



# Rapport

## Sociaal Domein Index (SDI) 2021

Onderzoeksdocumentatie en kwaliteitsanalyse

Versie 1.1

**Josien Knops**

**Francis van der Mooren**

**Marko Budé**

CBS Heerlen  
CBS-weg 11  
6412 EX Heerlen  
Postbus 4481  
6401 CZ Heerlen  
+31 45 570 60 00

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

Dataverzameling | Sociaal-economische en Ruimtelijke Statistieken  
9 mei 2022

## Inhoudsopgave

1.	Inleiding .....	4
1.1	Doel van dit document .....	4
1.2	Definities, acroniemen en afkortingen .....	4
2.	Doel, achtergrond en geschiedenis van het onderzoek .....	5
2.1	Doel .....	5
2.2	Geschiedenis van het onderzoek .....	5
3.	Dataverzameling .....	6
3.1	Steekproef .....	6
3.2	Benaderingsstrategie .....	11
3.3	Vragenlijstontwikkeling .....	15
3.4	Instructie .....	16
3.5	Veldwerk .....	16
4.	Verwerken (data processing) .....	19
4.1	Inleiding .....	19
4.2	Verrijken .....	19
4.3	Afleiden .....	19
5.	Responsrapportage .....	21
5.1	Afbakenen respons .....	21
5.2	Steekproef- en responsverantwoording .....	21
5.3	Selectiviteit van de respons naar kenmerken .....	22
6.	Weging .....	26
6.1	Insluitgewichten .....	26
6.2	Weegmodellen .....	27
6.3	Resultaten van de weging .....	28
7.	Andere kwaliteitsindicatoren .....	30
7.1	Missings, geen antwoord op vragen en tellingen .....	30
7.2	Inconsistenties .....	30
8.	Plausibiliteitsanalyse .....	31
8.1	Vergelijking oudere jaren met dit onderzoek .....	31
8.2	Andere bronnen .....	32
9.	Conclusies en aanbevelingen .....	34
9.1	Conclusies .....	34
9.2	Aanbevelingen .....	34
10.	Opgeleverde producten .....	35
10.1	Microbestanden .....	35
10.2	Statistische beveiliging .....	35
11.	Referenties .....	36
12.	Bijlagen .....	37
	Bijlage I – Procedure toekenning incentives .....	37
	Bijlage II – Algemene beschrijving dataprocessing .....	38

# 1. Inleiding

## 1.1 Doel van dit document

In dit document wordt verslag gedaan van het onderzoek Sociaal Domein Index 2021. De voorbereiding, dataverzameling en verwerking van de gegevens worden beschreven. Daarnaast dient dit document als naslagwerk voor onderzoekers die meer inzicht wensen in de achtergronden van het onderzoek en in het tot stand komen van de analysebestanden. Hoofdstuk 2 beschrijft het doel, achtergrond en geschiedenis van het onderzoek. Vervolgens gaat hoofdstuk 3 uitgebreid in op de dataverzameling en komen in hoofdstuk 4 de uitzonderingen in de verwerkingsstraat die voor dit onderzoek zijn toegepast aan de orde. De standaard verwerkingsstraat is als bijlage toegevoegd.

Tevens gaat dit document in op enkele kwaliteitsaspecten van het onderzoek. In hoofdstuk 5 komt de weging aan de orde en vervolgens in hoofdstuk 6 de responsanalyse. De responsanalyse bestaat uit een korte beschrijving van de respons naar achtergrondkenmerken. Vervolgens gaat hoofdstuk 7 in op andere kwaliteitscriteria, zoals gegevens over het aantal keer dat geen antwoord is gegeven en eventuele inconsistenties, en hoofdstuk 8 op de plausibiliteit. Tot slot volgen enkele conclusies en aanbevelingen (hoofdstuk 9), de opgeleverde producten (hoofdstuk 10), de referenties (hoofdstuk 11) en de bijlagen (hoofdstuk 12).

## 1.2 Definities, acroniemen en afkortingen

Begrip	Omschrijving
BRP	Basisregistratie Personen
BSN	Burger Service Nummer
BZK	Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
CAPI	Computer Assisted Personal Interviewing
CATI	Computer Assisted Telephone Interviewing
CAWI	Computer Assisted Web Interviewing
CBS	Centraal Bureau voor de Statistiek
CCI	Contact Centre Inbound
CCN	CBS Communicatie en Nieuws
CKP	Centrale Koppelbestand Personen
DBC	Dataverzameling, team Berichtgevers Contact
DVS	CBS Vragenlijsten en Steekproeven
EBB	Enquête Beroepsbevolking
SAL	Sector Arbeid, Inkomen en Leefsituatie
SCP	Sociaal en Cultureel Planbureau
SER	Sociaal-economische en ruimtelijke statistieken

## 2. Doel, achtergrond en geschiedenis van het onderzoek

### 2.1 Doel

Het doel van het onderzoek is het leveren van informatie over het sociaal domein. In het sociaal domein hebben een aantal decentralisaties plaatsgevonden en het ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK) heeft aan deze decentralisaties een monitoringstraject gekoppeld. De monitoring moet de Tweede Kamer informeren over de behaalde resultaten en het kabinet inzicht bieden in het functioneren van het stelsel, zodat het kan worden aangepast indien het stelsel niet tot de juiste resultaten leidt.

De informatie op basis van SDI 2021 wordt door het Sociaal en Cultureel Planbureau (SCP) gebruikt om aan te kunnen geven in welke mate de situatie van burgers in het sociaal domein verbetert of verslechtert.

Het onderzoek wordt uitgevoerd in opdracht van het SCP.

### 2.2 Geschiedenis van het onderzoek

In 2015, 2016 en 2017 is het SDI-onderzoek door CBS en SCP uitgevoerd. In 2018 heeft het ministerie van BZK aangegeven een andere opzet van het onderzoek te willen, aangezien er in de tot dan toe gebruikte opzet te weinig mensen werden getroffen die wel problemen hebben, maar geen gebruik maken van voorzieningen. Er werden vooral de mensen geselecteerd die de weg naar de individuele voorzieningen al hebben gevonden. Voor SDI 2021 worden respondenten daarom op een andere manier te geselecteerd en wordt uitgegaan van de problemen die mensen mogelijk hebben. De problemen die mensen kunnen hebben worden geschat op basis van registraties. Op basis van een aantal registervariabelen wordt de kans geschat dat mensen een bepaald probleem hebben. Op basis van de grootte van deze kans wordt op basis van registraties een stratificatie gemaakt.

## 3. Dataverzameling

### 3.1 Steekproef

#### *Doelpopulatie*

De doelpopulatie bestaat uit alle in Nederland woonachtige personen van 18 jaar of ouder op 1 januari 2020 m.u.v. de institutionele bevolking, i.e. personen in inrichtingen, instellingen of tehuizen. De doelpopulatie wordt geoperationaliseerd als alle personen van 18 jaar of ouder die zijn ingeschreven in de Basisregistratie Personen (BRP) en die niet behoren tot de institutionele bevolking.

#### *Steekproefkader*

Voor het steekproefkader wordt het zogenaamde stapelingsbestand gebruikt. Het stapelingsbestand geeft inzicht in het gebruik van sociale regelingen en voorzieningen binnen huishoudens. Dit maatwerkbestand laat van iedereen die op 31 december 2019 ingeschreven stond in de BRP zien of hij participatievoorzieningen, jeugdvoorzieningen of WMO-voorzieningen heeft gebruikt. Hoewel de peildatum van het stapelingsbestand 31 december 2019 is, geldt die datum niet voor alle variabelen die in het bestand zitten. Voor elke variabele is de meest recente versie gebruikt. Ook geeft het bestand voor veel personen, onder wie alle minderjarige kinderen, de juridische ouders weer. Met deze informatie kan het gebruik van jeugdvoorzieningen door minderjarige kinderen gekoppeld worden aan ouders.

#### *Steekproefontwerp*

Voor SDI 2021 wordt een personensteekproef getrokken volgens gestratificeerd steekproefontwerp. De stratumindeling in 2021 wijkt af van de vorige editie (2017): stratum 5 is opgesplitst in verschillende substrata, die hieronder zijn aangegeven met 5a tot en met 5e. Stratum 1 tot en met 4 zijn qua indeling ongewijzigd gebleven:

1. Personen die zelf participatievoorzieningen gebruiken en die lid zijn van een huishouden dat alleen participatievoorzieningen gebruikt.
2. Personen die juridisch ouder zijn van een of meerdere minderjarige kinderen die jeugdvoorzieningen gebruiken en die lid zijn van een huishouden dat alleen jeugdvoorzieningen gebruikt. Het minderjarige kind hoeft niet bij de juridische ouder te wonen.
3. Personen die zelf WMO-thuisvoorzieningen gebruiken en die lid zijn van een huishouden dat alleen WMO-thuisvoorzieningen heeft.
4. Personen uit multiprobleemhuishoudens. Die bestaan uit:
  - Personen die zelf alleen participatievoorzieningen gebruiken en die lid zijn van een huishouden dat meer dan één soort publieke voorzieningen gebruikt.
  - Personen die juridisch ouder zijn van een of meerdere minderjarige kinderen die jeugdvoorzieningen gebruiken en die lid zijn van een huishouden dat meer dan één soort publieke voorzieningen gebruikt.
  - Personen die zelf alleen WMO-thuisvoorzieningen gebruiken en die lid zijn van een huishouden dat meer dan één soort publieke voorzieningen gebruikt.
  - Personen die zelf meer dan één soort publieke voorzieningen gebruiken.
  - Personen die zelf geen publieke voorzieningen gebruiken, maar die lid zijn van een huishouden dat meer dan één soort publieke voorzieningen gebruikt.
5. Personen die zelf geen publieke voorzieningen gebruiken en die lid zijn van een huishouden dat of geen publieke voorzieningen gebruiken of dat alleen één soort publieke voorzieningen gebruikt. Deze groep wordt op basis van een door SCP ontwikkeld schattingsmodel uitgesplitst naar:
  - a. Personen die op basis van het door SCP ontwikkelde schattingsmodel geen grote kans hebben op problemen met inkomen, overige problemen, persoonlijke problemen of

problemen met werk. (Inclusief de mensen waarvoor we door ontbrekende kenmerken geen voorspelling van de kansen op problemen kunnen maken.)

- b. Personen die op basis van het door SCP ontwikkelde schattingsmodel een grote kans hebben op problemen met inkomen.
- c. Personen die op basis van het door SCP ontwikkelde schattingsmodel een grote kans hebben op overige problemen en die geen grote kans hebben op problemen met inkomen.
- d. Personen die op basis van het door SCP ontwikkelde schattingsmodel een grote kans hebben op persoonlijke problemen en die geen grote kans hebben op problemen met inkomen of overige problemen.
- e. Personen die op basis van het door SCP ontwikkelde schattingsmodel een grote kans hebben op problemen met werk, en die geen grote kans hebben op problemen met inkomen, overige problemen of persoonlijke problemen.

Bij het bepalen van het aantal verschillende soorten publieke voorzieningen dat een huishouden gebruikt, wordt alleen gekeken naar de voorzieningen die de meerderjarige personen van het huishouden gebruiken. Telt een huishouden dus een minderjarig kind dat participatievoorzieningen gebruikt en een ouder met een minderjarig kind dat jeugdvoorzieningen gebruikt, dan kent het huishouden voor het afleiden van de stratumindeling maar het gebruik van één soort voorzieningen, namelijk het gebruik van de jeugdvoorzieningen.

Op basis hiervan is de stratumindeling in strata 1 tot en met 4 bekend. Het SCP heeft voor de personen die geen gebruik maken van sociaal domein voorzieningen, i.e. stratum 5, geschat of deze personen een grote kans hebben op een of meer van de onderscheiden problemen. Dit is gedaan op basis van koppeling van enquêtegegevens voor SDI 2015-2017 aan allerlei kenmerken van de respondenten uit het SCP-stapelingsbestand. Op basis hiervan heeft het SCP iedereen die op 31 december 2017 ingeschreven stond in de BRP ingedeeld in één van de strata 5a tot en met 5e.

Bij de bepaling van het aantal te benaderen personen wordt uitgegaan van:

- 500 te behalen responsen in elk van de strata 1 tot en met 4
- 750 te behalen responsen in elk van de strata 5b tot en met 5e
- 1000 te behalen responsen in stratum 5a
- een design met vaste omvang per maand voor de cati- en capi-uitzet
- een totale cati-uitzet van maximaal 4.935 en capi-uitzet van maximaal 3.045

In totaal gaat het om een te realiseren aantal van 6.000 responsen.

Hieronder wordt eerst de totstandkoming van de uit te zetten aantallen in strata 1 tot en met 4 besproken en vervolgens voor strata 5a tot en met 5e.

#### Strata 1 tot en met 4

De cawi-, cati- en capi-responskansen en de kansen op niet herbenaderbare non-respons worden geschat op basis van de gerealiseerde (non-)responspercentages van SDI 2017 en SDI 2020<sup>1</sup>. De kansen op telefoonbezit zijn geschat op basis van gerealiseerde percentages van SDI 2020. Zie tabel 1 voor een overzicht.

De benodigde cawi-uitzet en bijbehorende cati- en capi-uitzet is in eerste instantie per stratum zo berekend dat de verwachte respons gelijk is aan 500, dat wil zeggen dat de cawi-uitzet gelijk is aan 500 gedeeld door het overall responspercentage. De verwachte capi- en cati-uitzet zijn dan te berekenen met de schattingen van de responspercentages, weigerpercentage en het percentage personen waarvan het telefoonnummer bekend is. Met deze steekproefomvang zou in 50% van de keren dat we de steekproef uitzetten het gewenste responsaantal van 500 niet gehaald worden. Om meer zekerheid te hebben dat de gewenste respons van 500 per stratum wel gehaald wordt, is

---

<sup>1</sup> Realisaties afkomstig uit SDI 2020 zijn alleen beschikbaar voor cawi en cati



vervolgens de cawi-uitzet verhoogd zodanig dat de linker grens van het 95%-betrouwbaarheidsinterval van het verwacht aantal respondenten gelijk is aan 500.

De capi- en cati-uitzet vallen te hoog uit bij het volgen van bovenstaande procedure. Daarom is vervolgens in elk stratum de capi- en cati-uitzet geschaald met dezelfde factoren<sup>2</sup> zodanig dat de totale capaciteit capi (3.045) en cati (4.935) exact wordt benut. Ten slotte is de cawi-uitzet opnieuw zodanig verhoogd dat de linker grens van het 95%-betrouwbaarheidsinterval van het verwacht aantal respondenten gelijk is aan 500.

#### Stratum 5a tot en met 5e

Uit SDI 2017 zijn géén responspercentages bekend van strata 5a tot en met 5e. Wel zijn er responspercentages bekend van stratum 5 als geheel. Het indelen van de steekproefpersonen van stratum 5 van SDI 2017 op dezelfde manier als voor 2018 is gedaan op basis van de geschatte kansen op problemen in 2017 biedt geen uitkomst. Dit is heel complex en levert niets op vanwege te kleine aantallen om responskansen op te baseren.

Om toch een inschatting te kunnen maken is gekeken naar een onderzoek met vergelijkbare mixed-mode waarneming. Het onderzoek Sociale Samenhang en Welzijn kwam hiervoor in aanmerking. Alle steekproefpersonen van dit onderzoek uit de periode januari 2018 tot en met april 2019 zijn gekoppeld aan de strata 5a tot en met 5e van SDI. Vervolgens zijn responspercentages per stratum berekend. In eerste instantie zijn de responskansen gebaseerd op het onderzoek Sociale Samenhang en Welzijn responskansen vermenigvuldigd met een correctiefactor<sup>3</sup>. Ten slotte zijn de cawi- en cati-responskansen bijgesteld n.a.v. de realisaties van SDI 2020. De kansen op telefoonbezit zijn geschat op basis van gerealiseerde percentages van SDI 2020. Zie tabel 3.1.1 en 3.1.2 voor een overzicht.

De benodigde cawi-uitzet en bijbehorende cati- en capi-uitzet van strata 5a tot en met 5e is op dezelfde wijze berekend als vermeld voor strata 1 tot en met 4. In tabel 3.1.3 is per portie en per stratum aangegeven wat de uit te zetten aantallen voor cawi-, cati- en capi-waarneming zijn. Merk op dat de onder de maand juli vermelde uitzet capi in feite de uitzet capi bevat voor de maanden juli en augustus samen. De capi- en cati-uitzet staan in principe vast per stratum. Echter, als de cawi respons in een stratum slechter uitvalt dan verwacht en in een ander stratum beter dan verwacht dan kan ervoor worden gekozen om in het eerste stratum meer cati/capi uit te zetten en in het laatste stratum minder. Randvoorwaarde is dat de totale uitzet cati- en capi over alle strata samen gelijk blijven. Zo kunnen zonder extra kosten toch in alle strata de targets van responsen mogelijk gehaald worden.

Verwacht wordt dat 33,7% van de eenheden respondeert. Deze 33,7% is als volgt opgebouwd. Van de gehele steekproef respondeert naar verwachting 19,9% via cawi en 4,2% geeft aan niet herbenaderd te willen worden. De herbenaderbare non-respons van de internetwaarneming wordt naar verwachting met 44,5% uitgedund voorafgaand aan de uitzet voor cati- en capi-waarneming. Van de herbenaderbare non-respons die niet wordt uitgedund gaat 61,8% naar cati. Bij cati verwachten we een responskans van 33,0%. Voor de overige 38,2% waarbij geen telefoonnummer wordt gevonden respondeert naar verwachting 32,4% bij capi. De totale uitzet is 18.929 personen. De cati-uitzet is gelijk aan 4935 personen en de capi-uitzet is gelijk aan 3.045 personen. Verwacht wordt dat via cawi 3.760 responsen gehaald worden, via cati 1.627 en via capi 986. Totaal verwachten we dus 6.373 responsen.

---

<sup>2</sup> Schaalfactor cati = 0,825; schaalfactor capi = 0,764

<sup>3</sup> De correctiefactor is gelijk aan het quotiënt van de gerealiseerde SDI 2017 responskansen in stratum 5 en de gewogen responskansen van Sociale Samenhang waarbij gewogen is naar de populatieverdeling binnen strata 5a t/m 5e. Op die manier corrigeren we voor verschillen tussen onderzoeken. De cawi, cati, en capi correctiefactoren zijn respectievelijk 0.81, 1.03 en 0.92.



De steekproef wordt verdeeld over vijf porties volgens de verdeling 3.786, 3.786, 3.786, 3.786 en 3.785.

#### *Steekproefbewerking*

De steekproeven worden per portie getrokken. Aansluitend op de steekproeftrekkingen worden de steekproeven bewerkt. De steekproefbewerking begint met het opvragen van de meest recente NAW-gegevens van steekproefpersonen bij de BRP. Vervolgens worden uit de getrokken steekproef bepaalde personen verwijderd. Het proces waarin dit gebeurt, wordt screening genoemd. Personen die tijdens de screening uit de steekproef verwijderd worden, zijn personen van wie het adres gedurende de afgelopen twaalf maanden al eerder in een CBS-steekproef voorkwam, personen van wie het adres behoort tot de adressen van de institutionele bevolking en personen van wie geen bruikbaar adres wordt verkregen bij de BRP. Indien meerdere personen woonachtig op hetzelfde adres getrokken worden, blijft na screening maximaal één persoon woonachtig op dit adres over. De overige personen woonachtig op het adres vallen tijdens de screening uit, omdat hun adres dan gedurende de afgelopen twaalf maanden al eerder in een CBS-steekproef voorkwam. Het niet verkrijgen van een bruikbaar adres kan veroorzaakt worden doordat personen bij de BRP hebben aangegeven hun gegevens niet te willen delen met derden. Een andere reden voor het niet verkrijgen van een bruikbaar adres kan zijn dat de betreffende persoon inmiddels is overleden. Om na screening nog genoeg personen in de steekproef over te houden, zijn de getrokken steekproeven groter dan de uit te zetten steekproeven: te trekken omvang = uit te zetten omvang \* screeningsfactor. De screeningsfactoren kunnen per stratum verschillen en zijn nog nader te bepalen.

De na screening resterende personen worden per stratum gesorteerd op postcode, huisnummer, etc., waarna systematisch personen voor uitzet worden geselecteerd: in mei zijn dat 387 in stratum 1, 251 in stratum 2, etc. De adressen van personen die in SDI zijn getrokken worden niet in de screeningsbakken geplaatst. Het is dus mogelijk dat een persoon op een adres uit SDI voor een ander onderzoek wordt benaderd in de komende 12 maanden. In de steekproef voor SDI zijn namelijk adressen van mensen die gebruik maken van allerlei voorzieningen en mensen die een grote kans op problemen hebben oververtegenwoordigd. Wanneer deze in de screening worden meegenomen, kan dit leiden tot een ondervertegenwoordiging van dergelijke adressen in andere onderzoeken, en dit is niet wenselijk. Ten behoeve van cati- en capi-herbenadering in geval van cawi-nonrespons worden alle te benaderen steekproefpersonen ter telefoonnummerverrijking aan externe partners aangeboden.

Om te komen tot vaste cati- en capi-omvang per stratum wordt een deel van de herbenaderbare cawi-nonrespondenten random uitgedund. Dat gebeurt maandelijks via een systematisch steekproefontwerp waarbij per stratum en apart voor personen met een bekend en een onbekend telefoonnummer de herbenaderbare groep wordt gesorteerd op postcode en huisnummer. De stapgrootte wordt zodanig bepaald dat het benodigde aantal uit te zetten personen wordt behaald. Deze personen worden telefonisch benaderd als een telefoonnummer van voldoende kwaliteit beschikbaar is, anders aan-huis.

*Tabel 3.1.1: Responsfracties, weigerfractie en telefoonbezitfractie naar stratum, schattingen SDI 2021*

	Stratum 1	Stratum 2	Stratum 3	Stratum 4
	Participatie	Jeugdzorg	WMO	Combi
respons cawi t.o.v. uitzet cawi	0,16	0,29	0,19	0,15
weigering cawi t.o.v. uitzet cawi	0,02	0,01	0,07	0,03
telbezit t.o.v. herbenaderbare non-respons	0,42	0,80	0,72	0,45
respons cati t.o.v. uitzet cati	0,22	0,28	0,55	0,31
respons capi t.o.v. uitzet capi	0,34	0,41	0,56	0,39
overall respons	0,28	0,43	0,40	0,29

Tabel 3.1.2 Responsfracties, weigerfractie en telefoonbezitfractie naar stratum, schattingen SDI 2021

	Stratum 5a	Stratum 5b	Stratum 5c	Stratum 5d	Stratum 5e
respons cawi t.o.v. uitzet cawi	0,31	0,11	0,15	0,27	0,26
weigering cawi t.o.v. uitzet cawi	0,02	0,03	0,11	0,05	0,03
telbezit t.o.v. herbenaderbare non-respons	0,72	0,47	0,65	0,85	0,53
respons cati t.o.v. uitzet cati	0,33	0,17	0,42	0,41	0,33
respons capi t.o.v. uitzet capi	0,39	0,31	0,30	0,29	0,17
overall respons	0,47	0,22	0,30	0,43	0,38

Tabel 3.1.3 Uitzet naar stratum, maand en mode, SDI 2021

		Mei	Juni	Juli	Aug	Sep	Okt	Nov	Totaal
Stratum 1	Uitzet cawi	387	387	388	388	388			1938
	Uitzet capi			183		92	92	91	458
	Uitzet cati		73	72	72	72	72		361
Stratum 2	Uitzet cawi	251	250	250	250	250			1251
	Uitzet capi			43		21	21	22	107
	Uitzet cati		91	92	92	92	92		459
Stratum 3	Uitzet cawi	270	271	270	270	271			1352
	Uitzet capi			52		26	26	25	129
	Uitzet cati		74	73	73	73	73		366
Stratum 4	Uitzet cawi	374	374	374	374	374			1870
	Uitzet capi			158		79	79	79	395
	Uitzet cati		68	68	69	69	68		342
Stratum 5a	Uitzet cawi	447	448	448	448	447			2238
	Uitzet capi			105		53	53	53	264
	Uitzet cati		143	144	143	143	144		717
Stratum 5b	Uitzet cawi	733	733	733	733	732			3664
	Uitzet capi			321		160	160	161	802
	Uitzet cati		157	156	157	157	156		783
Stratum 5c	Uitzet cawi	538	538	538	538	539			2691
	Uitzet capi			138		69	69	69	345
	Uitzet cati		140	140	139	139	140		698
Stratum 5d	Uitzet cawi	365	365	365	364	364			1823
	Uitzet capi			45		22	23	22	112
	Uitzet cati		135	135	136	136	135		677
Stratum 5e	Uitzet cawi	421	420	420	421	420			2102
	Uitzet capi			173		87	86	87	433
	Uitzet cati		106	107	106	106	107		532
Totaal	Uitzet cawi	3786	3786	3786	3786	3785			<b>18929</b>
	Uitzet capi			1218		609	609	609	<b>3045</b>
	Uitzet cati		987	987	987	987	987		<b>4935</b>

#### Verhuizers

Hoewel steekproeven zo kort mogelijk voor aanvang van de waarneemperiode worden getrokken, is niet te voorkomen dat het adres in de Basisregistratie Personen (BRP) niet juist blijkt te zijn omdat de steekproefpersoon is verhuisd. De NAW-gegevens worden vlak voor de uitzet gecontroleerd. Tijdens de waarneemperiode worden verhuizers niet herbenaderd.

### *Dataverzameling in de Bijlmer en op de Waddeneilanden*

In heel Amsterdam Zuidoost wordt face-to-face waargenomen met uitzondering van sommige adressen in de postcodes 1102, 1103 en 1104. Maandelijks wordt door de regiomanager van regio Amsterdam beoordeeld welke steekproefadressen niet aan huis worden benaderd. Het criterium dat daarbij wordt gehanteerd is de veiligheid voor de interviewers. In de praktijk komt het erop neer dat sloopflats, gedeeltelijke sloopflats en een deel van de hoogbouw niet aan huis worden benaderd. In deze gebieden wordt wel via internet en telefonisch waargenomen. Omdat op de Waddeneilanden, Vlieland, Terschelling, Ameland, Schiermonnikoog en Texel geen interviewers wonen, wordt op deze eilanden niet face-to-face geënquêteerd. Wel wordt daar via internet en telefonisch waargenomen.

## 3.2 Benaderingsstrategie

Bij het ontwerp van het onderzoeksdesign is uitgegaan van de navolgende randvoorwaarden.

1. Het onderzoek wordt uitgevoerd via mixed mode waarneming: internetwaarneming (cawi) bij non-respons gevolgd door hetzij telefonische (cati) hetzij face-to-face-waarneming (capi). Steekproefpersonen ontvangen een aanschrijfbrief en een folder met het verzoek om de vragenlijst via internet in te vullen. In de brief zijn een inlogcode en een gebruikersnummer opgenomen. Daarnaast wordt in de brief ingegaan op het doel van het onderzoek. Enkele weken na de verzending van de aanschrijfbrief wordt een rappelbrief verstuurd naar degenen die op dat moment nog niet hebben gerespondeerd en die niet via het Contact Centre Inbound (CCI) hebben aangegeven niet te willen of kunnen deelnemen aan het onderzoek. Weer enkele weken later gebeurt hetzelfde. De aanschrijf- en rappelbrieven worden verstuurd met niet-kritische post. Dit is post die binnen 48 tot 72 uur wordt bezorgd.
2. Om de respons te verhogen worden conditionele kansincentives ingezet. Concreet maken respondenten kans op het winnen van een iPad bij deelname.
3. Voor de aanschrijving en rappelling van steekproefpersonen worden de brieven gebruikt die zijn ontwikkeld voor SDI 2020.
4. In de initiële planning liep de totale waarneming van maart tot en met september 2021. Door de maatregelen vanwege het COVID-19 is deze planning aangepast en loopt het veldwerk van mei tot en met november 2021.
5. De steekproef wordt in 5 porties uitgezet. De doorlooptijd van de waarneming voor elke van deze 5 porties is 3 maanden. Daarbij wordt opgemerkt dat de totaal waar te nemen steekproef niet evenredig wordt verdeeld over de 5 steekproefporties.
6. Voor het onderzoek wordt uitgegaan van een te behalen aantal responsen van 500 in strata 1 tot en met 4, 750 responsen in strata 5b tot en met 5e en 1.000 te behalen responsen in stratum 5a. In totaal worden daarmee 6.000 responsen gerealiseerd. Er is een resultaatverplichting.

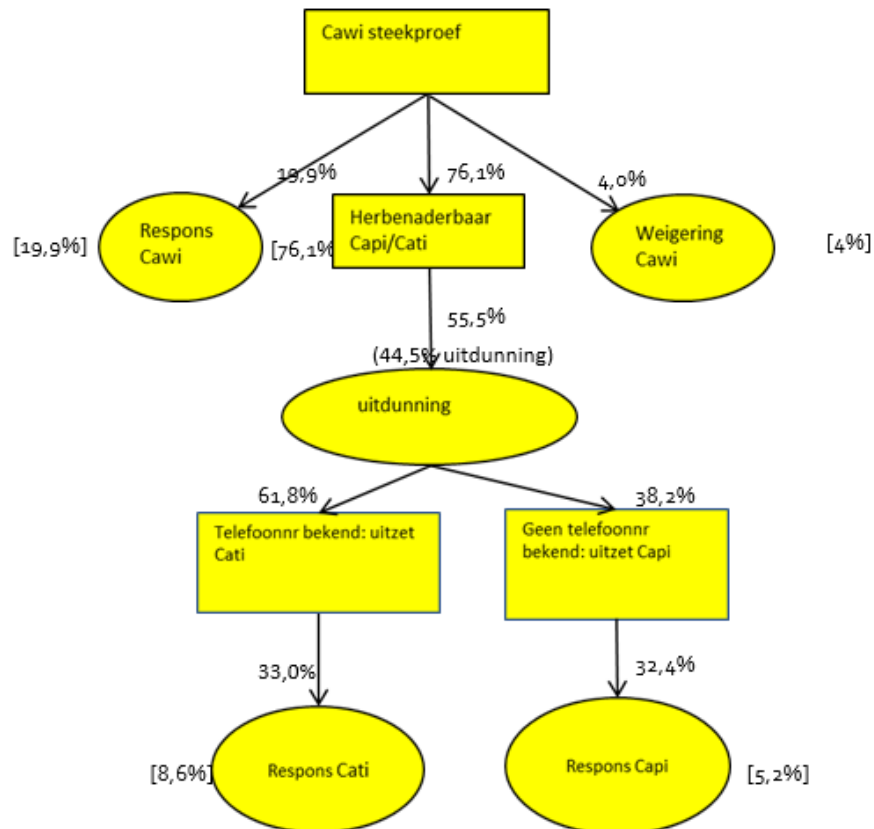
Dit heeft geleid tot de volgende benaderingsstrategie:

- Steekproefpersonen ontvangen een aanschrijfbrief met daarin het verzoek om via internet deel te nemen aan het onderzoek en de bijbehorende inloggegevens. Daarnaast wordt in de brief ingegaan op het doel van het onderzoek en gewezen op de verrijkingen. Bij de aanschrijfbrief wordt een folder bijgevoegd die specifiek is samengesteld voor SDI.
- De aankondigingsbrief (cawi) wordt gevolgd door twee schriftelijke rappels.
- Personen die naar aanleiding van één van de brieven via de CBS-helpdesk deelname hebben geweigerd worden niet telefonisch of face-to-face herbenaderd. De resterende steekproefpersonen worden telefonisch of face-to-face herbenaderd. Op het moment dat de herbenaderingsporties worden geprepareerd, wordt voor respondenten de mogelijkheid afgesloten om de vragenlijst via internet in te vullen. Hiermee wordt voorkomen dat respondenten gelijktijdig via internet en telefonisch dan wel face-to-face reageren.

In figuur 3.2.1 wordt de benaderstrategie schematisch weergegeven. Per benaderingsstap is de geschatte responskans vermeld, waarbij zowel de responskans is vermeld uitgaande van de

betreffende deelsteekproef (zonder vierkante haken) als van de uitgezette steekproef (tussen vierkante haken). Deze kansen zijn gebaseerd op realisatiecijfers van het Schoolverlatersonderzoek 2019 en de inzet van conditionele incentives.

Figuur 3.2.1: Benaderstrategie, SDI 2021



Uit figuur 3.2.1 is de totaal te verwachten responskans berekend op basis van de onderscheiden benaderingsstappen. Tabel 3.2.1 geeft de responskans per benaderingsstap en de cumulatieve responskans op de uitgezette steekproef.

Tabel 3.2.1: Responskans per benaderstap en totale responskans, SDI 2021

Benaderingsstap	Respons
Cawi	0,199
Cati	$0,761 \times 0,555 \times 0,618 \times 0,330 = 0,086$
Capi	$0,761 \times 0,555 \times 0,382 \times 0,324 = 0,052$
Totaal	0,337

Uit tabel 3.2.1 blijkt dat naar verwachting 33,7 procent van de uitgezette steekproef tot respons leidt. Het totale verwachte cawi-responspercentage bedraagt 19,9, het verwachte cati-responspercentage 33,0 en het verwachte capi-responspercentage 32,4.

#### Incentive

De toekenning van de beloning in de vorm van een iPad wordt uitgevoerd met door CBS ontwikkelde software (manipula) die standaard gebruikt wordt voor de steekproefbewerking. Vooraf wordt in de steekproef vastgelegd welke steekproefpersonen de iPad's krijgen toegekend, zie bijlage I Procedure toekenning Incentives. De prijs wordt daadwerkelijk vergeven als een steekproefpersoon, waaraan op voorhand een prijs is toegekend in de steekproef, volledig respondeert (dus de vragenlijst tot en met de verzendpagina invult).

### Contact Center Inbound (CCI)

CCI ontvangt vragen en reacties (waar onder klachten) van externe relaties c.q. berichtgevers/respondenten over de, door CBS, aan hen toegestuurde vragenlijsten. Deze vragen kunnen via verschillende kanalen binnenkomen: per telefoon, mail, web en brief. Het afhandelen van deze vragen gebeurt door IHD (Inhoudelijke HelpDesk), THD (Technische HelpDesk) of KA (Klachtenafhandeling)/Backoffice dan wel door de statistische sectoren.

Bij het formuleren van de antwoorden op de vragen wordt gebruik gemaakt van de zogenaamde kennisbank. Hierin is inhoudelijke informatie over het onderzoek opgenomen, zoals de data van de uitzendingen, het doel van het onderzoek en de omvang van het onderzoek. Daarnaast staat omschreven hoe respondenten die geen medewerking aan het onderzoek willen verlenen (met reden van non-respons) dienen te worden afgeboekt. Tot slot wordt ook gebruik gemaakt van het enquêtemateriaal (brieven, brochures).

Voor het registreren van vragen en reacties (waar onder klachten) en het afboeken van non-respons van externe relaties c.q. berichtgevers/respondenten, maakt CCI gebruik van Contactmanagement.

### Planning en logistieke uitwerking benaderstrategie

De steekproef voor SDI bedraagt 18.929 steekproefelementen. De eerste waarneemperiode start op woensdag 28 april 2021 met de verzending van de aanschrijfbrieven voor de internetbenadering. Tabel 3.2.2 bevat de datums waarop de aanschrijf- en rappelbrieven worden verstuurd voor de verschillende cawi-porties. De aanschrijf- en rappelbrieven worden telkens op dinsdag verstuurd en op donderdag of vrijdag bezorgd, zodat de steekproefpersonen ze voor het weekend ontvangen. Alle brieven worden verzonden met niet- tijdkritische post. Dit is post die binnen 48 tot 72 uur wordt bezorgd.

In de tabellen 3.2.3 en 3.2.4 zijn de datums waarop respectievelijk de telefonische als face-to-face waarneming wordt uitgevoerd weergegeven.

Tabel 3.2.2 Verzenddatums brieven internetwaarneming, SDI 2021

Cawi-portie*	Uitzendingen				aantallen		
	verzend		bezorgd			Cawi afsluiting	waarneemperiode
	dag	datum	dag	datum			
c1	wo	28-4-2021	vrij/za	30/4; 1-5-2021	3786		
c1.1	di	11-5-2021	do/vrij	14/15-5-2021	3404		
c1.2	di	18-5-2021	do/vrij	20/21-5-2021	3201	1-6-2021	mei
c2	di	25-5-2021	do/vrij	27/28-5-2021	3786		
c2.1	di	8-6-2021	do/vrij	10/11-6-2021	3404		
c2.2	di	15-6-2021	do/vrij	17/18-6-2021	3201	1-7-2021	juni
c3	di	29-6-2021	do/vrij	1/2-7-2021	3786		
c3.1	di	13-7-2021	do/vrij	15/16-7-2021	3404		
c3.2	di	20-7-2021	do/vrij	22/23-7-2021	3201	2-8-2021	juli
c4	di	27-7-2021	do/vrij	29/30-7-2021	3786		
c4.1	di	10-8-2021	do/vrij	12/13-8-2021	3403		
c4.2	di	17-8-2021	do/vrij	19/20-8-2021	3200	1-9-2021	aug
c5	ma	24-8-2021	wo/do	26/27-8-2021	3785		
c5.1	di	7-9-2021	do/vrij	9/10-9-2021	3403		
c5.2	di	14-9-2021	do/vrij	16/17/-9-2021	3200	1-10-2021	Sept

\* c = cawi waarneming

Vlak vóór het moment dat belporties worden gemaakt voor de herbenadering, worden de internet non-respondenten hiervoor geselecteerd en worden de internetporties afgesloten. Voorafgaand aan de telefonische herbenadering wordt geen nieuwe aanschrijfbrief verstuurd.

De startdatum voor het 65+ deel van de porties t1 en t2 ligt een dag na de startdatum van het 64-deel van betreffende porties. Op de dag dat de internetwaarneming wordt afgesloten, kan alleen een avonddienst worden gebeld. Voor 65+ personen geldt dat het eerste contact 's ochtends of 's middags wordt gepland.

Tabel 3.2.3 Gegevens over de cati-porties, SDI 2021

Portie *	startdatum	einddatum	totaal aantal	per portie	waarneem- periode
t1.1 64-	1-6-2021	30-6-2021	987	718	juni
t1.2 65+	2-6-2021	30-6-2021		269	
t2.1 64-	1-7-2021	30-7-2021	987	718	juli
t2.2 65+	2-7-2021	30-7-2021		269	
t3.1 64-	2-8-2021	31-8-2021	987	718	aug
t3.2 65+	3-8-2020	31-8-2021		269	
t4.1 64-	1-9-2021	30-9-2021	987	718	sept
t4.2 65+	2-sep-21	30-9-2021		269	
t5.1 64-	1-okt-21	29-okt-21	987	718	okt
t5.2 65+	4-okt-21	29-okt-21		269	

\* t = cati waarneming

Vanwege de vakantieperiode worden steekprofelementen die in juli en augustus aan-huis worden waargenomen, begin juli tegelijkertijd aan interviewers toegekend. Interviewers hebben vervolgens twee maanden om deze elementen af te handelen. Dit wordt de twee-maands uitzet genoemd. Voorafgaand aan de face-to-face herbenadering wordt geen nieuwe aanschrijfbrief verstuurd.

Tabel 3.2.4 Gegevens over de capi-porties, SDI 2021

Portie*	startdatum	einddatum	aantal	waarneem- periode
f1	1-jul-21	31-aug-21	609	juli tweemaands uitzet
f2	8-jul-21	31-aug-21	609	augustus
f3	1-sep-21	30-sep-21	609	september
f4**	1-okt-21	31-okt-21	609	oktober
f5	1-nov-21	30-nov-21	609	november

\* f = capi waarneming

\*\* kan pas starten op 8 juli, gaat mee met de 2<sup>e</sup> uitzet voor de twee-maandsportie

#### *Communicatiemiddelen en -kanalen*

Voor SDI is een communicatiestrategie opgesteld. Binnen het CBS is de shared service Berichtgeverscommunicatie van de sector Ontwerp van Dataverzameling (DVO) verantwoordelijk hiervoor.

Het opleidingsniveau van de doelgroep is verschillend. De communicatieproducten moeten zodanig opgesteld worden zodat de gehele doelgroep zich aangesproken voelt. Concreet vertaalt zich dat in de inzet van de volgende communicatiemiddelen en -kanalen:

#### Aanschrijfbrieven

In de aanschrijf- en rappelbrieven zijn de volgende zaken relevant:

- In de brieven wordt vermeld dat het om een gezamenlijk onderzoek van SCP en het CBS gaat en dat het CBS de verzamelde gegevens aanvult met al bij het CBS aanwezige gegevens. In de

brieven wordt een formulering voor de “informed consent” opgenomen, waarbij in algemene zin wordt opgenomen welke variabelen het betreft.

- De aanschrijf- en rappelbrieven maken gebruik van de standaardlinea's.
- In de aanschrijf- en rappelbrieven wordt opgenomen dat respondenten kans maken op een conditionele beloning in de vorm van een iPad. Tevens wordt expliciet het aantal iPads genoemd dat naar verwachting wordt gewonnen (9 iPads). In de aanschrijf- en rappelbrieven wordt ook een beeldkenmerk van de conditionele incentive (foto iPad) opgenomen.
- In de aanschrijf- en rappelbrieven wordt de standaardpassage met betrekking tot de zogenoemde opt-out optie opgenomen. Dit houdt in dat een mogelijke winnaar kan aangeven geen aanspraak te willen maken op de beloning.
- In de aanschrijf- en rappelbrieven wordt ook de link naar de vragenlijst inclusief inloggegevens opgenomen. De algemene URL die op de brieven komt te staan en die de respondenten moeten intypen is 'antwoord.cbs.nl'. Deze URL komt automatisch op de loginpagina '<https://antwoord.cbs.nl/login/...>' terecht. Wanneer gebruikersnaam en wachtwoord kloppen, vindt er een verwijzing naar de vragenlijstomgeving plaats. Dit gebeurt automatisch, maar de gebruiker krijgt een andere URL te zien. Deze URL zal dan beginnen met: '<https://onderzoek.cbs.nl/...>'.
- In de aanschrijf- en rappelbrieven wordt een verwijzing opgenomen naar de onderzoekspagina voor SDI op de CSB-website.
- In de 2e rappelbrief moet de afsluitdatum worden opgenomen van de internetvragenlijst.
- De aanschrijf- en rappelbrieven worden in kleur geprint.
- Zowel de aanschrijf- als rappelbrieven worden op dinsdag verstuurd met 48/72-uurs niet-tijdkritische (NTK) post.

#### Folder

Voor SDI wordt een onderzoeksspecifieke folder gemaakt. In deze folder wordt een link naar twee actuele nieuwsberichten opgenomen.

#### CBS Website

Voor het onderzoek is een pagina gemaakt op de CBS-website ([www.cbs.nl/leven](http://www.cbs.nl/leven)). Hierbij worden de volgende uitgangspunten gehanteerd, de onderzoekspagina:

- bevat een link naar de inlogpagina van de vragenlijst;
- bevat een link naar de manier waarop het CBS omgaat met privacy;
- bevat een verwijzing naar de incentive pagina (gedragscode voor incentives).

### 3.3 Vragenlijstontwikkeling

Voor SDI 2021 is een elektronische vragenlijst gebruikt voor cawi, capi en cati. De vragenlijst is voor zover mogelijk voor alle modes hetzelfde, met dien verstande dat de lay-out en de aanwijzingen voor de interviewer (capi, cati) of respondent (cawi) per mode verschillend zijn.

Het CBS is verantwoordelijk voor de vragenlijsten. Over de vragenlijst vindt intensieve afstemming met SCP plaats. CBS zal de voorstellen van het SCP zoveel mogelijk overnemen, waarbij rekening gehouden wordt met de lengte van de vragenlijsten en de bruikbaarheid van de informatie.

Voor de vragenlijst geldt dat de gemiddelde vragenlijstduur via internet of telefoon maximaal 21 minuten mag bedragen. En de gemiddelde face-to-face vragenlijstduur maximaal 26 minuten.

Voor de vragenlijst voor SDI 2021 is de voor SDI 2020 ontwikkelde vragenlijst beperkt aangepast.

#### *Vragenlijstblokken*

De volgende blokken zijn in het onderzoek SDI 2021 opgenomen, dit zijn dezelfde blokken als in de editie van 2017 waren opgenomen:

- Huishoudbox
- Aanvulling huishoudbox
- Betaald werk



- Onderwijs
- Huisvesting
- Tevredenheid
- Gezondheid
- Algemene vaardigheden
- Sociale Contacten
- Maatschappelijke participatie
- Redzaamheid
- Financiën
- Functioneren van het huishouden
- Jeugd
- Arbeidsmarkt
- Voorzieningen
- Werving en afsluiting
- Incentive

#### *Testprocedure*

Respondenten krijgen niet alle vragen uit het onderzoek voorgelegd, omdat de vragenlijst zo gebouwd wordt dat respondenten alleen de vragen krijgen die voor hen relevant zijn. Soms wordt de respondent ook om vragen heen geleid, omdat uit eerdere antwoorden zijn antwoord op bepaalde vragen al kan worden afgeleid.

De in de computergestuurde vragenlijsten opgenomen routing kan complex zijn. Het is van belang de routing goed te controleren om te voorkomen dat tijdens de analysefase blijkt dat essentiële vragen ten onrechte niet zijn gesteld. Er wordt binnen CBS een standaard testmethode toegepast, waarmee kan worden nagegaan of de in het ontwerp van de vragenlijst vastgelegde routing op een juiste wijze is geprogrammeerd. De gehanteerde testmethode bestaat eruit dat op basis van de te testen vragenlijst een testopzet opgesteld wordt. In de testopzet wordt aangegeven welke vragenblokken op welke manier getest gaan worden. Vervolgens worden voor de te testen vragenblokken testprotocollen ontwikkeld. Deze beschrijven stap voor stap welke gegevens in de vragenlijst ingevoerd dienen te worden. De vragenlijstontwikkelaars testen aan de hand van deze protocollen de elektronische vragenlijst en rapporteren via een standaardformulier hun bevindingen aan de programmeur. Het testen is een iteratief proces, dat meestal na drie iteraties kan worden afgesloten.

### **3.4 Instructie**

De interviewers die voor SDI worden ingezet, hebben de standaard opleiding gevolgd. Dit betekent het bestuderen van de handleiding en het doorlopen van de e-learning, waarna de opleiding afgerond wordt met een toets en doorlopen van de oefenvragenlijst.

Voor het onderzoek zijn daarnaast de regiomanagers (capi) en de teamleider (cati) goed geïnformeerd, zodat zij de interviewers en agents voldoende kunnen instrueren in hun werkoverleg.

Ook zijn de medewerkers van CCI geïnstrueerd en is de kennisbank gevuld met actuele informatie over het onderzoek.

### **3.5 Veldwerk**

#### *Aanpassingen onderzoekopzet*

Het onderzoek is grotendeels conform opzet uitgevoerd. Vanwege achterblijvende respons is in september 2021 in overleg met het SCP besloten tot een extra cawi-portie. Voor deze extra portie zijn 1.655 steekproefelementen uitgezet. De aanschrijfbrief daarvoor is op 26 oktober 2021 verstuurd, het veldwerk is op 1 december afgesloten. Voor deze portie heeft geen cati of capi

benadering plaatsgevonden. Naast het uitzetten van de extra cawi-portie heeft ook een herverdeling van de capi- en cati-uitzet van de maanden oktober en november plaatsgevonden.

#### *Veldwerkresultaten*

In totaal zijn er 6.077 responsen gerealiseerd. Daarmee wordt voldaan aan de responswens van 6.000, zoals beschreven in paragraaf 3.1. In totaal zijn 3.992 responsen via internet gerealiseerd. Via de telefonische waarneming zijn er 1.256 responsen gerealiseerd en voor de face-to-face waarneming werden er 828 responsen behaald. Tabel 3.5.1 geeft de gerealiseerde responsresultaten weer voor het totaal, in tabel 3.5.2 zijn de gerealiseerde responsen naar veldwerkperiode opgenomen en in tabel 3.5.3 naar stratum.

Tabel 3.5.1 Totale respons, SDI 2021

	Totaal
Uitzet	20.584
Respons cawi	3.992
Respons cati	1.256
Respons capi	828
Totale respons	6.077
% Totale respons	29,5
% Target	33,7

Tabel 3.5.2 Totale respons naar veldwerkperiode, SDI 2021

	Veldwerkperiode					
	202105	202106	202107	202108	202109	202111*
Uitzet	3.786	3.786	3.786	3.786	3.785	1.655
Respons cawi	813	719	709	735	755	260
Respons cati	274	255	277	249	201	
Respons capi	185	173	173	145	154	
Totale respons	1.272	1.147	1.159	1.129	1.110	260
% Totale respons	33,5	30,3	30,6	29,8	29,3	15,7

\* alleen cawi waarneming in stratum 1, 2, 3, 5a, 5b en 5e

Tabel 3.5.3 Totale respons naar veldwerkperiode, SDI 2021

	Stratum								
	1	2	3	4	5a	5b	5c	5d	5e
Uitzet	1.958	1.331	1.465	1.870	2.362	4.804	2.691	1.823	2.280
Respons cawi	294	343	292	277	716	461	500	545	564
Respons cati	63	127	166	78	181	103	219	200	119
Respons capi	127	38	58	125	101	206	67	26	80
Totale respons	484	508	516	480	999	771	786	770	763
% Totale respons	24,7	38,2	35,2	25,7	42,3	16,0	29,2	42,3	33,5

#### *Gerealiseerde vragenlijstduur*

De verwachte gemiddelde vragenlijstduur via internet en telefonische waarneming bedroeg 21 minuten en de verwachte gemiddelde face-to-face vragenlijstduur 26 minuten.

Bij de bepaling van de gemiddelde vragenlijstduur worden alle responsen uitgesloten waarvoor geldt dat de vragenlijst op een andere datum is gestart dan geëindigd. De resterende vragenlijsten worden vervolgens olopend gesorteerd op basis van vragenlijstduur. Van deze gesorteerde lijst wordt de onderste 2,5% (met de kortste vragenlijstduren) en de bovenste 2,5% (met de langste vragenlijstduren) buiten beschouwing gelaten.

Uit tabel 3.5.4 blijkt dat de gerealiseerde gemiddelde vragenlijstduur voor cawi en cati werd overschreden. Voor capi geldt dat dit ook het geval was voor de eerste twee porties, maar niet voor de overige drie.

Tabel 3.5.4 Gemiddelde vragenlijstduur (minuten) naar mode, SDI 2021

	CAWI	CATI	CAPI
202105	25,52	22,50	27,01
202106	24,08	23,59	27,54
202107	24,18	22,24	25,77
202108	23,79	25,67	23,22
202109	24,22	24,37	24,39
202111	23,73		

## 4. Verwerken (data processing)

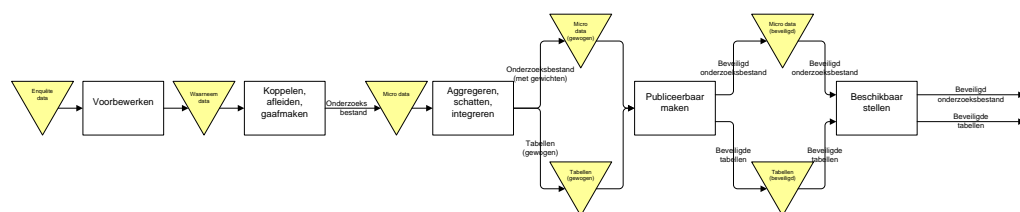
De term dataprocessing wordt gebruikt om de werkzaamheden samen te vatten die nodig zijn om aan de antwoorden die respondenten geven op de vragen in de vragenlijst plausibele statistische informatie te kunnen ontleen. De verwerking van de data is zoveel mogelijk geautomatiseerd.

Het verwerken is uitgevoerd door statistische afdeling Dynamisering verwerking, sector van de Divisie Sociaal-economische en Ruimtelijke Statistieken.

### 4.1 Inleiding

Onderstaande figuur geeft het verwerkingsproces van de onderzoeksdata in hoofdlijnen aan. De vijf deelprocessen (weergegeven in de vierkante blokken) worden in bijlage IV in meer detail beschreven.

**Figuur 4.1.1: Procesmodel Verwerkingsproces**



Hieronder wordt de verwerking van SDI 2021 beschreven en wordt aandacht besteed aan de keuzes die hierin gemaakt zijn.

### 4.2 Verrijken

Doordat CBS bij de steekproeftrekking aan iedere respondent een uniek volgnummer toekent, is het mogelijk om aan de gegevens uit de enquête informatie uit andere bronnen (registraties) te koppelen. Het databestand van SDI is bij oplevering verrijkt met de zogeheten standaard verrijkingen.

De definitieve waarden voor geslacht, leeftijd en burgerlijke staat in het bestand zijn niet uitsluitend gebaseerd op registergegevens. De informatie uit de registraties wordt gebruikt om vast te stellen of de juiste persoon de vragenlijst invult. Voor elke respondent zijn er vervolgens in het verwerkingsproces twee waarden voor beschikbaar. Ten eerste de gegevens die bij de steekproeftrekking uit de BRP zijn vastgelegd en ten tweede de gegevens uit de vragenlijst. Als de gegevens niet uit de vragenlijst bekend zijn, worden de steekproefgegevens de definitieve gegevens. In alle andere gevallen worden de gegevens uit de vragenlijst gebruikt. Deze gegevens worden gebruikt om te publiceren.

Verder is stratum (zowel grof als fijn) als kenmerk opgenomen en informatie over migratieachtergrond, het type huishouden, de plaats in het huishouden, het gestandaardiseerd besteedbaar huishoudinkomen, de sociaaleconomische categorie waar de respondent toe behoort en regionale informatie. Deze informatie wordt ook voor het corrigeren van de non-respons gebruikt.

### 4.3 Afleiden

CBS werkt in het verwerkingsproces met afleidingen. Het is efficiënt om in het verwerkingsproces vast te leggen hoe een bepaalde afbakening er uit moet zien zodat niet ieder voor zich dat hoeft te bepalen. Gevolg daarvan is ook dat cijfers eenvoudig te reproduceren zijn en gebruikers van het bestand ook dezelfde afbakening gebruiken en daarmee tot dezelfde uitkomsten.

Voor SDI gaat het om een beperkt aantal afleidingen voor publicatievariabelen. Hierbij is zoveel mogelijk aangesloten bij de situatie van voorgaande jaren.

## 5. Responsrapportage

Doordat verschillende groepen personen in de getrokken steekproef meer of juist minder vaak responderen, kan de representativiteit van de steekproef verminderd worden. In dit hoofdstuk wordt de respons, en daarmee ook de non-respons, naar achtergrondkenmerken beschreven. Daarbij gaat het om een vergelijking van de responspercentages per kenmerk. Ook wordt de verdeling van de respons per kenmerk in beeld gebracht.

### 5.1 Afbakenen respons

De statistische respons bestaat uit onderzoekspersonen van wie de vragenlijsten administratief en technisch correct binnengekomen zijn.

In paragraaf 5.2 is het steekproefkader en de steekproef beschreven en in paragraaf 5.3 de selectiviteit in de gerealiseerde respons. In totaal bedroeg de uitgezette steekproef 20.584 steekproefeenheden. Hiervan hebben 6.077 personen de vragenlijst ingevuld, wat neerkomt op een aandeel van 29,5 procent. Dit is een lagere respons dan in 2017. Opgemerkt moet worden dat het steekproefontwerp is gewijzigd en dat de beide edities daarom niet goed te vergelijken zijn.

### 5.2 Steekproef- en responsverantwoording

Niet alle personen die in een steekproef getrokken worden, kunnen ook echt worden benaderd. Het CBS respecteert het wanneer personen bij hun gemeente hebben aangegeven dat NAW-gegevens niet aan derden mogen worden verstrekt, personen kunnen zijn overleden en ook wordt de institutionele bevolking uitgesloten van benadering.

Om de enquêtedruk op de bevolking te spreiden, worden voor veel onderzoeken personen niet aangeschreven indien op dit adres in de voorafgaande twaalf maanden al iemand voor CBS-onderzoek is benaderd. Voor SDI is hier geen rekening mee gehouden gezien het specifieke steekproefkader en de oververtegenwoordiging van mensen met voorzieningen.

Tabel 5.2.1 laat het screeningsresultaat zien van de steekproeven per steekproefportie en tabel 5.2.2 de screeningsuitval naar stratum en gerealiseerde respons, naar mode en stratum.

Tabel 5.2.1 Screeningsresultaat van de steekproeven SDI 2021, naar steekproefportie

	mei	jun	jul	aug	sep	nov	totaal
Getrokken	4.216	4.202	4.208	4.240	4.277	1.840	22.983
Niet beschikbaar voor uitzet, waarvan:	429	411	395	409	398	185	2.227
- in eerdere CBS steekproef	268	268	237	255	209	73	1.310
- institutionele bevolking	4	4	4	5	5	1	23
- met onbekende straatnaam/NAW-gegevens	157	139	154	149	184	111	894
- buiten waarneming SDV (CAPI)	0	0	0	0	0	0	0
Beschikbaar voor uitzet	3.787	3.791	3.813	3.831	3.879	1.655	20.756
Uitdunning	1	5	27	45	94	0	172
Uitzet	3.786	3.786	3.786	3.786	3.785	1.655	20.584

Tabel 5.2.2 Screeningsuitval naar stratum en gerealiseerde respons, naar mode en stratum

Stratum	1	2	3	4	5a	5b	5c	5d	5e	Σ
getrokken steekproef	2.192	1.491	1.617	2.189	2.572	5.483	2.942	1.991	2.506	22.983
screeningsuitval	221	150	144	304	188	640	225	153	202	2.227
steekproefreductie	13	10	8	15	22	39	26	15	24	172
cawi-uitzet	1.958	1.331	1.465	1.870	2.362	4.804	2.691	1.823	2.280	20.584
volledige cawi-respons	294	343	292	277	716	461	500	545	564	3.992
als percentage	15,0%	25,8%	19,9%	14,8%	30,3%	9,6%	18,6%	29,9%	24,7%	19,4%
cati-uitzet	383	495	397	342	764	850	558	542	566	4.897
volledige cati-respons	63	127	166	78	181	103	219	200	119	1.256
als percentage	16,4%	25,7%	41,8%	22,8%	23,7%	12,1%	39,2%	36,9%	21,0%	25,6%
capi-uitzet	496	116	140	395	286	869	207	67	469	3.045
volledige capi-respons	127	38	58	125	101	206	67	26	80	828
als percentage	25,6%	32,8%	41,4%	31,6%	35,3%	23,7%	32,4%	38,8%	17,1%	27,2%
volledige respons	484	508	516	480	999	771	786	770	763	6.077
als percentage t.o.v. cawi-uitzet	24,7%	38,2%	35,2%	25,7%	42,3%	16,0%	29,2%	42,3%	33,5%	29,5%

Bovenstaande informatie tot het volgende totaal overzicht:

Tabel 5.2.3 Steekproef- en responsverantwoording

	totaal
Getrokken steekproef	22.983
Beschikbaar voor uitzet	20.756
Uitzette steekproef	20.584
Kaderfouten	1.293
Netto steekproef	19.291
Non-respons	13.214
Respons	6.077

### 5.3 Selectiviteit van de respons naar kenmerken

Voor de analyse naar de selectiviteit van de respons is gekeken naar het statistische responspercentage en de verdeling van de respondenten en de populatie naar persoons- en achtergrondkenmerken. Het responspercentage voor de analyses in deze paragraaf is gedefinieerd als de statistische respons minus de kaderfouten en cases verwijderd bij verwerking gedeeld door de uitgezette steekproef.

Tabel 5.3.1 Respons SDI naar persoons- en achtergrondkenmerken

	2017	2021
	%	
<b>Totaal</b>	49,2	31,5
<b>Geslacht</b>		
Vrouw	50,4	32,7
Man	47,7	30,1
<b>Leeftijd</b>		
18 tot 25 jaar	37,0	21,9
25 tot 35 jaar	42,1	23,4
35 tot 45 jaar	47,2	27,7



45 tot 55 jaar	50,8	28,8
55 tot 65 jaar	54,9	32,0
65 tot 75 jaar	55,3	45,0
75 jaar en ouder	53,3	37,1
<b>Migratieachtergrond</b>		
Nederlandse achtergrond	54,7	38,8
Westerse migratieachtergrond	32,9	17,6
Niet-westerse migratieachtergrond	48,0	29,0
<b>Positie in huishouden</b>		
Thuiswonend kind	39,7	21,7
Alleenstaande	48,4	30,5
Partner in paar met kinderen	55,2	29,9
Partner in paar zonder kinderen	51,9	40,6
Ouder in eenouderhuishouden	44,0	23,6
Overig lid van een huishouden	30,0	18,8
<b>Landsdeel</b>		
Noord-Nederland	51,6	35,6
Oost-Nederland	53,1	34,7
West-Nederland	45,5	27,8
Zuid-Nederland	52,3	36,0
<b>Stedelijkheid</b>		
Zeer sterk stedelijk	40,2	24,8
Sterk stedelijk	49,3	32,9
Matig stedelijk	56,4	34,9
Weinig stedelijk	54,0	38,0
Niet stedelijk	55,3	36,4
<b>Inkomen</b>		
1 <sup>e</sup> kwintiel	40,7	24,4
2 <sup>e</sup> kwintiel	51,2	32,4
3 <sup>e</sup> kwintiel	55,3	35,6
4 <sup>e</sup> kwintiel	57,4	41,2
5 <sup>e</sup> kwintiel	59,2	42,6

In tabel 5.3.1. zijn de responspercentages opgenomen voor 2021 en 2017. Het is niet zondermeer mogelijk om een vergelijking tussen beide jaren te maken omdat het steekproefontwerp en de strata zijn gewijzigd. Wel blijkt dat het beeld in beide jaren voor de meeste kenmerken vergelijkbaar is.

Vrouwen responderen iets beter dan mannen en de deelname loopt op met de leeftijd, met uitzondering van de groep van 75 jaar en ouder. Migratieachtergrond differentieert sterk naar respons: personen met een Nederlandse achtergrond responderen vaker dan personen met een migratie-achtergrond. Vooral mensen met een niet-westerse migratieachtergrond responderen minder. Thuiswonende kinderen en ouders in eenoudergezinnen responderen minder dan personen die deel uitmaken van een paar en alleenstaanden. Paren zonder kinderen hebben het hoogste responspercentage. Dit wijkt af van het beeld in 2017, toen hadden de paren met kinderen het hoogste responspercentage. De respons is het laagst in het westen van Nederland en in de zeer sterk stedelijke gebieden. Hier is sprake van een overlap, met name in West-Nederland zijn veel zeer sterk stedelijke gebieden. Personen in de hoogste inkomenskwintielen responderen vaker dan personen in de groepen met een lager huishoudinkomen. Deelname neemt tot het derde inkomenskwintiel toe en tussen de hoogste twee inkomensgroepen is er geen verschil.

Tabel 5.3.2 Verdeling respons SDI naar persoons- en achtergrondkenmerken

	2017		2021	
	Steekproef	Respons	Steekproef	Respons
	%			
<b>Totaal</b>	100,0	100,0	100,0	100,0
<b>Geslacht</b>				
Vrouw	55,3	56,6	54,1	56,2
Man	44,7	43,4	45,9	43,8
<b>Leeftijd</b>				
18 tot 25 jaar	7,8	5,8	6,5	4,5
25 tot 35 jaar	12,6	10,8	13,8	10,3
35 tot 45 jaar	20,5	19,6	14,0	12,4
45 tot 55 jaar	22,5	23,1	16,9	15,5
55 tot 65 jaar	15,0	16,7	17,3	17,6
65 tot 75 jaar	8,8	9,9	11,5	16,5
75 jaar en ouder	12,9	13,9	19,8	23,3
<b>Migratieachtergrond</b>				
Nederlandse achtergrond	68,1	75,7	58,5	72,1
Westerse migratieachtergrond	9,9	9,7	13,2	12,1
Niet-westerse migratieachtergrond	22,0	14,7	28,3	15,8
<b>Positie in huishouden</b>				
Thuiswonend kind	6,2	5,0	6,8	4,7
Alleenstaande	35,0	34,3	32,2	31,1
Partner in paar met kinderen	17,9	20,1	22,8	21,6
Partner in paar zonder kinderen	28,5	30,0	26,9	34,6
Ouder in eenouderhuishouden	10,4	9,3	8,5	6,4
Overig lid van een huishouden	2,1	1,3	2,7	1,6
<b>Landsdeel</b>				
Noord-Nederland	11,3	11,9	9,5	10,8
Oost-Nederland	21,0	22,7	18,7	20,6
West-Nederland	46,3	42,7	51,3	45,3
Zuid-Nederland	21,4	22,7	20,5	23,4
<b>Stedelijkheid</b>				
Zeer sterk stedelijk	26,4	21,6	34,3	27,0
Sterk stedelijk	33,6	33,6	29,4	30,7
Matig stedelijk	15,2	17,4	12,5	13,8
Weinig stedelijk	17,9	19,6	17,6	21,3
Niet stedelijk	6,9	7,7	6,2	7,2
<b>Inkomen</b>				
1 <sup>e</sup> percentiel	39,6	32,8	37,0	28,3
2 <sup>e</sup> percentiel	22,7	23,6	26,1	26,5
3 <sup>e</sup> percentiel	15,5	17,4	16,7	18,7
4 <sup>e</sup> percentiel	12,6	14,6	11,4	14,8
5 <sup>e</sup> percentiel	9,6	11,6	8,8	11,7

Tabel 5.3.2 geeft de verdeling naar achtergrondkenmerken voor de netto-steekproef en de respons. Ook dit geeft een beeld van de selectiviteit naar achtergrondkenmerken. Het ligt voor de hand dat

de groepen met een hogere respons iets oververtegenwoordigd zijn ten opzichte van de netto-steekproef.

## 6. Weging

Het onderzoek SDI is een mixed-mode onderzoek in 9 strata. De doelpopulatie bestaat uit alle in Nederland woonachtige personen van 18 jaar of ouder op 1 januari 2020 m.u.v. de institutionele bevolking, i.e. personen in inrichtingen, instellingen of tehuizen, die overigens ook zijn ingeschreven in de Basisregistratie Personen (BRP) met peildatum 13 augustus 2021 (Bkader). Voor SDI wordt het stapelingsbestand als uitgangspunt genomen om de doelpopulatie af te leiden. Het stapelingsbestand geeft inzicht in het gebruik van sociale regelingen en voorzieningen binnen huishoudens. Dit maatwerkbestand laat van iedereen die op 31 december 2019 ingeschreven stond in de BRP zien of hij participatievoorzieningen, jeugdvoorzieningen of WMO-voorzieningen heeft gebruikt. Vervolgens wordt voor het weegkader de doorsnede genomen van het stapelingsbestand en het Bkader.

In paragraaf 6.1 worden de insluitgewichten berekend. In paragraaf 6.2 wordt de werkwijze bij de ontwikkeling van het weegmodel uitgelegd. In paragraaf 6.3 worden de resultaten van de weging beschreven.

### 6.1 Insluitgewichten

Insluitgewichten corrigeren voor verschillen in kansen dat bepaalde groepen in de steekproef worden getrokken. Insluitgewichten corrigeren niet voor vertekeningen die veroorzaakt worden door selectieve non-respons.

Bij de steekproeftrekking van SDI 2021 hebben personen in de verschillende strata verschillende insluitkansen. De insluitkans  $\pi_i^h$  voor persoon  $i$  in stratum  $h$  in de doelpopulatie is gelijk aan

$$\pi_i^h = \frac{n^h}{N^h}$$

waarbij  $n^h$  de steekproefomvang is in stratum  $h$  en  $N^h$  de populatieomvang van stratum  $h$  in het steekproefkader. Het insluitgewicht, de reciproke van de insluitkans, is voor steekproefpersoon  $i$  in stratum  $h$ :

$$D_i^h = \frac{1}{\pi_i^h} = \frac{N^h}{n^h}$$

Een gewenste eigenschap voor de insluitgewichten van de respondenten is dat zij optellen tot het populatietotaal van het steekproefkader. Horvitz-Thompson schattingen van deelpopulaties tellen dan op tot de populatieomvang en kunnen zo vergeleken worden met bekende verdelingen in de populatie. Ook zorgt het ervoor dat de correctiegewichten rond de waarde 1 verdeeld zijn en dit maakt het eenvoudiger in te schatten hoe sterk de weging de correctiegewichten spreidt. Omdat niet elke persoon die in de steekproef zit respondeert, tellen de gewichten gesommeerd over de respondenten nog niet op tot het populatietotaal. Daarom worden voor de respondenten de insluitgewichten gecorrigeerd, zodat ze per stratum optellen tot de bekende populatieomvang. Het gecorrigeerde insluitgewicht van respondent  $i$  in stratum  $h$  is

$$d_i^h = \frac{N^h}{n^h} * \frac{n^h}{r^h} = D_i^h * \frac{n^h}{r^h} = \frac{N^h}{r^h}$$

Tabel 6.1.1 populatieaantallen, steekproefaantallen, responsen en insluitgewichten naar stratum, SDI 2021

Stratum	<i>N</i>	<i>n</i>	$D_i^h$	<i>r</i>	$d_i$
1	609.225	1.958	311,1	484	1258,7
2	521.271	1.331	391,6	508	1026,1
3	272.385	1.465	185,9	516	527,9
4	341.671	1.870	182,7	480	711,8
5a	10.370.528	2.362	4.390,6	999	10.380,9
5b	192.323	4.804	40,0	771	249,4
5c	294.351	2.691	109,4	786	374,5
5d	229.462	1.823	125,9	770	298,0
5e	561.858	2.280	246,4	763	736,4
Totaal	13.393.074	20.584		6.077	

## 6.2 Weegmodellen

### Algemeen

Correctiegewichten corrigeren voor scheefheid als gevolg van nonrespons. Een reden om naar een bepaalde variabele te wegen, is dus dat deze variabele over- of ondervertegenwoordigd is in de respons t.o.v. de doelpopulatie. Om naar zo'n variabele te kunnen wegen, moet deze variabele wel bekend zijn voor iedere respondent en moet het totaal aantal personen in de doelpopulatie bekend zijn voor iedere categorie van de variabele. Bovendien is het gewenst dat de variabele correleert met een of meerdere doelvariabelen. Indien dit niet het geval is, heeft het wegen naar deze variabele namelijk geen effect op de uitkomsten van het onderzoek. Een andere reden om naar een bepaalde variabele te wegen, is dat er naar deze variabele wordt uitgesplitst in publicatietabellen. Door de variabele in dezelfde categorieën op te splitsen in de weging als in de publicatietabellen, wordt ervoor gezorgd dat de randtotalen in de publicatietabellen juist zijn.

### Weegmodel SDI

Het weegmodel dat is gebruikt bij de weging van SDI 2021:

Weegmodel =

$$\text{Leeftijd3} * \text{Herkomst3} + \text{Leeftijd15} * \text{Geslacht2} + \text{TypeHuishouden4} + \text{Inkomen10} + \text{Provincie12} + \text{Stratum9}$$

Het getal achter de variabele geeft het aantal categorieën van de variabele aan.

Voor de weging zijn de achtergrondkenmerken uit het weegkader overgenomen, behalve voor geslacht, dat kenmerk is overgenomen uit het responsbestand. Als voor bepaalde variabelen de waarden in het weegkader onbekend zijn (of ontbreken) is de waarde overgenomen uit het responsbestand (indien de persoon in de respons zat). Als de waarde dan nog altijd onbekend is of ontbreekt dan is die als volgt geïmputeerd. Herkomst wordt "niet-westers". Inkomen wordt gezet op het laagste inkomensdecil. Type huishouden wordt "overig".

De indeling van de variabelen is als volgt:

Leeftijd15: in vijfjaarsklassen (18-19, 20-24, 25-29, ..., 80-84, 85+).

Leeftijd3: 18-34, 35-64, 65+.

Herkomst3: geen migratieachtergrond, niet-westerse migratieachtergrond, westerse migratieachtergrond.

Geslacht2: man, vrouw.

TypeHuishouden4: eenpersoonshuishouden, paren zonder kinderen/overig, paren met kinderen, eenouderhuishoudens.

Inkomen10: gestandaardiseerd besteedbaar huishoudinkomen in decielen.

Provincie12: provincie, indeling van het jaar 2021.

Stratum9: zie paragraaf 3.1.

Het weegmodel is tot stand gekomen door een verzameling kandidaat-weegvariabelen aan te wijzen. Deze bevatte een aantal beschikbare demografische variabelen waarnaar de respons meestal selectief is, zoals leeftijd. Overige kandidaat-weegvariabelen waren variabelen die mogelijk met een of meer van de belangrijke doelvariabelen zouden kunnen correleren en die, uiteraard, beschikbaar waren. De variabelen die in het weegmodel zaten van de vorige editie van SDI (2016) zijn ook allemaal onderzocht, m.u.v. opleidingsniveau. Bij elke weging is er sprake van een beperkt aantal vrijheidsgraden, daarom moet vrijwel altijd een keuze gemaakt worden uit de beschikbare kandidaat-weegvariabelen.

Deze keuze is o.a. gebaseerd op de mate van over- of ondervertegenwoordiging in de respons t.o.v. de doelpopulatie. Na het toevoegen van de variabele met de grootste scheefheid aan het weegmodel, wordt gekeken naar de effecten van het toevoegen van die variabele. Die effecten zijn o.a. (veranderingen in) het designeffect, de minimale celvulling, de spreiding van de gewichten, de populatieschattingen en de nieuwe rangschikking gemaakt van (nog) niet opgenomen kandidaat-weegvariabelen naar scheefheid. Zo'n rangschikking wordt altijd bepaald o.b.v. een statistische toets. Vervolgens wordt een nieuwe variabele toegevoegd aan het model en wordt opnieuw geëvalueerd. De nieuwe rangschikking naar scheefheid is bij die keuze een belangrijke indicator.

### 6.3 Resultaten van de weging

De weging is uitgevoerd met het R-pakket *Brascula*. Er zijn responsgegevens van 6.077 personen gebruikt. De lineaire weegmethode kan negatieve gewichten geven en dat is hier inderdaad het geval. Er zijn drie negatieve gewichten. Negatieve eindgewichten zijn onwenselijk en daarom is het Huang-Fuller algoritme toegepast en dit zorgt ervoor dat alle eindgewichten positief zijn. Idealiter wordt dit algoritme alleen gebruikt bij een laag aantal negatieve gewichten. Daar is hier zeker sprake van. Na begrenzing is het kleinste eindgewicht 48,07. In tabel 2 staan enkele kenmerken van de verdeling van de correctie- en eindgewichten. De minimale celvulling is gelijk aan 48, behorend bij de weegterm *Leeftijd15\*Geslacht2*.

Tabel 6.3.1 Verdeling van de correctie- en eindgewichten

	Correctiegewicht	Eindgewicht
Minimum	0,19	48
Eerste kwartiel	0,71	318
Mediaan	0,94	561
Gemiddelde	1	2204
Derde kwartiel	1,22	1177
Maximum	2,62	27164

Kish<sup>4</sup> 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. Daarom wordt bij een weging vaak de vuistregel gehanteerd dat het maximum van de eindgewichten niet groter moet zijn dan drieënhalf keer de mediaan van de eindgewichten. De getallen onderin tabel 3 laten zien dat voor de weging van SDI dit het geval is in alle strata.

Tabel 6.3.2 Wortel van het maximale designeffect en Kish-factor per stratum

	1	2	3	4	5a	5b	5c	5d	5e
$\sqrt{(1 + V_w^2)}$	1,10	1,04	1,05	1,10	1,06	1,11	1,10	1,05	1,11
Kish factor <sup>5</sup>	0,74	0,77	0,79	0,76	0,78	0,72	0,80	0,77	0,81

<sup>4</sup> Kish, L., 1992. *Journal of Official Statistics*, 183-200.

<sup>5</sup> Hier gedefinieerd als maximale eindgewicht gedeeld door (3,5 x de mediaan van de eindgewichten). Bij een waarde kleiner dan 1,0 wordt dus voldaan aan de vuistregel van Kish.

Het designeffect is maximaal  $1 + V_w^2$ , waarbij  $V_w$  de variatiecoëfficiënt van de eindgewichten is. Deze coëfficiënt is het quotiënt van de standaarddeviatie van de eindgewichten en het gemiddelde van de eindgewichten. De wortel van het maximale designeffect per stratum is opgenomen bovenin tabel 6.3.2. Deze factor ligt tussen de 1,04 en 1,11. Dat is vergelijkbaar met andere wegenen uitgevoerd door het CBS.

Door toepassing van de weging verandert de variantie van de regressieschatter voor doelvariabelen ongeveer met een factor  $(1 + V_w^2)(1 - R^2)$ , waarbij  $R$  een correlatiecoëfficiënt is tussen de betreffende doelvariabele en de hulpvariabelen van het weegmodel, met  $|R| \leq 1$ . Als er geen verband is tussen de doelvariabele en de hulpvariabelen van het weegmodel, dan is  $R$  gelijk aan nul. In dit geval neemt de variantie van de regressieschatter voor de doelvariabele toe met factor  $1 + V_w^2$ . De bijbehorende marge neemt toe met factor  $\sqrt{1 + V_w^2}$ . Als er wel samenhang is tussen de doelvariabele en de hulpvariabelen van het weegmodel, dan is  $R$  ongelijk aan nul. Hoe groter de samenhang, hoe groter  $|R|$ , en dus hoe kleiner de variantie van de regressieschatter. Bij doelvariabelen waarvoor  $(1 + V_w^2)(1 - R^2)$  kleiner dan 1 is, leidt de weging tot variantiereductie.

Na de ontwikkeling van de weegmodellen is van iedere respondent de grootte van het eindgewicht bekend. Daarmee kunnen vervolgens schattingen van doelvariabelen worden gemaakt. In paragraaf 8 is te zien wat het effect van de weging is op de doelvariabelen.



## 7. Andere kwaliteitsindicatoren

Behalve door naar de respons en de verdeling van de respons naar achtergrondkenmerken te kijken, kan ook op andere manieren een indruk gekregen worden van de kwaliteit van het onderzoek.

### 7.1 Missings, geen antwoord op vragen en tellingen

De standaardinstelling van de vragenlijst is dat er geen lege antwoorden, geen “weet niet” en geen “weigert” zijn toegestaan. Dit kan echter wel verschillen per vraag en worden deze categorieën waar relevant alsnog toegestaan. Voor SDI 2021 is per vraag nagegaan wat het percentage onbekenden is, dat wil zeggen dat gekozen is voor de antwoordcategorie “weet niet” of “weigert”.

In het algemeen is het percentage onbekenden laag. Wel zijn er een aantal vragen die respondenten moeilijk kunnen beantwoorden of die ze niet willen beantwoorden. Zo zijn vragen uit het blok Sociale contacten die relatief vaak niet worden beantwoord, namelijk Ik ervaar een leegte om me heen (3,2 procent weigert), Ik mis mensen om me heen (2,1 procent weigert) en Vaak voel ik me in de steek gelaten (2,5 procent weigert). Ook vragen uit het blok Schulden worden relatief vaak niet beantwoord. Op de vraag of het huishouden kan rondkomen van het totale netto huishoudinkomen heeft 5 procent de vraag niet beantwoord en voor de vragen over Schulden geldt hetzelfde.

Voor de vragen over werkondersteuning geldt dat niet alle respondenten deze vragen voorgelegd krijgen. Door routing krijgen alleen degenen die werkzaam zijn en die aangeven een lichamelijke of psychische beperking te hebben deze vragen. Van hen geeft 9,1 procent aan dat ze niet weten of ze een of meerdere vormen van werkondersteuning hebben ontvangen het afgelopen jaar.

Tot slot blijkt ook de vraag of men contact heeft gehad met het Sociaal wijkteam moeilijk te beantwoorden, 10 procent geeft aan dit niet te weten.

### 7.2 Inconsistenties

Hoewel de vragenlijst uitgebreid wordt getest, kan het voorkomen dat er toch inconsistenties binnen de vragenlijst voorkomen of in de afleidingen die op basis van de data worden gemaakt. Ook kan het bijvoorbeeld voorkomen dat er overlap is tussen vragen, wat tot ergernis kan leiden bij respondenten. Door hier op terug te kijken kan de vragenlijst worden verbeterd voor de toekomst. Er zijn in de vragenlijst van SDI 2021 geen inconsistenties geconstateerd.

## 8. Plausibiliteitsanalyse

Om na te gaan of de uitkomsten van een onderzoek plausibel zijn, kan een vergelijking worden gemaakt met de data van eerdere edities en, indien beschikbaar, andere databronnen. In onderstaande worden de resultaten van SDI 2021 vergeleken met de resultaten van de editie van 2017.

Voor deze plausibiliteitsanalyse zijn de belangrijkste doelvariabelen bepaald. Hierbij is gebruik gemaakt van de gewogen resultaten van de antwoorden op de volgende vraagstellingen:

- Kunt u zelfstandig uw huishouden doen?
- Ik heb veel mensen op wie ik volledig kan vertrouwen.
- Deelnemen aan activiteiten van een vereniging, zoals een sport-, zang-, muziek-, toneel-, of hobbyvereniging.
- Hoe tevreden bent u met het leven dat u nu leidt?
- Hoe vaak heeft u contact met één of meer familieleden die niet bij u in huis wonen?
- Hoe vaak heeft u contact met vrienden, vriendinnen of echt goede kennissen?
- Hoe vaak heeft u contact met burenen?
- Bent u in staat om zelfstandig uw sociale contacten te onderhouden?
- Hoe is over het algemeen uw gezondheid?

De antwoorden zijn vergeleken met de resultaten van de vorige editie van dit onderzoek.

### 8.1 Vergelijking oudere jaren met dit onderzoek

In Tabel 8.1.1 worden de gewogen cijfers voor de belangrijkste doelvariabelen voor SDI 2021 van vergeleken met de editie van 2017. Uit deze tabel is af te lezen dat er op een aantal variabelen wat kleine verschuivingen zijn tussen 2017 en 2021. Met name de wekelijkse deelname aan activiteiten van een sport-, zang-, muziek-, toneel-, of hobbyvereniging is lager dan in 2017. Ook het wekelijkse contact met vrienden is iets afgenomen, terwijl het maandelijkse contact juist iets is toegenomen. Bij het contact met burenen is ten opzichte van 2017 een verschuiving te zien van het dagelijkse contact naar wekelijkse contact. Daarbij is ook een kleine afname te zien van het aandeel mensen dat zelden of nooit contact heeft met burenen.

In het licht van de coronapandemie en de daarbij horende maatregelen zijn bovenstaande verschuivingen plausibel.

Tabel 8.1.1 Uitkomsten belangrijkste doelvariabelen SDI 2021 en SDI 2017

		2017	2021
Kunt u zelfstandig uw huishouden doen?	Ja	90,5	91,1
	Nee	9,5	8,9
Ik heb veel mensen op wie ik volledig kan vertrouwen.	Ja	66,7	68,2
	Min of meer	25,1	24,3
	Nee	8,2	7,5
Deelnemen aan activiteiten van een vereniging, zoals een sport-, zang-, muziek-, toneel-, of hobbyvereniging.	Minstens 1 keer per week	43,4	34,5
	Minstens 1 keer per maand, maar niet wekelijks	11,1	9,2
	Minder dan 1 keer per maand	5,3	7,4
	Zelden of nooit	40,2	49,0
Hoe tevreden bent u met het leven dat u nu leidt?	Gemiddelde score (schaal 1-10)	7,7	7,7
	Dagelijks	32,0	32,9

Hoe vaak heeft u contact met één of meer familieleden die niet bij u in huis wonen?	Minstens 1 keer per week, maar niet dagelijks	51,7	50,5
	Minstens 1 keer per maand, maar niet wekelijks	11,1	10,6
	Minder dan 1 keer per maand	2,7	3,0
	Zelden of nooit	2,5	3,0
Hoe vaak heeft u contact met vrienden, vriendinnen of echt goede kennissen?	Dagelijks	26,3	26,5
	Minstens 1 keer per week, maar niet dagelijks	49,6	46,6
	Minstens 1 keer per maand, maar niet wekelijks	16,2	19,0
	Minder dan 1 keer per maand	4,1	4,8
	Zelden of nooit	2,0	1,7
	Ik heb geen vrienden, vriendinnen of echt goede kennissen	1,7	1,5
Hoe vaak heeft u contact met burens?	Dagelijks	19,9	15,4
	Minstens 1 keer per week, maar niet dagelijks	42,0	48,4
	Minstens 1 keer per maand, maar niet wekelijks	18,7	19,3
	Minder dan 1 keer per maand	7,3	7,3
	Zelden of nooit	12,2	9,6
Bent u in staat om zelfstandig uw sociale contacten te onderhouden?	Ja	97,7	97,5
	Nee	2,3	2,5
Hoe is over het algemeen uw gezondheid?	Zeer goed	20,6	20,4
	Goed	53,4	54,1
	Gaat wel	20,2	21,1
	Slecht	4,8	3,9
	Zeer slecht	1,0	0,6

## 8.2 Andere bronnen

Voor enkele variabelen is, net als bij de vorige editie van dit onderzoek, een vergelijking gemaakt met gegevens uit andere CBS-onderzoeken. Daarbij gaat het om algemene gezondheid en contacten met familie, vrienden en burens. Omdat het hoogst behaalde onderwijsniveau een belangrijk persoonskenmerk is, is voor deze variabele ook een vergelijking gemaakt.

In 2021 lag het percentage personen van 18 jaar of ouder met een goede of zeer goede gezondheid op 74,5 procent. Dit aandeel is iets kleiner dan op basis van de gezondheidsenquête, op basis waarvan 76,9 procent aangeeft dat een gezondheid goed of zeer goed is.

Het aandeel dat minstens wekelijks (som dagelijks en minstens wekelijks) contact had met familie is in 2021 vergelijkbaar met dat van de enquête Sociale Samenhang & Welzijn 2020; respectievelijk 83,4 en 83,0 procent van de Nederlanders van 18 jaar of ouder.

De vraag naar het contact met vrienden kent vergeleken met de enquête Sociale Samenhang & Welzijn in SDI een extra categorie, namelijk *Ik heb geen vrienden, vriendinnen of echt goede kennissen*. Het aandeel dat minstens wekelijks (waaronder dagelijks) contact met vrienden had is hierdoor niet goed vergelijkbaar met Sociale Samenhang & Welzijn. Kijk je naar degenen die minstens wekelijks, waaronder dagelijks contact hebben met vrienden dan ligt dit in SDI 2021 iets lager dan op basis van Sociale Samenhang & Welzijn.

Het aandeel dat frequent (minstens wekelijks, waaronder dagelijks) contact met burens had is in SDI 2021 met 63,8 procent groter dan dat op basis van de enquête Sociale Samenhang & Welzijn 2020 (55,5 procent). In de editie van 2017 was het contact met burens op basis van beide bronnen nog wel vergelijkbaar.

Het aandeel met een hbo of universitair diploma is op basis van SDI 2021 groter dan op basis van de Enquête Beroepsbevolking (EBB) 2021. Ook het aandeel met alleen basisonderwijs, vmbo of mbo1 is groter, net als het aandeel onbekend. Het grotere aandeel onbekenden heeft te maken met de relatief grote groep Geen van deze. Op basis van de EBB is het aandeel middelbaar opgeleiden groter, namelijk 38,4 procent vergeleken met 31,0 procent.

De verdere interpretatie en beoordeling van de plausibiliteit wordt verder overgelaten aan de inhoudelijk deskundigen van het SCP.

## 9. Conclusies en aanbevelingen

In dit document is het onderzoek Sociaal Domein Index 2021. Voor zowel de voorbereidingen en het veldwerk als voor de kwaliteit van de uitkomsten worden de belangrijkste conclusies beschreven en aanbevelingen gedaan.

### 9.1 Conclusies

Op basis van onderliggende onderzoeksverantwoording worden de volgende conclusies getrokken:

- De responstargets zijn in absolute aantallen in de meeste strata gehaald, maar helaas niet overal. Ook de extra steekproef en de herverdeling van capi en cati in oktober en november heeft dit niet kunnen voorkomen.
- De gerealiseerde gemiddelde vragenlijstduur werd voor met name cawi en cati overschreden. Het verschilde iets per portie en ging om 2 tot 4 minuten.
- Vanwege de coronamaatregelen is de initiële planning met twee maanden opgeschoven om zo te voorkomen dat het CAPI veldwerk net als in 2020 stopgezet zou moeten worden. Dit heeft goed uitpakkt, het CAPI veldwerk heeft gedurende de beoogde veldwerk doorgang kunnen vinden.
- Gezien het relatief grote aandeel onbekenden (weet niet, weigert of empty) hebben respondenten wat moeite met het beantwoorden van de vragen over het ervaren van een leegte om zich heen, het missen van mensen om zich heen (2,1 procent weigert) en het zich in de steek gelaten voelen. Ook vragen over inkomen, schulden en werkondersteuning worden relatief vaak niet beantwoord. De meeste moeite lijkt men te hebben met de vraag of men contact heeft gehad met het Sociaal wijkteam moeilijk te beantwoorden.
- De analyses laten voor de belangrijkste doelvariabelen veelal plausibele uitkomsten zien. Zeker in het licht van de coronapandemie en de maatregelen die daarbij golden.
- Het hoogste behaalde onderwijsniveau laat wel een ander beeld zien dan de cijfers op basis van de EBB. Met name het verschil voor degenen met een middelbaar onderwijsniveau is groot.

### 9.2 Aanbevelingen

Gezien bovenstaande conclusies worden de volgende aanbevelingen gedaan:

- Communicatie verbeteren om de CAWI-respons te verhogen. Onderzocht zou kunnen worden of een gedifferentieerde aanpak zou kunnen bijdragen.
- Verkorten vragenlijst: hoe korter de vragenlijst hoe groter de kans dat de respondent de vragenlijst volledig zal invullen.
- Nadere analyses van de vragen waar respondenten moeite mee lijken te hebben en nagaan of nog iets veranderd kan worden om het aantal weet niet/wil niet zeggen te verkleinen.
- Overwegen om voor de vraagstelling naar het hoogst behaalde onderwijsniveau aan te sluiten bij een van de standaard CBS vraagstellingen.

## 10. Opgeleverde producten

### 10.1 Microbestanden

Aan het SCP is een tussenbestand opgeleverd op basis waarvan zij onder meer hun voorbereidingen voor hun analyses kunnen uitvoeren. Vervolgens is het definitieve bestand opgeleverd, dit bestand komt ook voor externe onderzoekers beschikbaar via Remote Access (RA). Bij dit bestand komt een codeboek beschikbaar, conform de standaard werkwijze van het CBS. Ook is een zogeheten CCF bestand opgeleverd aan het SCP.

### 10.2 Statistische beveiliging

De microbestanden voor de verschillende afnemers worden vaak nog statistisch beveiligd. De wijze van beveiliging kan verschillen.

Onder statistische beveiliging verstaan we hier het voorkómen dat er inhoudelijke conclusies over herkenbare eenheden kunnen worden getrokken op basis van gepubliceerd of anderszins beschikbaar gesteld CBS-materiaal. Uit de statistische publicaties van het CBS (StatLine-tabellen, nieuwsberichten, wetenschappelijke artikelen) mogen zulke conclusies niet getrokken kunnen worden. Maar ook als het CBS microdata beschikbaar stelt voor wetenschappelijke analyse, moet deze grondregel van de statistiek overeind blijven.

## 11. Referenties

Interne CBS-documentatie:

Budé, M., Steekproefverantwoording SDI 2021, maart 2021.

Budé, M., Weging Sociaal Domein Index 2021, maart 2022.

Janssen, B. en T. Tol, Projectvoorstel Sociaal Domein Index 2021, september 2020.

Knops, J. Onderzoeksdesign Leven 2021 Herstart, maart 2021



## 12. Bijlagen

### Bijlage I – Procedure toekenning incentives

Het verloten van beloningen onder steekproefpersonen gebeurt via een systematisch steekproefontwerp. Hierbij heeft elke steekproefpersoon dezelfde kans om een beloning toegewezen te krijgen. Laat  $n$  de omvang van de te benaderen steekproef zijn. Als voor een onderzoek meerdere steekproefporties worden uitgezet, bijvoorbeeld maandporties, dan is  $n$  de som van de omvang van alle te benaderen porties. Soms is  $n$  niet op voorhand bekend, bijvoorbeeld als tijdens de dataverzameling wordt besloten een extra steekproefportie uit te zetten, of de laatste steekproefportie niet uit te zetten. Dan wordt voor  $n$  een schatting gemaakt. Uitgangspunt voor onderstaande procedure is dat de kans op het winnen van een beloning voor elke steekproefpersoon  $1 / 2.000$  is.

De steekproefpersonen die aanspraak maken op een beloning worden bepaald met door het CBS ontwikkelde software (Manipula) die standaard wordt gebruikt bij steekproefbewerking. Manipula kent standaardfuncties als sorteren (sort) en selecteren van een willekeurig getal uit de uniforme verdeling op het interval (0, 1) (random). Het toekennen van beloningen aan steekproefpersonen gebeurt zonder gebruik te maken van persoons- en adreskenmerken. Elke steekproefportie wordt als volgt behandeld:

1. Van alle voor benadering bestemde steekproefpersonen wordt een bestand met alleen de variabele persoonsidentificatie gemaakt. Dit wordt bestand A genoemd en bevat dus geen persoons- en adreskenmerken.
2. Aan ieder steekproefelement van bestand A wordt een random getal uit de uniforme verdeling op het interval (0, 1) toegekend.
3. Bestand A wordt olopend gesorteerd op dit random getal.
4. Er wordt een volgnummer toegevoegd aan elke steekproefpersoon.
5. Er wordt een stapgrootte bepaald:  $stap = \min(2.000, n)$ . Hierbij is  $n$  de exacte omvang van de te benaderen steekproef of een schatting voor de te benaderen steekproefomvang of een afgesproken minimaal te benaderen omvang.
6. Indien het de eerste steekproefportie van het onderzoek betreft, wordt het startgetal  $k$  random gekozen uit de getallen  $1, 2, \dots, stap$ . Met dit startgetal  $k$  en stapgrootte  $stap$  wordt een systematische steekproef getrokken. Het hoogste volgnummer van de portie wordt opgeslagen voor de eventuele volgende steekproefportie.
7. Indien het niet de eerste steekproefportie van het onderzoek betreft, wordt het hoogste volgnummer van de vorige portie bij het volgnummer in bestand A opgeteld. De systematische steekproef uit de vorige steekproefportie wordt voortgezet. Dit betekent dat personen worden geselecteerd met een volgnummer gelijk aan  $k + i \times stap$  met  $i = 0, 1, 2, \dots$
8. Alle getrokken steekproefelementen van het gesorteerde bestand A krijgen in een hulpvariabele SE\_Winnaar de waarde 1. Voor alle overige steekproefelementen krijgt deze variabele de waarde 0 (nul). Alleen de steekproefelementen met SE\_Winnaar = 1 ontvangen bij deelname aan het onderzoek een beloning, tenzij de respondent heeft aangegeven niet te willen meeloten (Meeloten = nee [Nietloten]).
9. Bestand A wordt aan de hand van de variabele persoonsidentificatie gekoppeld aan de steekproef die de persoons- en adreskenmerken bevat.
10. Na deze koppeling zijn NAWT-gegevens van de winnaars bekend.

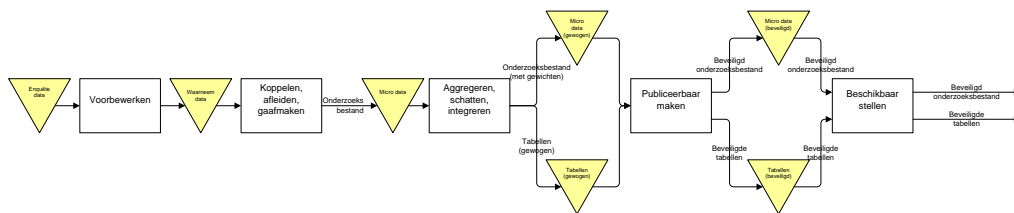
## Bijlage II – Algemene beschrijving dataprocessing

De term dataprocessing wordt gebruikt om de werkzaamheden samen te vatten die nodig zijn om aan de antwoorden die respondenten geven op de vragen in de vragenlijst, plausibele statistische informatie te kunnen ontleen. De verwerking van de data is zoveel mogelijk geautomatiseerd.

Het verwerken is uitgevoerd door de Sector Arbeid, Inkomen en Leefsituatie (SAL) van de Divisie Sociaal-economische en ruimtelijke statistieken (SER) van het CBS.

Het onderstaande figuur 1 geeft het verwerkingsproces van de onderzoeksdata in hoofdlijnen aan. De vijf deelprocessen (weergegeven in de vierkanten blokken) worden in hierna in detail beschreven.

**Figuur 1: Procesmodel Verwerkingsproces**



### 1. Voorbewerken: Van enquête data naar waarneemdata

Voordat enquête data geschikt is om te verwerken zijn er een aantal activiteiten nodig als voorbereiding. Kort gezegd gaat het erom de juiste data (qua cases en variabelen) in het juiste formaat beschikbaar te maken. Dit resulteert in waarneemdata. De waarneemdata is de grondstof voor het 'echte' verwerken.

Binnen het subproces "Voorbewerken" kunnen de volgende processtappen worden onderkend:

#### 1.1. Controleren enquête data: range en routing controle

Met de range- en routingcontrole wordt gekeken of het enquêtebestand voldoet aan de eisen voor de verdere verwerking van het bestand.

Het enquêtebestand wordt hier gedefinieerd als:

- alle data uit een vragenlijst;
- van een of meerdere respondenten;
- van een specifieke mode.

De range- en routingcontrole is specifiek:

- per onderzoek;
- per mode;
- per versie van de vragenlijst.

De eisen van de controles zijn overwegend gebaseerd op de definitie van de vragenlijsten zoals vastgelegd in de gecompileerde vragenlijsten en op het ontwerp van het uniforme datamodel.

Range controles zijn o.a.:

- formaat van variabele;
- variabele heeft juiste naam;
- waardebereik van variabele conform afspraak;
- dubbele sleutels.

Routing controles zijn o.a.:

- indien een blok niet op de route ligt, mogen de variabelen ook geen waarde hebben;

- indien een blok wel op de route ligt moeten deze in principe (tenzij veld niet verplicht) ook een waarde hebben.

## 1.2 Uniformeren

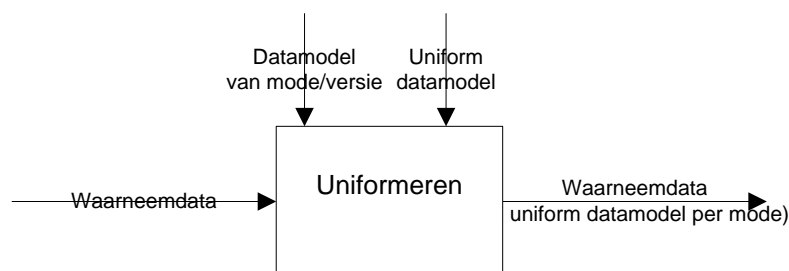
De recordstructuur van enquêtedata kan mode specifiek zijn. Dit komt omdat de vragenlijst mode specifiek is. Daarmee is ook de metadata mode specifiek. Het kan ook voorkomen dat gedurende het onderzoek er voor een mode meerdere versies van een vragenlijst worden gebruikt (eventueel ook met een verschil in variabelen). Dus ook per versie kunnen er verschillen qua recordstructuur zijn.

Doel bij uniformeren is om tot één uniforme recordstructuur te komen. De uniforme recordstructuur is onderzoekspecifiek en geldt in principe voor de duur van het hele onderzoek. Het bepalen van de uniforme recordstructuur<sup>6</sup> en het beschrijven van de bijbehorende metadata is een ontwerpactiviteit. Dit betekent dat de metadata van het uniforme datamodel vooraf gedefinieerd is; dus voordat het uniformeren daadwerkelijk plaats vindt. Bij het uniformeren van de data hoeft er in principe dus geen metadata meer te worden aangepast; als je de data in het uniforme model zet is de metadata automatisch correct.

Pre-conditie:

- het mode-specifieke (en vragenlijstversie specifieke) datamodel moet bekend zijn;
- het uniforme datamodel moet bekend zijn.

**Figuur 2: Procesmodel Uniformeren**

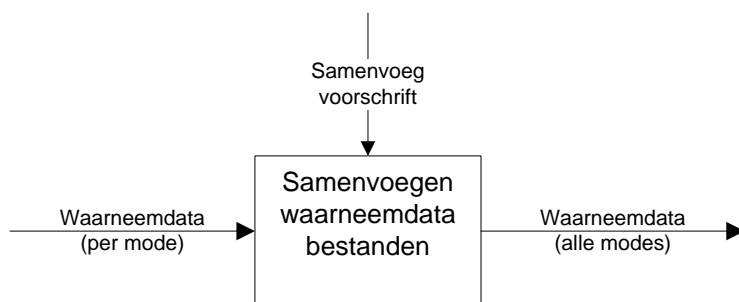


## 1.3 Controleren enquêtedata: range en routing controle

De waarneemdatabestanden van de diverse modes worden samengevoegd tot één fysiek bestand. Per case moet wel duidelijk blijven uit welke mode het record komt.

<sup>6</sup> en trouwens ook de mode-specifieke datamodellen

**Figuur 3: Procesmodel Samenvoegen waarneemdata**



## 2. Koppelen, afleiden en gaafmaken

Bij het koppelen, afleiden en gaafmaken wordt de data verrijkt met o.a. data uit de steekproef, registerdata en andere bronnen. Tevens wordt de respons afgebakend, vindt imputeren en gaafmaken plaats en worden variabelen afgeleid (inclusief coderen). Dit resulteert in een onderzoeksbestand.

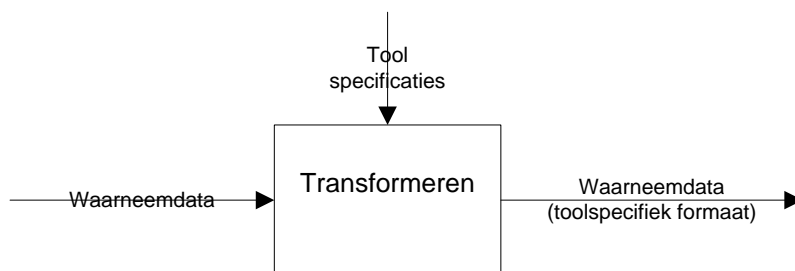
### 2.1 Transformeren

Doel van deze procesactiviteit is de waarneemdata transformeren naar een vorm die gebruikt kan worden bij de vervolg activiteiten. De waarneemdata is in een bepaald technisch formaat, in dit geval Blaise. Bij de vervolgstappen van het verwerkingsproces wordt SPSS als verwerkingstool gebruikt. Daartoe moet het technische formaat van de data worden aangepast (van Blaise via ASCII naar SPSS).

T.b.v. SPSS is transformeren bijvoorbeeld:

- Dichitomiseren (men werkt in het onderzoeksbestand niet met meervoudige antwoorden).
- Labels aanbrengen: Op basis van de metadata uit de vragenlijst worden variabelen en value labels gegenereerd, die worden gecombineerd met dit SPSS-systeembestand. Deze labels vormen de beschrijving van de data.
- Imputatie routing: Respondenten hoeven in de vragenlijst alleen die vragen te beantwoorden die op hun situatie van toepassing zijn. Vragen die door de respondent niet beantwoord hoefden te worden, gaan als blanco naar ASCII en krijgen vervolgens in SPSS de waarde SYSMIS. Dit betekent dat het SPSS-systeem de waarden als "n.v.t." beschouwd bij de uitvoering van statistische analyses.

**Figuur 4: Procesmodel Transformeren**



## 2.2 Verrinnen

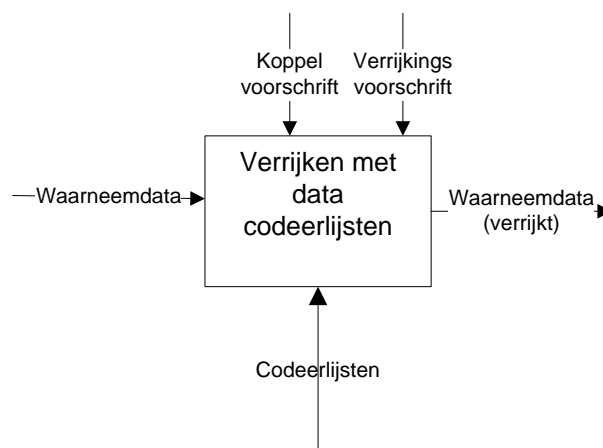
Binnen de verwerkingsprocessen gebeurt “het verrinnen” alleen t.b.v. het koppelen met registerdata en dus niet om met geanonimiseerde data in het verwerkingsproces te werken. Om waarneemdata te kunnen verrijken met registerdata dient de waarneemdata eerst verrind te worden. Dit betekent dat voor iedere persoon in de waarneemdata een betekenisloze identificerende variabele wordt bepaald (genaamd “RINPersoon”). Dit nummer is gebaseerd op data uit het BRP.

De te koppelen waarneemdata wordt daartoe geleverd aan CBK. Deze koppelen de data aan het Centrale Koppelbestand Personen (CKP). Sector CBK levert de verrinde data vervolgens terug aan het verwerkingsproces. Voor het leggen van een koppeling zijn het Burger Service Nummer (BSN) en/of de combinatie van geboortedatum, geslacht en adres nodig. Een geslaagde koppeling betekent in concreto dat aan het originele record RINPersoon en RINPersoonVolgNr uit het CKP worden toegevoegd. Hiermee is de desbetreffende persoon in het CKP identificeerbaar. Naast de CKP persoonidentificatie wordt nog informatie over de koppeling aan het record toegevoegd.

## 2.3 Verrijken met data codeerlijsten

De waarneemdata wordt hier gekoppeld aan verschillende codeerlijsten. Hier wordt bijvoorbeeld op basis van de viercijferige postcode de gemeentecode bepaald. Op basis van de gemeentecode worden de bovengemeentelijke regionale indelingen bepaald.

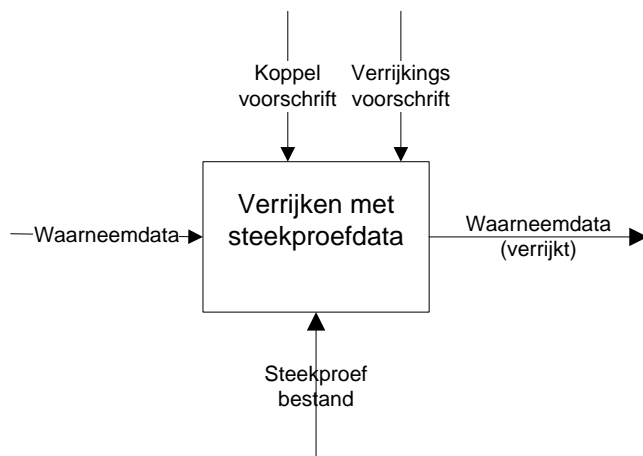
**Figuur 5 : Procesmodel Verrijken met data codeerlijsten**



## 2.4 Verrijken met steekproefdata

T.b.v. non-respons analyses kan de waarneemdata worden verrijkt met de complete steekproef met voor elk element de voor uitdunning van de adressensteekproef gebruikte variabelen (alle adresgegevens bijvoorbeeld), het startgewicht en een eindresultaat (bijvoorbeeld: uitgedunde GBA65plus, niet uitgezet door regiomanager, geen woonadres, leegstand, niemand thuis, taalbarrière, weigering, enzovoorts). Op deze manier kan een betere (uitgebreidere) non-respons analyse naar allerlei achtergrondkenmerken gemaakt worden. Bij het koppelen kan het voorkomen dat er een steekproefeenheid is waarvoor (nog) geen waarneemdata is. En tevens waarneemdata waarvoor geen steekproefeenheid is. In het laatste geval is blijkbaar een verkeerde respondent bevestigd.

**Figuur 6: Procesmodel Verrijken met steekproefdata**

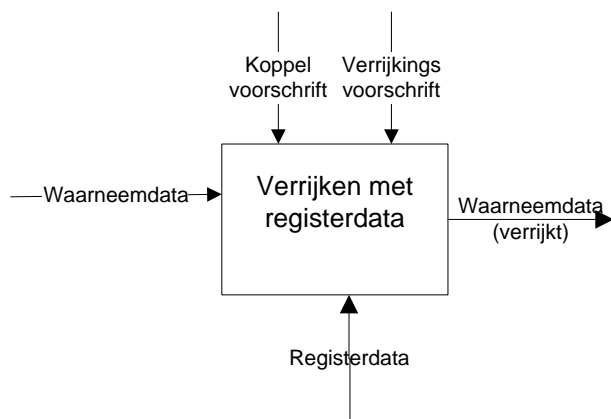


## 2.5 Verrijken met registerdata

Het koppelen met de registerdata gebeurt op basis van het RIN-nummer. Veel gebruikte registers zijn de GBA, de Polisadministratie en het UWV-Werkbedrijf.

De waarneemdata wordt uit de registers verrijkt met o.a. type huishouden, geboorteland persoon en van diens vader en moeder en afleidingen daarop (BRP), bron inkomen en hoogte inkomen (Polisadministratie), provincie, inschrijfduur (is een afleiding), werkend (UWV-Werkbedrijf).

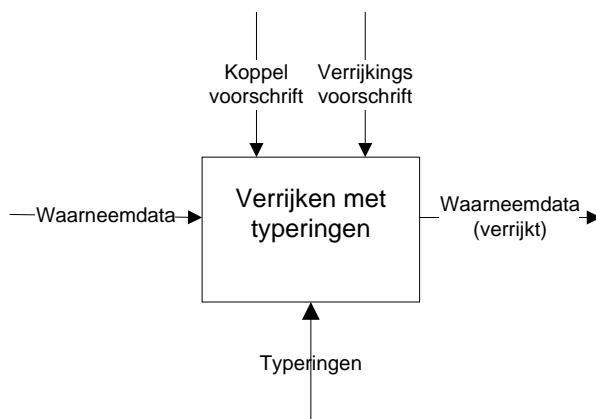
**Figuur 7: Procesmodel Verrijken met registerdata**



## 2.6 Verrijken met typeringen

Deze stap kan pas worden uitgevoerd als de typeringen beschikbaar zijn. Afhankelijk van het specifieke onderzoek heeft het typeren een bepaalde doorlooptijd, waardoor deze data (meestal) niet meteen beschikbaar is.

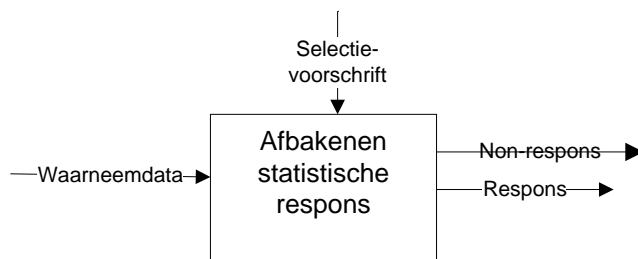
**Figuur 8: Procesmodel Verrijken met typeringen**



### 2.7 Afbakenen statistische respons

Alleen statistische respons wordt meegenomen in de verdere verwerking. Wat wel/niet tot respons behoort, staat in een voorschrift. In deze activiteit wordt o.b.v. het voorschrift de respons bepaald. Hóe de non-respons gescheiden wordt van de respons is vanuit logisch oogpunt niet relevant; dit kan bijvoorbeeld door de records fysiek van elkaar te scheiden, maar kan bv ook door met indicatoren te werken. Voorwaarde is dat de non-respons op een gegeven moment beschikbaar is ten bate van non-respons analyses.

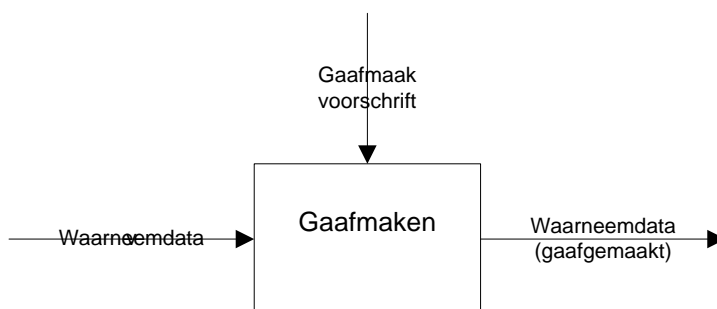
**Figuur 9: Procesmodel Afbakenen statistische respons**



### 2.8 Gaafmaken (micro)

Gaafmaken is het opsporen en corrigeren van foutieve gegevens in de waarneemdata. Bij micro gaafmaken vinden zowel de controles als de correcties plaats op microniveau. Voorbeelden van voorkomende fouten zijn: het geboortjaar klopt niet of is onwaarschijnlijk, een respondent rapporteert in euro's in plaats van in duizenden euro's (of omgekeerd), of de winst van een bedrijf is niet gelijk aan het verschil tussen baten en lasten. Gaafmaken gebeurt op basis van, bij het ontwerp bepaalde, voorschriften.

**Figuur 10: Procesmodel Gaafmaken (micro)**



## 2.9 Imputeren

Imputeren is het bepalen en introduceren van een (nieuwe) waarde op een plaats waar een waarde ontbreekt of op 'onbekend' (ontbrekend) is gezet.

Bij enquêtes komt het voor dat respondenten op één of meer vragen geen antwoord geven, terwijl dit wel van ze wordt verlangd. Men spreekt dan van item-nonrespons (of partiële nonrespons) en van (ten onrechte) ontbrekende waarden (missing values). Redenen om een vraag niet te beantwoorden zijn het niet kunnen of willen geven van het antwoord. Op ingewikkelde of moeilijk te begrijpen vragen kan men vaak geen antwoord geven, op gevoelige vragen wil men het dikwijls niet. Ook bij registers kunnen gegevens ontbreken die het CBS wel had willen hebben.

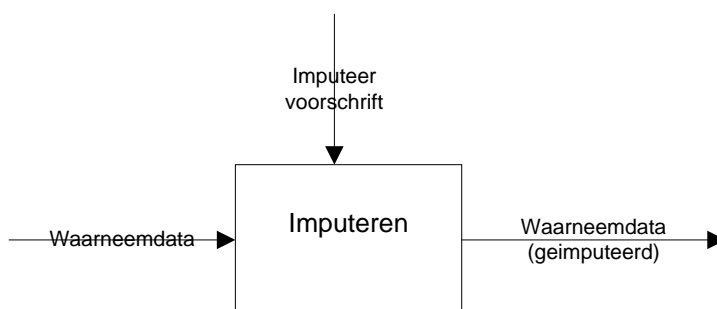
Er zijn een aantal manieren om met ontbrekende waarden om te gaan. Eén daarvan is imputeren van een geldige waarde voor de ontbrekende waarde in het databestand.

Een alternatief voor imputeren is om het achterwege te laten. De ontbrekende waarden blijven dan onbekend. Redenen om te imputeren, in plaats van het veld leeg te laten, zijn:

1. het verkrijgen van een 'volledig' (geheel gevuld) databestand;
2. verhoging van de kwaliteit van het micro-bestand en/of van de parameterschattingen.

We maken verder onderscheid tussen imputeren en afleiden. Bij het afleiden van variabelen worden nieuwe variabelen gecreëerd als functie van in het bestand reeds bestaande variabelen. Bij imputeren worden ontbrekende waarden op een bestaande variabele gecreëerd. Imputeren gebeurt op basis van, bij het ontwerp bepaalde, voorschriften. Steeds dient te worden vastgelegd dat een waarde is geïmputeerd.

**Figuur 11: Procesmodel Imputeren**





## 2.10 Afleiden

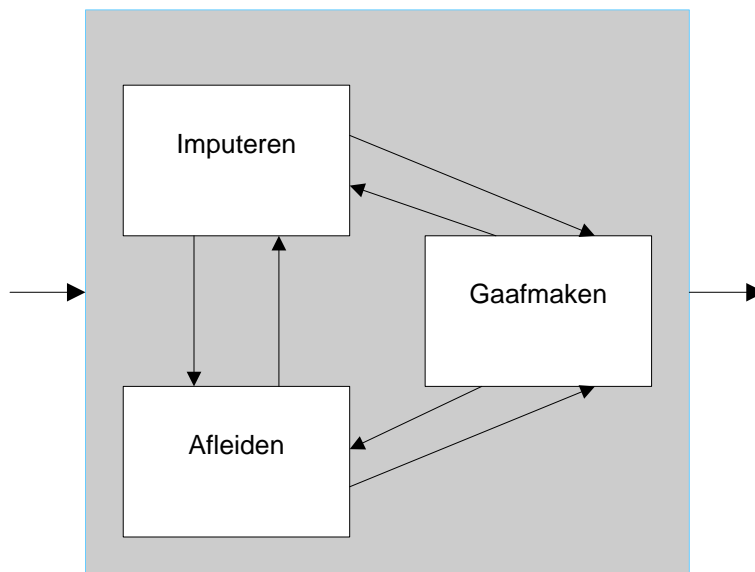
Met afleiden wordt hier bedoeld het creëren van nieuwe variabelen als functie van in het bestand reeds bestaande variabelen.

Coderen is ook een vorm van afleiden. Het coderen van een vraag is het (keuze)proces waarbij een beslissing wordt genomen om een antwoord te interpreteren in termen van een voor gedefinieerde verzameling mogelijke antwoorden. Een dergelijke keuze wordt, tijdens een interview of bij het invullen van een vragenformulier, vaak gedaan door respondenten, al of niet met hulp van een interviewer. Soms echter wordt deze keuze achteraf gedaan, op het CBS en zonder de aanwezigheid van een respondent, door codeurs of typeurs.

## 2.11 Gaafmaken, imputeren en afleiden: samenhang

De logische volgorde is eerst gaafmaken, dan imputeren en dan variabelen afleiden. Echter, gaafmaken, imputeren en afleiden zijn geen activiteiten die voor een dataset en individuele case sequentieel verlopen. De activiteiten worden per variabele of set van variabele doorlopen, waarna de activiteiten voor andere variabelen worden doorlopen. Eventueel kan de volgorde van de activiteiten ook anders zijn: bv eerst imputeren dan gaafmaken. Dit is afhankelijk van de specifieke regels die van toepassing zijn binnen het onderzoek.

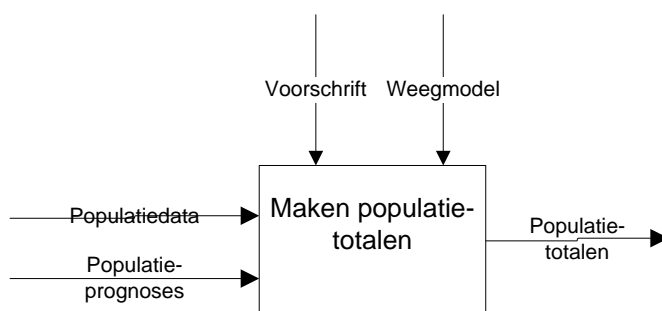
**Figuur 12: Procesmodel Gaafmaken, imputeren en afleiden: Samenhang**



## 2.12 Maken Populatietotalen

Voor het wegen zijn populatietotalen nodig. Het kan zijn dat de populatietotalen zelf geschat moeten worden aangezien de data bij de taakgroep Demografie niet altijd voldoende gedetailleerd en soms onvoldoende actueel zijn. Populatietotalen worden bepaald op het totale register, niet op de met register verrijkte waarneemdata.

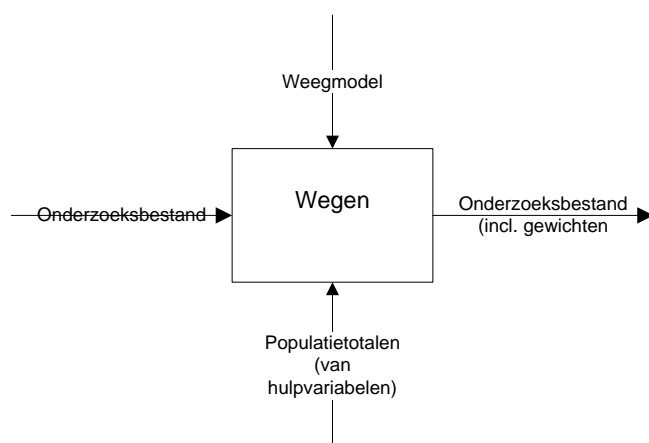
**Figuur 13: Procesmodel maken populatietotalen**



### 2.13 Wegen

Het wegen is de activiteit waarbij weegfactoren worden bepaald. In het weegmodel staat beschreven hoe het wegen moet plaatsvinden. Bij het wegen wordt de verdeling van variabelen in de steekproef in overeenstemming gebracht met de verdeling daarvan in de populatie. Daartoe wordt aan iedere case een gewicht toegekend. Ten bate van het wegen zijn populatietotalen (hulpvariabelen) nodig op persoonsniveau. Tevens kan gewogen worden naar verdelingen waaraan het responsproces idealiter moet voldoen, bijvoorbeeld: elke dag evenveel respons of een gelijk responspercentage per mode. De weging resulteert in één of meerdere ophoogfactoren (afhankelijk van het aantal entiteiten waarvoor gewogen wordt).

**Figuur 14: Procesmodel Wegen**



## 3. Publiceerbaar maken

Bij het publiceerbaar maken worden de tabellen en het gewogen onderzoeksbestand statistisch beveiligd. Binnen het subproces “publiceerbaar maken” kunnen de hieronder genoemde processtappen plaatsvinden.

### 3.1 Maken micro output

Op basis van het onderzoeksbestand wordt de output (microbestanden) voor bijvoorbeeld de externe klant, SAL-SAD/SIL, CvB, DANS en/of Eurostat gemaakt. De output voor de diverse afnemers bevatten meestal slechts een deelverzameling van de variabelen uit het onderzoeksbestand. Deze deelverzamelingen voor de verschillende afnemers worden in deze stap gemaakt. SAL-SAD/SIL

krijgen meestal het gehele onderzoeksbestand ten behoeve van het maken van publicaties. Hiervoor hoeven geen extra stappen gezet te worden.

### **3.2 Statistisch beveiligen microdata**

De microbestanden voor de verschillende afnemers worden vaak nog statistisch beveiligd. De wijze van beveiliging kan wel verschillen. Onder statistische beveiliging verstaan we hier het voorkómen dat er inhoudelijke conclusies over herkenbare eenheden kunnen worden getrokken op basis van gepubliceerd of anderszins beschikbaar gesteld CBS-materiaal. Uit de statistische publicaties van het CBS (StatLine-tabellen, web-artikelen, persberichten, wetenschappelijke artikelen) mogen zulke conclusies niet getrokken kunnen worden. Maar ook als het CBS microdata beschikbaar stelt voor wetenschappelijke analyse, moet deze grondregel van de statistiek overeind blijven.

### **3.3 Statistisch beveiligen standaardtabellen**

De tabellen (t.b.v. de Externe klant, ministeries, Statline, CvB, DANS en/of Eurostat) dienen ook statistisch beveiligd te worden.

## **4. Beschikbaar stellen**

De statistisch beveiligde producten worden vervolgens als output geleverd (beschikbaar gesteld) aan diverse belanghebbenden, waaronder externe opdrachtgevers, verschillende ministeries, Eurostat, DANS, Centrum voor Beleidsstatistiek. Binnen het subproces "Beschikbaar stellen" kunnen de hieronder genoemde processtappen plaatsvinden.

### **4.1 Leveren microbestanden**

Dit is de activiteit waarbij microdata daadwerkelijk wordt geleverd. Aan welke partijen geleverd wordt is afhankelijk van het betreffende onderzoek. Er worden voor het onderzoek CV twee bestanden opgeleverd: een responsbestand met data voor de ene vragenlijstvariant en een responsbestand met data voor de andere vragenlijstvariant.

### **4.2 Leveren standaardtabellen**

Dit is de activiteit waarbij de tabellen daadwerkelijk worden geleverd. Aan welke partijen geleverd wordt is afhankelijk van het betreffende onderzoek.

