



Discussion paper

Aanvullende analyses variabelen omtrent bewegen uit de Gezondheidsenquête

Yvonne Gootzen en
Jan van den Brakel

Augustus 2021

Inhoudsopgave

1. Inleiding 4
2. Methoden 5
3. Resultaten 7
4. Discussie 13

Referenties 14

Samenvatting

Bij de publicatie van de jaarcijfers van de Gezondheidsenquête over 2020 met betrekking tot bewegen en sporten zijn vragen gerezen omtrent het effect van corona(maatregelen) op de betrouwbaarheid van deze uitkomsten. In deze nota worden de resultaten beschreven van enkele aanvullende analyses die zijn uitgevoerd om de betrouwbaarheid van deze uitkomsten te evalueren.

Keywords

Coronacrisis, structureel tijdreeksmodel, sporten en bewegen

1. Inleiding

De coronacrisis resulteerde in 2020 naar een vraag om snellere cijfers omtrent het effect van corona(maatregelen) op cijfers over de gezondheid van de Nederlandse bevolking. Daarnaast moest een oplossing worden gevonden om te corrigeren voor het wegvallen van de face-to-facerespons (CAPI) gedurende de lockdown. In 2020 is om deze redenen een schattingsmethodiek ontwikkeld om kwartaalcijfers te maken voor acht belangrijke variabelen uit de Gezondheidsenquête (GE) waarbij tegelijkertijd gecorrigeerd wordt voor veranderingen in meetfouten door het ontbreken van CAPI gedurende de lockdown.

De kwartaalschattingen voor deze acht variabelen zijn opgenomen in het weegmodel van het jaarbestand. Hiermee wordt consistentie tussen jaarcijfers en eerder gepubliceerde kwartaalcijfers gewaarborgd. Tevens worden gerelateerde variabelen waarvoor geen tijdreeksmodellen zijn ontwikkeld zo goed mogelijk gecorrigeerd voor het wegvallen van CAPI gedurende de lockdown. De methodiek is beschreven in Smeets en Van den Brakel (2021).

Mede naar aanleiding van vragen vanuit het RIVM is er twijfel ontstaan of de hierboven beschreven weging afdoende is voor de variabele omtrent het wel of niet voldoen aan de richtlijn voor bewegen en de variabele over wekelijks sporten. Deze variabelen behoren niet tot de acht waarover kwartaalcijfers zijn gepubliceerd en zijn niet meegenomen in het uitgebreide weegmodel. Om deze reden zijn in deze notitie aanvullende analyses uitgevoerd. Er zijn daarbij afzonderlijke tijdreeksmodellen ontwikkeld voor de variabelen “Bewegen volgens de richtlijn” en “Wekelijks sporten”.

Deze nota is als volgt opgebouwd. In paragraaf 2 wordt de methodiek om kwartaalcijfers te schatten die corrigeren voor het wegvallen van CAPI door middel van een structureel tijdreeksmodel kort samengevat. In paragraaf 3 worden de resultaten op basis van de extra analyses voor “Bewegen volgens de richtlijn” en “Wekelijks sporten” beschreven. De nota wordt afgerond met een discussie in paragraaf 4.

2. Methoden

De gezondheidsenquête (GE) is een doorlopend steekproefonderzoek dat door het CBS wordt uitgevoerd met als doel de gezondheid, de leefstijl en het zorggebruik van de Nederlandse bevolking te beschrijven. Doordat de GE herhaaldelijk wordt uitgevoerd ontstaan er tijdreeksen die de ontwikkeling van de variabelen met betrekking tot gezondheid, leefstijl en zorggebruik door de tijd heen zo goed mogelijk beschrijven. De dataverzameling is gebaseerd op een sequentieel mixed-mode design waarbij gebruik wordt gemaakt van internetwaarneming (CAWI) en face-to-facewaarneming (CAPI). Tot en met 2019 zijn op basis van het onderzoek alleen jaarcijfers over de gezondheid, de leefstijl en het zorggebruik van de Nederlandse bevolking gepubliceerd.

Door het tijdelijk wegvallen van de CAPI-waarneming als gevolg van de coronamaatregelen in 2020 ontstaan er breuken in de uitkomsten van de gezondheidsenquête (GE). Via een tijdreeksmethode wordt zo goed mogelijk gecorrigeerd voor de verandering in meet- en selectiefouten t.g.v. het wegvallen van CAPI. Daarnaast wordt de tijdreeksmethode gebruikt om verbanden tussen de COVID19-pandemie en de uitkomsten van de GE zo goed mogelijk in kaart te brengen en snellere cijfers over het effect van corona op gezondheidsindicatoren te publiceren.

Het netto-effect van het wegvallen van CAPI is berekend op basis van responsbestanden van de voorgaande jaren. Dit is gedaan door de CAPI-respondenten uit de respons te verwijderen en de resterende responsen opnieuw te wegen. Daarmee worden twee schattingen voor een doelvariabele verkregen: één gebaseerd op de volledige respons (CAPI plus CAWI) en één gebaseerd op alleen internetrespons (CAWI). Voor de GE kunnen op deze manier tijdreeksen gemaakt worden vanaf het eerste kwartaal in 2014: de reguliere reeks gebaseerd op de volledige respons en de internetreeks gebaseerd op de internetrespons. De internetreeks is beschikbaar voor alle kwartalen, ook gedurende de lockdown. Voor de reguliere reeks ontbreken de schattingen voor het tweede kwartaal 2020. Via een bivariaat structureel tijdreeksmodel wordt een nowcast verkregen voor het ontbrekende cijfer in het tweede kwartaal 2020 op basis van de internetreeks.

Het bivariate structurele tijdreeksmodel wordt gegeven door

$$\begin{pmatrix} \hat{y}_t^R \\ \hat{y}_t^I \end{pmatrix} = \begin{pmatrix} \vartheta_t \\ \vartheta_t \end{pmatrix} + \begin{pmatrix} \lambda_t \\ \lambda_t \end{pmatrix} + \begin{pmatrix} e_t^R \\ e_t^I \end{pmatrix}, \text{ met} \quad (1)$$

$\vartheta_t = L_t + S_t$ en kwartaal t .

Hierin is

- $\hat{y}_t^R$: de directe schatting in kwartaal t gebaseerd op de volledige respons (reguliere reeks).
- $\hat{y}_t^I$: de directe schatting in kwartaal t gebaseerd op de internetrespons (internetreeks).
- ϑ_t : de onbekende populatieparameter in kwartaal t .
- L_t : de trend, gemodelleerd als smooth-trendmodel.
- S_t : de seizoenscomponent met kwartaal als periode, gemodelleerd met een trigonometrisch seizoen model.

- λ_t : het systematisch verschil tussen de internetreeks en de reguliere reeks, gemodelleerd als random walk.
- e_t^j : de meetfout van $\hat{y}_t^j$ voor $j \in \{R, I\}$, gemodelleerd als $e_t^j = \sqrt{\hat{V}(\hat{y}_t^j)} \tilde{e}_t^j$ waarbij $\hat{V}(\hat{y}_t^j)$ de geschatte variantie van $\hat{y}_t^j$ is en $\tilde{e}_t^j$ witte ruis.

De volledige uitwerking van dit model is beschreven in Smeets en Van den Brakel (2021). Voor 2020 zijn aan de hand van dit tijdreeksmodel kwartaalcijfers gemaakt voor de volgende variabelen:

- Ervaren gezondheid (0 jaar en ouder)
- Psychisch ongezond (12 jaar en ouder)
- Huisartscontact in de afgelopen vier weken (0 jaar en ouder)
- Dagelijks roken (vanaf 18 jaar)
- Overgewicht (vanaf 18 jaar)
- Overmatig alcoholgebruik (vanaf 18 jaar)
- Tandartsbezoek in de afgelopen vier weken (0 jaar en ouder)
- Specialistcontact in de afgelopen vier weken (0 jaar en ouder)

De keuze voor deze variabelen is gemaakt in overleg met enkele externe partijen, te weten VWS, RIVM en SCP. Van de geselecteerde variabelen was er de wens om tijdens de coronacrisis te kunnen monitoren of er op kwartaalbasis veranderingen zouden optreden in de uitkomsten. De gekozen variabelen bestrijken de drie hoofdthema's van de gezondheidsenquête: gezondheid (ervaren en psychisch), zorggebruik (huisarts, tandarts, specialist) en leefstijl (roken, alcohol, overgewicht).

De correcties voor het wegvallen van CAPI zijn in de jaarcijfers over 2020 verwerkt door in de weging van het jaarbestand een extra tabel op te nemen met de gecorrigeerde kwartaalcijfers over 2020 van de acht onderzochte doelvariabelen die geschat zijn via het tijdreeksmodel. De reguliere weging van het jaarbestand is beschreven in Boonstra (2019). Door de tabel met kwartaalcijfers op te nemen in de weging worden de jaarcijfers consistent gemaakt met de eerder gepubliceerde kwartaalcijfers en worden de correcties voor het deels wegvallen van CAPI zo goed mogelijk verwerkt in de jaarcijfers.

Jaarcijfers omtrent bewegen en sporten worden door het RIVM berekend op basis van het jaarbestand van de Gezondheidsenquête. Dit kon pas gebeuren in de loop van 2021 zodat het niet mogelijk was om gedurende 2020 ook voor deze variabelen tijdreeksmodellen te ontwikkelen. Bij de beschouwing van de jaarcijfers over 2020 is twijfel ontstaan over de betrouwbaarheid van deze uitkomsten. Ook de kwestie van vraagstellingen, die niet erg geschikt lijken voor coronaomstandigheden gaf reden om te twijfelen aan de uitkomsten. Een belangrijke vraag die rees was of het wegvallen van de CAPI-respons in het tweede kwartaal van 2020 resulteert in een overschatting van het percentage mensen dat voldoet aan de beweegrichtlijn. Om dit nader te onderzoeken is het hierboven beschreven tijdreeksmodel gebruikt om kwartaalcijfers te schatten voor de variabelen "Bewegen volgens de richtlijn" en "Wekelijks sporten". De resultaten zijn beschreven in paragraaf 3.

3. Resultaten

Tabel 1 geeft voor de variabele “Bewegen volgens de richtlijn” de kwartaalschattingen voor 2020 weer. In kolom 2 en 3 staan de schattingen op basis van de volledige respons en alleen de webresponse (CAWI) op basis van het oorspronkelijk weegmodel zonder de extra kwartaalcijfers verkregen via het tijdreeksmodel. Samen met de voorgaande jaren zijn dit in feite de input reeksen voor het tijdreeksmodel. In de vierde kolom staan de kwartaalcijfers en het jaarcijfer geschat via het weegmodel voor 2020, dat wil zeggen de weging uitgebreid met de kwartaalcijfers voor de acht variabelen waarvoor tijdreeksen in 2020 zijn ontwikkeld. In de zevende kolom staan de schattingen op basis van het tijdreeksmodel ontwikkeld voor “Bewegen volgens de richtlijn”. In Tabel 2 staan dezelfde cijfers voor de variabele “Wekelijks sporten”.

Tabel 1 laat zien dat de schatting voor het tweede kwartaal voor “Bewegen volgens de richtlijn” op basis van de uiteindelijke weging volgens het uitgebreide weegmodel circa 9% punten hoger ligt ten opzichte van de andere kwartalen. Het verschil tussen het cijfer voor het tweede kwartaal op basis van het uitgebreide weegmodel en het cijfer op basis van het tijdreeksmodel is 7,3% punten. Het jaarcijfer wordt door het tijdreeksmodel met 2,9% punten naar beneden bijgesteld. Een vergelijkbaar maar minder extreem effect is te zien voor de variabele “Wekelijks sporten” in Tabel 2. Het tijdreeksmodel stelt de schatting voor het tweede kwartaal bij met 1.6% punten en het jaarcijfer met 1,7% punten.

Tabel 1: Analyse resultaten “Bewegen volgens de richtlijn”

Kwartaal	Oorspronkelijke weging		Weging volgens uitgebreid weegmodel 2020			Tijdreeksmodel		
	Volledige respons	Web respons	Schatting	SF	95% BI	Schatting	SF	95% BI
2020(1)	50,1	58,6	49,2	1,10	47,0 - 51,4	48,4	1,23	46,0 – 50,8
2020(2)	-	58,9	59,1	1,19	56,8 – 61,4	51,8	1,19	49,5 – 54,1
2020(3)	46,6	55,5	51,5	1,34	48,9 – 54,1	50,5	1,13	48,3 – 52,7
2020(4)	50,7	56,1	51,1	1,29	48,6 – 53,6	48,5	1,24	46,1 – 50,9
Jaarcijfer			52,7	0,61	51,5 – 53,9	49,8	0,61	48,6 – 51,0

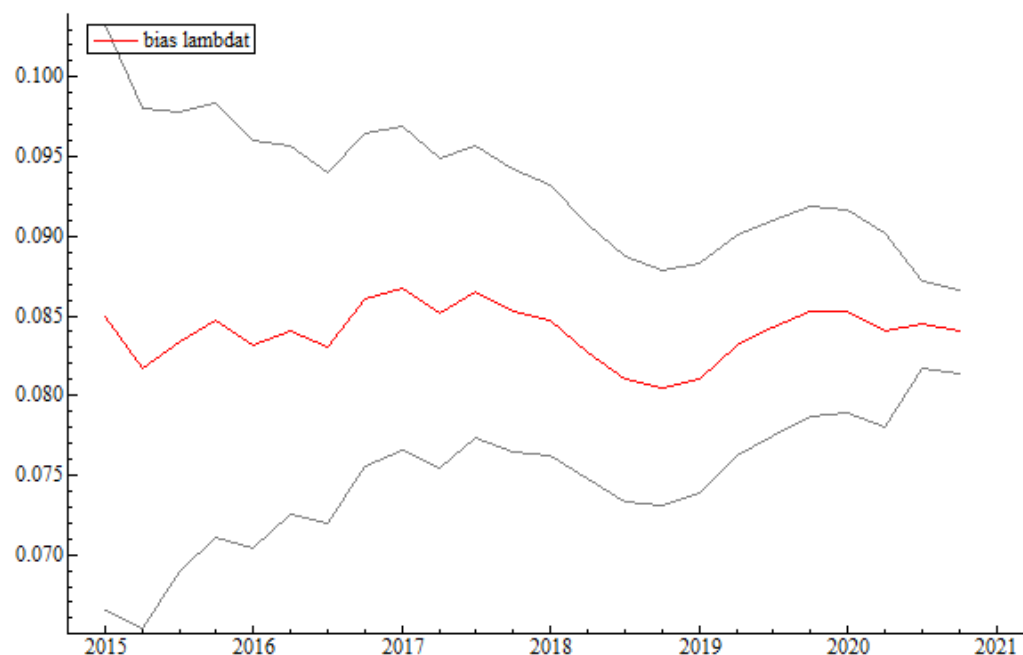
SF: Standaardfout, BI: Betrouwbaarheidsinterval

Tabel 2: Analyse resultaten “Wekelijks sporten”

Kwartaal	Oorspronkelijke weging		Weging volgens uitgebreid weegmodel 2020			Tijdreeksmodel		
	Volledige respons	Web respons	Schatting	SF	95% BI	Schatting	SF	95% BI
2020(1)	54,1	57,9	54,6	1,11	52,4 – 56,8	54,0	0,88	52,3 – 55,7
2020(2)	-	55,4	55,4	1,21	53,0 – 57,8	53,8	0,89	52,1 – 55,5
2020(3)	52,3	57,2	54,1	1,35	51,5 – 56,7	51,0	0,34	50,3 – 51,7
2020(4)	53,6	55,4	54,9	1,30	52,4 – 57,4	53,4	1,07	51,3 – 55,5
Jaarcijfer			54,8	0,62	53,6 – 56,0	53,1	0,62	51,9 – 54,3

SF: Standaardfout, BI: Betrouwbaarheidsinterval

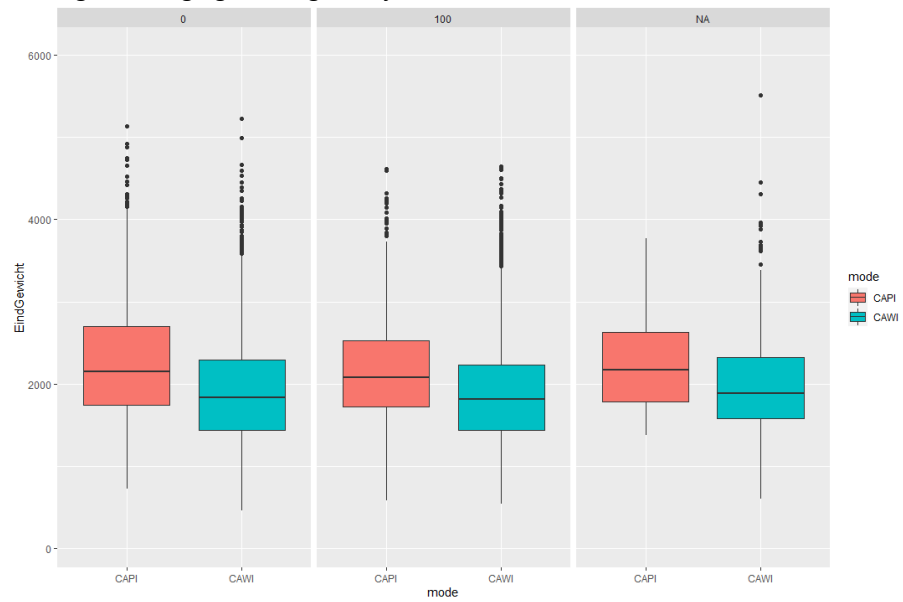
De vraag is of de kwartaalschattingen op basis van het tijdreeksmodel voor de variabele “Bewegen volgens de richtlijn” in de weging voor het jaarbestand van 2020 had moeten worden opgenomen. Daarmee zou het gepubliceerde cijfer op landelijk niveau met 2,9% dalen. Het is niet zeker of dit een verbetering is omdat de onderliggende vraagstelling gedurende de coronacrisis verwarrend is. De vraagstelling luidt namelijk “hoeveel beweegt u in een normale week?” Het is onduidelijk hoe deze vraag door respondenten in de lockdown wordt geïnterpreteerd. Sommige respondenten zullen refereren naar een week voor de lockdown. Andere zullen refereren naar een normale week in de lockdown periode. Daarbij komt dat het afleiden en berekenen van de sport- en beweegvariabelen wordt uitgevoerd door het RIVM na afloop van het meetjaar. Om deze reden was het bij de start van het project om kwartaalcijfers te maken niet mogelijk om deze variabele op te nemen in het uitgebreide weegmodel. Het tijdreeksmodel laat over de periode 2015 tot eind 2020 wel een redelijk stabiel verschil van circa 8% punten tussen de volledige reeks en de reeks op basis van alleen CAWI-responsen zien, zoals blijkt uit Figuur 1.



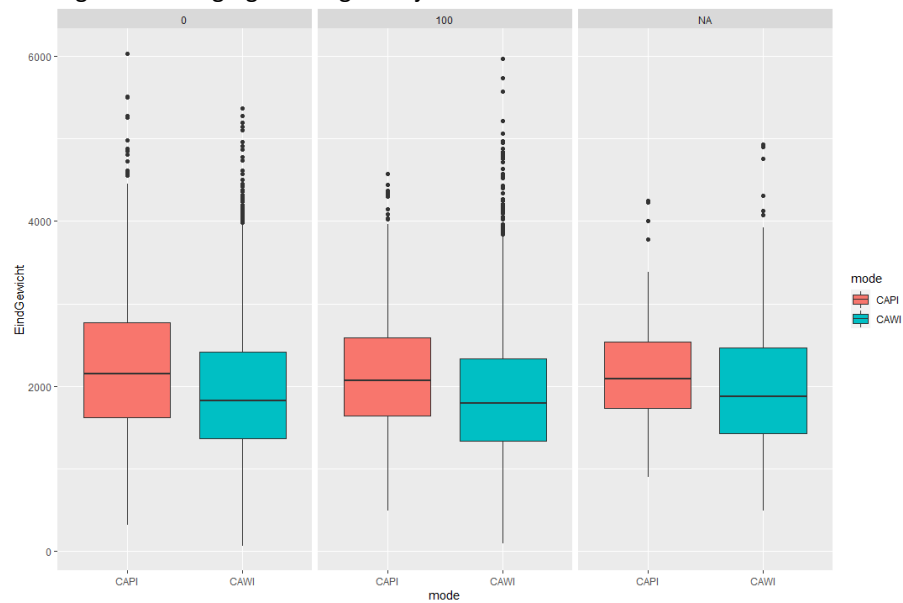
Figuur 1: Verschil tussen kwartaalcijfers gebaseerd op volledige respons en CAWI respons voor “Bewegen volgens de richtlijn”.

In Figuren 2 en 3 worden de gewichten voor de uitkomstcategorieën van “Bewegen volgens de richtlijn” vergeleken voor de CAPI- en CAWI-respondenten. In de bovenste plot (Figuur 2a en 3a) zijn de gewichten van het oorspronkelijk weegmodel gebruikt (dus zonder de kwartaalcijfers van de acht variabelen waarvoor in 2020 tijdreeksmodellen zijn ontwikkeld). In de onderste plot (Figuur 2b en 3b) zijn de gewichten van het uitgebreide weegmodel gebruikt (dus met de kwartaalcijfers van de acht variabelen waarvoor in 2020 tijdreeksmodellen zijn ontwikkeld). Wat opvalt is dat de gewichten voor CAPI gemiddeld groter zijn dan voor CAWI. Het effect van het opnemen van de kwartaalcijfers van de acht variabelen waarvoor tijdreeksmodellen zijn ontwikkeld, is dat de spreiding van de gewichten toeneemt. Er is geen duidelijke verschuiving van het gemiddelde van de gewichten te zien. Een dergelijk beeld wordt bevestigd in de Tabellen 3 tot en met 8. Tabellen 3 en 6 geven de responsaantallen voor CAPI en CAWI naar de antwoordcategorieën van respectievelijk “Bewegen volgens de richtlijn” en “Wekelijks sporten”. Tabellen 4 en 5 geven het gemiddeld gewicht voor CAPI en CAWI naar de antwoordcategorieën van respectievelijk “Bewegen volgens de richtlijn” voor respectievelijk het oorspronkelijk weegmodel en het uitgebreide weegmodel. Tabellen 7 en 8 geven het gemiddeld gewicht voor CAPI en CAWI naar de antwoordcategorieën van respectievelijk “Wekelijks sporten” voor respectievelijk het oorspronkelijk weegmodel en het uitgebreide weegmodel.

2a Originele weging Beweegrichtlijn

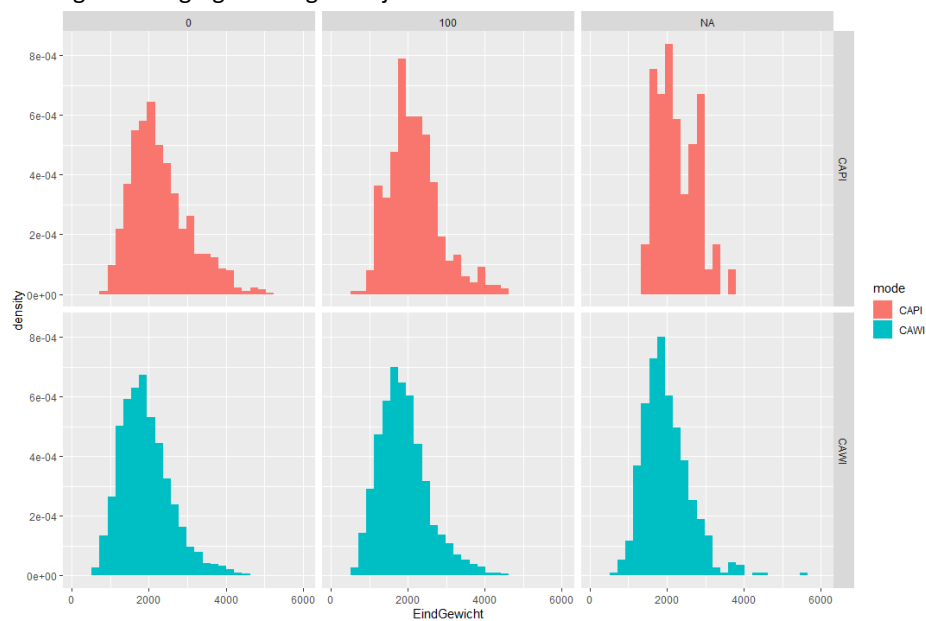


2b Uitgebreide weging Beweegrichtlijn

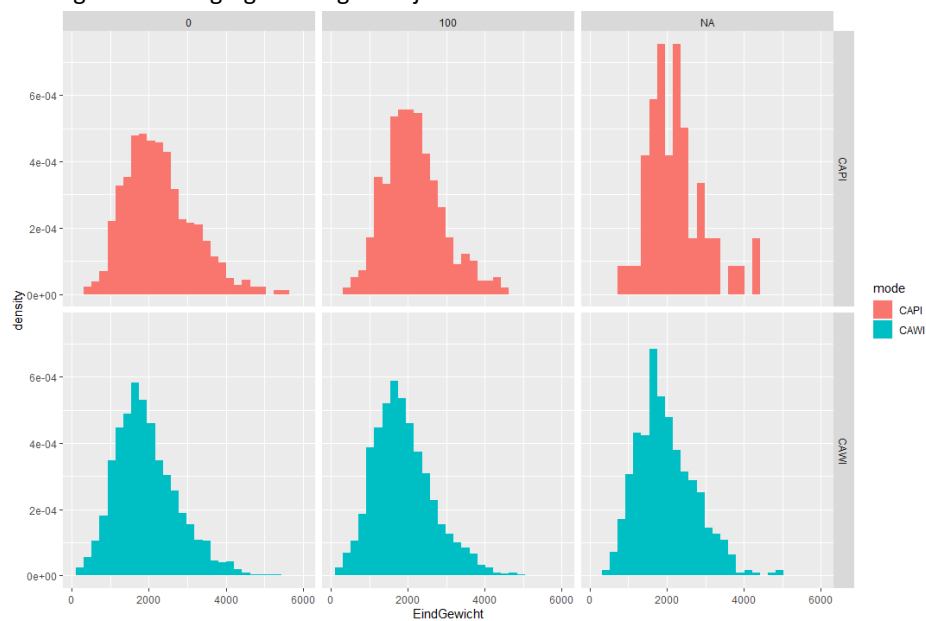


Figuur 2: Box-Whisker plots eindgewichten voor de twee categorieën van Bewegen volgens de richtlijn (0: onder de richtlijn, 100: boven de richtlijn, NA: missing) voor de CAPI en CAWI respons. Bovenste plot: gewichten op basis van de originele weging, onderste plot gewichten op basis van de uitgebreide weging.

3a Originele weging Beweegrichtlijn



3b Uitgebreide weging Beweegrichtlijn



Figuur 3: Histogrammen eindgewichten voor de twee categorieën van Bewegen volgens de richtlijn (0: onder de richtlijn, 100: boven de richtlijn, NA: missing) voor de CAPI en WI respons. Bovenste plot: gewichten op basis van de originele weging, onderste plot gewichten op basis van de uitgebreide weging.

	CAPI	CAWI
BeweegRichtlijn 0	906	2850
BeweegRichtlijn 100	481	3902

Tabel 3: Responsaantallen voor de variabele BeweegRichtlijn (0: onder de richtlijn, 100: boven de richtlijn)

Originele weging	CAPI	CAWI
BeweegRichtlijn 0	2290,6	1920,6
BeweegRichtlijn 100	2172,0	1893,6

Tabel 4: Gemiddelde gewichten in de originele weging voor de variabele Beweegrichtlijn (0: onder de richtlijn, 100: boven de richtlijn)

Uitgebreide weging	CAPI	CAWI
BeweegRichtlijn 0	2266,8	1037,6
BeweegRichtlijn 100	2149,9	1890,2

Tabel 5: Gemiddelde gewichten in de uitgebreide weging voor de variabele BeweegRichtlijn (0: onder de richtlijn, 100: boven de richtlijn)

	CAPI	CAWI
WekelijksSporter 0	730	2850
WekelijksSporter 100	657	3883

Tabel 6: Responsaantallen voor de variabele WekelijksSporter (0: niet wekelijks sporter, 100: wekelijks sporter)

Originele weging	CAPI	CAWI
WekelijksSporter 0	2305,1	1021,2
WekelijksSporter 100	2187,6	1893,0

Tabel 7: Gemiddelde gewichten in de originele weging voor de variabele WekelijksSporter (0: onder de richtlijn, 100: boven de richtlijn)

Uitgebreide weging	CAPI	CAWI
WekelijksSporter 0	2302,6	1943,2
WekelijksSporter 100	2141,5	1885,9

Tabel 8: Gemiddelde gewichten in de uitgebreide weging voor de variabele WekelijksSporter (0: wekelijks sporter, 100: niet wekelijks sporter)

4. Discussie

Er zijn twee redenen om de weging niet uit te breiden. Zoals hiervoor aangegeven kan de vraagstelling leiden tot verwarring en extra meetfouten in de antwoorden van respondenten. Dat blijft een doorslaggevende reden om voor deze variabelen geen correctie door te voeren. Daarbij komt dat het toevoegen van variabelen aan het weegmodel meer restricties oplegt waaraan voldaan moet worden bij het zoeken van een oplossing voor het weegmodel. Dit leidt tot een toename in de spreiding van de gewichten. Hoe meer variabelen worden toegevoegd aan het weegmodel, hoe eerder er negatieve gewichten ontstaan. Uiteindelijk kan het zelfs gebeuren dat er geen oplossing voor het weegmodel kan worden gevonden. Naast deze inhoudelijke argumenten is het nadeel van het toevoegen van deze variabelen aan de weging dat de publicatie van alle op de Gezondheidsenquête 2020 gebaseerde cijfers moet worden herzien. De vraag is of een aanpassing van 2,9% punten op het landelijkcijfer over het voldoen aan de beweegnorm deze extra kosten rechtvaardigt.

Referenties

Boonstra, H.J. (2019). Weging Gezondheidsenquête 2014. Discussion paper CBS, Centraal Bureau voor de Statistiek, Heerlen. URL: <https://www.cbs.nl/nl-nl/achtergrond/2019/18/weging-gezondheidsenquête-2014>

Smeets, M.J.E. en J.A. van den Brakel (2021). Toelichting berekening kwartaal- en jaarcijfers Gezondheidsenquête 2020, Interne CBS-nota, 2 februari 2021, Centraal Bureau voor de Statistiek, Heerlen.

Verklaring van tekens

Niets (blanco)	Een cijfer kan op logische gronden niet voorkomen
.	Het cijfer is onbekend, onvoldoende betrouwbaar of geheim
*	Voorlopig cijfer
**	Nader voorlopige cijfer
2017–2018	2017 tot en met 2018
2017/2018	Het gemiddelde over de jaren 2017 tot en met 2018
2017/'18	Oogstjaar, boekjaar, schooljaar enz., beginnend in 2017 en eindigend in 2018
2013/'14–2017/'18	Oogstjaar, boekjaar, enz., 2015/'16 tot en met 2017/'18

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

Prepress

CCN Creatie en visualisatie

Ontwerp

Edenspiekermann

Inlichtingen

Tel. 088 570 70 70, fax 070 337 59 94
Via contactformulier: www.cbs.nl/infoservice

© Centraal Bureau voor de Statistiek, Den Haag/Heerlen/Bonaire, 2018.
Verveelvoudigen is toegestaan, mits het CBS als bron wordt vermeld.