

# Weging Gezondheidsenquête 2009 van de GGD regio Hart voor Brabant

*Mattijn Morren en Harm Jan Boonstra*

De in dit rapport weergegeven opvattingen zijn die van de auteurs en komen niet noodzakelijk overeen met het beleid van het Centraal Bureau voor de Statistiek.

**Statistische Methoden (201103)**



## Verklaring van tekens

.	=	gegevens ontbreken
*	=	voorlopig cijfer
**	=	nader voorlopig cijfer
x	=	geheim
–	=	nihil
–	=	(indien voorkomend tussen twee getallen) tot en met
o (o,o)	=	het getal is kleiner dan de helft van de gekozen eenheid
niets (blank)	=	een cijfer kan op logische gronden niet voorkomen
2010–2011	=	2010 tot en met 2011
2010/2011	=	het gemiddelde over de jaren 2010 tot en met 2011
2010/'11	=	oogstjaar, boekjaar, schooljaar enz., beginnend in 2010 en eindigend in 2011
2008/'09–		
2010/'11	=	oogstjaar, boekjaar enz., 2008/'09 tot en met 2010/'11

In geval van afronding kan het voorkomen dat het weergegeven totaal niet overeenstemt met de som van de getallen.

## Colofon

### Uitgever

Centraal Bureau voor de Statistiek  
Henri Faasdreef 312  
2492 JP Den Haag

### Prepress

Centraal Bureau voor de Statistiek -  
Grafimedia

### Omslag

TelDesign, Rotterdam

### Inlichtingen

Tel. (088) 570 70 70  
Fax (070) 337 59 94  
Via contactformulier: [www.cbs.nl/infoservice](http://www.cbs.nl/infoservice)

### Bestellingen

E-mail: [verkoop@cbs.nl](mailto:verkoop@cbs.nl)  
Fax (045) 570 62 68

### Internet

[www.cbs.nl](http://www.cbs.nl)

ISSN: 1876-0333

© Centraal Bureau voor de Statistiek,  
Den Haag/Heerlen, 2011.  
Vereenvoudiging is toegestaan,  
mits het CBS als bron wordt vermeld.

# Weging Gezondheidsenquête 2009 van de GGD regio Hart voor Brabant

*Samenvatting: Het CBS en de GGD werken samen in een pilot om de meerwaarde te onderzoeken van verrijking van de gegevens uit de GGD-gezondheidsenquêtes met gegevens van het CBS. Als onderdeel hiervan wordt in dit rapport een weging uitgevoerd op de Gezondheidsenquête 2009 van de GGD regio Hart voor Brabant, nadat deze verrijkt is met gegevens van het CBS. Het resulterende weegmodel en totaalschattingen van geselecteerde doelvariabelen worden vergeleken met het weegmodel van de GGD. De verschillen tussen beide weegmodellen zijn klein, hoewel niet kan worden uitgesloten dat uitsplitsing naar achtergrondvariabelen grotere verschillen aan het licht brengt. Wanneer men het gevonden weegmodel wil toepassen op landelijke enquêtegegevens, is het van belang eerst te onderzoeken in hoeverre er regionale en provinciale verschillen in non-respons in het weegmodel verdisconteerd moeten worden.*

*Trefwoorden: gezondheidsenquête, GGD, selectieve non-respons, wegen, Bascula*

## 1. Inleiding

Het ministerie van Volksgezondheid, Welzijn en Sport (VWS) heeft verzocht om te onderzoeken in hoeverre de gezondheidsenquêtes van de GGD'en en de enquête POLS-Gezondheid van het CBS (GEZO) geïntegreerd kunnen worden. Hiertoe is onder andere een pilot uitgevoerd, waarin gegevens van het CBS werden gekoppeld aan een GGD-gezondheidsenquête (verrijking) om enkele baten van een dergelijke samenwerking in kaart te kunnen brengen (van den Brink, van Hilten & van den Berg, 2010). In dit rapport wordt in aanvulling daarop onderzocht in hoeverre deze verrijking kan leiden tot een betere weging van het databestand. Hierbij wordt gebruik gemaakt van gezondheidsenquête gegevens van de GGD Hart voor Brabant.

Dit rapport is als volgt opgebouwd. In paragraaf 2 worden de gebruikte populatiebestanden van de GGD en van het CBS beschreven en in paragraaf 3 het steekproefontwerp en het responsbestand. Vervolgens wordt in paragraaf 4 de weging besproken en de wijze waarop variabelen voor het weegmodel zijn geselecteerd. Paragraaf 5 beschrijft de resultaten van de weging en in paragraaf 6 wordt afgesloten met een conclusie.

## **2. Populatie**

De populatie van de Gezondheidsenquête omvat alle personen in de leeftijd van 19 t/m 64 jaar woonachtig in de 29 gemeenten in het werkgebied van de GGD Hart voor Brabant. Een overzicht van deze gemeenten staat in Bijlage 1.

### **2.1 GGD populatiebestand**

Het populatiebestand van de GGD is samengesteld op basis van een afschrift van de GBA van de deelnemende gemeentes met als peildatum 4 juni 2009. Dit bestand omvat 638.710 personen in de leeftijd van 19 t/m 64 jaar; personen uit instellingen zijn uit het bestand verwijderd. In het bestand zijn naast de NAW gegevens geboortjaar en geslacht opgenomen.

De GGD heeft voor de weging die in dit rapport beschreven wordt de populatietabel beschikbaar gesteld die zij voor haar eigen weging heeft gebruikt. Hierin is voor elke gemeente en voor een aantal wijken binnen enkele gemeentes waarvoor extra steekproeven zijn getrokken (oversampling) het aantal mannen en vrouwen weergegeven in drie leeftijdscategorieën (19-34, 35-49 en 50-64 jaar).

### **2.2 CBS populatiebestand**

Voor onderhavig onderzoek naar een meer uitgebreide weging is een populatiebestand samengesteld op basis van het Sociaal Statistisch Bestand (SSB). Dit was noodzakelijk om ook andere achtergrondvariabelen dan gemeente, leeftijd en geslacht in het weegmodel te kunnen opnemen. Het GGD populatiebestand bevatte namelijk alleen deze drie achtergrondvariabelen.

Het CBS populatiebestand bevat alle per 4 juni 2009 in de GBA ingeschreven personen in de gemeenten die onder de GGD Hart voor Brabant vallen in de leeftijd van 19 t/m 64 jaar, in totaal 636.045. De omvang van het populatiebestand van het CBS valt 0,26% lager uit dan dat van de GGD. De exacte reden van dit verschil is niet bekend, maar het heeft waarschijnlijk te maken met de manier waarop beide bestanden tot stand zijn gekomen.

Naast gemeente, geslacht en leeftijd bevat dit bestand een groot aantal socio-demografische achtergrondkenmerken. Tevens zijn enkele variabelen toegevoegd ter identificatie van de oversamplede wijken, die via een syntax van de GGD Hart voor Brabant zijn afgeleid uit de postcode. De reden dat de wijkindeling van de GGD is gebruikt, is dat deze op een aantal punten afwijkt van de door het CBS gehanteerde wijkindeling.<sup>1</sup>

---

<sup>1</sup> Dit betrof vooral de buurt Fatima, die in de indeling van het CBS niet voorkomt.

### 3. Steekproef en responsbestand

#### 3.1 Steekproeftrekking en steekproefontwerp

De steekproef voor de Gezondheidsenquête is getrokken uit het populatiebestand van de GGD (zie paragraaf 2.1). Vervolgens zijn hieraan in de eerste week van september 2009 de NAW-gegevens uit het GBA opgevraagd. Dit betekent dat personen die tussen begin juni en begin september uit de regio zijn verhuisd of zijn overleden niet in de steekproef zitten. Het populatiebestand waarop de steekproef was gebaseerd bevatte 637.677 personen.

Er is sprake van een gestratificeerde steekproef naar gemeente, waarbij Oss in twee gebieden is opgedeeld. Daarnaast zijn er uit drie gemeenten extra steekproeven getrokken in een aantal buurten dat in de vorm van het zogenaamde wijkgericht werken extra aandacht krijgt. Er waren twee extra steekproeven in Oss, twee in 's Hertogenbosch en vier in Tilburg (zie Bijlage 1 voor het steekproefontwerp).

Het benodigde aantal respondenten per stratum  $r_b$  is op basis van een standaardformule vastgesteld op 384:

$$r_b = \frac{p(1-p)}{d^2} 1,96^2 = 384,$$

waarbij voor  $p$ , de fractie van een doelkenmerk, de waarde 0,5 is gehanteerd bij een van tevoren vastgestelde marge  $d$  en bijbehorende  $z$  waarde van 1,96.

Op dit aantal is vervolgens een eindigheidscorrectie toegepast om te komen tot het definitieve aantal benodigde respondenten per stratum:

$$\tilde{r}_{bj} = \frac{1}{1 + \frac{r_b}{N_j}} r_b.$$

Hierbij is  $r_b$  de vereiste respons in elk stratum en  $N_j$  de populatie van het stratum. Als gevolg van deze correctie valt het aantal benodigde respondenten lager uit naarmate de populatieomvang kleiner is. Op basis van dit getal is vervolgens de vereiste steekproefomvang berekend, uitgaande van een verwachte respons van 50% in de reguliere steekproef en van 45% in de extra steekproeven:

$$n_{j_r} = \frac{r_{bj}}{0,50} \text{ en } n_{j_e} = \frac{r_{bj}}{0,45}.$$

Dit heeft geresulteerd in een steekproef van 28.670, waarvan 28.474 zijn uitgestuurd.

#### 3.2 GGD databestand

Het databestand dat op basis van de enquête is samengesteld bevat informatie van en over 28.474 personen. De antwoorden op de enquêtevragen zijn beschikbaar voor 12.057 respondenten (42,3%). Voorts is het bestand gekoppeld aan het Sociaal Sta-

tistisch Bestand (SSB), zodat voor zowel respondenten als non-respondenten een groot aantal achtergrondvariabelen beschikbaar zijn (indien relevant voor onderzoeksjaar 2009). Deze variabelen staan in Bijlage 2; het CBS populatiebestand bevat overeenkomstige variabelen.

## 4. Weging

### 4.1 Insluitgewichten

Het totaal van de doelvariabele  $Y$  kan worden geschat als gewogen som van de steekproefwaarnemingen  $y_i$  aan  $r$  respondenten:

$$\hat{Y} = \sum_{i=1}^r w_i y_i, \quad (1)$$

waarbij gewicht  $w_i$  geïnterpreteerd kan worden als het aantal personen in de doelpopulatie vertegenwoordigd door respondent  $i$ . De som van de gewichten is gelijk aan de populatieomvang op 4 juni 2009 van  $N$  inwoners van de regio Hart voor Brabant in de leeftijd van 19 t/m 64 jaar:

$$\sum_{i=1}^r w_i = N = 638.710.^2 \quad (2)$$

De institutionele bevolking maakt geen deel uit van de doelpopulatie en zit dus niet in de steekproef. Het gewicht  $w_i$  is het product van een insluitgewicht  $d_{j[i]}$  van persoon  $i$  in steekproefstratum  $j$  en correctiegewicht  $c_i$ :

$$w_i = d_{j[i]} c_i. \quad (3)$$

Het insluitgewicht  $d_j$  is de inverse van de insluitkans  $\pi_j$ :

$$d_j = \frac{1}{\pi_j}. \quad (4)$$

Hierbij is de insluitkans  $\pi_j$  de fractie van de doelpopulatie van elk stratum dat in de steekproef is opgenomen:

$$\pi_j = \frac{n_j}{N_j}, \quad (5)$$

waarbij  $n_j$  de steekproefomvang in stratum  $j$  is en  $N_j$  de populatieomvang in stratum  $j$ . Deze steekproeffractie wordt vaak vermenigvuldigd met de responsfractie  $r_j/n_j$  en omvat daarmee het deel van de doelpopulatie, dat per stratum respondeert:

---

<sup>2</sup> Wanneer gebruik gemaakt wordt van het CBS populatiebestand geldt  $N = 636.045$ .

$$\pi_j = \frac{n_j}{N_j} \frac{r_j}{n_j} = \frac{r_j}{N_j}, \quad (6)$$

met  $r_j$  als de bruikbare respons in stratum  $j$ . Door de responsfractie op te nemen als onderdeel van de insluitkans, wordt gecorrigeerd voor selectieve non-respons naar stratum. De insluitkansen van alle personen binnen één stratum zijn gelijk. De steekproef omvat  $n = 28.474$  personen van wie er  $r = 12.057$  hebben gerepsondeerd.

Het insluitgewicht is het aantal personen in de populatie dat wordt vertegenwoordigd door een respondent. Als schattingen gebaseerd zouden zijn op alleen de insluitgewichten, zou de aanname daarbij zijn dat de non-respons binnen strata niet selectief is. Omdat het echter waarschijnlijk is dat non-respons ook binnen strata selectief is, worden de insluitgewichten nog aangepast via een weging. Hiertoe worden correctiegewichten  $c_i$  berekend die corrigeren voor overige selectieve non-respons via het gebruik van hulpvariabelen die samenhangen met de respons en de doelvariabelen, en waarvan op populatieniveau gegevens bekend zijn.

## 4.2 Wegen

Als insluitgewichten worden de inverse insluitkansen gebruikt zoals beschreven in formule (6) in de vorige paragraaf. De eindgewichten of ophoogfactoren zijn dan:

$$w_i = \frac{1}{\pi_{j[i]}} c_i,$$

waarbij  $c_j$  staat voor de correctiegewichten. Kortom, het insluitgewicht vermenigvuldigd met het correctiegewicht geeft het eindgewicht. Afhankelijk van het weegmodel worden de correctiegewichten berekend door poststratificatie of door lineair wegen (Särndal, Swensson & Wretman, 1992). Zowel de correctiegewichten als de eindgewichten worden berekend met behulp van het softwarepakket Bascula (Nieuwenbroek & Boonstra, 2002).

## 4.3 Relevante weegvariabelen

De keuze van het uiteindelijke weegmodel wordt bepaald door een aantal zaken (Burger, Gouweleeuw & Buelens, 2010). Ten eerste dienen de populatietotalen voor de hulpvariabelen bekend te zijn. Ten tweede worden hulpvariabelen opgenomen die sterk samenhangen met belangrijke doelvariabelen. Ten derde worden hulpvariabelen waarvoor de meeste selectieve non-respons optreedt eerder opgenomen in het weegmodel. Verder is de invloed van de hulpvariabele op de verdeling van correctiegewichten van belang; negatieve gewichten en gewichten  $> 3$  zijn niet wenselijk (hoewel wel beperkt toelaatbaar is). Wanneer een groter aantal termen in het weegmodel wordt opgenomen, betekent dit een grotere consistentie met de gegevens uit het register. Anderzijds wordt de schatting bij een te groot weegmodel minder stabiel, doordat het aantal respondenten per weegcel – de celvulling – kleiner wordt.

Om tot het uiteindelijke weegmodel te komen is de volgende werkwijze gehanteerd:

1. Als eerste is de samenhang van de beschikbare hulpvariabelen (zie Bijlage 2) met de respons onderzocht, met uitzondering van aantminzorgzv (Aantal minuten zorg zonder verblijf) en oplniv (Hoogst behaalde opleiding) vanwege het hoge percentage missende waarden (>65%). Hiertoe is een stapsgewijze logistische regressieanalyse uitgevoerd met als afhankelijke variabele de respons.

Omdat de resultaten van de gezondheidsenquête per gemeente worden gepresenteerd, wordt de variabele gemeente a priori in het weegmodel opgenomen. Om die reden is de variabele gemeente in de eerste stap van de regressie standaard in de analyse opgenomen. In de navolgende stappen werd steeds in iedere stap die variabele aan het model toegevoegd die de grootste verklarende waarde had ten opzichte van de respons en waarvan de  $p$ -waarde van de regressiecoëfficiënt kleiner was dan 0,05. Variabelen die eerder in het model waren opgenomen, maar na toevoeging van één of meer andere variabelen niet langer een significante bijdrage leverden aan het verklaren van de respons ( $p > 0,1$ ), werden uit het model verwijderd. De analyse werd afgerond wanneer er geen variabelen meer waren die een significante bijdrage konden leveren aan het regressiemodel.

Een overzicht van mogelijke weegvariabelen staat in Bijlage 2. In de bijlage is te zien dat een aantal van deze weegvariabelen continue variabelen zijn. Voor deze variabelen zijn verschillende indelingen uitgetoetst (2, 3, 4 en 5 categorieën), waarna de indeling die het sterkst samenhangt met de respons is gebruikt voor het vervolg van de weging (indien van toepassing).

Verder is voor leeftijd de door de GGD gehanteerde indeling in 3 categorieën aangehouden. Daarnaast is een indeling in 9 categorieën gemaakt, wat bij het leeftijdsbereik van de doelpopulatie overeenkomt met leeftijd in 5-jaars klassen.

2. De samenhang van de hulpvariabelen en een zestal belangrijke doelvariabelen werd op vergelijkbare wijze onderzocht. De volgende doelvariabelen zijn hiervoor geselecteerd:
  - Ervaren gezondheid slecht;
  - Ervaren psychische gezondheid slecht;
  - Contact met de huisarts in het afgelopen jaar;
  - Contact met de medische specialist in het afgelopen jaar;
  - Roker;
  - Drinkt alcohol;
  - Overmatig drankgebruik (mannen: >21 glazen/week; vrouwen: >14 glazen/week).
3. De op basis van stap 1 en 2 geselecteerde hulpvariabelen zijn in het weegmodel opgenomen. Dit weegmodel is vervolgens aangepast: De variabele sociaal-economische categorie zorgde voor een groot aantal negatieve gewichten en is daarom uit het model verwijderd. In plaats daarvan is de variabele arbeidsongeschiktheid aan het model toegevoegd. Negatieve gewichten zijn soms acceptabel voor het schatten van doelvariabelen die sterk met de betreffende hulpvariabele correleren. Omdat in deze weging de gewichten voor alle doelvariabelen moeten kunnen worden toegepast, is ervoor gekozen negatieve gewichten te vermijden.



## 5. Resultaten

### 5.1 Insluitgewichten

De insluitgewichten  $d_j$  worden bepaald door het steekproefontwerp. Dit steekproefontwerp, inclusief de extra steekproeven heeft  $J = 38$  strata (zie Bijlage 1).

- De reguliere steekproef bestaat uit 29 gemeenten, waarbij de gemeente Oss is opgedeeld in kern en buitengebied ( $J = 30$  strata);
- Er zijn extra steekproeven getrokken in 8 wijken (oversampling) in de gemeente Oss kern, 's-Hertogenbosch (beide 2 wijken) en Tilburg (4 wijken) (zie Bijlage 1). Deze wijken zijn in het steekproefontwerp opgenomen als aparte strata ( $J = 8$  strata).

Uit dit ontwerp volgt dat personen in de extra wijksteekproeven, die niet in de reguliere steekproef zaten, twee keer een kans hadden om in de steekproef getrokken te worden: één keer in de reguliere steekproef en één keer in de extra steekproef. Tevens waren de populatieaantallen van de drie betreffende gemeenten niet verminderd met de wijkpopulaties. Om strata te krijgen die onderling geen overlap hebben en te zorgen dat de insluitgewichten optellen tot het populatietotaal, zijn de insluitgewichten in de weging als volgt gecorrigeerd:

- de populatieomvang van de drie gemeenten is verminderd met de populatieomvang van de oversampled wijken;
- de steekproefaantallen van de gemeenten zijn verminderd met het aantal personen uit oversampled wijken dat was getrokken in de reguliere steekproef;
- de steekproefaantallen van de oversampled wijken zijn vermeerderd met het aantal personen uit oversampled wijken dat in de reguliere steekproef zat.

### 5.2 Missende waarden

Een aantal voor het weegmodel geselecteerde variabelen in het CBS populatiebestand bevatte missende waarden, met een maximum van 4,7% voor huishoudgrootte (zie tabel 5.1). Missende waarden zijn op de volgende wijze geïmputeerd:

- Voor de variabelen burgerlijke staat, woningtype (huur/koop) en huishoudgrootte zijn de missende waarden naar rato willekeurig verdeeld over de categorieën van de betreffende variabele.
- Missende waarden van huishoudinkomen (onder/boven modaal) konden voor een deel worden geïmputeerd op basis van informatie in de variabelen mndbedragpers (Fiscaal bruto maandbedrag persoon) en mndbedragsec (Fiscaal bruto maandbedrag sociaal-economische categorie). Wanneer de waarde van beide variabelen in de categorie boven dan wel onder modaal van huishoudinkomen viel, of één van de variabelen geen waarde bevatte en de andere wel, werd de overeenkomstige waarde in huishoudinkomen geïmputeerd. Vervolgens werden de resterende missende waarden geïmputeerd zoals beschreven in de vorige alinea.

- Bij etniciteit betroffen alle missende waarden allochtonen, zo bleek uit kruising met de variabele waarin de generatie allochtoon. De missende waarden zijn daarom naar rato willekeurig verdeeld over de categorieën westerse en niet-westerse allochtoon.
- Bij sociaal economische categorie zijn de missende waarden toegevoegd aan de categorie ‘overige’.

Geslacht, leeftijd en gemeente waren voor alle personen bekend.

**Tabel 5.1: Percentages missende waarden van kandidaat variabelen voor het definitieve weegmodel**

Variabele	CBS-populatiebestand	GGD-databestand*
Gemeente	0,0	0,0
Leeftijd	0,0	0,0
Geslacht	0,0	0,0
Etniciteit	1,0	0,7
Burgerlijke staat	0,4	0,2
Sociaal-economische categorie	0,3	0,2
Huishoudgrootte	4,7	0,3
Huishoudinkomen	0,9	2,0
Woningtype	1,5	0,7
Arbeidsongeschiktheid	0,0	0,0

\* Alleen van respondenten.

### 5.3 Weegmodellen

De huidige uitgebreide weging is niet direct vergelijkbaar met de weging van de GGD. Dit komt ten eerste doordat de populatietabellen van de GGD iets afwijken van die van de huidige weging. Daarnaast is in het weegmodel van de GGD geen rekening gehouden met de verhoogde trekkingskans van personen in de extra steekproeven. Omdat deze factoren naast het weegmodel zelf van invloed zijn op de resulterende eindgewichten, zijn een aantal weegmodellen opgesteld om het effect van deze factoren in kaart te brengen. Onderstaand worden de overeenkomsten en verschillen tussen de verschillende modellen toegelicht.

Model 1 is het eigen weegmodel van de GGD: gemeente (38)  $\times$  leeftijd (3)  $\times$  geslacht (2). De weegfactoren zijn door de GGD berekend door voor elke cel van het weegmodel het populatieaantal te delen door het aantal respondenten (poststratificatie). Model 2 repliceert dit model op basis van de informatie uit het door de GGD beschikbaar gestelde databestand en populatietabellen. In beide modellen is niet gecorrigeerd voor de populatieoverlap. Model 3 is identiek aan model 2, zij het dat de populatieaantallen verschillen, omdat is gecorrigeerd voor de dubbele trekkingskans/populatieoverlap gerelateerd aan de oversampled wijken (zie paragraaf 5.1). Model 4 is identiek aan model 3, maar maakt gebruik van het populatiebestand van het CBS in plaats van dat van de GGD. In alle vier deze modellen worden de weegfactoren berekend door middel van poststratificatie. De schattingen van de geselecteerde doelvariabelen worden verkregen met (1) en de bijbehorende standaardfouten worden berekend via Taylor linearisatie.

Model 5 is het definitieve weegmodel, waar rekening wordt gehouden met de ongelijke insluitkansen via de insluitgewichten. Omdat naar meerdere populatietabellen tegelijkertijd wordt gewogen, is poststratificatie hier niet mogelijk. Daarom is gebruik gemaakt van lineair wegen<sup>3</sup>. De volgende variabelen zijn in het definitieve weegmodel opgenomen (aantal categorieën tussen haakjes):

Leeftijd (9) +  
 Gemeente (38) × Geslacht (2) +  
 Gemeente (38) × Leeftijd (3) +  
 Geslacht (2) × Leeftijd (3) +  
 Geslacht (2) × Burgerlijke staat (3) +  
 Leeftijd (3) × Burgerlijke staat (3) +  
 Etniciteit (3) +  
 Arbeidsongeschikt (2) +  
 Huishoudgrootte (5) +  
 Huishoudinkomen (2) +  
 Woningtype (2).<sup>4</sup>

Merk op dat ter voorkoming van negatieve gewichten het minimum correctiegewicht werd vastgelegd op 0; zonder deze limiet werden er drie negatieve gewichten gevonden. De variabelen in dit definitieve weegmodel worden uitgebreid beschreven in Bijlage 3. Al deze variabelen hingen afzonderlijk significant samen met de respons, met uitzondering van Arbeidsongeschikt die sterk met de geselecteerde doelvariabelen samenhangt. De variabele sociaal-economische categorie was ook een voorspeller van respons, maar is uit het weegmodel verwijderd om negatieve gewichten te vermijden (zie paragraaf 4.3).

#### 5.4 Resultaten van de weging

Model 2 was een poging om de door de GGD uitgevoerde weging te repliceren en zo als uitgangspunt te dienen voor model 3 t/m 5. Dit is echter niet gelukt: de eindgewichten van model 2 bleken niet in alle gevallen overeen te komen met die van model 1 (zie tabel 5.3). Deze discrepantie vindt zijn oorsprong in kleine verschillen in (de kruising van) de variabelen gemeente, leeftijd en geslacht zoals de GGD voor haar weging heeft gebruikt (model 1) en zoals opgenomen in het voor deze weging beschikbaar gestelde databestand. In de tabel is echter ook te zien dat het totaalplaatje vrijwel gelijk is, met als grootste verschil een hoger maximum gewicht in model 2. Verder is het opvallend dat het gemiddelde eindgewicht en de bijbehorende standaard deviatie van model 3, 4 en 5 substantieel lager zijn dan van model 1 en 2.

---

<sup>3</sup> Feitelijk is poststratificatie een specifieke vorm van lineair wegen, waarbij er maar één populatietabel is.

<sup>4</sup> De variabelen in de interactietermen worden in Bascula automatisch meegenomen als hoofdeffecten in het weegmodel.

Dik kan worden toegeschreven aan correctie van de populatie voor de oversamplende wijken. Het kleine verschil in het gemiddelde eindgewicht tussen enerzijds model 3 en anderzijds model 4 en 5 kan verklaard worden door het kleine verschil in het populatietotaal van de GGD en CBS populatiebestanden.

Tabel 5.2 toont de verdeling van de correctiegewichten van de verschillende modellen en tabel 5.3 de verdeling van de eindgewichten. Er zijn weinig verschillen tussen de modellen onderling, wat suggereert dat de correctie voor dubbele trekkingskans/populatieoverlap (model 3) en het gebruik van CBS populatietabellen (model 4) de verdeling van correctie- en eindgewichten weinig beïnvloedt. Merk op dat er voor model 1 geen correctiegewichten beschikbaar zijn. De reden hiervoor is dat de GGD de eindgewichten in één stap heeft berekend, namelijk door voor elk stratum het populatieaantal te delen door het responsaantal.

**Tabel 5.2: Verdeling van de correctiegewichten**

	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4</b>	<b>Model 5</b>
Minimum	0,63	0,45	0,44	0,17
1 <sup>e</sup> kwartiel	0,81	0,81	0,80	0,73
Mediaan	0,91	0,91	0,91	0,93
3 <sup>e</sup> kwartiel	1,09	1,08	1,08	1,21
Maximum	2,42	2,46	2,44	3,54
M	1,00	1,00	1,00	1,00
SD	0,29	0,30	0,30	0,40

Wegen met het definitieve weegmodel (model 5) levert een andere verdeling van correctie- en eindgewichten dan wegen met de andere modellen. De conclusie is daarom dat de verschillen met het oorspronkelijke weegmodel van de GGD zijn toe te schrijven aan het weegmodel en niet aan verschillen in de populatietabellen. Bij vergelijking tussen model 2 en 5 valt op dat de spreiding van de gewichten bij model 5 hoger is en dat de extremen verder uit elkaar liggen. Er zijn echter geen extreem hoge correctiegewichten; het maximum is 3,5.

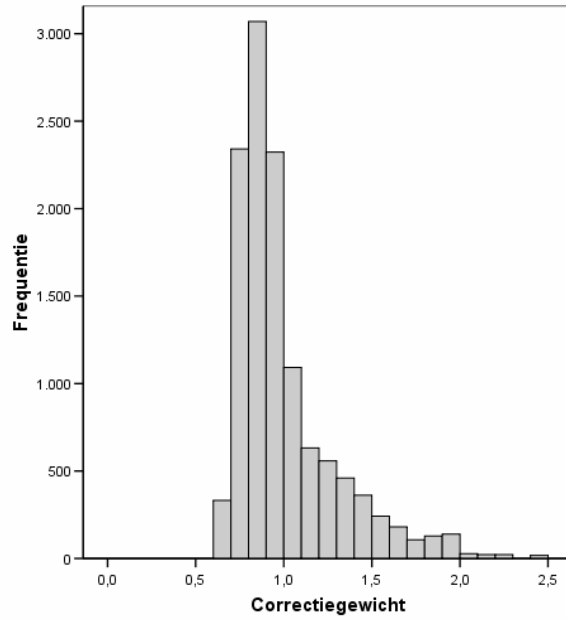
**Tabel 5.3: Verdeling van de eindgewichten**

	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4</b>	<b>Model 5</b>
Minimum	10,2	10,0	10,0	9,7	2,6
1 <sup>e</sup> kwartiel	21,0	21,3	21,3	21,6	20,5
Mediaan	35,4	35,5	35,5	35,3	33,7
3 <sup>e</sup> kwartiel	54,7	52,7	52,7	52,4	53,5
Maximum	850,9	911,7	788,6	787,6	987,8
M	56,2	56,4	53,0	52,8	52,8
SD	84,0	85,0	73,3	73,0	76,8

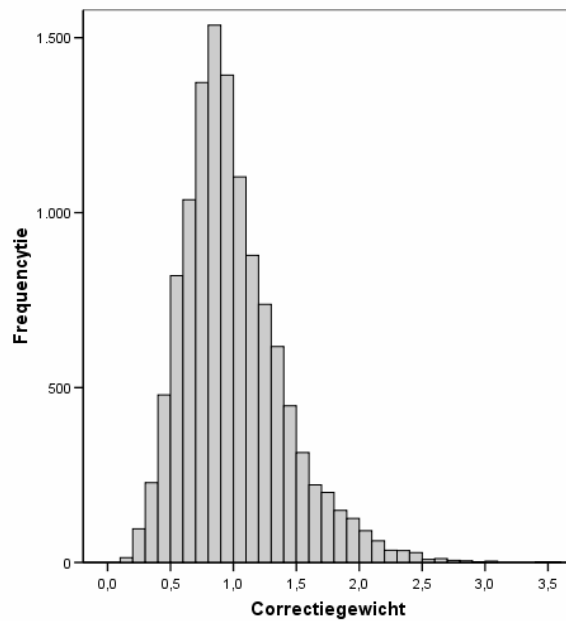
Figuur 5.1 toont de correctiegewichten van model 2 en figuur 5.2 van model 5. De figuren bevestigen de grotere spreiding van correctiegewichten na wegen met model 5 in vergelijking tot model 2. Daarnaast wordt duidelijk dat het aantal verschillende gewichten resulterend uit model 5 groter is; niet vreemd gezien het aanzienlijk grotere aantal weegtermen in dit model dan in model 2. De eindgewichten van model 2 en 5 worden tot slot weergegeven in figuur 5.3 en 5.4. De rechter grafiek in deze

twee figuren is steeds een uitvergroting van de rechter staart van de linker grafiek en toont alleen eindgewichten groter dan 100.

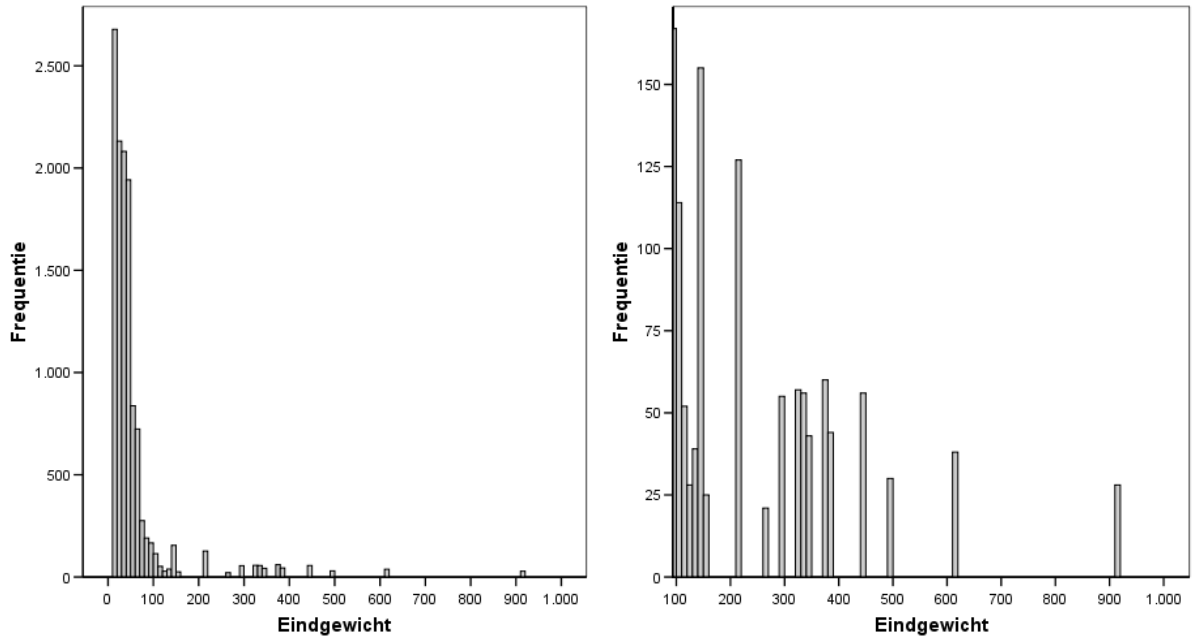
**Figuur 5.1: Verdeling van de correctiegewichten van model 2**



**Figuur 5.2: Verdeling van de correctiegewichten van model 5**

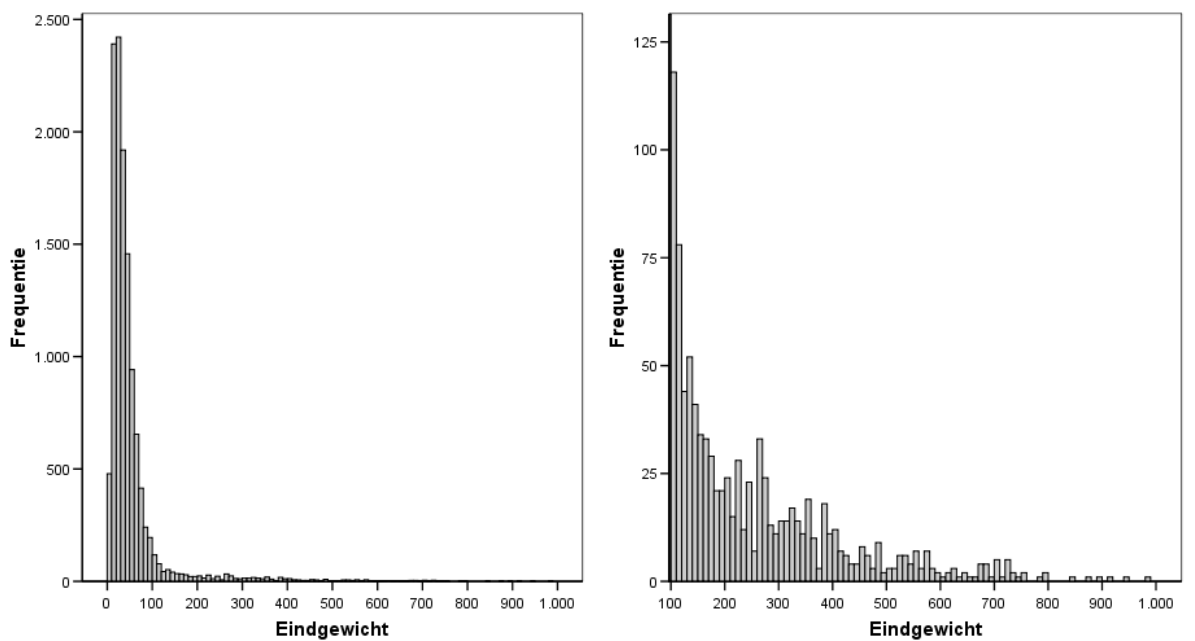


**Figuur 5.3: Verdeling van de eindgewichten van model 2**



*Noot. De rechter grafiek is een uitvergroting van de eindgewichten >100 in de linker grafiek.*

**Figuur 5.4: Verdeling van de eindgewichten van model 5**



*Noot. De rechter grafiek is een uitvergroting van de eindgewichten >100 in de linker grafiek.*

De uitkomsten van de zes belangrijke doelvariabelen staan in tabel 5.4, met indien beschikbaar tussen haakjes de standaardfouten. Hier is te zien dat er verschillen zijn tussen de schattingen van de populatiegemiddelden wanneer wel en niet gewogen is, wat aangeeft dat wegen zinvol is. De verschillen tussen de schattingen op basis van model 1 t/m 4 zijn echter verwaarloosbaar klein. Net als de verdeling van de correctie- en eindgewichten worden blijkbaar ook de schattingen weinig beïnvloed door verschillen in de populatietabellen. Wel dalen de standaardfouten tussen model 2 en 3 consequent met ruim 5%; dit kan wederom worden toegeschreven aan de correctie van de populatie voor de oversampled wijken. Het weegmodel zelf heeft wel invloed. De verschillen tussen schattingen met het uitgebreide weegmodel (model 5) en het oorspronkelijke GGD model (model 2) zijn namelijk groter.

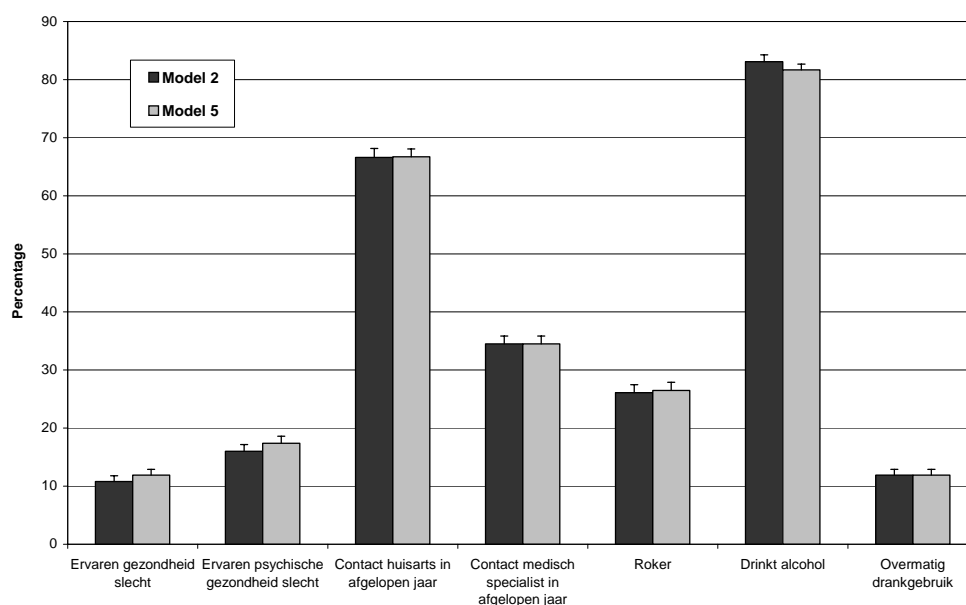
**Tabel 5.4: Schattingen van populatiegemiddelden op basis van weegmodellen**

	Geen weging	Model 1	Model 2	Model 3	Model 4	Model 5
Ervaren gezondheid slecht	11,6	10,8	10,8 (0,49)	10,8 (0,46)	10,7 (0,46)	11,9 (0,48)
Ervaren psychische gezondheid slecht	15,3	16,0	16,0 (0,62)	15,8 (0,58)	15,8 (0,58)	17,4 (0,62)
Contact huisarts in afgelopen jaar	67,1	66,6	66,6 (0,77)	66,6 (0,73)	66,6 (0,73)	66,7 (0,74)
Contact medisch specialist in afgelopen jaar	36,9	34,4	34,5 (0,74)	34,6 (0,70)	34,6 (0,70)	34,5 (0,71)
Roker	25,1	26,2	26,1 (0,73)	26,1 (0,69)	26,1 (0,69)	26,5 (0,71)
Drinkt alcohol	81,8	83,2	83,1 (0,56)	83,2 (0,53)	83,2 (0,53)	81,7 (0,54)
Overmatig drankgebruik	11,5	12,0	11,9 (0,53)	11,9 (0,50)	11,9 (0,50)	11,9 (0,51)

*Cijfers zijn percentages met standaardfouten tussen haakjes.*

In figuur 5.5 worden de schattingen op basis van het weegmodel van de GGD (model 2) en het definitieve weegmodel (model 5) naast elkaar weergegeven met hun

**Figuur 5.5: Vergelijking tussen de schattingen op basis van weegmodel 2 en 5**



standaardfouten. Hoewel de verschillen over het algemeen niet erg groot zijn, geldt voor ervaren gezondheid, ervaren psychische gezondheid en drankgebruik dat de schattingen van model 4 en 5 buiten elkaars 95% betrouwbaarheidsinterval vallen. Daarbij zij opgemerkt dat de schattingen van de populatiegemiddelden van de verschillende variabelen zijn weergegeven. Wanneer de schattingen worden uitgesplitst naar achtergrondvariabelen, zoals leeftijd, geslacht, etniciteit of burgerlijke staat, kunnen mogelijk grotere verschillen optreden ten opzichte van het GGD-weegmodel.

## 6. Conclusie

In dit rapport wordt de weging van de Gezondheidsenquête van de GGD regio Hart voor Brabant beschreven. Deze weging is uitgevoerd in het kader van een pilot waarin bekeken wordt wat de extra mogelijkheden zijn voor GGD'en als de gegevens uit hun gezondheidsenquêtes verrijkt worden met gegevens van het CBS.

Het rapport begint met een beschrijving van de bestanden die gebruikt zijn voor de uitgebreide weging: het respons- en populatiebestand van de GGD en het populatiebestand voor de uitgebreide weging. Vervolgens wordt ingegaan op de steekproef-trekking en op de wijze waarop de weging is uitgevoerd. Er is niet direct overgegaan tot het opstellen van een uitgebreid weegmodel, maar met het weegmodel van de GGD als uitgangspunt zijn enkele tussenstappen gemaakt. Dit was nodig om het effect op de weging van een aantal verschillen tussen de populatietabellen van de GGD en het CBS in kaart te brengen. Het gaat om de volgende verschillen:

- De GGD heeft alleen gewogen naar gemeente, leeftijd en geslacht. Om additionele variabelen in het weegmodel te kunnen opnemen is een populatiebestand samengesteld waarin een aantal kandidaat achtergrondvariabelen voor de weging waren opgenomen.
- De GGD heeft in een aantal wijken extra steekproeven getrokken, maar de populatieaantallen van de betreffende gemeentes niet gecorrigeerd. Hierdoor zijn de wijkpopulaties dubbel geteld, zodat personen in deze wijken twee keer kans hadden in de steekproef te worden getrokken: één keer als inwoner van de gemeente en één keer als inwoner van de wijk.

Het effect van deze verschillen is onderzocht door de populatieaantallen van het weegmodel van de GGD (gemeente  $\times$  leeftijd  $\times$  geslacht) aan te passen (model 3 en 4).

De resultaten van de weegmodellen die zijn gebruikt om deze tussenstappen zichtbaar te maken, laten zien dat de extra steekproeven en het gebruikte populatiebestand weinig effect hebben op de resulterende verdeling van de correctie- en eindgewichten en op de schattingen van de populatiegemiddelden. Daarom kan geconcludeerd worden dat de invloed van bovengenoemde verschillen in de populatietabellen klein is. Dit impliceert dat het definitieve weegmodel (model 5) rechtstreeks kan worden vergeleken met het weegmodel van de GGD (model 2). Hiervoor is de



replicatie van de CBS weging gebruikt, omdat in verband met het beschikbaar zijn van correctie- en eindgewichten en standaardfouten van de schattingen het beter mogelijk is beide modellen te vergelijken.

De vergelijking tussen deze twee weegmodellen laat zien dat het uiteindelijke, uitgebreide weegmodel invloed heeft op de schattingen van zeven geselecteerde doelvariabelen. Dit zijn ervaren gezondheid, ervaren psychische gezondheid, contact met de huisarts, contact met de medische specialist, roken, alcoholgebruik en overmatig alcoholgebruik. De verschillen zijn over het algemeen klein, behalve voor ervaren gezondheid, ervaren psychische gezondheid en alcoholgebruik.

Het is belangrijk om bij de interpretatie van deze uitkomsten in de gaten te houden dat de schattingen zijn gemaakt op gemiddelden van de doelvariabelen voor de gehele doelpopulatie. De doelstelling van de GGD regio Hart voor Brabant is echter om schattingen te maken per gemeente. De bevinding dat de totaalschattingen van de twee modellen op regioniveau weinig verschillen, betekent echter niet dat dit ook op gemeenteniveau het geval is. Meer algemeen geldt dat wanneer verdere uitsplitsing plaatsvindt naar bijvoorbeeld leeftijd, etniciteit en/of burgerlijke staat, het goed mogelijk is dat de bevindingen op basis van het uitgebreide weegmodel méér afwijken van het oorspronkelijke GGD-weegmodel.

Ook wanneer besloten zou worden het uitgebreide weegmodel toe te passen op een grotere populatie of zelfs landelijk, is voorzichtigheid vereist. Het is belangrijk om dan eerst te controleren of het weegmodel hiervoor wel geschikt is. In concreto dient te worden nagegaan of de verdeling van de non-respons in andere regio's van Nederland afwijkt van die in de regio Hart voor Brabant. Een voorstudie op de weging van de landelijke gezondheidsenquête van het CBS (GE 2010) suggereert dat dit het geval is (Buelens, 2010). In het weegmodel van deze enquête zijn namelijk de variabelen provincie en landsdeel opgenomen en wordt daarnaast rekening gehouden met stedelijkheidsgraad en de maand waarin de enquête is afgenomen. Dit is een sterke aanwijzing dat aanpassing van het weegmodel nodig is indien men landelijke responsgegevens van de GGD gezondheidsenquête wil wegen.

Tot slot nog één opmerking. In een poging de weging van de GGD te repliceren op het CBS populatiebestand, bleek dat de wijkindeling van de GGD en het CBS niet volledig overeenkomen. Dit bleek het geval te zijn voor de Tilburgse wijken; zo komt de wijk Fatima niet voor in de wijkindeling van het CBS. Het wordt aanbevolen dat met identieke indelingen gewerkt wordt.

## **7. Dankwoord**

De auteurs danken Jaap van den Berg voor de prettige samenwerking en zijn scherpe commentaar, Marc Aaldijk, Fred Gast en Alex Hellenthal voor hun hulp bij het tot stand komen van het populatiebestand en Gemma Smulders voor het aanleveren van voor deze weging relevante informatie vanuit de GGD Hart voor Brabant.

## 8. Referenties

- Buelens, B. (2010). *Voorstudie weging Gezondheidsenquête 2010*. CBS interne nota. BPA-nummer: DMH-2010-11-16-BBUS.
- Burger, J., Gouweleeuw, J.M., & Buelens, B. (2010). *Weging Woononderzoek Nederland 2009, basismodule Woningmarkt met oversampling*. Heerlen: Centraal Bureau voor de Statistiek. BPA-nummer DMH-2010-01-210-JBUR.
- Nieuwenbroek, N. & Boonstra, H.J. (2002). *Bascula 4.0 reference manual*. Heerlen: Centraal Bureau voor de Statistiek.
- Särndal, C.E., Swensson, B. & Wretman, J. (1992). *Model assisted survey sampling*. New York: Springer-Verlag.
- Van den Brink, C., van Hilten, O. & van den Berg, J. *Pilot integratie Lokale en Nationale Monitor en CBS POLS-Gezondheid*.

## Bijlage 1. Gemeenten in de steekproef

Gemeente	Gemeentecode	Extra steekproeven
Bernheze	1721	
Boekel	0755	
Boxmeer	0756	
Boxtel	0757	
Cuijk	1684	
Dongen	0766	
Gilze en Rijen	0784	
Goirle	0785	
Grave	0786	
Haaren	0788	
's-Hertogenbosch	0796	Kruiskamp en Schutskamp Boschveld en Deuteren
Heusden	0797	
Hilvarenbeek	0798	
Landerd	1685	
Lith	0808	
Loon op Zand	0809	
Maasdonk	1671	
Mill en Sint Hubert	0815	
Oisterwijk	0824	
Oss kern	0828	Schadewijk Ruwaard
Oss buitengebied		
Schijndel	0844	
Sint Anthonis	1702	
Sint-Michielsgestel	0845	
Sint-Oedenrode	0846	
Tilburg	0855	Fatima, Jeruzalem en Broekhoven Groenewoud Kruidenbuurt Stokhasselt
Uden	0856	
Veghel	0860	
Vught	0865	
Waalwijk	0867	

## Bijlage 2. Achtergrondvariabelen in het CBS populatiebestand

Variabele	Uitleg	
GEMJJJJ	Gemeentecode	N
GESLACHT		N
GEBJAAR	Geboortejaar	O
Leeftijd	Leeftijd afgeleid uit GEBJAAR en peildatum	O
ETNGROEP	Herkomst-groep	N
ETNGEN	Herkomst-generatie	N
BURGSTPBL	Burgerlijke staat	N
MNDBEDRAGPERS	Fiscaal bruto maandbedrag persoon	C
MNDBEDRAGSEC	Fiscaal bruto maandbedrag SEC	C
SEC	Sociaal-economische categorie	N
BAAN	Wel/geen baan	N
ZLF	Wel/geen zelfstandige	N
ABW	Wel/geen bijstandsuitkering	N
AO	Wel/geen arbeidsongeschiktheidsuitkering	N
WW	Wel/geen WW-uitkering	N
OVU	Wel/geen sociale uitkering	N
PNS	Wel/geen pensioen	N
WSF	Wel/geen student	N
OARB	Wel/geen inkomen uit overige arbeid	N
ARBBUIT	Wel/geen inkomen uit arbeid uit buitenland	N
UITKBUIT	Wel/geen uitkering uit het buitenland	N
WSF	Wel/geen studiefinanciering	N
CWI	Wel/geen inschrijving bij het CWI	N
AANTMINZORGZV*	Aantal minuten zorg zonder verblijf	C
PLHHN	Plaats van persoon in het huishouden	N
TYPHHN	Type huishouden	N
AANTPERSHH	Aantal personen in het huishouden	C
AANTKINDHH	Aantal kinderen in het huishouden	C
SOORTVERDIENER	Aantal verdiemers per huishouden	N
MNDBEDRAGHHKERN	Fiscaal bruto maandbedrag binnen huishouden	C
WOZWaarde	WOZ-waarde woning	C
HuurKoop	Huur- of koopwoning	N
STEDGEM	Stedelijkheid van de gemeente	O
OPLNIV*	Hoogst behaalde opleiding	O

N = nominale variabele, O = ordinale variabele, C = continue variabele.

\* Variabele heeft veel missende waarden en is daarom geen kandidaat voor het weegmodel.

### **Bijlage 3. Weegvariabelen in het definitieve weegmodel**

#### Gemeente(29)

1. Bernheze
2. Boekel
3. Boxmeer
4. Boxtel
5. Cuijk
6. Dongen
7. Gilze en Rijen
8. Goirle
9. Grave
10. Haaren
11. 's-Hertogenbosch
12. Heusden
13. Hilvarenbeek
14. Landerd
15. Lith
16. Loon op Zand
17. Maasdonk
18. Mill en Sint Hubert
19. Oisterwijk
20. Oss
21. Schijndel
22. Sint Anthonis
23. Sint-Michielsgestel
24. Sint-Oedenrode
25. Tilburg
26. Uden
27. Veghel
28. Vught
29. Waalwijk

#### Geslacht(2)

1. man
2. vrouw

#### Leeftijd(3)

1. 19-34 jaar
2. 35-49 jaar
3. 50-64 jaar

#### Leeftijd(9)

1. 19-24 jaar
2. 25-29 jaar
3. 30-34 jaar
4. 35-39 jaar
5. 40-44 jaar
6. 45-49 jaar
7. 50-54 jaar
8. 55-59 jaar
9. 60-64 jaar

#### Burgerlijke staat(3)

1. gehuwd
2. nooit gehuwd geweest
3. gescheiden of verweduwd

Etniciteit(3)

1. autochtoon
2. westers allochtoon
3. niet-westers allochtoon

Huishoudinkomen(2)

1. onder modaal (€ 3790,-)
2. boven modaal (€ 3790,-)

Woningtype(2)

1. huurwoning
2. koopwoning

Huishoudgrootte(5)

1. éénpersoonshuishouden
2. tweepersoonshuishouden
3. driepersoonshuishouden
4. vierpersoonshuishouden
5. huishouden bestaande uit vijf of meer personen

Arbeidsongeschikt(2)

1. wel arbeidsongeschikt
2. niet arbeidsongeschikt