inko- men 4									42				
PCA 1 oude kenmerken: 1										-32			
PCA 1 oude kenmerken: 2										-4			
PCA 1 oude kenmerken: 3										-15			
PCA 1 oude kenmerken: 4										-14			
PCA 1 oude kenmerken: 5										-2			
PCA 1 oude kenmerken: 6										5			
PCA 1 oude kenmerken: 7										7			
PCA 1 oude kenmerken: 8										18			
PCA 1 oude kenmerken: 9										0			
PCA 1 oude kenmerken: 10										36			
PCA 2 oude kenmerken: 1										8			
PCA 2 oude kenmerken: 2										24			
PCA 2 oude kenmerken: 3										24			

<b>Verevenings- criterium</b>	M00	M0	M0*	M1	M1*	M2	M2*	M3	M3*	M4	M4*	M5	M5*
PCA 2 oude kenmerken: 4										6			
PCA 2 oude kenmerken: 5										1			
PCA 2 oude kenmerken: 6										-5			
PCA 2 oude kenmerken: 7										-12			
PCA 2 oude kenmerken: 8										-20			
PCA 2 oude kenmerken: 9										-14			
PCA 2 oude kenmerken: 10										-13			
PCA 1 nieuwe kenmerken: 1											-29		
PCA 1 nieuwe kenmerken: 2											-17		
PCA 1 nieuwe kenmerken: 3											-7		
PCA 1 nieuwe kenmerken: 4											-7		
PCA 1 nieuwe kenmerken: 5											1		
PCA 1 nieuwe kenmerken: 6											6		
PCA 1 nieuwe kenmerken: 7											7		
PCA 1 nieuwe kenmerken: 8											9		
PCA 1 nieuwe kenmerken: 9											10		
PCA 1 nieuwe kenmerken: 10											27		
PCA 2 nieuwe kenmerken: 1											-32		
PCA 2 nieuwe kenmerken: 2											-17		
PCA 2 nieuwe kenmerken: 3											-20		
PCA 2 nieuwe kenmerken: 4											-15		
PCA 2 nieuwe kenmerken: 5											1		
PCA 2 nieuwe kenmerken: 6											-10		
PCA 2 nieuwe kenmerken: 7											16		
PCA 2 nieuwe kenmerken: 8											6		
PCA 2 nieuwe kenmerken: 9											25		
PCA 2 nieuwe kenmerken: 10											47		

Bron: OT2018-bestand, berekening SEO Economisch Onderzoek



## Bijlage E Verevenende werking varianten M0 en M3

In de hoofdtekst staat een aantal varianten beschreven van model M0 (het huidige regiomodel) en M3 (model op individuniveau). Onderstaande tabel toont de verevenende werking van deze varianten, en van de modellen M0/M0\* en M3/M3\* die al zijn behandeld in de hoofdtekst.

Model M0\* schat het huidige regiomodel met een groot aantal variabelen, namelijk alle variabelen met een significante (maar niet noodzakelijk logische) relatie met het residu op pc4-niveau.

Model M3<sup>vol</sup> is een model op individuniveau waarin alle huidige regiokenmerken zijn opgenomen (westerse migratieachtergrond, niet-westerse migratieachtergrond, regionaal klantenpotentieel, OAD en percentage huurwoningen). Per kenmerk zijn steeds drie klassen gecreëerd met een (ongeveer) gelijk aantal verzekerdenjaren. Uitzonderingen zijn OAD (is al in vijf klassen) en niet-westerse migratieachtergrond (inhoudelijke grenzen van 0-10 procent, 10-20 procent en >20 procent aangehouden omdat in grote steden vaak postcodegebieden voorkomen met zelfs wel 40 procent personen met een migratieachtergrond).

M3\*\* is gelijk aan M3\* (westerse migratieachtergrond, niet-westerse migratieachtergrond en lage SES) en voegt daar het kenmerk OAD aan toe.

M3<sup>3kl</sup> is gelijk aan M3\*, maar met een indeling in drie klassen met een gelijk aantal verzekerdenjaren in plaats van vier klassen met een ongelijk aantal verzekerdenjaren. Voor niet-westerse migratieachtergrond zijn geen groepen gecreëerd met een gelijk aantal verzekerdenjaren maar zijn inhoudelijke grenzen van 0-10 procent, 10-20 procent en >20 procent aangehouden omdat in grote steden vaak postcodegebieden voorkomen met zelfs wel 40 procent personen met een migratieachtergrond.

M3<sup>27kl</sup> is gelijk aan M3\*\*\*, maar dan met een volledige interactie van de drie variabelen in ieder drie klassen, resulterend in  $3 * 3 * 3 = 27$  verschillende klassen.

M3<sup>10kl</sup> is gelijk aan M3\*, maar met een indeling in tien klassen met een gelijk aantal verzekerdenjaren in plaats van vier klassen met een ongelijk aantal verzekerdenjaren.

Tabel E. 1 Verevenende werking aanvullende varianten op M0 en M3

Maatstaf	M00	M0	M0*	M0**	M3	M3 <sup>vol</sup>	M3*	M3**	M3 <sup>3kl</sup>	M3 <sup>27kl</sup>	M3 <sup>10kl</sup>
<b>Individu</b>											
R <sup>2</sup> x 100%	32,1	32,1	32,1	32,1	32,1	32,1	32,1	32,1	32,1	32,1	32,1
CPM x 100%	31,9	31,9	31,9	31,9	31,9	31,9	31,9	31,9	31,9	31,9	31,9
<b>Subgroep</b>											
GGAA (€/verjaar)	808	808	808	808	808	808	808	808	808	808	808
Resultaat 15% laagste kosten t-3 (€/verjaar)	106	104	105	104	104	104	105	105	105	105	105
Resultaat 15% hoogste kosten t-3 (€/verjaar)	-134	-134	-134	-134	-134	-134	-134	-134	-134	-134	-134
Resultaat VGG>0 (€/verjaar)	0	0	0	0	0	0	0	0	0	0	0
<i>Resultaat regionaal niveau (€/verjaar)</i>											
G4	-25,3	-1,9	-3	-0,9	2,7	-5,9	-4,8	-8,6	-6,7	-6,8	-4,8
G21	-13,3	2,5	-0,4	-1,4	4,1	3,7	-2,1	-3,6	-2	-2,2	-2,1
Rest van NL	5,7	-1,1	0,2	0,6	-1,7	-1,6	0,9	1,5	0,9	0,9	0,9
Achterstandswijken	-34,8	-4,6	3,3	-3,4	-6,4	-12,1	-2,7	-5,1	-4,4	-6	-2,7
Randstad	-4,3	-3,6	-4,4	-5	1,5	-0,9	-3,1	-3,4	-3	-3,1	-3,1
Intermediaire zone	10,6	6,4	8,5	5,9	6,6	5	8,0	8,1	8,9	8	8
Periferie	-10,4	-4,1	-6,4	-0,2	-15,4	-7,4	-8,2	-7,7	-9,9	-8,2	-8,2
<i>Resultaat PC4-niveau (€/verjaar)</i>											
Gemiddeld	0	0	0	0	0	0	0	0	0	0	0
Bandbreedte	93.314	93.216	93.217	93.237	93.290	93.244	93.247	93.294	93.256	93.291	93.253
GGARV	24		18,8	22,5	19,2	16	20,0	18,2	19,7	19,9	19,7
<i>Resultaat gemeenteniveau (€/verjaar)</i>											
Gemiddeld	0	0	0	0	0	0	0	0	0	0	0
Bandbreedte	554	567	597	615	554	575	584	588	578	581	547
GGARV (€/verjaar)	16,6		9	12,5	12,0	9,4	10,1	9,6	10,5	10,6	12,4
<b>Risicodragers</b>											
GGAA (€/verjaar)	24	23	23	21	23	23	23	23	23	23	23
Bandbreedte	213	213	216	215	216	208	215	214	215	216	215
GGARV (€/verjaar)	3,1		1,6	2,4	1,8	1,9	0,9	1,4	1,3	1,6	0,9

Bron: OT2018-bestand, berekening SEO Economisch Onderzoek



## Bijlage F Data en databewerking

Het analysebestand dat voor dit onderzoek is gebruikt is samengesteld uit de volgende bronnen:

1. OT2018-bestand met kenmerken risicovereveningsmodel voor de somatische zorg;
2. Postcodebestand van Vektis met per verzekerden de postcode (4-cijferig);
3. Regiokenmerken uit de OT2018;
4. Nieuwe regiokenmerken van Atlas voor Gemeenten;
5. Bestand met postbusnummer en gemeentecode 2015;
6. Bestand met achterstandswijken op basis van NZA-definitie.

Daarnaast zijn ook de OT2017- en OT2016-bestanden gebruikt om de kaartjes met residuen op viercijferig postcodeniveau en gemeenteniveau te maken - zie Bijlage A. Deze OT-bestanden zijn in de rest van het onderzoek niet gebruikt.

### 1. OT2018-bestand

Het OT2018-bestand bevat alle gegevens die nodig zijn om het risicovereveningsmodel voor de somatische zorg te schatten: somatische zorgkosten uit 2015 en alle kenmerken uit het risicovereveningsmodel, ingedeeld in klassen. Daarnaast bevat het een indicator die aangeeft of de verzekerde behoorde tot de 15 procent met de laagste kosten in t-3 (gegevensjaar 2012) en een indicator die aangeeft of de verzekerde behoorde tot de 15 procent met de hoogste kosten in t-3.

Vanaf vereveningsjaar 2018 (gegevensjaar 2015) wordt de viercijferige postcode niet meer opgenomen in het OT-bestand. Om toch de postcode van de verzekerde te kunnen koppelen is gebruikgemaakt van een postcodebestand, zie hieronder.

### 2. Postcodebestand

Het Postcodebestand is geleverd door Vektis en bevat per verzekerde/uzovi combinatie een postcode. Ongeldige postcodes worden verwijderd. Het gaat om ongeveer 17.000 verzekerden met postcode 0, en postcode 9999 met een landcode ongelijk aan Nederland. De ontbrekende postcodes vullen we vervolgens aan met postcodegegevens uit de OT2017. Dan resulteren nog 1.416 ontbrekende postcodes. In de berekening van de normbedragen 2018 is van deze verzekerden wel een postcode bekend. Dit komt omdat voor de berekening van de normbedragen de ontbrekende postcodes in de gegevensfase zijn aangevuld met gegevens uit BASIC. In BASIC zitten meer verzekerden dan in de OT-bestanden. Bijvoorbeeld niet-ingezetenen zitten wel in BASIC maar worden in de OT verwijderd. Aanvullen op basis van de OT2017 resulteert daarom in meer ontbrekende postcodes dan aanvullen op basis van BASIC. De ontbrekende postcodes worden niet meegenomen in de analyse.

Er zijn vervolgens nog 570 verzekerde/uzovi combinaties waarbij meerdere verschillende postcodes geregistreerd zijn. Voor deze gevallen houden we per verzekerde/uzovi alleen de meest recente gegevens. Dit is een beslisregel die Vektis ook gebruikt. Dan resulteren nog 74 verzekerden met verschillende postcodes bij dezelfde uzovi. In die gevallen selecteren we de postcode met de hoogste 'fractie' (i.e. de fractie van de tijd verzekerd). Dan zijn er geen dubbele postcodes meer voor dezelfde verzekerde/uzovi combinatie.

### 3. Regiokenmerken uit de OT2018

Dit bestand bevat per viercijferige postcode de regiokenmerken uit de OT2018: percentage personen met een niet-westerse migratieachtergrond, percentage huurwoningen, luchtvervuiling PM10, OAD en regionaal klantenpotentieel.

Bij de koppeling met het OT2018-bestand blijkt dat voor een aantal postcodegebieden geen regiokenmerken beschikbaar zijn. Het gaat om 9.326 verzekerden/uzovi observaties. Dit zijn vrijwel allemaal postbusnummers. Het huidige regiomodel wordt alleen geschat op postcodes met minimaal 500 inwoners, en neemt postbusnummers niet mee. Verzekerden met een postbusnummer worden ingedeeld op basis van het geschatte residu *op gemeenteniveau*. Voor het huidige regiomodel maakt het dus niet uit dat voor verzekerden die een postbusnummer invullen soms regiokenmerken ontbreken. Bij schatten van het model op individuniveau maakt dit wel uit, omdat dan alle regiokenmerken direct meelopen in het risicovereveningsmodel. Missende regionale kenmerken op postcodeniveau worden daarom aangevuld met het gewogen gemiddelde van het regionale kenmerk *op gemeenteniveau*.

### 4. Nieuwe regiokenmerken van Atlas voor Gemeenten

Dit bestand bevat per viercijferige postcode de nieuwe regiokenmerken zoals geïdentificeerd in het literatuuronderzoek (zie Tabel 3.1 en Tabel 3.2). Daarnaast bevat het bestand de gemeentecode (in gemeentelijke indeling 2015), een indicator die Nederland indeelt in de Randstad, een intermediaire zone en de periferie, en het aantal inwoners in het postcodegebied.

Bij de koppeling met het OT2018-bestand blijkt dat voor een aantal postcodegebieden geen regiokenmerken beschikbaar zijn. Het gaat om 9.406 verzekerden/uzovi observaties. Dit zijn vrijwel allemaal postbusnummers. Het huidige regiomodel wordt alleen geschat op postcodes met minimaal 500 inwoners, en neemt postbusnummers niet mee. Verzekerden met een postbusnummer worden ingedeeld op basis van het geschatte residu *op gemeenteniveau*. Voor het huidige regiomodel maakt het dus niet uit dat voor verzekerden die een postbusnummer invullen soms regiokenmerken ontbreken. Bij schatten van het model op individuniveau maakt dit wel uit, omdat dan alle regiokenmerken direct meelopen in het risicovereveningsmodel. Missende regionale kenmerken op postcodeniveau worden daarom aangevuld met het gewogen gemiddelde van het regionale kenmerk *op gemeenteniveau*.

Het aantal verzekerden waarvoor de nieuwe regio-informatie ontbreekt is iets hoger dan het aantal verzekerden waarvoor de regiokenmerken uit de OT2018 ontbreken. In de volgende postcodegebieden ontbreken wel de nieuwe regiokenmerken, maar niet de regiokenmerken uit de OT2018: 1327, 1384, 2212, 7732. Het gaat in totaal om 151 verzekerden. Anderzijds ontbreken in de volgende postcodegebieden wel de regiokenmerken uit de OT2018, maar niet de nieuwe regiokenmerken: 1362, 7928, 8466. Het gaat in totaal om 74 verzekerden.

### 5. Bestand met postbusnummer en gemeentecode 2015

Dit bestand bevat per postcodegebied waarvoor geen nieuwe regiokenmerken beschikbaar zijn het postcodegebied plus de gemeentecode 2015. Het gaat vaak om postbusnummers.

In het huidige regiomodel worden de verzekerden in kleine postcodegebieden met minder dan 500 inwoners ingedeeld op basis van het gemiddelde residu op gemeenteniveau. Dit geldt ook voor

verzekerden die een postbusnummer opgeven. In de modellen op individuniveau moeten de ontbrekende regionale kenmerken worden aangevuld met het gemiddelde op gemeenteniveau.

Dit betekent dat voor alle verzekerden de gemeentecode bekend moet zijn. Het bestand met de nieuwe regiokenmerken van Atlas voor Gemeenten bevat de gemeentecode voor vrijwel alle bestaande postcodegebieden. Het bevat echter niet de gemeentecode van postbusgebieden. Deze gemeentecode wordt daarom aangevuld op basis van bestand (4) met postbusnummer en gemeentecode.

#### **6. Bestand met achterstandswijken op basis van NZA-definitie**

Dit bestand bevat per viercijferige postcode een indicator of deze postcode door de NZA is aangewezen als achterstandswijk. De NZA gebruikt deze tabel voor het vaststellen van tarieven voor huisartsen en verloskundigen.



## Bijlage G Reproductie regiomodel M0

Bij schatten van het uitgangsmodel M0 is het regiokenmerk gereproduceerd zoals ook berekend in de OT2018 (WOR 875). De schattingsresultaten komen vrijwel overeen, zie Tabel G. 1. Een uitzondering is het percentage huurwoningen, de puntschatting op deze variabele is in de OT2018 100 keer groter dan in de reproductie. Vermoedelijk is in de OT2018 niet het percentage huurwoningen maar de fractie huurwoningen gebruikt als kenmerk.

De regioklasse-indeling op basis van de reproductie is ook vrijwel gelijk aan de regioklasse-indeling uit de OT2018, zie Tabel G. 2. Kleine verschillen kunnen zijn ontstaan doordat in het analysebestand dat wordt gebruikt voor de reproductie een kleine 1.500 verzekerden niet zijn meegenomen omdat hun postcode miste.

**Tabel G. 1** Reproductie resulteert in vrijwel gelijke schattingen in vergelijking met OT2018

	Reproductie		OT2018 (WOR 875)	
	Coëfficiënt	T-waarde	Coëfficiënt	T-waarde
Normkosten	0,04	8,01	0,05	8,37
OAD1	-72,21	-2,67	-78,04	-2,89
OAD2	-47,42	-1,82	-52,85	-2,04
OAD3	-63,97	2,51	-69,67	-2,74
OAD4	-47,46	-1,88	-53,26	-2,11
OAD5	-45,41	-1,83	-51,03	-2,07
Huurwoningen (%)	-0,43	-2,12	-42,43	-2,12
Luchtvervuiling PM10	-3,25	-3,09	-3,21	-3,06
Regionaal klantenpotentieel	0,21	5,67	0,21	5,76
Niet-westerse migratieachtergrond (%)	1,65	7,81	1,65	7,80

Bron: OT2018-bestand, berekening SEO Economisch Onderzoek

**Tabel G. 2** Kleine verschuivingen in regioklasse-indeling tussen OT2018 en reproductie

Verschuiving (# regioklassen)	Aantal verzekerden	Percentage verzekerden
0	16.082.809	94,6
1	919.350	5,4
2	836	0
3 of meer	237	0
Totaal	17.003.232	100

Bron: OT2018-bestand, berekening SEO Economisch Onderzoek



**seo** economisch onderzoek

Roetersstraat 29 . 1018 WB Amsterdam . T (+31) 20 525 16 30 . F (+31) 20 525 16 86 . [www.seo.nl](http://www.seo.nl)