

HANDLEIDING CBS MICRODATA

HANDLEIDING VOOR HET GEBRUIK VAN CBS MICRODATA
VOOR REPRESENTATIVITEITSONDERZOEK

HANDLEIDING

seo • economisch onderzoek

AUTEURS

JUSTUS VAN KESTEREN, MENNO POMP

IN OPDRACHT VAN

MINISTERIE VAN SOCIALE ZAKEN EN WERKGELEGENHEID (SZW)

AMSTERDAM, 1 FEBRUARI 2023

SEO-notitie nr. 2023-09

Informatie & Disclaimer

SEO Economisch Onderzoek heeft op de verkregen informatie en data geen onderzoek uitgevoerd dat het karakter draagt van een accountantscontrole of due diligence. SEO is niet verantwoordelijk voor fouten of omissies in de verkregen informatie en data.

Copyright © 2022 SEO Amsterdam.

Alle rechten voorbehouden. Het is geoorloofd gegevens uit dit rapport te gebruiken in artikelen, onderzoeken en collegesyllabi, mits daarbij de bron duidelijk en nauwkeurig wordt vermeld. Gegevens uit dit rapport mogen niet voor commerciële doeleinden gebruikt worden zonder voorafgaande toestemming van de auteur(s). Toestemming kan worden verkregen via secretariaat@seo.nl.

Inleiding

De handleiding bevat een stappenplan voor het gebruik van CBS Microdata voor representativiteitsonderzoek. Hierbij is onder andere ingegaan op de benodigde bewerkingen van de data, het berekenen van het representativiteitscijfer, het doorvoeren van correcties en het presenteren, exporteren en bewaren van de resultaten. Verder bevat de notitie een indicatie van de kosten en de doorlooptijd van representativiteitsonderzoek gebaseerd op CBS Microdata. De handleiding is voornamelijk bedoeld voor onderzoeksbureaus die de analyse gaan uitvoeren, zodat het representativiteitscijfer consistent berekend kan worden op basis van de CBS Microdata.

Stappenplan gebruik CBS Microdata

Het is op basis van de CBS Microdata en de daarin vermelde cao-codes mogelijk om het representativiteitscijfer voor een cao-sector te berekenen. Een cao-partij kan doorgaans zelf geen gebruikmaken van de CBS Microdata. De cao-partij kan wel een onderzoeksbureau met toegang tot CBS Microdata inschakelen om het representativiteitscijfer te laten berekenen. Hierbij kan het onderzoeksbureau onderstaande stappen doorlopen.

Stap 1: Aanvragen CBS Microdataproject

Dien een projectaanvraag in bij CBS door een 1) onderzoeksvoorstel 2) publicatieplan en 3) formulier met de AVG-grondslag van het onderzoek te mailen naar microdata@cbs.nl. De formulieren zijn te vinden en downloaden via

Onderzoeksvoorstel

Geef in het onderzoeksvoorstel aan welke bestanden en bestandsonderwerpen nodig zijn voor het onderzoek. Een bestand is bijvoorbeeld Spolisbus met banen en lonen van werknemers of Betab met informatie over de bedrijven waar werknemers werken, en het bestandsonderwerp is in dat geval bijvoorbeeld de Spolis. CBS brengt geen kosten in rekening per bestand, maar per bestandsonderwerp. Het kost dus geen extra geld om een extra bestand aan te vragen binnen een bestandsonderwerp waar al toegang tot is verleend. Let er verder op dat er bij het aanvragen van (extra) bestanden via microdata@cbs.nl altijd een controle plaatsvindt door de relatiebeheerder bij het CBS of het bestand wel binnen de doelbinding van het project valt. Beargumenteer daarom altijd goed waarom het bestand nodig is.

Vraag voor het representativiteitsonderzoek in ieder geval de polisadministratie (Spolis) en het Algemeen Bedrijvenregister (ABR) aan. De polisadministratie bevat gegevens over banen en lonen van werknemers bij Nederlandse bedrijven, waaronder de cao-code die is ingevuld door de werkgever bij de loonaangifte. Daarnaast bevat de polisadministratie een bedrijfsidentificatienummer (BEID), waardoor het mogelijk is om het bedrijf te identificeren waarvoor een werknemer werkt. Het ABR bevat ook het bedrijfsidentificatienummer, waardoor het mogelijk is om via dit nummer beide bestanden aan elkaar te koppelen. Dit is noodzakelijk voor het berekenen van het representativiteitscijfer (zie Stap 4).

Vraag verder bestanden aan die nodig zijn om te corrigeren voor registratiefouten van werkgevers (zie Tabel 1 voor een overzicht). Laat de keuze in bestanden afhangen van de werknemers- en werkgeversdefinities uit de cao. Kent de cao bijvoorbeeld een leeftijdsgrens van 67 jaar, gebruik dan de personenadministratie (Gbapersonktab) om de leeftijd te achterhalen van werknemers waarvoor de cao-code is geregistreerd. Hierdoor is het mogelijk om werknemers die ouder dan 67 jaar zijn, maar waarvoor de cao-code toch is ingevuld, te verwijderen uit de bestanden. Meer informatie over de verschillende bestanden is te vinden in de [online catalogus](#) van CBS.

Tabel 1 Relevante bestanden uit de CBS Microdata

Beschikbare informatie voor verdere afbakening/correctie werkingsfeer	Bestanden in CBS Microdata	Relevante variabelen in bestanden
Leeftijd	Gbapersonktab	Gbageboortejaar, rinpersoon, rinpersoons
Contractvorm (regulier, oproep, uitzend, stagiair, DGA, wsw-er)	Spolisbus	Soortbaan, rinpersoon, rinpersoons
Student/scholier	Onderwijsinschrtab	Rinpersoon, rinpersoons ¹
Inkomen	Spolisbus	Sbasisloon, rinpersoon, rinpersoons
Sector / SBI-code	Betab	Sbi2008vjjj, BEID
Bedrijfs grootte	Betab	Gksbs, BEID
Regio/woonplaats	Vslgtab & Gbaadresobjectbus	Soortobjectnummer, rinobjectnummer, gemjjjj, rinpersoon, rinpersoons ²

Bron: CBS Microdata, bewerking SEO Economisch Onderzoek (2023)

AVG-grondslag

Onderbouw in het AVG-formulier de AVG-grondslag van het onderzoek. De opdrachtgever (de cao-partij) is in dit geval gezamenlijk met CBS verantwoordelijk voor de verwerking van gegevens. De verwerkingsverantwoordelijke dient een geldige grondslag voor de verwerking te hebben volgens artikel 6 lid 1 AVG. Voor onderzoek op basis van CBS Microdata kunnen alleen grondslag e of f worden gebruikt. Voor de cao-partij geldt dat de verwerking noodzakelijk is voor een taak van algemeen belang (grondslag e), omdat zij wettelijk verplicht zijn om aan het meerderheidsvereiste te voldoen. Dit maakt verwerking van gegevens over de omvang van de sector en de omvang van het werknemersbestand bij gebonden werkgevers noodzakelijk. Verwijs in de onderbouwing naar de [Wet op het algemeen verbindend en het onverbindend verklaren van bepaling van cao's](#), het [Besluit aanmelding van collectieve arbeidsovereenkomsten](#) en het verzoek om algemeenverbindendverklaring en het [Toetsingskader algemeenverbindendverklaring AVV](#).

Publicatieplan

Stel een publicatieplan op waarin wordt ingegaan op het moment en de wijze waarop de onderzoeksresultaten openbaar worden gemaakt. Informeer CBS gedurende het onderzoek als wordt afgeweken van de publicatiedatum. Het CBS ontvangt binnen vier weken na publicatie (de link naar) de betreffende publicatie(s).

Stap 2: Ledenbestand uploaden bij CBS

Upload daarna de KvK-registraties van de direct aan de cao gebonden werkgevers in de beveiligde uploadfaciliteit van het CBS. Na de upload versleutelt het CBS de KvK-nummers tot onherleidbare versleutelde codes die het

¹ Het bestand 'Onderwijsinschrtab' kan gekoppeld worden aan Spolisbus o.b.v. rinpersoon & rinpersoons. Hieruit wordt duidelijk of een werknemer op dat moment ook ingeschreven staat bij een opleiding.

² Vslgtab kan gekoppeld worden aan gbaadresobjectbus op basis van soortobjectnummer en rinobjectnummer. De variabele gemjjjj geeft dan de woonplaats weer.

onderzoeksbureau vervolgens kan koppelen aan CBS-microdatabestanden. Box 1 biedt praktische informatie over het uploaden van de KvK-registratie bij CBS.

Verzeker hierbij vóór de upload bij CBS dat de KvK-registratie compleet en betrouwbaar is. Door de versleuteling van CBS is het immers niet goed mogelijk om in de CBS-omgeving vast te stellen bij welke werkgever de koppeling niet goed gaat, en bij welke werkgever de koppeling juist wel goed gaat.

Gebruik hiervoor de volgende methoden:

1. Controleer de juistheid van de KvK-registratie

Controleer op basis van de bedrijfsnaam en adresgegevens de juistheid van het bijbehorende KvK-nummer in de ledenlijst. Websites als www.openkvk.nl bieden mogelijkheden om dit snel en eenvoudig te doen. Ook is het mogelijk een zelfgemaakt geautomatiseerd script te gebruiken om op basis van de bedrijfsnaam en adresgegevens bijbehorende KvK-nummers te vinden. Hierdoor is het mogelijk om de lijst met KvK-nummers te schonen van foutieve, ontbrekende of gedateerde KvK-nummers.

2. Controleer de volledigheid van de KvK-registratie

Controleer of alle KvK-nummers van één lid zijn opgenomen. Dit is vooral relevant voor grote werkgevers, omdat zij vaak meerdere KvK-nummers hanteren. Mogelijk zijn niet alle KvK-nummers te koppelen aan de polisadministratie, bijvoorbeeld doordat een KvK-nummer niet gebruikt wordt voor economische activiteiten, maar als financiële holding of intern transport. Hierdoor kan er sprake zijn van koppelverlies. Beperk het koppelverlies door alle KvK-nummers van ieder lid op te nemen in de KvK-registratie. Let wel: het gaat hier nadrukkelijk *niet* om KvK-vestigingsnummers. KvK-vestigingsnummers zijn niet te koppelen aan de CBS Microdata.

Box 1 Uploaden KvK-registratie

Lever het bestand aan als CSV- of SAV-bestand. Bij een CSV-extensie dient bijgaand een volledige bestandsomschrijving te worden aangeleverd met daarin het gebruikte scheidingsteken,³ het decimaalscheidingsteken, welke variabelen numeriek dan wel tekstueel zijn, en wat de variabelen betekenen en eventueel waarom ze nodig zijn. Variabelen die gemakkelijk herleidbaar zijn (zoals bedrijfsnaam, e-mailadres of de naam van de respondent), dienen verwijderd te worden. Het vormt geen probleem om KvK-nummers zonder zogeheten 'voorloopnul' te uploaden.⁴ Observaties zonder KvK-nummer kunnen niet gekoppeld worden en kunnen voor de upload al verwijderd worden. Mochten er meerdere KvK-nummers van hetzelfde lid zijn, voeg dan extra variabelen in het bestand toe voor het tweede KvK-nummer, derde KvK-nummer etc. Vul tot slot het formulier 'Upload informatie' (XLSX) in. Het gehele voorschrift van het bestandsformaat en de uploadprocedure van het CBS is [hier](#) te vinden

Upload vervolgens de KvK-registratie, de bijbehorende beschrijvingen en het formulier als gecomprimeerd ZIP-bestand in de volgende omgeving: <https://databestanden.cbs.nl>. Inlogcodes zijn te verkrijgen door een mail te versturen naar Microdata@cbs.nl met daarin bij het onderwerp: projectnummer, inlogcodes en @doc (voorbeeld "1001 inlogcodes @doc"). Stuur tegelijkertijd met de upload een mail naar Microdata@cbs.nl met daarin de reden voor het uploaden van het bestand (korte omschrijving over het onderzoek) en de verwachtingen van het CBS (de KvK-nummers versleutelen en het bestanden uploaden in de Microdata-omgeving). In de mail bij het onderwerp "projectnummer en @doc" vermelden.

³ Let er daarbij op dat het scheidingsteken (bijv. ";") niet als teken voorkomt in een cel.

⁴ In CSV-bestanden worden numerieke cellen beginnend met een nul automatisch ingekort. In het geval van KvK-nummers verdwijnt deze 'voorloopnul' dan, zoals bij bijvoorbeeld '02345678' dat '2345678' wordt. Het CBS corrigeert dit.

Stap 3: Databestand gereed maken

Doorloop de volgende stappen om een databestand te construeren voor het berekenen van het representativiteitscijfer:

1. Inladen polisadministratie

Laad de banen uit de polisadministratie (Spolisbus) in voor het gewenste peiljaar, de gewenste peilmaand⁵ en de gewenste cao-code. Kies daarbij de peilmaand die overeenkomt met de peilmaand van het ledenbestand. Stel bijvoorbeeld dat men het representativiteitscijfer wil berekenen voor de maand juli 2021 voor een bedrijfstak-cao met cao-code 0001. Een overzicht van cao-codes is te vinden via de [website](#) van CBS. Laad dan alle banen in waarvoor cao-code 0001 is geregistreerd (met gebruik van de variabele BedrijfstakCAO) uit de polisadministratie (Spolisbus) uit 2021, onder de restrictie dat de aanvang van de inkomstenopgave minimaal 1 juli is (met gebruikmaking van variabele Sdatumaanvangiko) en het einde van de inkomstenopgave maximaal 31 juli is (m.g.v. Sdatumeindeio). Vergeet niet andere relevante variabelen mee te nemen bij het inladen van de polisadministratie, waaronder in ieder geval het bedrijfsidentificatienummer (Sbeid), de versleutelde BSN-nummers (Rinpersoon, Rinpersoons) en de benodigde variabelen uit Tabel 1. Sla het resterende bestand op. Dit bestand wordt hierna aangeduid met *werknemersbestand*.

2. Koppelen ledenbestand aan polisadministratie

Maak vervolgens onderscheid tussen werknemers die wel en niet werken voor direct door de cao gebonden werkgevers. Gebruik hiervoor de versleutelde KvK-registratie, het algemeen bedrijvenregister (ABR) en het werknemersbestand. Doorloop daarbij de volgende stappen:

- Koppel de versleutelde KvK-registratie aan het ABR-bestand CBS_persoon_KvK_FIN2021V1 (via de variabele KvkNummer) en haal hieruit de overeenkomstige waarde van de variabele CbsPersoonIdentificatie. Sla een bestand op met de CbsPersoonIdentificatie van alle gekoppelde leden;
- Koppel dit bestand aan het ABR-bestand ABR_BE_PERSOON2021V1 (via de variabele CbsPersoonIdentificatie) en haal hieruit de overeenkomstige waarde van de variabele BE_ID. Dit is het bedrijfsidentificatienummer. Sla een bestand op met het bedrijfsidentificatienummer van alle gekoppelde leden;
- Koppel nu het bestand met het bedrijfsidentificatienummer van alle leden aan het werknemersbestand (op basis van de variabele BE_ID). Maak een nieuwe variabele aan die weergeeft of een werknemer werkt voor een werkgever die lid is van een werkgeversorganisatie (gekoppeld bedrijfsidentificatienummer) en die geen lid is van een werkgeversorganisatie (niet-gekoppeld bedrijfsidentificatienummer).

Let wel: de bestands- en variabelenamen kunnen verschillen van jaar tot jaar. Alle bestanden en variabelen in het voorbeeld hierboven corresponderen met het jaar 2021. De corresponderende namen in andere jaren zijn opgenomen in Bijlage A.

3. Optioneel: beperken koppolverlies

Bij een onvolledige koppeling is er sprake van koppolverlies. Er zijn twee manieren om het koppolverlies te beperken:

⁵ Een peilmaand heeft de voorkeur boven een exacte peildatum. Neemt men bijvoorbeeld de peildatum 31 december, dan ontbreken personen die tot en met december een baan hadden en onder zogeheten vierwekenloners vallen (per vier weken loon uitbetaald krijgen) en per 1 januari van het jaar erna een nieuwe baan starten. De vierwekenperiode eindigt namelijk precies op 30 december (dit verschilt per kalenderjaar) en de nieuwe baan begint op 1 januari. Doordat de vierwekenperiode toevallig in december 2019 eindigt op de 30e, komt men uit op een lager aantal banen, terwijl dit eigenlijk wel mensen zijn die meegeteld moeten worden als vallend onder de werkingssfeer van de cao. Daarom is het beter om een peilmaand te hanteren in analyses op basis van CBS Microdata.

- *Verbeteren KvK-registratie* Controleer de KvK-registratie nogmaals op foutieve, gedateerde of ontbrekende KvK-nummers (zie Stap 2). Upload vervolgens de aangepaste KvK-registratie bij CBS en doorloop bovenstaande stappen opnieuw met de aangepaste KvK-registratie;
- *Uitbreiding koppeling* Voeg een extra koppelronde met een bestand uit een ander jaar toe voor de direct gebonden werkgevers die nog niet zijn gekoppeld. Maak bijvoorbeeld een koppeling met het ABR-bestand uit 2020 in plaats van het ABR-bestand uit 2021. Dit biedt in sommige gevallen uitkomst, bijvoorbeeld bij gedateerde KvK-nummers.

4. Optioneel: overige relevante data koppelen

Verrijk het werknemersbestand met benodigde informatie om te corrigeren voor registratiefouten in de berekeningen. Het kan dan gaan om informatie over inkomen, leeftijd, contractvorm, maar ook of iemand als scholier/student actief is naast het werk in loondienst. Tabel 1 geeft aan in welke microdatabestanden informatie te vinden is over deze kenmerken. Gebruik de versleutelde BSN-nummer (variabelen Rinpersoon en Rinpersoons) voor het koppelen van deze variabelen aan het werknemersbestand.

5. Ontdubbelen

Maak het databestand uniek op persoonsniveau. Het Toetsingskader stelt immers dat het aantal personen, en niet het aantal banen, gebruikt moet worden in de representativiteitsberekening (zie ook Box 1). Het is echter mogelijk dat één persoon meerdere keren voorkomt in het werknemersbestand, omdat elke baan een eigen registratie kent in de polisadministratie. 'Ontdubbel' daarom het databestand. Dat houdt in: verwijder alle dubbele BSN-nummers (Rinpersoonnummers).

Hierna is het databestand gereed voor het berekenen van het representativiteitscijfer.

Stap 4: Berekening representativiteitscijfer

Bereken zowel het representativiteitscijfer vóór als na correctie voor registratiefouten. Hierdoor zijn de bewerkingen reproduceerbaar.

Resultaten vóór correctie

Bereken de relevante aantallen vóór correctie als volgt:

- Aantal direct gebonden werknemers: optellen van werknemers met de relevante cao-code werkzaam voor een direct door de cao gebonden werkgever;
- Aantal direct gebonden werkgevers: optellen aantal direct aan de cao gebonden werkgevers met minimaal één werknemer in dienst met de eigen cao-code;⁶
- Aantal werknemers in cao-sector: optellen van alle werknemers met de relevante cao-code;
- Aantal werkgevers in de cao-sector: optellen aantal werkgevers dat minimaal één werknemer in dienst heeft met de eigen cao-code.

Bepaal het representativiteitscijfer door het aantal direct gebonden werknemers te delen door het aantal werknemers in de cao-sector. Noteer daarbij de peilmaand uit de polisadministratie.

Merk op dat het bij werknemersaantallen gaat om het optellen van unieke BSN-nummers (rinpersoonnummers), en bij werkgevers om het optellen van unieke bedrijfsidentificatienummers. Door zowel voor de direct als indirect gebonden werkgevers en werknemers deze aanpak te hanteren, zijn de berekeningen intern consistent.

⁶ Immers, de werkgever mag alleen de cao-code voeren voor een werknemer als de werkgever voldoet aan de werkgeversdefinitie van de cao.

Resultaten na correctie

Verwijder vervolgens werkgevers en werknemers uit het bestand waarvoor ten onrechte de cao-code is opgevoerd. In sommige cao's zijn bijvoorbeeld scholieren en studenten uitgezonderd, in andere cao's weer werknemers die meer verdienen dan een bepaalde inkomensgrens. In de meeste cao's zijn stagiaires en DGA's uitgezonderd, dus verwijder hen in ieder geval uit het bestand. Houd bij wat de omvang is van de verwijderde groepen (bijv. aantal DGA's), zodat de bewerkingen zijn te reproduceren (zie ook Stap 6).

Bereken vervolgens het representativiteitscijfer op basis van het aangepaste databestand:

- Aantal direct gebonden werknemers: optellen van werknemers met de relevante cao-code werkzaam voor een gekoppeld lid van een gebonden werkgeversorganisatie (excl. uitgezonderde werknemers of werkgevers);
- Aantal direct gebonden werkgevers: optellen aantal leden van een gebonden werkgeversorganisatie die minimaal één werknemer in dienst heeft met de eigen cao-code (excl. uitgezonderde werknemers of werkgevers);
- Aantal werknemers onder werkingssfeer: optellen van alle werknemers met de relevante cao-code (excl. uitgezonderde werknemers of werkgevers);
- Aantal werkgevers onder werkingssfeer: optellen aantal werkgevers dat minimaal één werknemer in dienst heeft met de eigen cao-code (excl. uitgezonderde werknemers of werkgevers).

Corrigeer daarna voor het eventuele resterende koppelverlies (na stap 3). Dat kan door een veronderstelling te maken van het aantal werknemers dat werkt bij niet-gekoppelde werkgevers. Breng hierbij een bandbreedte aan in de berekeningen, waarmee de onzekerheidsmarge en beperking van de gemaakte veronderstelling worden onderkend. Veronderstel bijvoorbeeld dat bij de niet-gekoppelde werkgevers (i) evenveel werknemers werken als gemiddeld bij direct gebonden werkgevers (ii) evenveel werknemers werken als gemiddeld bij indirect gebonden werkgevers. Bereken vervolgens het representativiteitscijfer onder veronderstelling (i) en (ii) om tot een bandbreedte te komen van het representativiteitscijfer.

Stap 5: Exporteren bestanden

Maak de uitkomsten van Stap 4 vervolgens gereed voor export uit de Microdata-omgeving. Hiervoor dienen de gegevens aan een aantal richtlijnen te voldoen die gecontroleerd worden door het CBS. De richtlijnen en het complete stappenplan voor het exporteren van de output zijn te vinden via de [website](#) van CBS. Zo mogen de gegevens van individuele observaties niet onthuld worden en dienen alle gegevens gebaseerd te zijn op minstens 10 observaties.

Exporteer vervolgens de bestanden uit de CBS-omgeving. Voeg een ingevuld 'Standaardformulier Outputcontrole' toe. In dit formulier wordt onder andere aangegeven of voldaan wordt aan de voorwaarden, welke bestanden gebruikt zijn, wat het doel van de analyses is, en waar de output uit bestaat. Mail naar Microdata@cbs.nl als alles klaarstaat, en vermeld dat er een output klaarstaat. Vermeld daarbij ook het projectnummer. CBS deelt vervolgens de geëxporteerde gegevens met de onderzoekers.

Stap 6: Bewaren bestanden

Bewaar de resultaten van het representativiteitsonderzoek en voorzie de resultaten van toelichting over de gemaakte keuzes en de bewerkingen van de bestanden:

- **Toelichting afbakening werkgevers en werknemers in bron**
Licht toe waarom de grenzen van het domein waarop de CBS Microdata betrekking hebben in lijn zijn met de werkingssfeer van de cao en hoe rekening wordt gehouden met eventuele registratiefouten;

- **Ledenbestand**
Bewaar het ledenbestand dat gebruikt is voor de koppeling. Hiermee kan de (volledigheid van de) koppeling gecontroleerd worden. Noteer daarbij de peildatum van het ledenbestand;
- **De resultaten vóór correcties**
Bewaar de resultaten vóór correcties. Splits de resultaten uit naar verschillende kenmerken van de werkgevers en werknemers zodat zichtbaar wordt in welke mate er registratiefouten zijn gemaakt. Noteer de gehanteerde peilmaand van de polisadministratie. Dit is de maand waarop de cijfers betrekking hebben;
- **Toelichting op gemaakte koppeling**
Licht toe of het gelukt is het gehele ledenbestand te koppelen aan de CBS Microdata. Zijn alle leden gekoppeld aan de registratiebestanden, of zijn er ook leden die niet gekoppeld konden worden? Als de koppeling niet volledig was, dan is het van belang om dit toe te lichten en de gevolgen voor het representativiteitscijfer te beschrijven;
- **Toelichting op correcties**
Laat zien wat de omvang en oorsprong zijn van de doorgevoerde correcties. In andere woorden: laat zien welke groepen werkgevers en werknemers zijn verwijderd ten opzichte van het bestand vóór correcties. Laat ook zien welke veronderstellingen zijn gebruikt (bijv. om te corrigeren voor eventueel koppelverlies). Onderbouw waarom elk van de veronderstellingen of vuistregels aannemelijk is;
- **De resultaten na correcties**
Bewaar de resultaten na correcties. Splits de resultaten uit naar verschillende kenmerken van de werkgevers en werknemers zodat de veranderingen ten opzichte van het bestand vóór correcties zichtbaar worden. Noteer de gehanteerde peilmaand van de polisadministratie. Dit is de maand waarop de cijfers betrekking hebben;
- **Toelichting op onzekerheden**
Onderbouw in het geval van onzekerheden wat de gevolgen zijn van deze onzekerheid op het representativiteitscijfer. Laat zien wat de gevolgen zijn van deze onzekerheden op het representativiteitscijfer door verschillende veronderstellingen met betrekking tot de onzekerheden te maken. Het kan helpen om een kleine bandbreedte aan te brengen, waarmee de onzekerheidsmarge rondom de schatting wordt weergegeven;
- **Syntax analyses**
Bewaar de syntaxbestanden (zoals Do-files of R-scripts), zodat een derde partij het onderzoek op basis van CBS Microdata kan reproduceren. Het is mogelijk syntaxbestanden kosteloos te exporten vanuit de CBS-omgeving.

Deel deze bestanden en toelichtingen vervolgens met de opdrachtgever (de cao-partij).

Stap 7: Opstellen representativiteitsopgave en bewaren bestanden

Vervolgens kan de cao-partij de informatie uit stap 6 gebruiken voor het opstellen van de representativiteitsopgave. De cao-partij dient in de representativiteitsopgave de resultaten na correctie op te nemen, evenals de toelichting op de gemaakte koppeling, de correcties en onzekerheden in de berekeningen. Bovendien zijn cao-partijen verplicht de informatie uit stap 6 te bewaren voor het periodieke onderzoek naar de kwaliteit van representativiteitsopgaven.

Kosten en doorlooptijd

De kosten voor het gebruik van CBS-data zijn sterk afhankelijk van het aantal bestanden dat nodig is, het aantal (ervaren⁷) onderzoekers en de doorlooptijd van het onderzoek. Bij iedere aanvraag maakt CBS daarom een kostenoverzicht op maat met de kosten voor het desbetreffende project. Op dit moment kost het opstarten van een CBS-Microdataproject 1.800 euro (zie Tabel 2). Daar komen nog kosten bij van 180 euro per bestandsonderwerp. In het geval van representativiteitsonderzoek zijn in ieder geval twee bestandsonderwerpen nodig: het algemeen bedrijvenregister (ABR) en de polisadministratie (Spolisbus). Daarnaast bedragen de kosten van het uploaden en versleutelen van KvK-nummers 250 euro. Gedurende het onderzoek zijn er nog lopende kosten voor ondersteuning. Het gaat dan om maandelijkse kosten van 18 euro per bestandsonderwerp en 122 euro per onderzoeker. Tot slot zijn de kosten voor het aanbieden van output 250 euro per output. Een uitgebreid kostenoverzicht is te vinden op de [website](#) van CBS. De kosten kunnen ieder jaar per 1 januari gewijzigd worden.

CBS streeft ernaar de projectaanvraag binnen één maand na aanlevering van de benodigde gegevens goed te keuren.⁸ Zo nodig houdt het CBS een wachtlijst aan. Houd er rekening mee dat hieraan een periode van voorbereiding voorafgaat binnen de eigen organisatie, bijvoorbeeld voor het aanleveren van het onderzoeksvoorstel, de AVG-grondslag en het publicatieplan (zie stap 1). Na opening van het project kan de KvK-registratie worden aangeleverd (zie stap 2). Vervolgens heeft CBS maximaal twee weken nodig om de versleutelde bestanden beschikbaar te stellen in de beveiligde CBS-omgeving. Daarna kunnen de onderzoekers de analyses uitvoeren. De duur hiervan hangt af van de ervaring van de onderzoekers, de kwaliteit van de KvK-registraties en de benodigde correcties die uitgevoerd moeten worden. Na afronding van de analyses kunnen de onderzoekers de output aanbieden. Het kost CBS doorgaans één werkdag om de output te bewerken en vrij te geven.

Tabel 2 Kosten opstarten nieuw project voor representativiteitsonderzoek met CBS Microdata

Kostenposten	Tarieven 2023
Opstarten project	
Opstartkosten nieuw project	€ 1800
Kosten voor per bestandsonderwerp	€ 180
Uploaden en versleutelen KvK-nummers	€ 250
Autorisatie nieuwe onderzoekers	€ 375
Lopende kosten	
Ondersteuningskosten per bestandsonderwerp (per maand)	€ 18
Ondersteuningskosten per onderzoeker (per maand)	€ 122
Bestanden expertoren vanuit CBS-omgeving (per output)	€ 250
Totaal	
Totale kosten (2 onderzoekers, 2 maanden, 2 bestandsonderwerpen)	€ 3.220
Totale kosten (2 onderzoekers, 3 maanden, 4 bestandsonderwerpen)	€ 3.968

Bron: CBS Dienstencatalogus, bewerking SEO Economisch Onderzoek (2023)

⁷ Heeft de onderzoeker nog niet eerder gewerkt met de CBS Microdata, dan komen er extra kosten bij voor autorisatie van deze nieuwe onderzoeker.

⁸ Houd er rekening mee dat de startdatum van een project altijd op de eerste dag van een maand is. Deze datum staat ook vermeld in het contract.

Bijlage A Bestandsnamen

	Bestand 1	Bestand 2	Variabele 1	Variabele 2
2017	CBS_persoon_KvK_FIN_ABR 2017V1	BE_persoon_ABR 2017V1	VEP_IDENTIFICATIE	PERSOON_IDENTIFICATIE
2018	CBS_persoon_KvK_FIN_ABR 2018V1	BE_persoon_ABR 2018V1	VEP_IDENTIFICATIE	PERSOON_IDENTIFICATIE
2019	CBS_persoon_KvK_FIN_ABR 2019V1	BE_PERSOON_ABR 2019V1	VEP_IDENTIFICATIE	PERSOON_IDENTIFICATIE
2020	CBS_persoon_KvK_FIN_ABR 2020V1	BE_PERSOON_ABR 2020V1	VEP_IDENTIFICATIE	PERSOON_IDENTIFICATIE
2021	CBS_persoon_KvK_FIN2021V1	ABR_BE_PERSOON2021V1	CbsPersoonIdentificatie	CbsPersoonIdentificatie

Bron: CBS Microdata, ABR, bewerking SEO Economisch Onderzoek (2023)

Koppel de versleutelde KvK-registratie aan het ABR-bestand [Bestand 1] (via de variabele KvkNummer) en haal hieruit de overeenkomstige waarde van [variabele 1]. Sla een bestand op met de [variabele 1] van alle gekoppelde leden.

Koppel dit bestand aan het ABR-bestand [Bestand 2] (via de variabele [Variabele 2]) en haal hieruit de overeenkomstige waarde van de variabele BE_ID. Dit is het bedrijfsidentificatienummer. Sla een bestand op met het bedrijfsidentificatienummer van alle gekoppelde leden.