



Paper

# Weging gezondheids- enquête 2014

Harm Jan Boonstra

April 2019

# Inhoud

- 1. Inleiding 4**
  
- 2. Weging 5**
  - 2.1 Gewichten 6
  - 2.2 Schattingen 6
  
- 3. Conclusie en aanbeveling 11**
  
- 4. Naschrift aanpassing weging vanaf statistiekjaar 2018 12**
  - Referenties 13
  - A Categorie-indeling van de weegvariabelen 14

Het ontwerp van de Gezondheidsenquête 2014 is op een aantal punten gewijzigd. In deze nota wordt een weging voorgesteld voor het nieuwe onderzoek waarbij het weegmodel van de Gezondheidsenquête 2013 als uitgangspunt dient. Een aantal uitbreidingen van dit weegmodel worden vergeleken aan de hand van gewichtenverdelingen en cijfers voor een aantal belangrijke doelvariabelen. Het uiteindelijke voorstel is om het weegmodel uit te breiden met gestandaardiseerd inkomen en vermogen, beide in kwintielen zoals ook gebruikt voor het detailleren van enkele gezondheidscijfers op StatLine. Deze variabelen zijn zowel gerelateerd aan de respons als aan enkele van de onderzochte doelvariabelen en zorgen voor een kleine extra selectiviteitscorrectie.

*Trefwoorden: Gezondheidsenquête, weging, Bascula*

# 1. Inleiding

De Gezondheidsenquête (GE) is in 2014 op een aantal punten veranderd ten opzichte van eerdere jaren. De sequentiële benaderstrategie is veranderd van een CAWI-CAPI/CATI strategie naar een CAWI-CAPI strategie. Dat wil zeggen dat de volledige non-respons op CAWI nu alleen nog via CAPI benaderd wordt, en niet meer via CATI (Goris, 2013). Ook wordt er vanaf 2014 nog maar één vragenlijst afgenomen, waar eerder sprake was van een tweede vragenlijst als vervolgmodule. De extra weging voor de vervolgmodule (Bruggink en van Hertem, 2012) om te corrigeren voor tussentijdse uitval is dus niet meer nodig.

Verspreid over het jaar wordt een aselechte steekproef onder alle personen van de niet-institutionele bevolking uitgezet. In totaal zijn 9516 responsen gerealiseerd in 2014. In deze nota wordt de weging van deze respons beschreven. In een vooronderzoek, uitgevoerd in het eerste kwartaal van 2014 zijn een aantal potentieel belangrijke achtergrondkenmerken aangemerkt om mogelijk in de weging te gebruiken (Boonstra, 2014). Uiteindelijk is de keuze gemaakt om uitbreiding van het bestaande weegmodel met inkomen en vermogen te onderzoeken. Deze sociaal-economische achtergrondvariabelen zouden een goede aanvulling kunnen zijn op de demografische en regionale achtergrondvariabelen waaruit het tot nu toe gebruikte GE weegmodel is opgebouwd.

Inkomen wordt tegenwoordig vaker in wegingen van personensteekproeven gebruikt, onder meer bij de EBB, WoON en OViN (Smeets, 2011). Vooral lagere inkomensgroepen zijn vrijwel altijd ondervertegenwoordigd in de respons. In Bruggink et al. (2013) is informatie uit het inkomensregister gebruikt om kleine-domeinschattingen voor GGD-regio's gebaseerd op de GE 2010 te verbeteren. Hierbij is het gestandaardiseerd besteedbaar inkomen gebruikt. Dat is het besteedbare inkomen waarbij rekening gehouden wordt met grootte en samenstelling van het huishouden. Deze is gebruikt in 5 groepen van gelijke omvang (kwintielen) qua aantallen huishoudens. Voor een aantal belangrijke doelvariabelen van de GE, zoals roken en overgewicht, is gevonden dat er een duidelijke samenhang bestaat met inkomen. De indeling van gestandaardiseerd huishoudinkomen in kwintielen wordt ook gebruikt om GE cijfers op StatLine naar uit te splitsen. Het ligt daarom voor de hand om dezelfde gegevens met dezelfde indeling in de weging te gebruiken. In Buelens et al. (2013) is gestandaardiseerd huishoudinkomen in kwintielen al meegenomen bij het wegen van de Gezondheidsmonitor 2012.

Ook vermogen wordt gebruikt in de StatLine tabellen over gezondheidscijfers: cijfers worden uitgesplitst naar 5 kwintielen. In Bruggink et al. (2013) is ook een relatie gevonden van vermogen met GE doelvariabelen, vooral met roken, ook als al gecorrigeerd is voor inkomen. Vermogen zou daarom in dezelfde kwintiel-indeling in de weging opgenomen kunnen worden. Net als inkomen is vermogen beschikbaar uit het inkomensregister. De weging wordt met Bascula uitgevoerd (Nieuwenbroek en Boonstra, 2002). Bascula kan ook varianties (en daarmee onzekerheidsmarges) berekenen die rekening houden met het steekproefontwerp en de weging. In deze nota worden schattingen met onzekerheidsmarges gegeven voor een aantal kerncijfers. Deze geven inzicht in de invloed van de extra weegvariabelen inkomen en vermogen.

## 2. Weging

Uitgangspunt voor de bepaling van het nieuwe weegmodel is het weegmodel dat gebruikt is voor het GE hoofdonderzoek in 2010 tot en met 2013:

$$\text{model0} = \text{Geslacht2} \times \text{Leeftijd16} + \text{BurgerlijkeStaat4} + \text{Stedelijkheidsgraad5} + \\ \text{ProvinciePlus16} + \text{Huishoudgrootte5} + \text{Geslacht2} \times \text{Leeftijd3} \times \text{BurgerlijkeStaat2} + \\ \text{Landsdeel4} \times \text{Leeftijd3} + \text{HerkomstGeneratie4} + \text{Seizoen4}$$

Elke variabelenaam eindigt op een getal dat het aantal categorieën aangeeft; zie Bijlage A voor de categorie-indelingen. In dit weegmodel zitten een aantal gebruikelijke demografische variabelen waarnaar de respons meestal selectief is zoals geslacht, leeftijd en herkomst. Ook zijn een aantal persoonskenmerken opgenomen waarnaar GE schattingen worden uitgesplitst: geslacht, leeftijd, stedelijkheidsgraad, provincie en herkomst.

In model0 is al één kleine wijziging aangebracht ten opzichte van het oorspronkelijke GE weegmodel. De variabele Leeftijd16 in 16 klassen in de kruisterm met geslacht is in de plaats gekomen van een indeling in 17 klassen. De indeling in 17 klassen had drie kleine klassen 12-14, 15-17 en 18-19 jaar. De nieuwe variabele Leeftijd16 voegt deze drie klassen samen tot twee klassen 12-15 en 16-19, zodat minder snel problemen met lage celvulling zullen ontstaan. De nieuwe indeling sluit bovendien beter aan bij de indeling die voor de publicatie wordt gebruikt.

We vergelijken de resultaten van drie wegingen. Er wordt gewogen met het oorspronkelijke GE weegmodel (model0), een uitbreiding hiervan met alleen inkomen, en een uitbreiding met inkomen en vermogen. Voor inkomen wordt de variabele Inkomen5 gebruikt, waarbij de populatie van huishoudens gesorteerd op gestandaardiseerd huishoudinkomen in vijf groepen van gelijke omvang wordt ingedeeld. Vervolgens worden deze huishoudkwintielen toegekend aan alle huishoudleden. Zo ontstaan op persoonsniveau vijf inkomensgroepen van ongelijke omvang. Voor vermogen wordt de variabele Vermogen5 gebruikt, die op analoge wijze is gedefinieerd in termen van huishoudvermogen.

Ter vergelijking worden ook ongewogen schattingen gemaakt om de invloed van het gehele weegmodel te kunnen zien. Omdat de GE een aselechte personensteekproef is zijn de insluitgewichten allemaal gelijk. Ongewogen schattingen voor populatiegemiddelden zijn dus gelijk aan schattingen gebaseerd op de insluitgewichten. We nemen de insluitgewichten gelijk aan  $N/n$  met  $N$  de populatieomvang en  $n = 9516$ , de totale responsomvang, zodat ze over alle respondenten optellen tot de populatieomvang.

De weging van de respons wordt uitgevoerd met Bascula. De weging past het insluitgewicht  $d_i$  van respondent  $i$  aan met een correctiegewicht  $g_i$  tot het eindgewicht

$$w_i = d_i g_i. \tag{1}$$

De correctiegewichten corrigeren voor verschillen tussen steekproef en populatie wat betreft de variabelen in het weegmodel. In paragraaf 2.1 worden de gewichten besproken en in paragraaf 2.2 de schattingen voor een aantal belangrijke GE doelvariabelen.

## 2.1 Gewichten

Bij een weging is het belangrijk om de verdeling van de gewichten te onderzoeken. Daarbij moet een redelijke balans tussen vertekening en variantie gevonden worden: spreiding van de gewichten is nodig om verschillen in samenstelling tussen respons en populatie op te heffen, maar kan ook voor een toename van de variantie van schattingen zorgen. In hoeverre de varianties van schattingen toenemen door spreiding van de gewichten hangt af van de relatie tussen de weegvariabelen en de doelvariabelen. Is die sterk dan kunnen de varianties zelfs kleiner worden door weging. In het slechtste geval, als er nauwelijks een relatie bestaat tussen de weegvariabelen en een doelvariabele dan zal de variantie toenemen met een factor  $d_{eff}$ , het design-effect door weging (Kish, 1987; Gabler et al., 1999),

$$d_{eff} = 1 + \frac{var(w)}{\bar{w}^2} \quad (2)$$

met  $var(w)$  de variantie van de eindgewichten  $w_i$  en  $\bar{w}$  het gemiddelde. In tabel 2.1.1 worden de gewichten onder de drie weegmodellen vergeleken wat betreft een aantal spreidingsmaten. Hierin is ook de wortel van het design-effect opgenomen, de factor waarmee standaardfouten van schattingen maximaal groter worden als gevolg van de weging. De tabel laat verder het minimum en maximum van de correctiegewichten zien, het aantal negatieve gewichten, en het aantal gewichten tussen 0 en 1 (corresponderende respondenten zouden voor minder dan zichzelf meetellen in populatieschattingen).

### 2.1.1 Spreidingsmaten voor de gewichten onder de drie weegmodellen

	Model0	Model0 + Inkomen5	Model0 + Inkomen5 + Vermogen5
min $g$	0,52	0,46	0,42
max $g$	2,05	2,06	2,11
aantal $w < 0$	0	0	0
aantal $0 < w < 1$	0	0	0
$\sqrt{d_{eff}}$	1,02	1,03	1,03

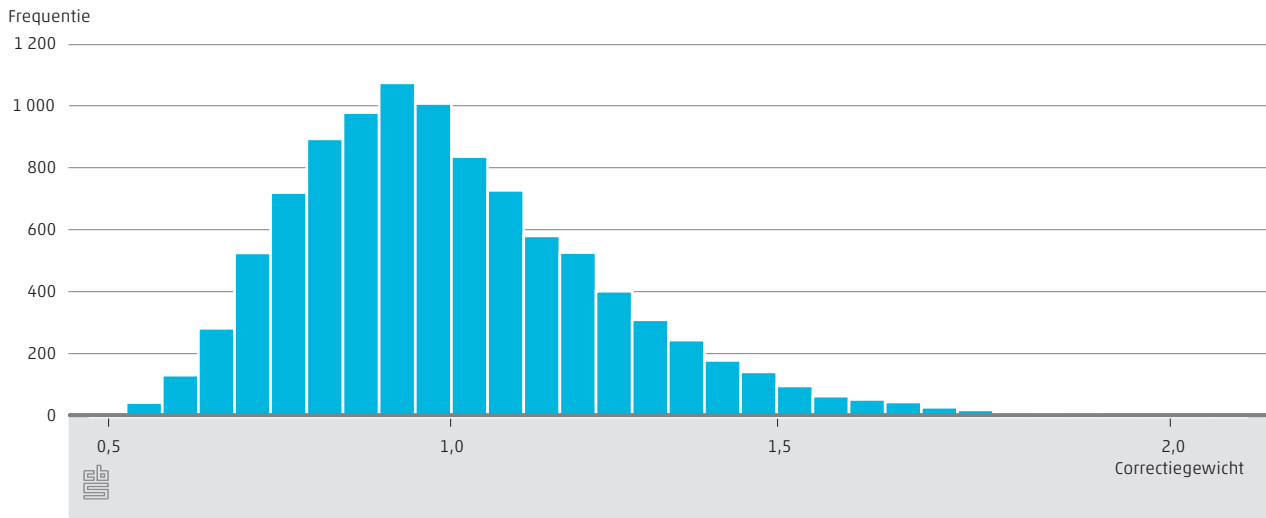
Bron: CBS.

Geen van de drie wegingen leidt tot een ongebruikelijk grote spreiding van de gewichten. Er zijn geen eindgewichten kleiner dan 1, en  $\sqrt{d_{eff}}$  wijkt niet veel van 1 af. Deze laatste neemt maar weinig toe door uitbreiding van het weegmodel met Inkomen5 en Vermogen5. Figuren 2.2.1–2.2.3 laten histogrammen zien van de gewichten. Ook daaraan is te zien dat de spreiding van de gewichten nauwelijks toeneemt door uitbreiding van het weegmodel met Inkomen5 en Vermogen5.

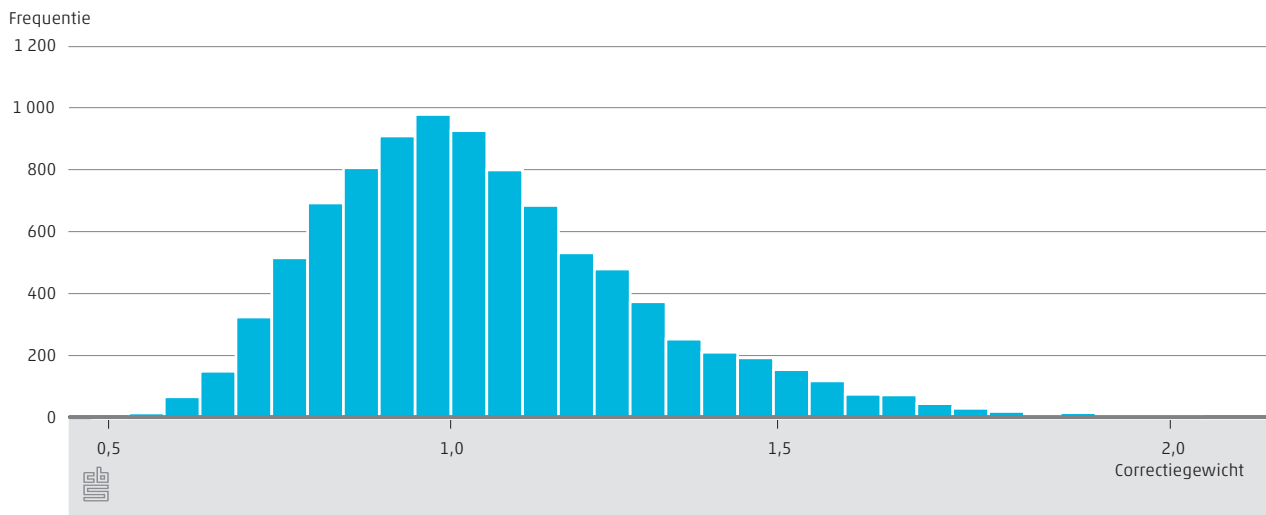
## 2.2 Schattingen

We onderzoeken de invloed van het weegmodel op een aantal kerncijfers gebaseerd op de gewogen GE respons. Daarbij gebruiken we de drie eerder genoemde weegmodellen en vergelijken ook nog met schattingen gebaseerd op de insluitgewichten, dat wil zeggen ongewogen schattingen.

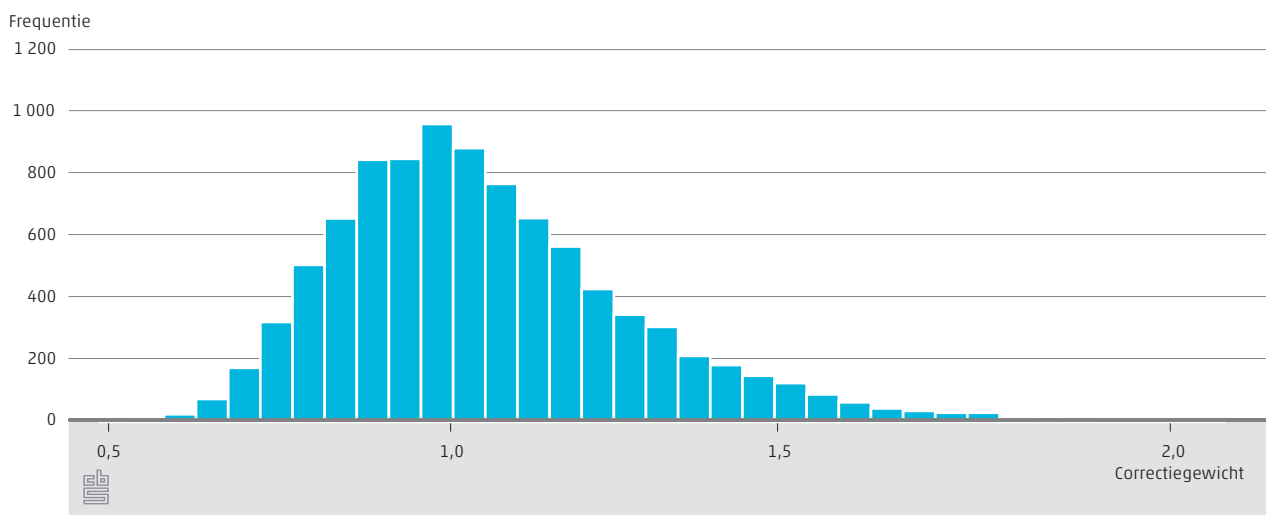
### 2.2.1 Gewichtenverdeling model0



### 2.2.2 Gewichtenverdeling model0 + Inkomen5



### 2.2.3 Gewichtenverdeling model0 + Inkomen5 + Vermogen5



De doelvariabelen die worden onderzocht zijn:

- ErvarenGezondheid (Ja/Nee): het antwoord op een vraag naar de algemene gezondheidstoestand. Ja betekent zeer goed/goed en Nee betekent minder dan goed.
- RookstatusRoker (Ja/Nee): het antwoord op de vraag 'rookt u wel eens?'. Alleen van toepassing op personen vanaf 12 jaar.
- Specialist (Ja/Nee): wel of niet een medisch specialist bezocht in de afgelopen 12 maanden.
- Suikerziekte (Ja/Nee).
- Minstens1OESOBep (Ja/Nee): of er sprake is van 1 of meer OESO beperkingen. Alleen van toepassing op personen vanaf 12 jaar. Deze OESO-indicator (Organisatie voor Economische Samenwerking en Ontwikkeling) is gebaseerd op een aantal vragen over vaardigheden.
- PsychischOngezonderMHI (Ja/Nee): of er sprake is van psychische ongezondheid. Alleen van toepassing op personen vanaf 12 jaar. MHI staat voor Mental Health Inventory, een internationale standaard voor het meten van de psychische ongezondheid gebaseerd op een aantal vragen.

Voor de meeste van deze doelvariabelen ontbreken enkele waarden. Omdat de gewichten betrekking hebben op alle respondenten wordt deze item non-respons geïmputeerd. Dit is gedaan door per klasse van Leeftijd16 ontbrekende waarden in te vullen met aselect gekozen beschikbare waarden.

In tabel 2.2.4 staan de schattingen voor deze doelvariabelen onder de verschillende weegmodellen. De opgenomen schattingen hebben allemaal betrekking op de minder gezonde deelpopulatie. Te zien is dat wegen de schattingen voor minder goed ervaren gezondheid, roken en psychische ongezondheid verhoogt en dat toevoegen van Inkomen5 en Vermogen5 dit effect nog iets versterkt. Voor de variabelen bezoek aan een specialist, suikerziekte en minimaal 1 beperking dalen de schattingen iets door weging met het oorspronkelijke GE weegmodel. Ze stijgen echter weer licht door uitbreiding van het weegmodel met Inkomen5 en Vermogen5. Bij de variabele roken is het effect van de weging relatief het grootst: het meest uitgebreide weegmodel geeft een schatting die bijna drie standaardfouten hoger ligt dan de ongewogen schatting.

#### 2.2.4 Ongewogen schattingen en schattingen gebaseerd op de gewichten onder de drie weegmodellen (in %)<sup>1)</sup>

	Ongewogen	Model0	Model0 + Inkomen5	Model0 + Inkomen5 + Vermogen5
ErvarenGezondheid: nee	19,17 (0,40)	19,27 (0,39)	19,52 (0,39)	19,57 (0,39)
RookstatusRoker: ja	22,78 (0,47)	23,89 (0,48)	24,12 (0,48)	24,17 (0,48)
Specialist: ja	39,80 (0,50)	39,32 (0,49)	39,38 (0,49)	39,43 (0,49)
Suikerziekte: ja	4,58 (0,21)	4,47 (0,21)	4,50 (0,21)	4,49 (0,21)
Minstens1OESOBep: ja	12,82 (0,37)	12,54 (0,34)	12,71 (0,34)	12,77 (0,34)
PsychischOngezonderMHI: ja	9,93 (0,33)	10,47 (0,35)	10,68 (0,36)	10,71 (0,36)

Bron: CBS.

<sup>1)</sup> Tussen haakjes staan geschatte standaardfouten. Een aantal van deze schattingen hebben alleen betrekking op de populatie van personen van 12 jaar en ouder.



Het toevoegen van Inkomen5 en Vermogen5 zorgt voor een kleine extra correctie voor selectieve non-respons. Tabellen 2.2.5 en 2.2.6 laten dit zien. Voor inkomen zijn de laagste klassen ondervertegenwoordigd en de hoogste oververtegenwoordigd. De weging met het oorspronkelijke GE model kan de selectiviteit naar inkomen maar voor een klein deel wegnemen. Voor vermogen is het tweede kwintiel ondervertegenwoordigd en het vijfde kwintiel het meest oververtegenwoordigd. De wegingen met het oorspronkelijke GE weegmodel en met het model inclusief Inkomen5 kunnen de selectiviteit naar vermogen maar deels wegnemen.

## 2.2.5 Schattingen voor omvangen van inkomenskwintielen (in %) onder de verschillende wegingen<sup>1)</sup>

Inkomenskwintiel	Ongewogen	Model0	Model0 + Inkomen5	Model0 + Inkomen5 + Vermogen5
1	12,6	13,8	15,7	15,7
2	17,2	17,2	17,7	17,7
3	22,1	21,7	21,4	21,4
4	23,7	23,4	22,3	22,3
5	24,4	24	22,9	22,9

Bron: CBS.

<sup>1)</sup> De laatste twee kolommen bevatten de populatieverdeling waar de wegingen met Inkomen5 op kalibreren.

Tabellen 2.2.7 en 2.2.8 laten cijfers zien voor ervaren gezondheid voor de deelpopulaties gedefinieerd door de inkomens en vermogenskwintielen, respectievelijk. In tabel 5 is te zien dat toevoegen van Inkomen5 en Vermogen5 aan het weegmodel niet veel doet met de schattingen per inkomenskwintiel. In tabel 2.2.8 is te zien dat toevoegen van Inkomen5 aan het weegmodel de schattingen per vermogenskwintiel licht doet stijgen. Toevoegen van Vermogen5 doet vervolgens niet veel met de schattingen per vermogenskwintiel. De schattingen dalen zelfs licht in alle 5 categorieën wat schijnbaar tegenstrijdig is met het feit dat de schatting voor de gehele populatie licht stijgt door toevoeging van Vermogen5, zie tabel 2.2.4. De verklaring voor dit geval van Simpson's paradox is dat het tweede vermogenskwintiel waarin het percentage personen met een minder dan goede gezondheid het hoogst is, het sterkst ondervertegenwoordigd is in de respons, zelfs na weging met model0 + Inkomen5, zie tabel 2.2.6.

In tegenstelling tot het laagste inkomenskwintiel bevat het laagste vermogenskwintiel relatief veel huiseigenaren. Het gaat hier waarschijnlijk in veel gevallen om mensen die recent een (eerste) woning hebben gekocht, en een hypotheekschuld hebben die hoger is dan de waarde van de woning. Hieronder bevinden zich waarschijnlijk veel relatief jonge mensen met een redelijk inkomen, wat het begrijpelijk maakt dat deze groep een goede gezondheid rapporteert. Tot de tweede vermogenscategorie, die ondervertegenwoordigd is in de respons, blijken juist nauwelijks mensen met een eigen woning te behoren.<sup>1)</sup>

Omdat de GE 2014 een andere waarneemstrategie hanteert met alleen nog maar CAWI en CAPI waarneming, bekijken we de gewogen en ongewogen aandelen CAWI en CAPI, zie tabel 2.2.9. Te zien is dat het aandeel CAPI slechts licht stijgt met het groter worden van het weegmodel. Zolang de waarneemstrategie hetzelfde blijft en de mode-samenstelling niet

<sup>1)</sup> Zie de Statline tabel [Vermogensklassen; particuliere huishoudens naar diverse kenmerken](#).

te sterk van jaar op jaar schommelt is er geen reden om te wegen naar een vaste mode-verhouding.

Ten slotte blijkt dat er wat schommelingen zijn in het aantal responsen per maand, met vooral een oververtegenwoordiging van februari. Deze worden nauwelijks weggenomen door de besproken wegingen. Door maand in plaats van seizoen in het weegmodel op te nemen wordt dit wel rechtgetrokken. Voor de hier onderzochte doelvariabelen had dit echter een verwaarloosbaar effect op de schattingen. Bovendien wordt niet naar maand gepubliceerd. Het lijkt daarom niet nodig om maand in het weegmodel op te nemen.

## 2.2.6 Schattingen voor omvangen van vermogenskwintielen (in %) onder de verschillende wegingen.

Vermogenskwintiel	Ongewogen	Model0	Model0 + Inkomen5	Model0 + Inkomen5 + Vermogen5
1	21,8	22,3	22,3	22,0
2	13,1	14,6	15,1	16,1
3	18,9	19,2	19,2	17,9
4	22,5	21,6	21,5	22,0
5	23,8	22,3	21,9	22,1

Bron: CBS.

<sup>1)</sup> De laatste kolom bevat de populatieverdeling waar de weging met Vermogen5 op kalibreert.

## 2.2.7 Schattingen voor ErvarenGezondheid:Nee per inkomenskwintiel<sup>1)</sup>

Inkomenskwintiel	Ongewogen	Model0	Model0 + Inkomen5	Model0 + Inkomen5 + Vermogen5
1	22,91 (1,21)	23,89 (1,26)	23,75 (1,24)	23,92 (1,25)
2	28,84 (1,12)	28,75 (1,12)	28,78 (1,11)	28,77 (1,11)
3	19,37 (0,86)	19,35 (0,87)	19,39 (0,87)	19,47 (0,87)
4	16,19 (0,77)	15,81 (0,76)	15,87 (0,77)	15,85 (0,77)
5	13,11 (0,70)	13,14 (0,71)	13,17 (0,71)	13,21 (0,71)

Bron: CBS.

<sup>1)</sup> Ongewogen schattingen en schattingen gebaseerd op de gewichten onder de drie weegmodellen (in %). Tussen haakjes staan geschatte standaardfouten.

## 2.2.8 Schattingen voor ErvarenGezondheid:Nee per vermogenskwintiel<sup>1)</sup>

Vermogenskwintiel	Ongewogen	Model0	Model0 + Inkomen5	Model0 + Inkomen5 + Vermogen5
1	12,55 (0,73)	13,14 (0,77)	13,37 (0,78)	13,32 (0,78)
2	29,77 (1,30)	29,82 (1,31)	30,03 (1,31)	29,93 (1,30)
3	19,91 (0,94)	19,77 (0,94)	19,94 (0,94)	19,82 (0,94)
4	18,93 (0,85)	18,48 (0,83)	18,71 (0,84)	18,66 (0,84)
5	19,04 (0,82)	18,82 (0,82)	18,97 (0,82)	18,95 (0,82)

Bron: CBS.

<sup>1)</sup> Ongewogen schattingen en schattingen gebaseerd op de gewichten onder de drie weegmodellen (in %). Tussen haakjes staan geschatte standaardfouten.

## 2.2.9 Gewogen aandelen CAWI en CAPI (in %)

Mode	Ongewogen	Model0	Model0 + Inkomen5	Model0 + Inkomen5 + Vermogen5
CAWI	51,8	51,0	50,7	50,6
CAPI	48,2	49,0	49,3	49,4

Bron: CBS.

# 3. Conclusie en aanbeveling

Omdat de Gezondheidsenquête 2014 een nieuw ontwerp heeft gekregen is onderzocht hoe de weging hierop aangepast en eventueel verbeterd kan worden. Meer specifiek is onderzocht of de weging verbetert door uitbreiding met de variabelen inkomen en vermogen in kwintielen. Zoals bekend van andere onderzoeken zijn de lage inkomensgroepen minder goed vertegenwoordigd in de respons. Ook naar vermogen is de respons scheef. Het bestaande GE weegmodel blijkt maar ten dele in staat de scheefheid naar deze variabelen te verminderen. Daarnaast vertonen een aantal belangrijke GE doelvariabelen een duidelijke samenhang met inkomen en vermogen. Omdat inkomen en vermogen voldoende verschillend zijn, en beide gebruikt worden om naar te publiceren stellen we voor het oorspronkelijke weegmodel uit te breiden met deze twee weegvariabelen. Door het toevoegen van deze weegvariabelen neemt de spreiding van de gewichten maar heel weinig toe, zodat het variantieverhogend effect daarvan klein is. De vermoedelijke reductie van vertekening weegt daar meer dan tegenop. Het voorgestelde weegmodel wordt daarmee

Geslacht2 x Leeftijd16 + BurgerlijkeStaat4 +  
Stedelijkheidsgraad5 + ProvinciePlus16 + Huishoudgrootte5 +  
Geslacht2 x Leeftijd3 x BurgerlijkeStaat2 +  
Landsdeel4 x Leeftijd3 + HerkomstGeneratie4 + Seizoen4 +  
Inkomen5 + Vermogen5.

## 4. Naschrift aanpassing weging vanaf statistiekjaar 2018

Vanaf 2018 wordt niet meer de volledige non-respons op de CAWI-benadering herbenaderd. Dat wil zeggen dat er een gestratificeerde steekproef wordt getrokken uit de CAWI nonrespondenten naar bekende achtergrondkenmerken, en alleen de getrokken eenheden worden via CAPI herbenaderd. De trekkingskansen in deze zogenaamde doelgroepenbenadering zijn zodanig gekozen dat de uiteindelijke respons naar verwachting een betere afspiegeling van de gehele bevolking is. Doelgroepen die relatief al goed responderen via CAWI worden dus met kleinere kans geselecteerd voor herbenadering. De doelgroepen zijn bepaald op basis van achtergrondkenmerken leeftijd, stedelijkheidsgraad, inkomen en migratieachtergrond, zie van den Broek et al. (2018).

De trekkingskansen voor de doelgroepenbenadering worden vooraf zo goed mogelijk geschat op basis van responsgegevens van CAPI herbenaderingen uit het verleden. De gerealiseerde respons kan qua samenstelling naar doelgroep echter altijd afwijken van de verwachtingen. Om er zeker van te zijn dat de doelgroepenbenadering geen vertekening introduceert in de cijfers, worden de doelgroepkenmerken ook in de weging meegenomen. Het aangepaste weegmodel wordt daarmee

$$\begin{aligned} & \text{Geslacht2} \times \text{Leeftijd16} + \text{BurgerlijkeStaat4} + \\ & \text{Stedelijkheidsgraad5} + \text{ProvinciePlus16} + \text{Huishoudgrootte5} + \\ & \text{Geslacht2} \times \text{Leeftijd3} \times \text{BurgerlijkeStaat2} + \\ & \text{Landsdeel4} \times \text{Leeftijd3} + \text{HerkomstGeneratie4} + \text{Seizoen4} + \\ & \text{Inkomen5} + \text{Vermogen5} + \text{Doelgroep9}. \end{aligned}$$

Hierin is Doelgroep9 de indeling naar de 9 doelgroepen gedefinieerd op basis van leeftijd, stedelijkheidsgraad, inkomen en migratieachtergrond, zie Bijlage A voor de precieze indeling. Deze kenmerken apart komen al voor in het bestaande weegmodel. Bij de keuze voor doelgroepen is rekening gehouden met de indeling van de weegtermen in het bestaande weegmodel, zodat er geen kleine restgroepen ontstaan die de weging eventueel zouden kunnen verstoren.

# Referenties

Boonstra, H. J. (2014). *Weging Gezondheidsenquête vanaf 2014*. Interne memo, februari 2014.

van den Broek, B., van der Doef, S., en Paulissen, R. (2018). *Steekproefontwerp Gezondheidsenquête 2018*. Interne CBS-nota, januari 2018.

Bruggink, J.-W., Smeets, M., en Tennekes, M. (2013). *Kleinedomeinschatters bij de Gezondheidsenquête*. BPA nr. PPM-2013-01-29-JBRK.

Bruggink, J.-W. en van Herten, M. (2012). *Weging Gezondheidsenquête, vanaf 2010*. Interne memo, mei 2012.

Buelens, B., Meijers, R., en Tennekes, M. (2013). *Weging Gezondheidsmonitor 2012*. BPA nr. PPM-2013-06-14-BBUS.

Gabler, S., Haeder, S., en Lahiri, P. (1999). *A model-based justification of Kish's formula for design effects for weighting and clustering*. *Survey Methodology* 25 (1).

Goris, G. (2013). *Opzet Design GEZO 2014*.

Kish, L. (1987). *Weighting in Deft2*. *The Survey Statistician* June 1987.

Nieuwenbroek, N. en Boonstra, H. J. (2002). *Bascula 4.0 Reference Manual*. BPA nr. 279-02-TMO.

Smeets, M. (2011). *Harmonisatie Weegmodellen bij SAH-onderzoeken*. BPA nr. DMH-2011-06-16-MSET.

# A Categorie-indeling van de weegvariabelen

## Geslacht2

- Man
- Vrouw

## Leeftijd3

- 0-34 jaar
- 35-54 jaar
- 55 jaar en ouder

## Leeftijd16

- 0-3 jaar
- 4-11 jaar
- 12-15 jaar
- 16-19 jaar
- 20-24 jaar
- 25-29 jaar
- 30-34 jaar
- 35-39 jaar
- 40-44 jaar
- 45-49 jaar
- 50-54 jaar
- 55-59 jaar
- 60-64 jaar
- 65-69 jaar
- 70-74 jaar
- 75 jaar en ouder

## BurgerlijkeStaat2

- Gehuwd
- Niet-gehuwd

## BurgerlijkeStaat4

- Gehuwd
- Gescheiden
- Verweduwd
- Nooit gehuwd geweest

## Landsdeel4

- Noord
- Oost
- West
- Zuid

#### ProvinciePlus16

- Groningen
- Friesland
- Drenthe
- Overijssel
- Gelderland
- Utrecht (exclusief gemeente Utrecht)
- Noord-Holland (exclusief gemeente Amsterdam)
- Zuid-Holland (exclusief gemeenten Den Haag en Rotterdam)
- Zeeland
- Noord-Brabant
- Limburg
- Flevoland
- Gemeente Amsterdam
- Gemeente Rotterdam
- Gemeente Den Haag
- Gemeente Utrecht

#### Stedelijkheidsgraad5

- Zeer sterk stedelijk
- Sterk stedelijk
- Matig stedelijk
- Weinig stedelijk
- Niet stedelijk

#### Huishoudgrootte5

- 1 persoon
- 2 personen
- 3 personen
- 4 personen
- 5 of meer personen

#### Seizoen4

- Winter (december, januari, februari)
- Lente (maart, april, mei)
- Zomer (juni, juli, augustus)
- Herfst (september, oktober, november)

#### HerkomstGeneratie4

- Autochtoon
- Westers allochtoon
- Niet-westers allochtoon, eerste generatie
- Niet-westers allochtoon, tweede generatie

#### Inkomen5 (gebaseerd op gestandaardiseerd huishoudinkomen) kwintielen afgeleid op huishoudniveau

- Populatiekwintiel 1
- Populatiekwintiel 2
- Populatiekwintiel 3

- Populatiekwintiel 4
- Populatiekwintiel 5

#### Vermogen5

- kwintielen afgeleid op huishoudniveau
- Populatiekwintiel 1
- Populatiekwintiel 2
- Populatiekwintiel 3
- Populatiekwintiel 4
- Populatiekwintiel 5

#### Doelgroep9 (indeling 2018)

- Westerse allochtonen en autochtonen, 12 t/m 24 jaar en 65 jaar of ouder, inkomen in Populatiekwintiel 2,3,4 of 5
- Westerse allochtonen en autochtonen, 12 t/m 24 jaar en 65 jaar of ouder, inkomen in Populatiekwintiel 1
- Westerse allochtonen en autochtonen, 25 t/m 64 jaar, inkomen in Populatiekwintiel 2,3,4 of 5
- Westerse allochtonen en autochtonen, 25 t/m 64 jaar, inkomen in Populatiekwintiel 1
- Westerse allochtonen en autochtonen, 0 t/m 11 jaar, niet zeer sterk stedelijke gemeente
- Westerse allochtonen en autochtonen, 0 t/m 11 jaar, zeer sterk stedelijke gemeente
- Niet-westerse allochtonen, 25 jaar of ouder
- Niet-westerse allochtonen, 12 t/m 24 jaar
- Niet-westerse allochtonen, 0 t/m 11 jaar



## Verklaring van tekens

Niets (blanco)	Een cijfer kan op logische gronden niet voorkomen
.	Het cijfer is onbekend, onvoldoende betrouwbaar of geheim
*	Voorlopige cijfers
**	Nader voorlopige cijfers
2018-2019	2018 tot en met 2019
2018/2019	Het gemiddelde over de jaren 2018 tot en met 2019
2018/'19	Oogstjaar, boekjaar, schooljaar enz., beginnend in 2018 en eindigend in 2019
2016/'17-2018/'19	Oogstjaar, boekjaar, enz., 2016/'17 tot en met 2018/'19

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

### *Prepress*

Centraal Bureau voor de Statistiek

### *Ontwerp*

Edenspiekermann

### *Inlichtingen*

Tel. 088 570 70 70  
Via contactformulier: [www.cbs.nl/infoservice](http://www.cbs.nl/infoservice)

© Centraal Bureau voor de Statistiek, Den Haag/Heerlen/Bonaire, 2019.  
Verveelvoudigen is toegestaan, mits het CBS als bron wordt vermeld.