



Centraal Bureau voor de Statistiek

Tijdsbestedingsonderzoek 2011/2012

Onderzoeksdocumentatie

Versie 1.0

Inhoudsopgave

1	Inleiding.....	3
2	Doel, achtergrond en geschiedenis van het onderzoek	4
2.1	Doel	4
2.2	Geschiedenis van het onderzoek	4
3	Dataverzameling	5
3.1	Registers.....	5
3.2	Steekproefontwerp	5
3.3	Benaderingsstrategie	10
3.4	Vragenlijsten	15
3.5	Veldwerkresultaten.....	17
3.6	Coderen en vertoetsen	19
3.7	Bestandsopbouw, controles en oplevering	20
4	Verwerken (data processing)	22
4.1	Proces.....	222
4.2	Controleren en uniformeren waarneemdata	299
4.3	Verrijken.....	30
4.4	Afbakenen statistische respons	30
4.5	Gaafmaken (micro)	31
4.6	Imputeren	34
4.7	Afleiden	34
4.8	Controleren en corrigeren	334
4.9	Wegen	39
4.10	Maken tabellen, aggregeren	39
4.11	Gaafmaken (macro)	39

4.12	Controleren tabellen: plausibiliteitanalyse.....	39
4.13	Statistisch beveiligen.....	40
4.14	Output van verwerken.....	41
5	Publiceren.....	41

1 Inleiding

In dit document wordt verslag gedaan van het Tijdsbestedingsonderzoek (TBO) 2011/2012. De voorbereiding, dataverzameling en verwerking van de gegevens worden gedocumenteerd. Dit document dient als naslagwerk voor onderzoekers die meer inzicht wensen in de achtergronden van het onderzoek en in het tot stand komen van de analysebestanden.

2 Doel, achtergrond en geschiedenis van het onderzoek

2.1 Doel

Het Tijdsbestedingsonderzoek (TBO) is een 5-jarlijks onderzoek dat in 2011/2012 in een samenwerkingsproject is uitgevoerd door het Sociaal Cultureel Planbureau (SCP) en het Centraal Bureau voor de Statistiek (CBS). Het doel van het onderzoek is informatie te verzamelen over de wijze waarop mensen hun tijd doorbrengen, welke activiteiten ze verrichten, hoeveel tijd ze eraan besteden en met wie ze die activiteiten uitvoeren. Het TBO ligt aan de basis van veel onderzoek uit het werkprogramma van het SCP. Door het CBS worden de data gebruikt in het kader van de speerpunten Sociale Samenhang en Welzijn.

2.2 Geschiedenis van het onderzoek

Het TBO is door SCP en partners van 1975 tot 2005 met vijfjaarlijkse regelmaat uitgevoerd door onderzoeksbureau Intomart. In 2006 is het onderzoek voor het eerst geheel verricht conform de daartoe door Eurostat opgestelde 'Hetus' (Harmonized European Time Use Studies) richtlijnen. In dat jaar is het onderzoek gehouden onder respondenten van het Mobiliteitsonderzoek Nederland en werd het veldwerk uitgevoerd door onderzoeksbureau Social Data. In 2011 is het onderzoek voor het eerst uitgevoerd door het CBS.

3 Dataverzameling

3.1 Registers

Aan het TBO wordt informatie toegevoegd vanuit het Postaal WijkRegister (PWR), vanuit het CBS Statline tabel Gebieden in Nederland (GIR) 2011, vanuit de Gemeentelijke BasisAdministratie (GBA) en vanuit de integrale CBS inkomensbestanden 2010.

Deze verrijking van het TBO wordt uitgebreid beschreven in het document "Koppel_en_verrijgingsregels_TBO_2011_2012.pdf" en wordt in dit document kort beschreven in paragraaf 4.4.

3.2 Steekproefontwerp

3.2.1 Doelpopulatie en steekproefontwerp

De doelpopulatie voor het TBO bestaat uit de in Nederland woonachtige personen van tien jaar of ouder, die deel uitmaken van particuliere huishoudens en zijn ingeschreven in de Gemeentelijke Basisadministratie (GBA). Voor het TBO is de institutionele bevolking, dat zijn personen in inrichtingen, instellingen of tehuizen, niet geënquêteerd. Maandelijks is een gestratificeerde tweetrapssteekproef van personen getrokken uit de GBA. Hierbij is gestratificeerd naar Corop × interviewregio. In de eerste trap zijn deelgemeenten geselecteerd. Een deelgemeente is een gemeente zoals die bestond in 1990. Hierop zijn enkele uitzonderingen: gemeenten die in 1990 klein waren zijn samengevoegd met naburige gemeenten. Deelgemeenten die met kans 1 zijn geselecteerd, de zogenoemde zelfselecterende deelgemeenten, zijn elke maand in de steekproef terecht gekomen. Zij zijn dan ook gedurende het hele jaar waargenomen. Niet-zelfselecterende gemeenten zijn niet elke maand in de steekproef terecht gekomen. Hoe groter de deelgemeente, hoe vaker de deelgemeente in een maandsteekproef zit. Wel wordt altijd voor goede spreiding over het jaar gezorgd. Als een deelgemeente bijvoorbeeld zes keer in een maandsteekproef zit, dan zit deze deelgemeente alleen in de oneven maanden of alleen in de even maanden in de steekproef. De deelgemeenten zijn via een systematisch steekproefontwerp getrokken met kansen evenredig aan het aantal inwoners. In de tweede trap is een enkelvoudig aselechte steekproef van personen getrokken uit de deelgemeenten die in de eerste trap zijn geselecteerd.

Het gewenste aantal te trekken personen per deelgemeente, de zogenoemde clusteromvang, is vooraf bepaald en is voor iedere deelgemeente gelijk. Voor zelfselecterende deelgemeenten is het aantal te trekken personen aangepast tot het product van de steekproeffractie (quotiënt van steekproefomvang en

populatieomvang) en het aantal personen van tien jaar of ouder in de betreffende deelgemeente. Voor alle andere deelgemeenten kan het aantal getrokken personen iets afwijken van de clusteromvang. Dit komt door afrondingen.

Hoe groot de clusteromvang c wordt gekozen hangt af van de gewenste nauwkeurigheid en de kosten voor aan-huis enquêtering. Als c klein is, dan wordt het aantal te trekken deelgemeenten groot met als gevolg hogere reiskosten voor de interviewers. Bij een grote c kan er een cluster-effect optreden, waardoor de foutmarges groter worden. Dit is vooral het geval als er een samenhang bestaat tussen de antwoordpatronen van de bewoners van eenzelfde deelgemeente.

Bij een verwachte responskans voor de dagboekjes van $0,58 \times 0,78 \times 0,9 = 0,407$ op de uitgezette steekproef voor de startvragenlijst, was de geschatte omvang van de uit te zetten steekproef $2.000 / 0,407 = 4.914$ personen. Omdat de waarneemperiode voor de startvragenlijst startte op 14 maart 2011, hadden interviewers ongeveer twee derde van maart beschikbaar om de startvragenlijst af te nemen voor de maartsteekproef. Daarom is ervoor gekozen om de omvang van de maartsteekproef gelijk te nemen aan twee derde van elke andere maandsteekproef. Bij deze keuze was de geschatte omvang van de maartsteekproef 281 personen en van elke andere maandsteekproef 422 personen. De geschatte uit te zetten steekproef bevat dus 4.923 personen. In paragraaf 3.2.4. wordt de gerealiseerde steekproef verantwoord en is terug te vinden dat er uiteindelijk 4.976 personen zijn uitgezet.

3.2.2 Designeffect

Op basis van een gegeven clusteromvang kan voor bepaalde variabelen het designeffect worden berekend. Dit designeffect geeft aan met welke factor de variantie van de schatter voor de betreffende variabele toeneemt door het toegepaste steekproefontwerp vergeleken met een enkelvoudig aselect ontwerp. Daarbij wordt uitgegaan van hetzelfde aantal waarnemingen. Wanneer bijvoorbeeld een variabele bij het toegepast steekproefontwerp een designeffect van 1,25 heeft, dan hebben 2.000 responsen dezelfde precisie als $2.000 / 1,25 = 1.600$ responsen bij een enkelvoudig aselect steekproefontwerp. Het berekenen van het designeffect leidt dus tot een belangrijke bijdrage aan het bepalen van de nauwkeurigheid van schattingen.

Met enquêtedata van het TBO 2006 is het designeffect voor achttien variabelen geschat op basis van 1.873 respondenten. Hierbij is uitgegaan van de gemeentelijke indeling van 1990. Het aantal deelgemeenten dat in de uit te zetten steekproef voorkomt, het gemiddelde, minimale en maximale designeffect van de betreffende

variabelen zijn voor verschillende clusteromvang en weergegeven in tabel 1. Het totaal aantal deelgemeenten in Nederland is 652.

Tabel 1 Designeffect naar clusteromvang, bij 2.000 respondenten

c	deelgem	designeffect		
		gemiddeld	min	max
16	226	1,89	1,51	1,38
12	279	1,46	1,25	1,13
8	362	1,03	1,00	0,95
4	517	0,87	0,81	0,68

Uit tabel 1 volgt dat het designeffect voor de onderzochte variabelen ongeveer 1 is bij clusteromvang 8. In dat geval is het aantal deelgemeenten in de steekproef 362. Bij een kleinere clusteromvang neemt het designeffect af naar gemiddeld 0,87 bij c = 4. Hierbij hoort een effectieve responsomvang van $2.000 / 0,87 = 2.299$ personen. Bij een grotere clusteromvang neemt het designeffect toe naar gemiddeld 1,89 bij c = 16 met een effectieve responsomvang van $2.000 / 1,89 = 1.058$ personen. In onderling overleg tussen CBS en SCP is voor clusteromvang 8 gekozen.

3.2.3 Verhuizingen

Hoewel steekproeven zo kort mogelijk voor aanvang van de waarneemperiode worden getrokken is niet te voorkomen dat het adres in de GBA niet juist blijkt te zijn omdat de steekproefpersoon is verhuisd. Wordt door de interviewer aan-huis geconstateerd dat een steekproefpersoon niet meer woonachtig is op het opgegeven adres, dan wordt dit geregistreerd. Voor steekproefpersonen van wie een verhuizing binnen Nederland is vastgesteld, zijn opnieuw de NAW-gegevens opgevraagd bij de GBA. Als er een ander dan het oorspronkelijke adres is verkregen dan zijn de betreffende steekproefpersonen vervolgens opnieuw benaderd op het nieuwe adres.

Voor het TBO zijn alleen verhuizers die bij de benadering voor het startinterview geconstateerd werden, herbenaderd. Als er tussentijds een verhuizing plaatsvond, kon de interviewer het huishouden volgen. In verband met kleine hoeveelheden verhuizers, zijn deze zoveel mogelijk opgespaard. Om toch nog enige spreiding over de veldwerkperiode te bewerkstelligen zijn er op twee momenten verhuizers herbenaderd. In verband met de doorlooptijd van het opnieuw opvragen van

gegevens bij het GBA en de doorlooptijd binnen één hele portie voor het TBO, was er geen tijd om de verhuizers uit de 11^e en 12^e portie her te benaderen.

3.2.4 Steekproefverantwoording

In tabel 2 wordt de steekproefverantwoording weergegeven vanaf het moment van trekking tot het daadwerkelijk uitzetten van de steekproef. In de veldwerkperiodes 201110 en 201201 zijn hierbij ook de verhuizers uit de voorgaande periodes meegenomen. In totaal zijn er 4.976 personen uitgezet voor de waarneming.

Tabel 2 Steekproefverantwoording TBO per veldwerkperiode, 2011-2012

	<i>Totaal</i>	201103	201104	201105	201106	201107/08	201109	201110*	201111	201112	201201*	201202
Getrokken steekproef	5.546	308	460	460	460	920	460	475	460	528	509	506
Onbruikbare steekproef	569	27	38	38	38	76	38	38	38	106	48	84
In eerdere CBS-steekproef	241	4	18	10	17	32	14	21	22	43	30	30
IIT-bevolking	1	1	0	0	0	0	0	0	0	0	0	0
Met onbekende straatnaam	0	0	0	0	0	0	0	0	0	0	0	0
Adressen buiten gebied waarneming DV	14	0	1	9	0	0	1	1	0	0	2	0
Geen levering aan derden	111	8	6	6	6	17	6	9	6	15	16	16
Te veel getrokken adressen	215	14	13	13	15	27	17	7	10	48	0	51
Bruikbare steekproef	4.977	281	422	422	422	844	422	437	422	422	461	422
Uit te zetten steekproef	4.977	281	422	422	422	844	422	437	422	422	461	422
Uitgezet door AxIs	4.772	273	381	382	402	829	420	419	414	396	445	411
Gewijzigd door RM	1.106	80	105	103	85	182	86	100	72	91	117	85
Niet gewijzigd door RM	3.666	193	276	279	317	647	334	319	342	305	328	326
Uitgezet door RM	204	7	41	40	20	15	2	18	8	26	16	11
Niet uitgezet	1	1	0	0	0	0	0	0	0	0	0	0
Uitgezette steekproef	4.976	280	422	422	422	844	422	437	422	422	461	422

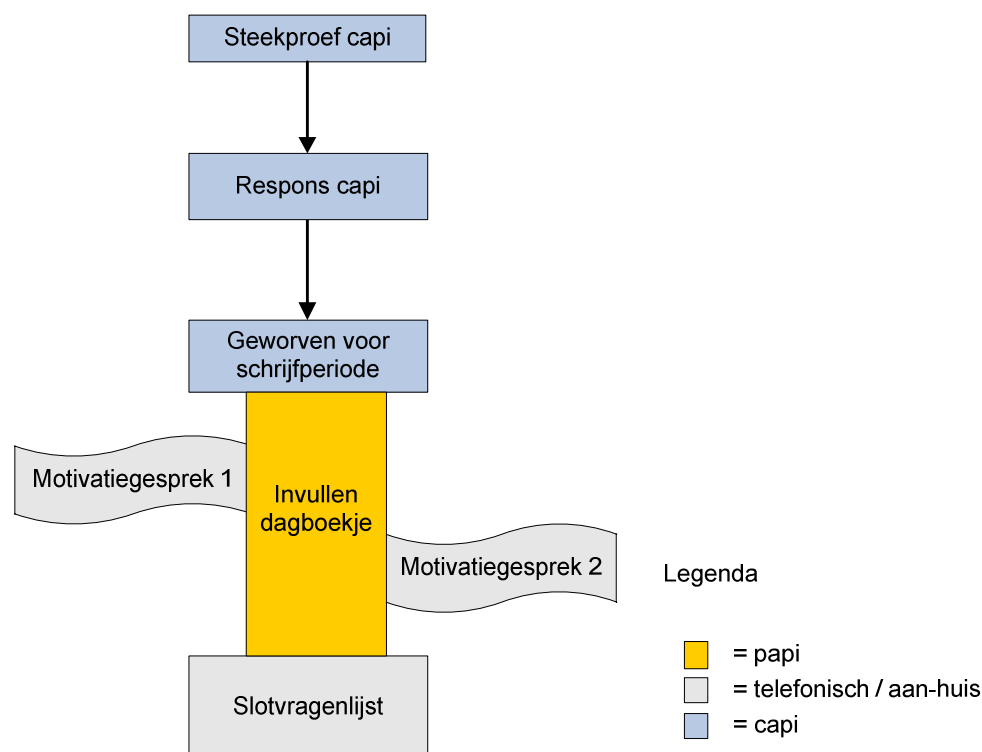
*incl. verhuizers

3.3 Benaderingsstrategie

Het TBO vraagt veel van respondenten en vergt om die reden een behoorlijke inspanning om de medewerking te verkrijgen en behouden van respondenten. Alle steekproefpersonen zijn voor de startvragenlijst aan-huis benaderd. In de startvragenlijst zijn steekproefpersonen geworven voor het dagboekje en eventueel voor het partnerboekje. De interviewer heeft deze achter gelaten bij de respondent. Voor de slotvragenlijst zijn steekproefpersonen bij voorkeur aan-huis benaderd behalve als respondenten aan hebben gegeven liever telefonisch te worden benaderd. De telefonische benadering is in dat geval uitgevoerd door de face to face interviewer die de startvragenlijst heeft afgenomen. Bij de aan-huis benadering voor de slotvragenlijst haalde de interviewer de dagboekjes en eventuele partnerboekjes op en had de interviewer de mogelijkheid om de boekjes te controleren op volledigheid. Bij de telefonische benadering ging de interviewer óók langs om de boekjes te controleren en op te halen. Er hebben motivatiegesprekken plaats gevonden door face to face interviewers vlak vóór en tijdens de schrijfperiode om te stimuleren dat respondenten tijdig begonnen met schrijven en om te voorkomen dat ze voortijdig afhaakten. Deze motivatiegesprekken waren telefonisch als een telefoonnummer bekend was. Was geen telefoonnummer bekend dan zijn respondenten tijdens de schrijfperiode aan-huis bezocht.

De benaderingsstrategie is schematisch weergegeven in Figuur 1.

Figuur 1 Benaderingsstrategie TBO



De benaderingsstrategie die bij de aan-huis benadering voor de startvragenlijst wordt toegepast heeft zich in de praktijk bewezen. Toegepast op het TBO zag de benaderingsstrategie er op hoofdlijnen als volgt uit (meer details zijn terug te vinden in het Onderzoeksdesign):

- Steekproefpersonen ontvingen voorafgaand aan het eerste aan-huis bezoek een aanschrijfbrief waarin het bezoek van de interviewer werd aangekondigd en het doel van het onderzoek werd toegelicht. Met de aanschrijfbrief is een folder meegestuurd en een postzegelboekje als incentive.
- De steekproefpersonen in een maandsteekproef zijn in de eerste helft van de betreffende maand minstens één keer bezocht. Alleen in maart 2011 is hiervan afgeweken aangezien TBO halverwege deze maand startte. De omvang van de maartsteekproef is aangepast op de kortere waarneemperiode.
- Interviewers zijn geïnstrueerd om de bezoeken voor TBO zo goed mogelijk te spreiden over de waarneemperiode.

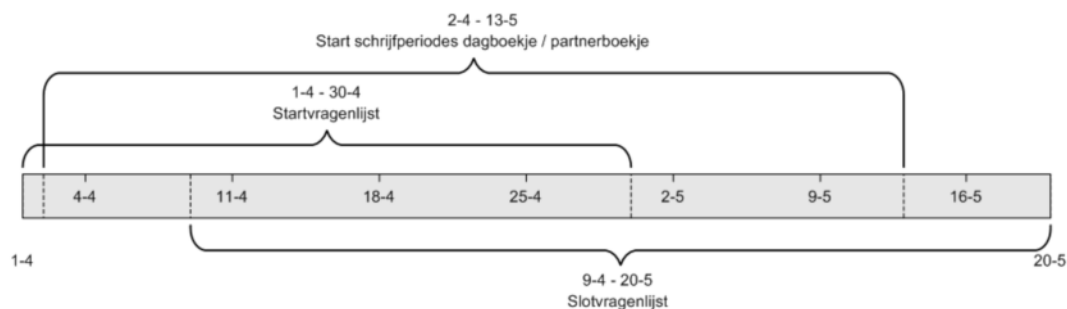
- Aan het eind van de startvragenlijst heeft de interviewer de respondent geworven voor het invullen van het dagboekje en het eventuele partnerboekje door de partner van de respondent. Hierbij maakte de interviewer gebruik van een incentive in de vorm van een kadobon. Gaf de respondent aan daartoe bereid te zijn, dan noteerde de interviewer de schrijfperiode op het dagboekje en het eventuele partnerboekje en overhandigde beide boekjes aan de respondent met uitleg over wat van hem of haar werd verwacht. De schrijfperiodes voor dagboekje en partnerboekje waren gelijk.
- Aan de steekproefpersonen is op voorhand in de steekproef een dag in de week toegekend waarop de schrijfperiode start. De ervaring bij TBO 2005/2006 heeft geleerd dat respondenten geneigd zijn om in de loop van de schrijfperiode minder gedetailleerd hun activiteiten bij te houden. Om toch activiteiten op alle dagen van de week met een zelfde mate van detail waar te nemen, is de start van de schrijfperiodes gespreid over de dagen van de week. Wanneer een schrijfperiode precies gestart is, is afhankelijk van de enquêtedag en -datum van de startvragenlijst. De schrijfperiode startte op de eerstvolgende toegekende weekdag na het startinterview. De interviewer heeft de schrijfperiode op het dagboekje genoteerd. Een steekproefpersoon die zijn of haar schrijfperiode op dinsdag moest starten en op maandag 14 maart 2011 gerepsondeerd heeft op de startvragenlijst kreeg als schrijfperiode dinsdag 15 maart tot en met maandag 21 maart 2011. Een steekproefpersoon met dezelfde startdag van de schrijfperiode die op dinsdag 15 maart 2011 respondeerde op de startvragenlijst heeft als schrijfperiode dinsdag 22 maart tot en met maandag 28 maart 2011 gekregen.
- Mocht de respondent aangeven niet in staat te zijn om te starten met de schrijfperiode op de aangegeven startdatum dan bestond de mogelijkheid om de schrijfperiode met 1 of maximaal 2 weken uit te stellen. Te denken valt aan een kortstondige ziekenhuisopname. De startdag bleef ongewijzigd. Een dergelijk uitstel is zoveel mogelijk voorkomen.
- Vlak vóór en tijdens de schrijfperiode zijn respondenten twee of drie keer telefonisch benaderd als een telefoonnummer bij CBS bekend was of de interviewer in het startgesprek een telefoonnummer heeft gekregen van de respondent. Deze telefonische motivatiegesprekken zijn uitgevoerd door de face to face interviewer en niet door een telefonische interviewer. Hiervoor is niet alleen gekozen voor de eenvoud van het logistieke proces van het onderzoek, maar vooral ook omdat de face to face interviewer al contact heeft gehad met de OP en op basis daarvan een inschatting kan maken van een passende benadering. Het volledig laten afhandelen van het waarneemproces bij een

steekproefelement door een en dezelfde interviewer werkt responsverhogend en kwaliteitsverbeterend. De motivatiegesprekken waren niet alleen bedoeld als motivatie voor de respondent, maar ook om na te gaan of het boekje op de juiste manier wordt ingevuld. In het eerste motivatiegesprek is daarnaast expliciet gecontroleerd of de respondent zich houdt aan de schrijfperiode zoals vermeld op de voorkant van het boekje.

- Respondenten van wie ook na het startinterview geen telefoonnummer bekend is zijn tijdens de schrijfperiode minstens één keer bezocht door de interviewer. Het is aan het oordeel van de interviewer overgelaten of een tweede bezoek wenselijk was.
- De interviewer heeft tijdens de schrijfperiode een afspraak gemaakt om de slotvragenlijst af te nemen. Dit gebeurde bij voorkeur aan-huis maar kon ook telefonisch als de respondent aangeeft voorkeur te hebben voor een telefonische benadering. De telefonische benadering is in dat geval uitgevoerd door de face to face interviewer die de startvragenlijst heeft afgenomen. Bij aan-huis benadering voor het slotinterview zijn de boekjes door de interviewer gecontroleerd op volledigheid, meegenomen en verstuurd naar het CBS. Bij telefonische benadering zijn de boekjes óók opgehaald door de interviewer, waardoor controle mogelijk bleef. De afspraak voor het slotinterview is zo kort mogelijk op de schrijfperiode gepland.

In Figuur 2 is het tijdsplan voor de aprilsteekproef van 2011 schematisch weergegeven. Een steekproefpersoon die op vrijdag 1 april reageerde op de startvragenlijst en zaterdag als startdag voor de schrijfperiode heeft, kreeg als schrijfperiode zaterdag 2 april tot en met vrijdag 8 april. Een steekproefpersoon die op zaterdag 30 april reageerde op de startvragenlijst en zaterdag als startdag voor de schrijfperiode heeft, kreeg als schrijfperiode zaterdag 7 mei tot en met vrijdag 13 mei. De schrijfperiodes op basis van de aprilsteekproef startten dus op zijn vroegst op 2 april en eindigden op zijn laatst op 13 mei. De afspraak voor de slotvragenlijst is zo kort mogelijk op de schrijfperiode gepland. Dit was op zijn vroegst 9 april. Er van uitgaande dat uiterlijk 1 week na de schrijfperiode de slotvragenlijst is afgenomen vond het slotinterview in bovenstaand voorbeeld uiterlijk 20 mei plaats.

Figuur 2 Schematische weergave planning aprilsteekproef



Tabel 3 bevat per maandsteekproef de mogelijke datums van start- en slotinterviews plus de eerst en laatst mogelijke datums van de bijbehorende invulweken.

Tabel 3 Gegevens over maandsteekproeven voor TBO, 2011-2012

Portie	Startinterview	Invulweek	Slotinterview
1	14 mrt t/m 31 mrt	15 mrt t/m 13 apr (27 apr)	22 mrt t/m 20 apr (4 mei)
2	1 apr t/m 30 apr	2 apr t/m 13 mei (27 mei)	9 apr t/m 20 mei (3 jun)
3	1 mei t/m 31 mei	2 mei t/m 13 jun (27 jun)	9 mei t/m 20 jun (4 jul)
4	1 jun t/m 30 jun	2 jun t/m 13 jul (27 jul)	9 jun t/m 20 jul (3 aug)
5	1 jul t/m 31 jul	2 jul t/m 13 aug (27 aug)	9 jul t/m 20 aug (3 sep)
6	1 aug t/m 31 aug	2 aug t/m 13 sep (27 sep)	9 aug t/m 20 sep (4 okt)
7	1 sep t/m 30 sep	2 sep t/m 13 okt (27 okt)	9 sep t/m 20 okt (3 nov)
8	1 okt t/m 31 okt	2 okt t/m 13 nov (27 nov)	9 okt t/m 20 nov (4 dec)
9	1 nov t/m 30 nov	2 nov t/m 13 dec (27 dec)	9 nov t/m 20 dec (3 jan)
10	1 dec t/m 31 dec	2 dec t/m 13 jan (27 jan)	9 dec t/m 20 jan (3 feb)
11	1 jan t/m 31 jan	2 jan t/m 13 feb (27 feb)	9 jan t/m 20 feb (5 mrt)
12	1 feb t/m 29 feb	2 feb t/m 13 mrt (27 mrt)	9 feb t/m 20 mrt (3 apr)

Er moest over 365 dagen worden waargenomen. Dat zou de periode 15 maart 2011 t/m 13 maart 2012 betekenen. Echter, in uitzonderingsgevallen is het toegestaan 1 of maximaal 2 weken te schuiven met de start van de invulweek. De uiterste datums die dit tot gevolg zouden hebben zijn in tabel 1 aangegeven tussen haakjes. Van de laatste portie zullen enkel invuldagen t/m 13 maart gebruikt worden.

3.4 Vragenlijsten

Voor de vragenlijsten van TBO 2011/2012 zijn de vragenlijsten van TBO 2005/2006 als uitgangspunt genomen. Voor het TBO 2011/2012 zijn de volgende vragenlijsten gebruikt: startvragenlijst, boekjes invulweek en slotvragenlijst. In samenspraak met het SCP is de exacte inhoud van de vragenlijsten vastgesteld. In Tabel 4 worden de verschillende blokken in de startvragenlijst weergegeven.

Aan het einde van de startvragenlijst zijn respondenten geworven. Hier is respondenten gevraagd gedurende zeven aaneengesloten dagen hun activiteiten in een boekje (papi) bij te houden. De volgende boekjes zijn daarbij onderscheiden: dagboekje voor volwassenen of kinderen (afhankelijk van leeftijd OP) en partnerboekje.

Tabel 4 Blokken startvragenlijst TBO

Bloknaam
1. Standaard huishoudbox
2. Type Woning
3. Bezit Goederen
4. Hobby en Muziek
5. Sport en Bewegen
6. Drukke
7. Belang Betaalde Arbeid
8. Betaald Werk
9. Werkzoeken
10. WerkzoekenA
11. Werktijden
12. Arbeidsplaats
13. Informatie Partner
14. Kinderopvang
15. Onderwijs Actueel

-
16. Opleidingsplaats
 17. Onderwijs Gevolgd
 18. Religie en Stemmen
 19. Opleiding Ouders
 20. Werving voor de boekjes
-

In het dagboekje houden de respondenten per tijdseenheid van tien minuten bij wat hun hoofdactiviteit geweest is en wat eventueel gedurende deze activiteit de belangrijkste nevenactiviteit was. Deze activiteiten konden de respondenten in hun eigen woorden omschrijven. Verder werd gevraagd of men tijdens de activiteit(en) alleen was of met iemand die de respondent kent. Wanneer de activiteit een verplaatsing was, is respondenten gevraagd om in de omschrijving van de activiteit ook het gebruikte vervoermiddel op te geven. Na afloop van elke dag werd nog een aantal aanvullende vragen gesteld over de begin- en eindlocatie van de respondent op die betreffende dag, eventuele bijzonderheden zoals ziekte of verlof, het gevoel van de respondent op de dag en verplaatsingen buiten de woonplaats voor werk/studie en verplaatsingen voor andere redenen buiten de plaats waar de respondent verbleef. Met de dagboekjes voor volwassenen en voor kinderen is dezelfde informatie verzameld, maar de toelichting is aangepast op de doelgroep. Ook is in het kinderdagboekje niet gevraagd naar verplaatsingen in het kader van werk of studie en kon voor deze groep de aard van de invuldag geen 'gewone werkdag' zijn. Indien respondenten met een volwassenendagboekje samenwonen met een partner, dan is aan deze partner gevraagd om voor dezelfde invulperiode ook een dagboekje bij te houden. In dit dagboekje kon de partner op het niveau van hoofdactiviteiten bijhouden wat zijn/haar activiteiten zijn geweest. Dit gebeurde aan de hand van gesloten activiteitscategorieën, in tegenstelling tot de vrije tekst die in het basisdagboekje kon worden ingevuld.

Na afloop van de schrijfperiode vond nog een slotinterview plaats. De slotvragenlijst duurde maximaal 35 minuten. In Tabel 5 worden de verschillende blokken in de slotvragenlijst weergegeven.

Tabel 5 Blokken slotvragenlijst TBO

Bloknaam

1. Standaard huishoudbox

2. Huishoud / Sociale contacten

3. Bezit Goederen

4. Mediagebruik

5. Onbetaald werk

6. Informele groepen

7. Uitgaan en eten

8. Vervoer

9. Vakanties

10. Openingstijden

11. Waarden

12. Gezondheid

13. Arbeid en Zorg

14. Politiek en veiligheid

3.5 Veldwerkresultaten

3.5.1 Respons

Het veldwerk is in de periode maart 2011-maart 2012 uitgevoerd. In de tabellen 6 (absoluut) en 7 (procentueel) worden de behaalde responsresultaten weergegeven van het startinterview, het dagboekje (invulweek) en het slotinterview.

Tabel 6 Responsresultaten TBO, absoluut

	Benaderstap		
	Start	Invulweek	Slot
Uitgezette steekproef			
	4.976	2.153	2.153
Kaderfouten	86	-	-
Nonrespons	2.476	86	143
wv. Weigering	1.422	-	37
wv. Geen gelegenheid, niet in staat	395	-	30
wv. Geen contact	230	-	11
wv. Taalbarriere	179	-	-
wv. Verhuisd binnen Nederland	160	-	-
wv. Overig	90	-	65
Volledige respons	2.395	2.008	2.004
wv. Werving is Ja	2.153	-	-
Onvolledig/afgebroken respons	19	59	6

Tabel 7 Responsresultaten TBO, procentueel

	Benaderstap		
	Start	Invulweek	Slot
Uitgezette steekproef			
	100,0%	100,0%	100,0%
Kaderfouten	1,7%	-	-
Nonrespons	49,8%	1,7%	2,9%
wv. Weigering	28,6%	-	0,7%
wv. Geen gelegenheid, niet in staat	7,9%	-	0,6%
wv. Geen contact	4,6%	-	0,2%
wv. Taalbarriere	3,6%	-	-
wv. Verhuisd binnen Nederland	3,2%	-	-
wv. Overig	1,8%	-	1,3%
Volledige respons	48,1%	40,4%	40,3%
wv. Werving is Ja	89,9%	-	-
Onvolledig/afgebroken respons	0,4%	1,2%	0,1%

Zoals te zien is in tabel 6 is de target van 2.000 ingevuld boekjes tijdens de dataverzameling behaald. Tijdens de dataverzameling bleek dat de verhouding tussen de targets van de verschillende benaderstappen anders is uitgevallen dan van tevoren bedacht. Zo was de respons op de startvragenlijst lager dan ingeschat, maar de werving voor de invulweek juist hoger. Per saldo leverde dit eenzelfde aantal personen op dat aan de invulweek is begonnen.

3.5.2 Issue tijdens de dataverzameling

In november 2011 is er een fout opgetreden bij het laden van de adressen in de laptops van de interviewers. De startdag die vanuit de steekproef bepaald wordt, was niet goed meegenomen en stond overal op zaterdag. Na de ontdekking van deze fout, is een extra instructie voor de interviewers uitgegaan dat zij de startdag moesten gebruiken zoals die op hun zgn. looplijst staat en dus niet de startdag zoals die in de vragenlijst verschijnt.

3.6 Coderen en vertoetsen

Het Dagboek OP en het Dagboek Partner zijn papieren vragenlijsten, die door de onderzoekspersoon respectievelijk diens partner zelf gedurende 7 dagen worden ingevuld.

Deze dagboekjes worden eerst gecodeerd en later vertoetst (in de computer ingevoerd). De toegekende codes worden in het dagboek opgeschreven. Dat vergemakkelijkt controles achteraf en is nodig omdat het coderen en vertoetsen afzonderlijk plaatsvonden.

In overleg met het SCP is de codeerlijst met aanvullende toelichting gemaakt, zie "TBO_2011_2012_codelijst_met_toelichting.pdf". Deze codeerlijst is gebaseerd op de HETUS richtlijnen.

Er heeft regelmatig overleg plaatsgevonden met de codeurs, om ervoor te zorgen dat ze allemaal op eenzelfde manier met de codelijst zouden omgaan. Bij onduidelijkheden of problemen is er met het SCP overlegd om de afspraken te verduidelijken.

Het coderen was niet eenvoudig, omdat uit de omschrijving in het boekje zowel een code voor een activiteit als een code voor de plaats waar die activiteit plaatsvond (lokatie) of een code voor een verplaatsing (vervoermiddel) afgeleid moest worden. Daarnaast varieerde de kwaliteit van de dagboekjes enorm. Er zijn respondenten die al hun activiteiten en nevenactiviteiten heel uitgebreid beschrijven. Er zijn echter ook respondenten die met hele grote stappen door hun dagelijkse bezigheden gaan, waardoor er benodigde informatie ontbreekt.

Voor het vertoetsen van de twee dagboekjes zijn twee invoerprogramma's gemaakt.

Het coderen en vertoetsen heeft veel meer tijd gevegd dan gepland was. Het vertalen van omschrijvingen naar activiteiten en lokaties of verplaatsingen bleek veel complexer en tijdrovender dan verwacht. In het begin hebben alle bewerkers zowel gecodeerd als ingevoerd. Later is het coderen bij een aantal medewerkers neergelegd. Er konden namelijk extra medewerkers ingezet worden voor het vertoetsen. Het vertoetsen zelf had ook efficiënter gekund. Dat vergt echter meer kennis van het onderzoek op het moment dat de invoerprogramma's gemaakt worden.

3.7 Bestandsopbouw, controles en oplevering

In het contract met het SCP is bepaald dat aan het SCP een aantal bestanden worden opgeleverd:

- Start- en Slotvragenlijst, exclusief de huishoudbox van de Slotvragenlijst
- Huishoudbox van de Slotvragenlijst
- Dagboek OP
- Dagboek Partner.

Aan het SCP worden alleen records geleverd waarvoor zowel een Startvragenlijst als een Dagboek OP van voldoende kwaliteit (dat wil zeggen minimaal 1 werkdag en 1 weekenddag met minder dan 1 uur onbekende hoofdactiviteit) beschikbaar is.

Over de precieze vorm van de bestanden zijn geen afspraken gemaakt. Voor de vragenlijsten is dat geen probleem. In die bestanden worden de vragenlijstvariabelen, aangevuld met de afleidingen en verrijkingen en de gewichten geleverd. In de verwerking van de dagboekjes wordt in verschillende stappen gewerkt met bestanden met records per dagboek, per dag en per episode. Daarom zijn van de dagboekjes ook drie bestanden geleverd.

Op 31-07-2012 zijn de volgende bestanden aan het SCP geleverd:

NAAM BESTAND	DATA	META	RECORDS	VARIABELEN
SCP_TBO_START_SLOT	*.sav	*.sps	2.005	784
SCP_TBO_SLOT_HHBOX	*.sav	*.sps	1.940	51
SCP_TBO_OP_dagboek	*.sav	*.sps	2.005	61
SCP_TBO_OP_dag	*.sav	*.sps	28.070	117
SCP_TBO_OP_episode	*.sav	*.sps	381.557	142
SCP_TBO_Partner_dagboek	*.sav	*.sps	632	55
SCP_TBO_Partner_dag	*.sav	*.sps	4.424	95
SCP_TBO_Partner_episode	*.sav	*.sps	60.710	103

Op 27-11-2012 zijn twee bestanden opnieuw geleverd, namelijk:

- SCP_TBO_START_SLOT
in verband met de vulling van variabelen uit het vragenblok WerkZoekA, die in de eerdere levering leeg bleken te zijn
- SCP_TBO_SLOT_HHBOX
in verband met een aanpassing van het label bij variabele WinkelC.

Op 14-03-2013 is (op verzoek van het SCP) een weging van het TBO 2005/2006 opgeleverd, die door het CBS is uitgevoerd op (min of meer) dezelfde wijze als de weging van het TBO 2011/2012:

- 2013-03-14_TBO2006_weging_2011.sav en *.txt.

4 Verwerken (data processing)

Het verwerken omvat de werkzaamheden die nodig zijn om de antwoorden van de respondenten te combineren met registerdata en om aan deze data vervolgens plausible statistische data te kunnen ontleen. Het verwerken is uitgevoerd door de Sector Arbeid, Inkomen en Leefsituatie (SAL) van de Divisie Sociaal-economische en ruimtelijke statistieken (SER) van het CBS.

4.1 Proces

De verwerking van het TBO is in stappen uitgevoerd. Deze stappen worden heel uitgebreid beschreven in "TBO_2011_2012_beschrijving_verwerkingsproces.pdf".

In **Stap 1** wordt de **Startvragenlijst** verwerkt:

- 01 ophalen enquêtedata
- 02 selectie respondenten en relevante variabelen
- 03 dichotomiseren
(WelkInstr, VervWerk, VastDag en VervOpl)
- 04 correctie identificerende variabelen
(toetsfouten, dubbele nummers, verkeerd dagboektype)
- 05 coderen open antwoorden sport
- 06 codering open antwoorden vervoer
(AndVerv, VervBrSch, VervHISch, AndVervOp)
- 07 bepalen afstanden vanaf woonadres
(WerkOp, WerkPartner, OplOP).

In de verwerking worden 4.919 Startvragenlijsten opgehaald, waarvan er 2.408 respons zijn. Daarvan zijn 2.163 respondenten geworven voor het invullen van een Dagboek OP, te weten 1.955 volwassenen (15 jaar en ouder) en 208 kinderen (10 t/m 14 jaar). Er zijn 1.281 partners aanwezig. Daarvan zijn er 808 geworven voor het invullen van een Dagboek Partner.

In **stap 2** wordt de **Slotvragenlijst** verwerkt:

- 01 ophalen enquêtedata,
inclusief selectie respondenten en relevante variabelen
- 02 dichotomiseren
(WaarComp).

Er zijn 2.434 Slotvragenlijsten opgehaald, daarvan waren er 2.023 respons.

In **stap 3** worden de **Start-** en de **Slotvragenlijst gekoppeld:**

- 01 koppeling Start- en Slotvragenlijst
- 02 afleidingen
(Enq_Datum, Enq_Maand, Enq_WeekDag, Invul_Datum, Invul_Maand, Invul_WeekDag, Lft_EnqStart, Lft_EnqSlot, PosWrkOP, Aant_Afwijk)
- 03 gemiddeld bezit goederen
(AantKamer_mean, AantAuto_mean, AantFiets_mean, AantBrom_mean, AantMotor_mean, BestAuto_1, BestAutoOP_1, BAuto_1, Lidgroep_1)
- 04 correctie bezit goederen
(Kamer_1, AantAuto_1, AantFiets_1, AantBrom_1, AantMotor_1)
- 05 afleiding opleidingsniveau
(OplNiveau, OplNiv5)
- 06 probleem startdag november
(de gevolgen van het ontbreken van de startdag wordt in kaart gebracht, zie "TBO_2011_2012_beschrijving_verwerkingsproces.pdf")
- 07 bepalen Start- en Slotvragenlijst volledig.

Er blijken 2 Slotvragenlijsten niet te koppelen aan de Startvragenlijst, zij vallen dus uit. Er worden 2 Startvragenlijsten als onvolledig afgekeurd (de steekproefpersoon komt in het huishouden niet voor). Er blijven 2.406 Startvragenlijsten en 2.021 Slotvragenlijsten over. 385 Startvragenlijsten hebben geen Slotvragenlijst.

Er worden twee nieuwe bestanden gemaakt:

- StartSlotvragenlijst exclusief de huishoudbox van de Slotvragenlijst
- Huishoudbox Slotvragenlijst.

In **stap 4** wordt het **Dagboek** van de **Partner** verwerkt:

- 01 ophalen data
- 02 correctie 1
(correctie identificerende variabelen)
& afleidingen
(Dagtype, Afl_HfdActDuur_epi, Duur_epi_mm)
- 03 correctie 2
(koppeling met Startvragenlijst
& afleiding Gem_Duur_hfdact_*_dag_mm)
- 04 correctie 3
(N_HfdAct_*_dagboek,
Duur_hfdact_*_dagboek_mm,
N_HfdAct_onbekend_dag,
Duur_onbekend_dag_mm,

- N_HfdAct_werkdag_onbekend,
 N_HfdAct_weekenddag_onbekend,
 Resultaat_Dagboek_Partner,
 Resultaat_Dagboek_Partner_7dagen).
- 05 correctie 4
 (Gem_duur_hfdact*_dagboek_mm)
- 06 omzetten bestand met records per dagboek
 van Blaise naar Ascii naar SPSS
- 07 omzetten bestand met records per dag
 van Blaise naar Ascii naar SPSS
- 08 omzetten bestand met records per episode
 van Blaise naar Ascii naar SPSS.

Er zijn 636 Dagboek Partner ingevoerd.

Voor het TBO 2011/2012 is een dagboek van voldoende kwaliteit als er minimaal 1 werk- en 1 weekenddag van voldoende kwaliteit is en dat is weer het geval als er maximaal 60 minuten per dag een onbekende hoofdactiviteit heeft. Er zijn 4 dagboekjes die afgekeurd worden, dus er blijven 632 dagboekjes over met Resultaat_Dagboek_Partner = 1 = ja (goedgekeurd).

Voor het TBO 2005/2006 is een dagboek van voldoende kwaliteit als alle 7 dagen van voldoende kwaliteit zijn (als er maximaal 60 minuten per dag een onbekende hoofdactiviteit heeft). Er zijn 36 dagboekjes die afgekeurd worden, er blijven 600 dagboekjes over met Resultaat_Dagboek_Partner_7dagen = 1 = ja (goedgekeurd).

In **stap 5** wordt het **Dagboek** van de **OP** (onderzoekspersoon) verwerkt:

- 01 ophalen data
- 02 correctie 1
 (correctie identificerende variabelen)
- 03 correctie 2
 (ingevoerde hoofdactiviteit, nevenactiviteit, locatie, HfdActPc)
 & afleidingen
 (Dagtype, Afl_HfdActDuur_epi, Duur_epi_mm)
- 04 correctie 3
 (koppeling met Startvragenlijst: controle leeftijd OP en dagboektype;
 controle en imputatie bij ontbrekende dagboekvragen:
 WerkStud, LokatieStart, LokatieEind, afleiding SoortDag;
 afleiden And_Plaats_werk_studie, Gemcode_and_werk_studie,
 Landcode_and_werk_studie, And_Plaats_anders, Gemcode_anders,
 Landcode_anders;
 controle aantal episodes en totale duur episodes)

- 05 correctie 4
(afleiding Gem_Duur_hfdact_*_dag_mm)
- 06 correctie 5
(N_HfdAct_*_dagboek,
Duur_HfdAct_*_dagboek_mm,
N_HfdAct_overig_dag,
Duur_overig_dag_mm,
N_HfdAct_onbekend_dag,
Duur_onbekend_dag_mm,
N_HfdAct_werkdag_onbekend,
N_HfdAct_weekenddag_onbekend,
Resultaat_Dagboek_OP,
Resultaat_Dagboek_OP_7dagen).
- 07 correctie 6
(Gem_duur_hfdact_*_dagboek_mm)
- 08 omzetten bestand met records per dagboek
van Blaise naar Ascii naar SPSS
- 09 omzetten bestand met records per dag
van Blaise naar Ascii naar SPSS
- 10 omzetten bestand met records per episode
van Blaise naar Ascii naar SPSS.

Er zijn 2.025 Dagboek OP ingevoerd. Er waren 4 Dagboek OP dubbel ingevoerd (foute WE_ID's kunnen in het invoerprogramma niet gecorrigeerd worden, deze dagboekjes waren daarom twee keer ingevoerd). Er worden dus 2.021 dagboekjes verwerkt.

Voor het TBO 2011/2012 is een dagboek van voldoende kwaliteit als er minimaal 1 werk- en 1 weekenddag van voldoende kwaliteit is en dat is weer het geval als er maximaal 60 minuten per dag een onbekende hoofdactiviteit heeft. Er zijn 15 dagboekjes die afgekeurd worden, dus er blijven 2.006 dagboekjes over met Resultaat_Dagboek_Partner = 1 = ja (goedgekeurd).

Voor het TBO 2005/2006 is een dagboek van voldoende kwaliteit als alle 7 dagen van voldoende kwaliteit zijn (als er maximaal 60 minuten per dag een onbekende hoofdactiviteit heeft). Er zijn 215 dagboekjes die afgekeurd worden, er blijven 1.806 dagboekjes over met Resultaat_Dagboek_Partner_7dagen = 1 = ja (goedgekeurd).

In **stap 6** wordt de **steekproef** van het TBO aangevuld:

- 01 ophalen steekproeven 2011 en 2012
(dit zogenaamde achtergrondbestand is al verrijkt met GBA informatie)
- 02 omzetten naar eigen datamodel

- (inclusief labels)
- 03 informatie uit de Statline tabel Gebieden in Nederland (GIR) 2011 wordt omgezet in een ander formaat en aangevuld met labels
 - 04 achtergrondbestand TBO verder verrijken met gegevens uit het PWR, GIR en uit het integrale inkomensbestand 2010
 - 05 koppelen aan overzichten Dataverzameling (bruikbare steekproef, Startvragenlijst respons, Dagboek OP afgeboekt) en aan opgehaalde Startvragenlijst (Stap1_02) en Slotvragenlijst (Stap2_01)
 - 06 koppeling uitzet Dataverzameling aan achtergrondbestand TBO (komen alle uitgezette steekproefelementen voor het TBO ook uit de steekproef van het TBO, soms worden steekproefelementen “geleend”, dat was hier niet het geval)
 - 07 koppeling Dagboek Partner (Stap4_04, vullen DagboekP_ingevoerd, overnemen Resultaat_Dagboek_Partner en Resultaat_Dagboek_Partner_7dagen in achtergrondbestand)
 - 08 koppeling Dagboek OP (Stap5_06, vullen DagboekOP_ingevoerd, overnemen Resultaat_Dagboek_OP en Resultaat_Dagboek_OP_7dagen in achtergrondbestand)
 - 09 koppeling StartSlot (Stap3_07, overnemen Volledig_Start en Volledig_Slot in achtergrondbestand & afleiden wegingsvariabelen in het achtergrondbestand zodat die eventueel bij de nonrespons analyse gebruikt kunnen worden: HHGrootte, PosHH, PosHH5, SamHH5, Lft6, Lft12, HerkGnt, HHInkomen_kl, Prov en ProvPlus)
 - 10 bepalen Resultaat_SCP.

In paragraaf 3.1 is al verwezen naar de beschrijving van de verrijking van het TBO in het document: “Koppel_en_verrijgingsregels_TBO_2011_2012.pdf”.

De steekproef van het TBO 2011/2012 bevat 5.546 personen. In het achtergrondbestand van het TBO 2011/2012 komen 5.492 personen voor. Het verschil van 54 personen, betreft verhuizers die twee keer zijn benaderd: op hun oude en op hun nieuwe adres. Voor deze verhuizers is alleen de 2^e benadering verder verwerkt. Dataverzameling heeft gegevens aangeleverd van de 4.977 uitgezette personen, van 2.395 Startvragenlijsten volledige respons en van 2.153 Dagboek OP respons. Deze gegevens kunnen later gebruikt worden bij de nonrespons analyse.

Volgens de verwerking van het TBO zijn er niet 2.395 maar 2.408 Startvragenlijsten (volledige) respons. Er zijn 5 Startvragenlijsten door SAL gewijzigd in volledige

respons, omdat de Startvragenlijst helemaal was beantwoord en er was toegezegd een Dagboek OP in te vullen. Er zijn door SAL 2 Startvragenlijsten verwijderd, omdat de steekproefpersoon niet in het huishouden voorkwam. Er zijn 10 Startvragenlijsten later door Dataverzameling opgeleverd. Van deze 10 gevallen kwamen de Startvragenlijsten bij het CBS binnen nadat de betreffende veldwerkperiode was afgesloten, waardoor deze Startvragenlijsten buiten de reguliere datalevering van Dataverzameling aan SAL vielen, terwijl er later wel dagboekjes werden ingestuurd.

Bij Dataverzameling zijn 2.153 Dagboekjes OP afgeboekt, daarvan hebben er 64 geen eindresultaat, 32 zijn weigeringen (Eindresultaat 20), 59 zijn onvolledige respons (Eindresultaat 31) en 1.998 zijn volledige respons (Eindresultaat 30). Daarnaast heeft Dataverzameling een lijst met 66 afgekeurde dagboekjes OP geleverd, die niet afgeboekt zijn, omdat deze dagboekjes te slecht waren ingevuld.

In tabel 6 worden 2.008 Dagboekjes OP respons opgevoerd. Dit zijn de hierboven genoemde 1.998 als volledige respons afgeboekte Dagboekjes OP plus de 10 dagboekjes waarvan de Startvragenlijsten te laat bij het CBS binnen zijn gekomen. Dagboekjes behorend bij een Startvragenlijst die bij Dataverzameling niet als respons afgeboekt is, kunnen zelf ook niet afgeboekt worden. Wel zijn ze meegeteld in tabel 6.

Nadat in stap6_07 t/m Stap6_09 de resultaten van de afzonderlijke dagboekjes en vragenlijsten aan het achtergrondbestand zijn toegevoegd, kan in het achtergrondbestand het eindresultaat afgeleid worden.

Voor het TBO 2011/2012 is vereist dat minimaal de Startvragenlijst en het Dagboek OP (met minimaal 1 werkdag en 1 weekenddag van voldoende kwaliteit, dat wil zeggen met maximaal 1 uur onbekende hoofdactiviteit) beschikbaar zijn (Resultaat_SCP). Voor het TBO 2005/2006 moesten van het Dagboek OP alle 7 dagen van voldoende kwaliteit zijn. Op verzoek van het SCP is ook dit criterium als een variabele afgeleid (Resultaat_SCP_7dagen).

Er zijn 2.005 respondenten waarbij het Resultaat_SCP voldoende is. Dat betekent dat er 2.005 Startvragenlijsten en 2.005 Dagboekjes OP opgeleverd kunnen worden. Daarnaast voldoen er 1.940 Slotvragenlijsten en 626 Dagboekjes Partner aan het criterium. Er zijn 632 Dagboekjes Partner aan het SCP geleverd. Er komen dus 6 Dagboekjes Partner in de SCP bestanden voor, waarvan de Startvragenlijst en het Dagboek OP wél goedgekeurd zijn, maar het Dagboek Partner niet (Resultaat_TBO = 2).

Voor het strengere criterium van het TBO 2005/2006 zouden deze aantallen 1.806 Startvragenlijsten en Dagboekjes OP, 1.748 Slotvragenlijsten en 555 Dagboekjes Partner zijn.

In **stap 7** wordt het eindresultaat vanuit het achtergrondbestand aan de dagboekjes en vragenlijsten toegevoegd, zodat de respondenten voor de bestanden voor het SCP geselecteerd kunnen worden en de weging uitgevoerd kan worden:

- 01 SCP_TBO_Dagb_Partner_dagboek
- 02 SCP_TBO_Dagb_Partner_dag
- 03 SCP_TBO_Dagb_Partner_episode
- 04 SCP_TBO_Dagb_OP_dagboek
- 05 SCP_TBO_Dagb_OP_dag
- 06 SCP_TBO_Dagb_OP_episode
- 07 SCP_TBO_Slot_HHBox
- 08 SCP_TBO_Start_Slot_Verrijken
(overnemen van verrijkingen uit het achtergrondbestand)
- 09 SCP_TBO_Start_Slot_Afleiden1
(ten behoeve van de weging worden een aantal variabelen afgeleid van de vragenlijstvariabelen, namelijk HHGrootte, SamHH, SamHH5, PosHH, PosHH5, Lft6, Lft12 en Werkzaamheid; deze variabelen komen twee keer voor in het SCP bestand: de 1^e keer uit de registratie in het blok Verrijkingen en de 2^e keer {indien mogelijk afgeleid van vragenlijstvariabelen, anders overgenomen uit de registratie} in het blok Weging, in sav bestanden worden deze variabelen dan genummerd, de variabele uit het blok Weging heet dan HHGrootte2)
- 10 SCP_TBO_Start_Slot_Afleiden2
(voor zowel Resultaat_SCP als Resultaat_SCP_7dagen wordt het aantal respondenten per maand berekend ten behoeve van het startgewicht)
- 11 SCP_TBO_Start_Slot_Afleiden3
(voor elke respondent worden de voor het startgewicht benodigde variabelen gevuld: N, D, Dm, Rm_2d, RM_7d)
- 12 SCP_TBO_Start_Slot_Afleiden4
(de twee startgewichten worden berekend, respectievelijk $\{N/Rm_2d * Dm/D\}$ en $\{N/Rm_7d * Dm/D\}$)
- 13 SCP_TBO_Start_Slot_Weegbestanden
(voor zowel Resultaat_SCP als Resultaat_SCP_7dagen wordt een weegbestand aangemaakt)
- 14 SCP_TBO_Start_Slot_Weging_2dagen
SCP_TBO_Start_Slot_Weging_7dagen
(de twee wegingen worden uitgevoerd)
- 15 SCP_TBO_Start_Slot_Gewichten
(de correctie- en eindgewichten worden aan de StartSlotvragenlijst toegevoegd)
- 16 SCP_TBO_Start_Slot_SCP

(selectie van variabelen voor levering van de StartSlotvragenlijst aan het SCP)

4.2 Controleren en uniformeren waarneemdata

Range- en routecontroles zijn niet nodig bij het TBO. De Start- en Slotvragenlijst worden via CAPI (Blaise Computer Assisted Personal Interviewing) waargenomen. De dagboekjes worden ingevoerd in een CADI (Blaise Computer Assisted Data Editing) programma. In Blaise datamodellen worden range- en routes bepaald. Resultaten van Blaise vragenlijsten of invoerprogramma's voldoen altijd aan deze datamodellen. Extra range- en routecontroles zijn achteraf niet nodig.

Bij uniformeren wordt ervoor gezorgd dat de data in één uniforme datastructuur terecht komt die de basis vormt voor het verwerken. Dit betekent dat:

- Data uit de diverse modes in één en dezelfde datastructuur komt. De TBO vragenlijsten zijn alleen via CAPI waargenomen, hier hoeven dus geen datamodellen van verschillende waarneem-modes geuniformeerd te worden. Wel wordt een selectie van variabelen uit de vragenlijst meegenomen in de verwerking. Vragenblokken die niet van toepassing zijn (zoals administratieve blokken voor het monitoren van telefonisch enquêteren), zijn ook niet nodig in de verwerking.
- De datastructuur voorzien is van metadata (volledige beschrijving). Het toevoegen van metadata is wel in verschillende stappen van het TBO gebeurd. Alle variabelen in de bestanden zijn voorzien van labels.

4.3 Verrijken

Verrijken is het overnemen in het onderzoeksbestand van variabelen uit andere bronnen. Daartoe wordt het onderzoeksbestand gekoppeld met de betreffende bronnen.

In paragraaf 3.1 is al aangegeven dat er gegevens uit andere bronnen aan het TBO zijn toegevoegd.

Met het Postaal Wijk Register (PWR) wordt aan het postcodenummer uit de steekproef de gemeentecode van 2011 gekoppeld.

Vervolgens wordt aan de gemeentecode bijbehorende regionale indelingen gekoppeld, gebruik makend van de Statline tabel Gebieden in Nederland (GIR) 2011. Overgenomen worden: landsdeel, provincie, gemeentegrootte en stedelijkheid. Vervolgens wordt provincieplus (provplus) afgeleid, deze variabele bestaat uit de 12 provincies aangevuld met de 4 grote steden.

Uit de Gemeentelijke BasisAdministratie (GBA) worden een aantal variabelen overgenomen, namelijk: AantalKinderen, AantalPersonenInHuishouden, ImputatieCode, PlaatsInHuishouden, ReferentiePersoon, TypeHuishouden, AflGeneratie, AflHerkomst, GBABurgerlijkeStaat, GBAGeslacht, GBALft. Hiervan worden een aantal variabelen afgeleid: HHGrootte, PosHH, PosHH5, SamHH5, Lft6, Lft12, Herk3 en HerkGnt.

Tenslotte wordt uit het CBS integrale inkomensbestand 2010 de variabelen besteedbaar huishoudinkomen (BestInkH), equivalentiefactor (Equi), aantal huishoudleden (AHL), belangrijkste inkomensbron van het huishouden (BBIHAlg1) en de samenstelling van het huishouden (SamHuishouden_R) toegevoegd.

4.4 Afbakenen statistische respons

Wat wel of geen statistische respons is, is onderzoekspecifiek.

De Start- en Slotvragenlijst zijn respons als het Eindresultaat van de vragenlijst volledige respons (Eindresultaat=30) is.

Er bleken 5 Startvragenlijsten voor te komen die niet voldoen aan dit criterium, maar die compleet beantwoord waren en ook geworven voor het invullen van dagboekjes. Deze zijn gecorrigeerd in volledige respons, zodat ze verwerkt kunnen worden. Er zijn 2 Startvragenlijsten met volledige respons afgekeurd (Volledig_Start=nee), omdat de steekproefpersoon niet voorkomt in het geïnterviewde huishouden.

De dagboekjes zijn van voldoende kwaliteit als er minimaal 1 werk- en 1 weekenddag van voldoende kwaliteit is, dat wil zeggen dat op die dagen maximaal 60 minuten onbekende hoofdactiviteit mag voorkomen.

4.5 Gaafmaken (micro)

Gaafmaken is het opsporen en corrigeren van foutieve gegevens in de verwerkingsdata. Bij het corrigeren wordt een waarde van een variabele vervangen door een andere waarde.

Er zijn een aantal identificerende variabelen (WE_ID en NrSchrif) gecorrigeerd. Dit betreft bijvoorbeeld toetsfouten.

Ook kwam het een aantal keer voor dat de interviewer een verkeerd dagboektype had afgegeven, dat aan volwassen onderzoekspersonen een Dagboek OP Kind en aan een jongere onderzoekspersonen een Dagboek OP Volwassene was uitgereikt. Om de dagboekjes correct te kunnen invoeren en verwerken is het nodig het dagboektype te bepalen en te wijzigen.

In de dagboekjes zijn de coderingen en met name de combinaties van activiteiten en lokaties/verplaatsingen en het eventuele gebruik van een pc gecontroleerd. Gevonden problemen zijn verder uitgezocht en automatisch gecorrigeerd.

De eerste onwaarschijnlijkheid was het stadsuitje. Bij plausibiliteitscontrole bleek de culturele participatie in 2011/2012 ten opzichte van 2005/2006 flink toe te nemen. Dat werd veroorzaakt door het stadsuitje. De codeurs bleken verplaatsingen in de stad zonder duidelijke omschrijving als stadsuitje gecodeerd te hebben. Deze gevallen zijn bekeken en waar nodig (geen culturele bezigheid) gecorrigeerd. De codeurs zijn beter geïnstrueerd.

4.6 Imputeren

Bij enquêtes komt het voor dat respondenten op één of meer vragen geen antwoord geven, terwijl dit wel van ze wordt verlangd (WerkStud). Ook kan een variabele worden gevuld met behulp van informatie uit andere variabelen en/of bronnen (Afstanden).

Afstanden

De TBO Startvragenlijst is waargenomen van half maart 2011 tot en met februari 2012. Gekozen is om voor het gehele TBO de gemeentelijke indeling van 2011 te gebruiken. Via de 4-cijferige postcode wordt de gemeentecode 2011 toegevoegd. Ook kunnen met de postcodes (op 6 dan wel 5 dan wel 4 digits) de X- en Y-coördinaten van een adres worden toegevoegd.

Van het woonadres van de respondent zijn zowel de postcode als de gemeentecode bekend (in de vragenlijst, namelijk overgenomen uit de steekproef).

Wanneer de respondent en/of diens partner werkt of de respondent een opleiding volgt, wordt ook van het adres van het werk of de opleiding informatie verzameld. Eerst wordt gevraagd naar de postcode. Wanneer de interviewer een volledige postcode (4 cijfers en 2 letters) invoert, wordt deze postcode geaccepteerd en tijdens het interview niet op correctheid gecontroleerd. Wanneer de interviewer alleen de 4 cijfers invoert, dan wordt nog naar de gemeente gevraagd en is ook een gemeentecode beschikbaar. Tevens wordt dan aan de respondent een (geschatte) afstand gevraagd (enkele reis in kilometers). Deze informatie zou het mogelijk moeten maken afstanden te berekenen.

Allereerst wordt tijdens de verwerking de postcode gecontroleerd op geldigheid.

Helaas bleek een deel van de volledige postcode (4 cijfers en 2 letters) geen geldige postcode te zijn (1). Dan wordt gekeken of de cijfers wel geldig zijn (2). Indien de postcode helemaal ontbreekt of incorrect is en er wel een gemeentecode is, kan ook

met de gemeentecode gewerkt worden. Van alle gemeentecodes is de postcode van het adressenzwaartepunt bekend. Deze postcode wordt dan geïmputeerd (3). Aan de variabelen PostcW_OP_ok, PostcW_P_ok en PostcOpl_ok (Postcode werkadres OP resp. werkadres Partner resp. opleiding OP gecontroleerd) is dit te zien:

- (1) "pc6 ok"
- (2) "pc4 ok"
- (3) "pc4 niet, gemc ok"
- (4) "pc en gemc niet ok".

Alleen bij (4) kan geen afstand toegevoegd worden.

Bij het OViN (Onderzoek Verplaatsingen in Nederland) is een imputatie methode gebruikt voor ontbrekende verplaatsingsafstanden. Deze methode maakt gebruik van de X- en Y-coördinaten van postcodes. Daartussen kan een rechtlijnige afstand tussen vertrek en aankomst berekend worden met de stelling van Pythagoras. Aangezien een af te leggen afstand vrijwel nooit een rechte lijn zal zijn, wordt deze gecorrigeerd met een factor die verschillend is per (hoofd)vervoermiddel en per afstandsklasse.

Deze methode is ook gebruikt voor het TBO. Voor de verplaatsing van de onderzoekspersoon, hetzij naar werk hetzij naar school, is in de Startvragenlijst ook het (hoofd)vervoermiddel gevraagd. Voor het woonwerkverkeer van de partner is dit niet gevraagd. Daarom wordt voor alle drie de afstanden de afstand "over de weg" als in "met de auto" bepaald (AfstWerk_OViN_auto, AfstWrkP_OViN_auto, AfstOpl_OViN_auto). Voor de onderzoekspersoon is daarnaast ook nog een afstand met het gebruikte vervoermiddel bepaald (AfstWerk_OViN_hvm, AfstOpl_OViN_hvm).

WerkStud

De vraag WerkStud is relatief slecht beantwoord. Blijkbaar is het de respondent niet duidelijk wat met de vraag bedoeld wordt. In de Startvragenlijst is al vastgesteld of de respondent werkt en/of een opleiding volgt. In het Dagboek OP voor volwassenen wordt elke dag gevraagd "Bent u werkzaam, scholier of student?". Vervolgens wordt gevraagd "Wat voor dag was dit voor u?" en daarbij is het mogelijk te antwoorden "een normale werkdag" of "een normale schooldag" of "een dag waarop ik in principe nooit naar werk of school ga".

Van de 2.021 Dagboekjes OP zijn er 203 Kinder en 1.818 Volwassene Dagboekjes. Deze volwassenen vullen 12.726 keer de dagboekvragen in. Op 532 dagen ontbreekt een antwoord bij WerkStud.

In eerste instantie is gekeken naar de antwoorden van deze respondenten op andere dagen op de vraag WerkStud. Daaruit bleek dat sommige respondenten op de andere

dagen altijd hetzelfde antwoord gaven, maar dat er ook respondenten waren die op verschillende dagen verschillende antwoorden gaven. Daarom is besloten dat de antwoorden van de respondent op andere dagen op de vraag WerkStud niet zonder meer gebruikt kan worden om de ontbrekende antwoorden te imputeren. Daarom zijn er een aantal hulpvariabelen aangemaakt.

Uit de Startvragenlijst:

Afl_Lft_OP	BetWerk.Afl_BetWerkNu	Ow_Act.Opleiding	hlp_WerkStud_start
< 15			3=nvt
>14	1		1=ja
	2	1	1=ja
		2, RF, DK	2=nee

Uit het Dagboek OP:

Dag_Soort	Verpl_buit	hlp_WerkStud_vragen
RESPONSE (1..6)		1=ja
EMPTY	1	1=ja
	2, EMPTY	2=nee

Op de dag dat de variabele WerkStud onbeantwoord is [ii], wordt gekeken naar de opgegeven activiteiten. Bij elke activiteit wordt aangegeven waar de respondent was of welke verplaatsing het betreft. Lokatie code 13 betekent "op school of op het werk". Dan wordt de hulpvariabele hlp_WerkStud_dag[ii] voor die dag "ja".

Er wordt ook gekeken naar de activiteiten en de lokatie van de respondent op andere dagen. Als op één van de andere dagen een Lokatie code 13 voorkomt, wordt de hulpvariabele hlp_WerkStud_dagen "ja".

dag[ii]	episode[jjj]	Lokatie	hlp_WerkStud_dag[ii]	hlp_WerkStud_dagen
1..7	1..144	13	1=ja	1=ja
		else	2=nee	2=nee

Vervolgens is met een combinatie tussen deze hulpvariabelen de ontbrekende waarde bij WerkStud geïmputeerd. In de variabele WerkStud_imp wordt vastgehouden op basis van welke informatie WerkStud geïmputeerd is.

hlp_WerkStud_*					WerkStud_imp	
_start	_vragen	_dagen	_dag[i]			
√	√	√	√	gelijk	1	Startvrgl, vragen, episodes dagen en dag gelijk
√		√	√	gelijk	2	Startvrgl, episodes dagen en dag gelijk
	√	√	√	gelijk	3	vragen, episodes dagen en dag gelijk
		√	√	gelijk	4	episodes dagen en dag gelijk

WerkStud_imp	aantal imputaties
1	300
2	62
3	165
4	5
	532

4.7 Afleiden

In “TBO_2011_2012_beschrijving_verwerkingsproces.pdf” en in “Afleidingen_TBO_2011_2012.pdf” staan de gebruikte afleidingen beschreven.

4.8 Controleren en corrigeren

In “Controle_correctie_TBO_2011_2012.pdf” worden de controles en correcties beschreven die in Stap3_03 en Stap3-04 plaatsvinden.

4.9 Wegen

TBO 2011/2012

Het weegmodel voor het TBO 2011/2012 is bepaald in een vooronderzoek. Het resultaat is beschreven in “Vooronderzoek Weging Tijdsbestedingsonderzoek 2011, Erwin Vondenhoff, juni 2012” (“TBO_2011_2012_vooronderzoek_weging.pdf”).

Dat levert het volgende weegmodel op:

*geslacht(2) × leeftijd(12) +
burgerlijke staat(4) +
stedelijkheidsgraad (5) +
geslacht(2) × werkzaamheid(3) +
samenstelling huishouden(5) +
huishoudensinkomen(6) +
huishoudensgrootte(5) +
positie in huishouden(5) +
invulmaand(12).*

Met insluitgewicht:

$$d_i^{II} = \frac{ND_{m(i)}}{r_{m(i)}D}$$

De populatie-aantallen zijn afgeleid uit het maandbestand 2011-09 van de Enquête Beroepsbevolking (EBB).

De weging van het TBO bij Resultaat_SCP (2 dagen) verloopt zonder problemen en levert het volgende resultaat op:

Correction weights information	
Minimum weight	0.03
Median	0.976
First quartile	0.789
Third quartile	1.204
Maximum weight	2.557
Standard deviation	0.353
Number of negative weights	0
Final weights information	
Minimum weight	217
Median	6908.962
First quartile	5376.658
Third quartile	8789.819
Maximum weight	24443.97
Standard deviation	2980.251
Number of negative weights	0

De weging van het TBO bij Resultaat_SCP_7dagen levert in eerste instantie (ongewenste) negatieve gewichten op:

Correction weights information	
Minimum weight	-0.27
Median	0.965
First quartile	0.753
Third quartile	1.233
Maximum weight	2.623
Standard deviation	0.389
Number of negative weights	4
Final weights information	
Minimum weight	-2148.92
Median	7625.92
First quartile	5750.502
Third quartile	9946.088
Maximum weight	28183.95
Standard deviation	3568.535
Number of negative weights	4

Daarom is bij deze weging de instelling van Bascula aangepast (BOUNDING), daarna komen er geen negatieve gewichten meer voor:

Correction weights information	
Minimum weight	0.16
Median	0.961
First quartile	0.751
Third quartile	1.232
Maximum weight	2.66
Standard deviation	0.389
Number of negative weights	0
Final weights information	
Minimum weight	945.72
Median	7590.032
First quartile	5692.563
Third quartile	9947.021
Maximum weight	28534.59
Standard deviation	3563.199
Number of negative weights	0

TBO 2005/2006

Voor het vergroten van de vergelijkingsmogelijkheden van de uitkomsten van het TBO 2011/2012 met het voorgaande TBO 2005/2006, heeft het SCP verzocht het TBO 2005/2006 te wegen op dezelfde wijze als het TBO 2011/2012.

De variabelen burgerlijke staat en inkomen ontbreken bij het TBO 2005/2006. Dat leverde problemen op bij de weging. Daarom is stapsgewijs uitgeprobeerd met welke van de wel beschikbare variabelen het beste resultaat bereikt kon worden. Behalve burgerlijke staat en inkomen, moest ook SamHH5 uit het weegmodel verwijderd worden.

De overige variabelen en insluit- of startgewichten konden op dezelfde wijze worden afgeleid. Er zijn twee wegingen uitgevoerd: met de regionale indelingen van 2006 en

die van 2011. Dat maakt een vergelijking van de uitkomsten mogelijk met en zonder veranderingen in die regionale indelingen.

Weegmodel:

Geslacht x Werkzaamheid

Geslacht x Lft12

HHGrootte

PosHH5

Invul_Maand

StedGem

Correction weights information	
Minimum weight	0.1
Median	0.87
First quartile	0.634
Third quartile	1.164
Maximum weight	6.507
Standard deviation	0.702
Number of negative weights	0
Final weights information	
Minimum weight	768.48
Median	6527.187
First quartile	4751.268
Third quartile	8813.179
Maximum weight	47232.9
Standard deviation	5344.344
Number of negative weights	0

4.10 Maken tabellen, aggregeren

Van het TBO 2011/2012 zijn geen tabellen gemaakt.

4.11 Gaafmaken (macro)

Op macro niveau zijn geen gaafmaak acties uitgevoerd.

4.12 Controleren tabellen: plausibiliteitanalyse

Niet van toepassing.

4.13 Statistisch beveiligen

Er is nog geen statistische beveiliging toegepast. Dat zal gebeuren op de DANS bestanden. De inhoud van deze bestanden moet nog met het SCP afgestemd worden.

4.14 Output van verwerken

Alle stappen die worden beschreven in paragraaf 4.1 hebben een uitvoer, die opgeslagen is. Aan het SCP zijn de bestanden zoals beschreven in paragraaf 3.7 opgeleverd.

5 Publiceren

Er zijn afspraken tussen het SCP en het CBS over publicaties. Een daarvan is dat het CBS pas mag publiceren nadat het SCP de eerste resultaten heeft gepubliceerd.