

Survey integratie migranten: steekproeftrekking en weging

Toelichting van het CBS

Bijlage A bij het rapport Survey integratie migranten 2020. Uitvoering van veldwerk ten tijde van corona

Norbert Mikolajewski
José Gouweleeuw

Het Centraal Bureau voor de Statistiek heeft voor de Survey integratie migranten de steekproef getrokken en de weging uitgevoerd. In deze notitie lichten CBS-auteurs Mikolajewski en Gouweleeuw de werkwijze toe.

	17 juni 2021
samenvatting	Voor het onderzoek Survey integratie migranten 2020 is een steekproef van 15.498 personen getrokken. Deze personen zijn via internet en face to face benaderd. Voor de 5323 mensen die hebben gerespondeerd, worden gewichten berekend met behulp van een calibratieschatter. Het weegmodel lijkt hierbij zoveel mogelijk op het model dat in de vorige editie (2015) is gebruikt. De respons wordt ook geanalyseerd met betrekking tot representativiteit, die wordt gemeten met R-indicatoren.
trefwoorden	R-indicatoren, steekproef, Survey integratie migranten, weging

Inhoud

1	Inleiding	4
2	Steekproeftrekking	5
3	Weging	9
4	R-Indicatoren	14
5	Aanvullende gegevens over het weegmodel	16

1 Inleiding

In 2020 heeft het Sociaal en Cultureel Planbureau (SCP) in samenwerking met het CBS het onderzoek Survey integratie migranten (SIM) gehouden. Dit onderzoek is eerder gehouden in 2015 en is een voortzetting van de onderzoeken Survey integratie minderheden uit 2010 en Survey integratie nieuwe groepen (SING) uit 2009. Het is een enquêteonderzoek dat wordt uitgevoerd onder zeven groepen personen met een migratieachtergrond:

- Caribische,
- Iraanse,
- Marokkaanse,
- Poolse (voor zover zij na of op 1 januari 2004 in Nederland zijn komen wonen),
- Somalische,
- Surinaamse,
- Turkse.

Daarnaast wordt er ook een groep van personen zonder een migratieachtergrond ondervraagd. Het onderzoek wordt alleen uitgevoerd onder personen die tijdens het veldwerk 15 jaar of ouder zijn.

In deze notitie worden de steekproeftrekking en de weging voor de verschillende groepen binnen de SIM besproken. Deze nota is als volgt opgebouwd: in hoofdstuk 2 wordt de steekproefselectie beschreven; hoofdstuk 3 gaat over het wegen van de respons; en hoofdstuk 4 gaat in op de representativiteit van de respons met behulp van de R-indicatoren. De bijlage die volgt bevat enkele van de grotere tabellen.

2 Steekproeftrekking

Om de steekproeven voor het onderzoek te kunnen trekken, is er voor iedere bevolkingsgroep een steekproefkader afgeleid met gegevens uit de Basisregistratie Personen (BRP). Het steekproefkader van een bepaalde bevolkingsgroep telt iedereen die

1. op 9 december 2019 ingeschreven staat in de BRP op een adres dat niet bij een instelling of tehuis hoort;
2. op 31 december 2019, een dag voor de geplande startdatum van het veldwerk, 15 jaar of ouder is; en
3. de juiste migratieachtergrond heeft.

Voor de leden van het steekproefkader van de personen met een Poolse migratieachtergrond geldt nog dat zij pas op 1 januari 2004 of later voor het eerst mogen zijn ingeschreven in de BRP. Voor de leden van het steekproefkader van de personen zonder migratieachtergrond geldt ook dat het zevende cijfer van hun A-nummer – het identificatienummer in de BRP – gelijk is aan 6. Dit kan worden gezien als een aselechte steekproef van 10% van alle personen zonder migratieachtergrond in Nederland. Door deze keuze overlapt dit steekproefkader niet met veel andere onderzoeken die het CBS in dezelfde periode heeft uitgevoerd.

Uit deze steekproefkaders zijn, net als voor het onderzoek Survey integratie migranten van 2015, geclusterde steekproeven getrokken. In de eerste trap van de trekking van zo'n geclusterde steekproef is een steekproef van gemeenten getrokken. De steekproef van gemeenten is gestratificeerd naar de grootte van de gemeente: één stratum van kleine gemeenten met minder dan 50.000 inwoners en één stratum van grote gemeenten met 50.000 inwoners of meer. Bij het tellen van de inwoners van een gemeente zijn alle inwoners in Nederland geteld, niet alleen de inwoners die lid zijn van het steekproefkader. Uit ieder stratum is een steekproef van gemeenten getrokken, waarbij de trekkans van een gemeente evenredig is met het aantal personen uit het steekproefkader dat ingeschreven staat in die gemeente. De eilanden Schiermonnikoog, Ameland, Terschelling, Vlieland en Texel werden uitgesloten uit het steekproefkader.

Vervolgens is in de tweede trap van de trekking van de geclusterde steekproef een steekproef van personen getrokken uit het deel van het steekproefkader dat ingeschreven staat in de gemeenten die in de eerste trap zijn getrokken. Uit iedere gemeente is een enkelvoudig aselechte steekproef getrokken, die bij alle steekproeven behalve de steekproef onder personen zonder migratieachtergrond vijf personen telt voor een kleine gemeente en tien personen voor een grote gemeente. Bij de steekproef onder personen zonder migratieachtergrond is er uit iedere gemeente, klein en groot, een steekproef van tien personen getrokken. Bij sommige migrantengroepen zijn kleine gemeenten waar minder dan vijf personen of grote gemeenten waar minder dan tien personen in het steekproefkader staan ingeschreven. Deze gemeenten worden samengevoegd met andere aangrenzende gemeenten binnen het stratum van kleine respectievelijk grote gemeenten, zodanig dat in de samengevoegde gemeenten wel minimaal vijf of tien personen van het steekproefkader staan ingeschreven.

Tabel 1 laat zien hoeveel personen de steekproeven in totaal moeten tellen. Om te bepalen hoeveel gemeenten er per stratum getrokken moeten worden, zijn deze aantallen evenredig met de grootte

van de steekproefkaders verdeeld over de strata met de kleine gemeenten en de strata met de grote gemeenten. Het aantal gemeenten dat uit een stratum getrokken moet worden is dan gelijk aan de verhouding van het aantal personen dat de steekproef in het stratum moet tellen en het aantal personen dat per gemeente getrokken wordt, afgerond op het meest nabije gehele getal.

Tabel 1

Aantal personen en aantal gemeenten in de steekproef

migratieachtergrond	# personen in steekproef			# gemeenten in steekproef	
	totaal	groot	klein	groot	klein
Caribische migratieachtergrond	2114	1760	354	58	65
Iraanse migratieachtergrond	2020	1610	410	60	78
Marokkaanse migratieachtergrond	2114	1814	300	56	53
Poolse migratieachtergrond	2020	1245	775	56	113
Somalische migratieachtergrond	1443	1067	376	58	72
Surinaamse migratieachtergrond	2114	1864	250	50	46
Turkse migratieachtergrond	2020	1679	341	55	59
zonder migratieachtergrond	1653	883	770	67	77

Dat de steekproeven per stratum niet altijd een veelvoud van vijf of tien personen tellen, komt omdat er uit sommige gemeenten niet precies vijf of tien personen getrokken worden, maar meer dan vijf of tien personen. Dit wordt gedaan om ervoor te zorgen dat iedere persoon in het steekproefkader dezelfde kans heeft om getrokken te worden. De kans dat een persoon i getrokken wordt in de steekproef is gelijk aan het product van de kans dat de gemeente $j(i)$ waarin persoon i staat ingeschreven in de eerste trap van de trekking getrokken wordt – $\pi_{j(i)}$ – en de kans dat persoon i in de tweede trap getrokken wordt als de gemeente in de eerste trap getrokken is – $\pi_{i|j(i)}$. De kans dat een gemeente in de eerste trap van de trekking getrokken wordt, is gelijk aan

$$\pi_{j(i)} = \frac{n_g \cdot N_{j(i)}}{\sum_{k \in G} N_k},$$

met G de gemeenten in het stratum, n_g het aantal gemeenten in het stratum dat in de steekproef getrokken moet worden en N_k het aantal personen in het steekproefkader dat ingeschreven staat in gemeente k . De kans dat persoon i in de tweede trap getrokken wordt als de gemeente in de eerste trap getrokken is, is gelijk aan

$$\pi_{i|j(i)} = \frac{m}{N_{j(i)}},$$

met m het aantal personen dat uit de gemeente getrokken moet worden. De kans dat persoon i in de steekproef getrokken wordt is dus gelijk aan

$$\pi_i = \pi_{j(i)} \cdot \pi_{i|j(i)} = \frac{n_g \cdot N_{j(i)}}{\sum_{k \in G} N_k} \cdot \frac{m}{N_{j(i)}} = \frac{n_g \cdot m}{\sum_{k \in G} N_k} = \frac{n}{N},$$

met n het aantal personen in het stratum dat getrokken moet worden in de steekproef en N het aantal personen dat het steekproefkader in het stratum telt.

De formule voor de kans dat een gemeente in de eerste trap van de trekking getrokken wordt, kan groter dan 1 zijn. Als dit het geval is, dan wordt de kans gelijkgesteld aan 1. Zou uit zo'n gemeente ook m personen getrokken worden, dan hebben deze personen een kans kleiner dan n/N om getrokken te worden in de steekproef. Om ervoor te zorgen dat deze personen toch een kans n/N hebben, worden uit zo'n gemeente $j(i)$ niet m personen getrokken, maar $N_{j(i)} \cdot n/N$.

Van de personen die getrokken zijn in de steekproef, zijn op 23 januari 2020 bij de Rijksdienst voor Identiteitsgegevens (RvIG) namen en adressen opgevraagd. Met de namen en adressen die RvIG heeft teruggestuurd, zijn de volgende extra bewerkingen uitgevoerd.

- Personen die overleden of geëmigreerd zijn, werden uit de steekproef verwijderd.
- Personen die bij de gemeente te kennen hebben gegeven dat hun adressen niet aan derden geleverd mogen worden, werden uit de steekproef verwijderd.
- Personen die ingeschreven staan op een adres van een instelling of een tehuis, werden uit de steekproef verwijderd.
- Als de steekproef meerdere personen telt die op hetzelfde adres wonen, dan werden op een na al die personen uit de steekproef verwijderd. Een uitzondering hierop zijn de personen met een Poolse migratieachtergrond. Voor deze doelgroep is het toegestaan dat meerdere personen die op hetzelfde adres wonen, in de steekproef komen.

Omdat het aantal personen in de steekproef door deze extra bewerkingen afneemt (zie tabel 2), is er uit iedere gemeente, indien mogelijk, drie keer het gewenste aantal personen getrokken. De steekproef die getrokken is, telt dus geen 15.498 personen, maar 46.440 personen. Na het uitvoeren van de extra bewerkingen telt de steekproef nog 42.829 personen. Uit deze steekproef is de definitieve steekproef van 15.498 personen getrokken, die geleverd is aan het veldwerkbureau. De definitieve steekproef is getrokken door uit iedere gemeente een aselechte steekproef te trekken van het gewenste aantal personen.

Tabel 2

Uitval door de extra bewerkingen

migratieachtergrond	overleden of geëmigreerd	niet aan derden	verhuisd naar instelling of tehuis	adres dubbel in onderzoek
Caribische migratieachtergrond	23	292	6	144
Iraanse migratieachtergrond	22	272	1	484
Marokkaanse migratieachtergrond	8	309	1	167
Poolse migratieachtergrond	108	37	0	0
Somalische migratieachtergrond	19	115	5	745
Surinaamse migratieachtergrond	10	337	4	86
Turkse migratieachtergrond	11	184	3	133
zonder migratieachtergrond	3	76	0	6

3 Weging

Het veldwerkbureau heeft de steekproef in 2020 en 2021 waargenomen. De waarneming heeft een langere doorlooptijd gekend vanwege de coronapandemie. Hierdoor heeft de waarneming aan huis enige tijd stilgelegen. Van de 15.498 personen in de steekproef hebben 5323 personen gerepondeerd (zie tabel 3). Voor deze 5323 respondenten zijn gewichten berekend, die gebruikt kunnen worden om met de responsgegevens populatie-eigenschappen te schatten. Zo'n gewicht bestaat uit het product van drie getallen: een insluitgewicht dat corrigeert voor de verschillende trekkansen die zijn gebruikt bij het trekken van de steekproef; een eerste correctiegewicht dat moet corrigeren voor de uitval bij het uitvoeren van de extra bewerkingen op de steekproef; en een tweede correctiegewicht dat moet corrigeren voor de uitval als gevolg van de selectieve non-respons.

Tabel 3

Aantal personen en aantal gemeenten in de steekproef

migratieachtergrond	# personen in respons		responspercentage
	steekproef	Respons	
Caribische migratieachtergrond	2114	592	28,0%
Iraanse migratieachtergrond	2020	855	42,3%
Marokkaanse migratieachtergrond	2114	570	30,0%
Poolse migratieachtergrond	2020	682	33,8%
Somalische migratieachtergrond	1443	519	36,0%
Surinaamse migratieachtergrond	2114	638	30,2%
Turkse migratieachtergrond	2020	696	34,5%
zonder migratieachtergrond	1653	771	46,6%

Het insluitgewicht van een respondent is gelijk aan de inverse van de trekkans, die zoals hiervoor staat beschreven, gelijk is aan de verhouding van het aantal personen in de steekproef en in het steekproefkader. Voor het tellen van het aantal personen in de steekproef worden de steekproeven gebruikt, waarvan de namen en adressen zijn opgevraagd bij RvIG, de steekproef die ongeveer drie keer groter is dan de steekproef die aan het veldwerkbureau geleverd is. Tabel 4 laat de waarden van de insluitgewichten zien. Omdat het aantal personen dat een steekproef moet tellen evenredig met de grootte van de steekproefkaders verdeeld is over het stratum met de kleine gemeenten en het stratum met de grootte gemeenten, verschillen de trekkansen tussen de strata niet veel van elkaar. Het eerste correctiegewicht corrigeert voor de uitval bij het uitvoeren van de extra bewerkingen op de steekproef. Deze uitval bestaat, zoals tabel 2 laat zien, voor een groot deel uit de personen die verwijderd zijn omdat zij met andere personen uit de steekproef op hetzelfde adres wonen. De kans dat een persoon verwijderd wordt omdat hij of zij met andere personen uit de steekproef op hetzelfde adres woont, is evenredig met het aantal personen in het steekproefkader dat op hetzelfde adres woont. Hoe meer personen er op dat adres wonen, hoe groter de kans dat meer dan één van die personen getrokken is in een steekproef.

Vanwege deze samenhang wordt de kans dat een persoon niet verwijderd wordt, gemodelleerd met een variabele die telt met hoeveel personen uit het steekproefkader iemand op hetzelfde adres woont. Als functie van deze variabele wordt het aantal personen geteld in de steekproef die aan het veldwerkbureau geleverd is en in de steekproef waarvan de namen en adressen zijn opgevraagd bij RvIG, de ongeveer drie keer zo grote steekproef. De kans dat iemand niet verwijderd wordt, is dan

gelijkgesteld aan de verhouding van deze aantallen. Het eerste correctiegewicht is gelijk aan de inverse van deze kans. Tabel B1 in hoofdstuk 5 laat de waarden van het correctiegewicht zien. Zoals te verwachten neemt de waarde van het correctiegewicht toe als de persoon met meer personen uit het steekproefkader op hetzelfde adres woont.

Tabel 4
Waarden van de insluitgewichten

migratieachtergrond	stratum	# personen		insluitgewicht
		steekproefkader	steekproef	
Caribische migratieachtergrond	grote gemeenten	106.633	5.280	20,2
	kleine gemeenten	21.333	1.058	20,2
Iraanse migratieachtergrond	grote gemeenten	29.761	4.830	6,2
	kleine gemeenten	7.605	1.202	6,3
Marokkaanse migratieachtergrond	grote gemeenten	261.409	5.442	48,0
	kleine gemeenten	43.621	896	48,7
Poolse migratieachtergrond	grote gemeenten	79.726	3.735	21,3
	kleine gemeenten	49.615	2.322	21,4
Somalische migratieachtergrond	grote gemeenten	18.995	3.201	5,9
	kleine gemeenten	6.750	1.113	6,1
Surinaamse migratieachtergrond	grote gemeenten	269.349	5.592	48,2
	kleine gemeenten	36.463	750	48,6
Turkse migratieachtergrond	grote gemeenten	282.674	5.037	56,1
	kleine gemeenten	57.488	1.023	56,2
zonder migratieachtergrond	grote gemeenten	5.864.050 ^a	2.649	2213,7
	kleine gemeenten	5.157.660 ^a	2.310	2232,8

a Om te corrigeren voor de extra selectie op het zevende cijfer van het A-nummer zijn de aantallen personen in het steekproefkader van de personen zonder migratieachtergrond met tien vermenigvuldigd.

Het tweede correctiegewicht, dat moet corrigeren voor de uitval als gevolg van de selectieve non-respons, is met de calibratieschatter berekend. Deze calibratieschatter heeft een weegmodel nodig, waarmee de responskansen van de personen in de steekproef gemodelleerd kunnen worden. In eerste instantie is het weegmodel van het onderzoek SIM2015 gebruikt:

$$\begin{aligned}
 & \textit{Migratieachtergrond} * \textit{Geslacht} + \\
 & \textit{Migratieachtergrond} * \textit{Leeftijd} + \\
 & \textit{Migratieachtergrond} * \textit{Generatie} + \\
 & \textit{Migratieachtergrond} * \textit{Type huishouden} + \\
 & \textit{Migratieachtergrond} * \textit{Grootte huishouden} + \\
 & \textit{Migratieachtergrond} * \textit{Gemeentegrootte} + \\
 & \textit{Migratieachtergrond} * \textit{Persoonlijk inkomen}.
 \end{aligned}$$

Een omschrijving van de hulpvariabelen uit het weegmodel is te vinden in tabel 5. Met de startgewichten, het product van de insluitgewichten en de eerste correctiegewichten berekent de calibratieschatter de tweede correctiegewichten zodanig dat de gewogen som van de hulpvariabelen uit het weegmodel gelijk is aan de randtotalen van de termen van het weegmodel. De variabelen in tabel 6 zijn de andere kandidaatvariabelen die ook in aanmerking zijn gekomen voor het weegmodel. Deze variabelen zijn gebruikt bij de analyse van de vertekening van het weegmodel. Voor alle variabelen in zowel tabel 5 als tabel 6 wordt de precieze indeling gegeven in hoofdstuk 5 met tabellen B2 en B3.

Tabel 5

Variabelen in het weegmodel van SIM2020, gebaseerd op het model van 2015

variabele	omschrijving
migratieachtergrond	migratieachtergrond van de persoon
geslacht	geslacht van de persoon
leeftijd	leeftijd van de persoon op 1 januari 2020
generatie	immigratiegeneratie van de persoon
type huishouden	type huishouden waar de persoon op 14 januari 2020 lid van was
grootte huishouden	grootte van het huishouden waar de persoon op 14 januari 2020 lid van was
grootte gemeente	grootte van de gemeente waar de persoon op 14 januari 2020 stond ingeschreven
persoonlijk inkomen	persoonlijk inkomen van de persoon in 2019

Tabel 6.

Andere kandidaatvariabelen voor het weegmodel

variabele	omschrijving
plaats in huishouden	plaats in het huishouden waar de persoon op 14 januari 2020 lid van was
burgerlijke staat	burgerlijke staat van de persoon op 14 januari 2020
eigendom woning	eigendom van de woning waarin de persoon in 2019 heeft gewoond
sociaal-economische categorie	sociaal-economische categorie van de persoon in 2019
gestandaardiseerd besteedbaar huishoudensinkomen	gestandaardiseerd besteedbaar inkomen van het huishouden in 2019
landsdeel	landsdeel waarin de gemeente ligt waar de persoon op 14 januari 2020 stond ingeschreven
stedelijkheid	graad van stedelijkheid van de gemeente waar de persoon op 14 januari 2020 stond ingeschreven
inschrijfduur	inschrijfduur in de Basisregistratie van Personen op 1 januari 2020

De randtotalen van de termen uit het weegmodel (te zien in tabel A2 in het Excelbestand dat als bijlage B te vinden is bij het rapport Survey integratie migranten. Uitvoering van veldwerk ten tijde van corona op www.scp.nl) zijn niet berekend met het steekproefkader, maar met een nieuw kader: het weegkader, dat op bijna dezelfde manier is afgeleid als het steekproefkader. Alleen zijn voor het weegkader niet de personen geselecteerd die op 9 december 2019 ingeschreven staan in de BRP, maar op 14 januari 2020, dus zo dicht mogelijk bij de peildatum van 1 januari 2020. Bovendien worden voor de groep personen zonder migratieachtergrond alle meerderjarigen geselecteerd, in plaats van diegenen wiens zevende cijfer van hun A-nummer gelijk is aan 6. Net als het steekproefkader telt het weegkader alleen personen die lid zijn van een particulier huishouden. Voor alle personen in het weegkader is in registers gezocht naar de juiste waarde van de hulpvariabelen. Mocht voor een persoon de waarde van een hulpvariabele niet in een register te vinden zijn, dan is met een eenvoudig model een schatting gemaakt van de waarde van de hulpvariabele. Ook voor de respondenten zijn de waarden van de hulpvariabelen opgezocht in de registers.

Het eindgewicht dat gebruikt moet worden om met de responsgegevens de populatie-eigenschappen te schatten, is voor persoon i gelijk aan

$$w_i = \pi_i^{-1} \cdot c_i^{(1)} \cdot c_i^{(2)},$$

met π_i^{-1} het insluitgewicht van persoon i , $c_i^{(1)}$ het eerste correctiegewicht en $c_i^{(2)}$ het tweede correctiegewicht. Om te zien of de eindgewichten en dus ook de correctiegewichten goed corrigeren voor de verschillende vormen van uitval, worden met de responsgegevens en de eindgewichten de populatietotalen van de hulpvariabelen uit tabel 6 geschat. Een vergelijking tussen deze geschatte

populatietotalen en de 'werkelijke' populatietotalen, afgeleid met het weegkader en de registers, laat zien waar de weging tekortschiet. Om het verschil te kwantificeren wordt de chikwadraattoets gebruikt. En ook al levert de toets geen echte p-waardes op, de p-waardes van de toets kunnen wel gebruikt worden om te achterhalen voor welke hulpvariabelen de schattingen van de populatietotalen beter zijn en voor welke hulpvariabelen de schattingen minder goed zijn.

Een blik op tabellen A1 en A2 (bijlage B) laat zien waar de weging tekortschiet. Er zijn acht p-waardes kleiner dan 0,05 maar zeven daarvan kunnen worden geaccepteerd met de volgende redenering. Het samenvoegen van de vergelijkbare categorieën *Ontvanger bijstand* en *Ontvanger overige sociale voorzieningen* voor *Sociaal-economische categorie* (bij de personen met een Caribische migratieachtergrond en personen met een Surinaamse migratieachtergrond) en *Niet-stedelijk* en *Weinig stedelijk* voor *Stedelijkheid* (bij de personen met een Iraanse migratieachtergrond) leidt tot relatief grote p-waardes voor de bijbehorende variabelen en laat de problemen dus verdwijnen. De kleine p-waardes voor *Landsdeel* worden toegelaten, omdat dit in de weging van 2015 ook het geval was. De p-waarde voor *Inschrijfduur* bij de personen met Iraanse migratieachtergrond baart ook niet veel zorgen, aangezien het een nieuwe populatie is en de p-waardes acceptabel waren voor alle andere chikwadraattoetsen die ermee verband houden. Het enige overgebleven probleem is de vertekening voor *Eigendom woning* bij de personen zonder migratieachtergrond die niet gemakkelijk kan worden verklaard. Om deze reden wordt het weegmodel aangepast door de term *Eigendom woning* toe te voegen, maar alleen voor de personen zonder migratieachtergrond. Voor de overige populatiegroepen blijft het model ongewijzigd.

Tabel 7

Weegmodel gebruikt in de weging van SIM2020

zonder migratieachtergrond	andere groepen
<i>geslacht + leeftijd + generatie + type huishouden + grootte huishouden + grootte gemeente + persoonlijk inkomen + eigendom woning</i>	<i>migratieachtergrond * geslacht + migratieachtergrond * leeftijd + migratieachtergrond * generatie + migratieachtergrond * type huishouden + migratieachtergrond * grootte huishouden + migratieachtergrond * grootte gemeente + migratieachtergrond * persoonlijk inkomen</i>

Voor de nauwkeurigheid van de schatters van de populatietotalen is het belangrijk dat de spreiding van de eindgewichten niet al te groot is. Kish¹ laat zien dat voor schatters van de populatietotalen van doelvariabelen die niet of nauwelijks samenhangen met de hulpvariabelen uit het weegmodel, de variantie van de schatters toeneemt evenredig met de variantie van de eindgewichten. Nu heeft het toevoegen van de extra term aan het weegmodel geleid tot een toename in de variantie van de eindgewichten voor de personen zonder migratieachtergrond (zie tabel B4 in hoofdstuk 5), maar deze toename lijkt niet te groot. Ook al is niet precies duidelijk wanneer de variantie van de eindgewichten te groot is. De eindgewichten van het aangepaste weegmodel voldoen, net zoals de eindgewichten van het weegmodel van het onderzoek SIM2015, ruim aan de vuistregel dat de verhouding tussen het maximale eindgewicht en de mediaan van de eindgewichten kleiner dan 3,5 moet zijn.

¹ Kish, L., 1992. Journal of Official Statistics, 183-200.

Het aangepaste weegmodel bij de groep van personen met een Caribische migratieachtergrond kent één negatief eindgewicht. Omdat negatieve eindgewichten lastig kunnen zijn bij het uitvoeren van de analyses, zijn de responsgegevens met het aangepaste weegmodel gewogen waarbij het waardenbereik van de tweede correctiegewichten begrensd is tot positieve waarden. Deze begrenzing van het waardenbereik leidt niet tot een toename in de variantie van de eindgewichten (zie tabel B4 in hoofdstuk 5). Daarom lijkt het geen kwaad te kunnen om niet de onbegrensde eindgewichten te gebruiken voor het schatten van populatie-eigenschappen, maar de begrensde. Enkele statistieken van de geleverde eindgewichten (verkregen met het aangepaste weegmodel uit tabel 7 en met begrenzing van de eindgewichten) zijn weergegeven in tabel 8. De statistieken voor de andere varianten van het weegmodel zijn opgenomen in tabel B4 in hoofdstuk 5.

Tabel 8

Minimum (MIN), eerste kwartiel (KW1), mediaan (MED), derde kwartiel (KW3), maximum (MAX) en standaarddeviatie (STD) van de eindgewichten

migratieachtergrond	MIN	KW1	MED	KW3	MAX	STD
aangepast weegmodel (zoals gegeven in tabel 7), begrenzing						
Caribische migratieachtergrond	22,8	150,6	214,2	280,2	535,6	97,9
Iraanse migratieachtergrond	8,0	36,5	43,5	50,9	83,1	11,7
Marokkaanse migratieachtergrond	85,2	416,2	521,1	644,0	1.106,4	165,2
Poolse migratieachtergrond	31,4	138,9	185,2	235,2	414,8	70,3
Somalische migratieachtergrond	11,4	35,8	47,9	60,4	123,1	17,8
Surinaamse migratieachtergrond	88,1	360,9	466,6	582,9	997,8	164,1
Turkse migratieachtergrond	98,1	371,8	471,0	583,4	1169,0	168,1
zonder migratieachtergrond	5.894,4	11.620,4	13.800,7	16.607,8	27.988,9	3.502,1

4 R-Indicatoren

De representativiteit van de respondenten wordt berekend met de zogeheten R-indicatoren bepaald in R op basis van de code van De Heij et al.² R-indicatoren meten in hoeverre de samenstelling van de respons op een enquête afwijkt van de oorspronkelijke steekproef. Als alle responskansen gelijk zijn, is de respons sterk representatief en zullen er geen systematische verschillen zijn tussen de samenstelling van de respons en de steekproef. Als de responskansen niet gelijk zijn, is het belangrijk om vast te stellen in hoeverre de samenstelling van de respons wordt beïnvloed. Dit wordt bereikt door een functie te definiëren die meet in hoeverre de individuele responskansen verschillen van de gemiddelde responskansen.

Als alle individuele responskansen van alle elementen in de populatie bekend zijn, dan kan de standaarddeviatie van de responskansen worden berekend. De waarden van de individuele responskansen zijn helaas vaak niet bekend en moeten geschat worden met een hiervoor gespecificeerd responsmodel. Stel dat $\rho_1, \rho_2, \dots, \rho_N$ de geschatte responskansen van de steekproefelementen zijn, dan kan de geschatte standaarddeviatie van de responskansen worden berekend door

$$S(\rho) = \sqrt{\frac{1}{N-1} \sum_{i=1}^n \frac{(\rho_i - \bar{\rho})^2}{\pi_i}},$$

waarbij π_i de eerste-ordetrekingskans is van steekproefelement i en waarbij $\bar{\rho} = \frac{1}{N} \sum_{i=1}^n \frac{\rho_i}{\pi_i}$ het gemiddelde van de geschatte responskansen is.

Deze (geschatte) standaarddeviatie is gelijk aan 0 als alle (geschatte) responskansen gelijk zijn, en de waarde van $S(\rho)$ neemt toe met variatie in de waarden van de responskansen. Het is niet moeilijk te bewijzen dat de maximale waarde van $S(\rho)$ gelijk is aan 0,5. Een R-indicator is dan gedefinieerd als

$$R(\rho) := 1 - 2 S(\rho) = 1 - 2 \sqrt{\frac{1}{N-1} \sum_{i=1}^n \frac{(\rho_i - \bar{\rho})^2}{\pi_i}}.$$

Deze R-indicator neemt waarden aan in het interval $[0, 1]$. Een waarde van 1 impliceert een sterke representativiteit. Hoe kleiner de waarde is, hoe meer de responsamenstelling afwijkt van de steekproefamenstelling.

De R-indicatoren en bijbehorende standaardfouten worden per migratieachtergrondgroep voor twee responsmodellen berekend. Het eerste responsmodel bevat alle variabelen in het weegmodel en het tweede alle variabelen die in de kandidatenlijst voor het weegmodel staan, als de data dit toelaten. Dus voor tabel 9 lijkt het responsmodel zoveel mogelijk op het weegmodel uit tabel 7. De enige uitzonderingen zijn variabelen die uit slechts één categorie bestaan, zoals *Generatie* voor de personen zonder migratieachtergrond en personen met Poolse migratieachtergrond. Deze variabele is niet opgenomen in de responsmodellen voor deze twee populatiegroepen.

² De Heij, V., Schouten, B., Shlomo, N., 2015. RISQ manual 2.1. Tools in SAS and R for the computation of R-indicators, partial R-indicators and partial coefficients of variation.

Tabel 9

R-indicatoren wanneer het responsmodel zo veel mogelijk overeenkomt met het weegmodel

migratieachtergrond	R	RSE
Caribische migratieachtergrond	0,795	0,021
Iraanse migratieachtergrond	0,864	0,023
Marokkaanse migratieachtergrond	0,859	0,021
Poolse migratieachtergrond	0,838	0,022
Somalische migratieachtergrond	0,832	0,026
Surinaamse migratieachtergrond	0,868	0,022
Turkse migratieachtergrond	0,857	0,022
zonder migratieachtergrond	0,837	0,026

Tabel 10 geeft de R-indicatoren weer die zijn verkregen met een uitgebreid responsmodel dat zoveel mogelijk variabelen uit tabellen 5 en 6 bevat. Net als voor de voorgaande tabel worden variabelen met slechts één mogelijke waarde verwijderd uit het responsmodel (naast *Generatie* voor de personen zonder migratieachtergrond en met Poolse migratieachtergrond, ook *Inschrijfduur* voor de personen zonder migratieachtergrond). Bovendien worden variabelen met lineair afhankelijke termen ook verwijderd. Hierdoor wordt *Grootte huishouden* uitgesloten voor alle migratieachtergrondgroepen en bovendien *Type huishouden* voor de personen zonder migratieachtergrond. De twee responsmodellen voor alle populaties zijn volledig uitgeschreven in tabel A3 in bijlage B.

Tabel 10

R-indicatoren wanneer het responsmodel zoveel mogelijk analysevariabelen bevat

migratieachtergrond	R	RSE
Caribische migratieachtergrond	0,760	0,021
Iraanse migratieachtergrond	0,793	0,022
Marokkaanse migratieachtergrond	0,827	0,021
Poolse migratieachtergrond	0,801	0,022
Somalische migratieachtergrond	0,779	0,026
Surinaamse migratieachtergrond	0,834	0,021
Turkse migratieachtergrond	0,833	0,022
zonder migratieachtergrond	0,819	0,026

5 Aanvullende gegevens over het weegmodel

Tabel B1

Waarden van de eerste correctiegewichten

migratieachtergrond	# personen op hetzelfde adres	# personen		eerste correctiegewicht
		te grote steekproef	steekproef	
Caribische migratieachtergrond	1	1443	503	2,87
	2	1618	575	2,81
	3	1251	399	3,14
	4	1074	346	3,10
	5	523	159	3,29
	6 of meer	429	132	3,25
Iraanse migratieachtergrond	1	1562	589	2,65
	2	1719	561	3,06
	3	1286	404	3,18
	4	986	315	3,13
	5	274	92	2,98
	6 of meer	205	59	3,47
Marokkaanse migratieachtergrond	1	679	239	2,84
	2	1016	351	2,89
	3	1045	338	3,09
	4	1184	402	2,95
	5	1184	398	2,97
	6	756	245	3,09
	7 of meer	474	141	3,36
Poolse migratieachtergrond	1 of meer	6057	2.020	3,00
Somalische migratieachtergrond	1	670	270	2,48
	2	584	220	2,65
	3	624	206	3,03
	4	641	208	3,08
	5	530	166	3,19
	6	462	156	2,96
	7	316	101	3,13
	8	200	52	3,85
	9 of meer	287	64	4,48
Surinaamse migratieachtergrond	1	1329	446	2,98
	2	1668	567	2,94
	3	1316	436	3,02
	4	1219	407	3,00
	5	515	164	3,14
	6 of meer	295	94	3,14
Turkse migratieachtergrond	1	509	174	2,93
	2	1114	360	3,09
	3	1212	388	3,12
	4	1553	530	2,93
	5	1046	382	2,74
	6	406	118	3,44
	7 of meer	220	68	3,24
zonder migratieachtergrond	1	938	316	2,97
	2	1766	617	2,86
	3 of meer	2255	720	3,13

Tabel B2

Variabelen in het weegmodel van SIM2015 met indeling

variabele	omschrijving	waardenbereik
migratieachtergrond	migratieachtergrond van de persoon	Caribische migratieachtergrond; Iraanse migratieachtergrond; Marokkaanse migratieachtergrond; Poolse migratieachtergrond; Somalische migratieachtergrond; Surinaamse migratieachtergrond; Turkse migratieachtergrond; zonder migratieachtergrond
geslacht	geslacht van de persoon	man; vrouw
leeftijd	leeftijd van de persoon op 1 januari 2020	indeling in klassen van vijf jaar
generatie	immigratiegeneratie van de persoon	zonder migratieachtergrond; eerste generatie migratieachtergrond; tweede generatie migratieachtergrond
type huishouden	type huishouden waar de persoon op 14 januari 2020 lid van was	eenpersoonshuishouden; niet-gehuwd paar zonder kinderen; gehuwd paar zonder kinderen; niet-gehuwd paar met kinderen; gehuwd paar met kinderen; eenouderhuishouden; overig type huishouden
grootte huishouden	grootte van het huishouden waar de persoon op 14 januari 2020 lid van was	eenpersoonshuishouden; meerpersoonshuishouden
grootte gemeente	grootte van de gemeente waar de persoon op 14 januari 2020 stond ingeschreven	minder dan 20.000 inwoners; van 20.000 tot 50.000 inwoners; van 50.000 tot 100.000 inwoners; van 100.000 tot 150.000 inwoners; van 150.000 tot 250.000 inwoners; Utrecht; Den Haag; Rotterdam; Amsterdam
persoonlijk inkomen	persoonlijk inkomen van de persoon in 2019	indeling in decielen

Tabel B3

Andere kandidaatvariabelen voor het weegmodel met indeling

variabele	omschrijving	waardenbereik
plaats in huishouden	plaats in het huishouden waar de persoon op 14 januari 2020 lid van was	thuiswonend kind; alleenstaande; partner niet-gehuwd paar zonder kinderen; partner gehuwd paar zonder kinderen; partner niet-gehuwd paar met kinderen; partner gehuwd paar met kinderen; ouder eenouderhuishouden; overig lid
burgerlijke staat	burgerlijke staat van de persoon op 14 januari 2020	ongetrouwd; getrouwd of partnerschap; gescheiden na huwelijk of na partnerschap; verweduwd na huwelijk of na partnerschap
eigendom woning	eigendom van de woning waarin de persoon in 2019 heeft gewoond	eigen woning; huurwoning
sociaal-economische categorie	sociaal-economische categorie van de persoon in 2019	zelfstandige; werknemer of overig actieve; bijstandsontvanger; ontvanger overige sociale voorzieningen; arbeidsongeschikte; student; overig niet-actieve; zonder inkomen; ontvanger Pensioen
gestandaardiseerd besteedbaar huishoudensinkomen	gestandaardiseerd besteedbaar inkomen van het huishouden in 2019	indeling in decielen
landsdeel	landsdeel waarin de gemeente ligt waar de persoon op 14 januari 2020 stond ingeschreven	Noord-Nederland; Oost-Nederland; West-Nederland; Zuid-Nederland
stedelijkheid	graad van stedelijkheid van de gemeente waar de persoon op 14 januari 2020 stond ingeschreven	zeer sterk stedelijk; sterk stedelijk; matig stedelijk; weinig stedelijk; niet stedelijk
inschrijfduur	inschrijfduur in de Basisregistratie van Personen op 1 januari 2020	van 0 tot 5 jaar; van 5 tot 10 jaar; 10 jaar of langer

Tabel B4

Minimum (MIN), eerste kwartiel (KW1), mediaan (MED), derde kwartiel (KW3), maximum (MAX) en standaarddeviatie (STD) van de eindgewichten

migratieachtergrond	MIN	KW1	MED	KW3	MAX	STD
weegmodel SIM2015, begrenzing						
Caribische migratieachtergrond	22,8	150,6	214,2	280,2	535,6	97,9
Iraanse migratieachtergrond	8,0	36,5	43,5	50,9	83,1	11,7
Marokkaanse migratieachtergrond	85,2	416,2	521,1	644,0	1.106,4	165,2
Poolse migratieachtergrond	31,4	138,9	185,2	235,2	414,8	70,3
Somalische migratieachtergrond	11,4	35,8	47,9	60,4	123,1	17,8
Surinaamse migratieachtergrond	88,1	360,9	466,6	582,9	997,8	164,1
Turkse migratieachtergrond	98,1	371,8	471,0	583,4	1169,0	168,1
zonder migratieachtergrond	6.294,3	12.147,3	13.972,4	16.293,1	28.520,7	2.998,2
aangepast weegmodel (zoals gegeven in tabel 7), geen begrenzing						
Caribische migratieachtergrond	-1,3	150,6	214,5	280,4	533,6	97,9
Iraanse migratieachtergrond	1,5	36,5	43,5	50,9	83,1	11,7
Marokkaanse migratieachtergrond	71,7	416,4	521,2	643,8	1.106,5	165,2
Poolse migratieachtergrond	18,6	139,0	185,2	235,1	414,6	70,3
Somalische migratieachtergrond	9,9	35,8	47,9	60,4	123,1	17,8
Surinaamse migratieachtergrond	76,2	361,0	466,6	582,9	997,6	164,1
Turkse migratieachtergrond	87,2	371,8	471,2	583,2	1.168,8	168,1
zonder migratieachtergrond	5.850,2	11.620,6	13.800,6	16.607,7	27.989,2	3.502,1
aangepast weegmodel (zoals gegeven in tabel 7), begrenzing						
Caribische migratieachtergrond	22,8	150,6	214,2	280,2	535,6	97,9
Iraanse migratieachtergrond	8,0	36,5	43,5	50,9	83,1	11,7
Marokkaanse migratieachtergrond	85,2	416,2	521,1	644,0	1.106,4	165,2
Poolse migratieachtergrond	31,4	138,9	185,2	235,2	414,8	70,3
Somalische migratieachtergrond	11,4	35,8	47,9	60,4	123,1	17,8
Surinaamse migratieachtergrond	88,1	360,9	466,6	582,9	997,8	164,1
Turkse migratieachtergrond	98,1	371,8	471,0	583,4	1169,0	168,1
zonder migratieachtergrond	5.894,4	11.620,4	13.800,7	16.607,8	27.988,9	3.502,1