

# BRUIKBAARHEID CBS MICRODATA

ONDERZOEK NAAR DE BRUIKBAARHEID VAN CBS  
MICRODATA VOOR HET UITVOEREN VAN  
REPRESENTATIVITEITSONDERZOEK

RAPPORT

**seo** • economisch onderzoek

---

## AUTEURS

JUSTUS VAN KESTEREN, MENNO POMP

## IN OPDRACHT VAN

HET MINISTERIE VAN SOCIALE ZAKEN EN WERKGELEGENHEID (SZW)

AMSTERDAM, 1 FEBRUARI 2023

# Samenvatting

CBS Microdata zijn onder een aantal voorwaarden een bruikbare bron voor het uitvoeren van representativiteitsonderzoek. Zo moet de cao al eerder algemeen verbindend verklaard zijn én moet de KvK-registratie van gebonden werkgevers betrouwbaar en volledig zijn.

Werkgevers- en werknemersorganisaties (cao-partijen) kunnen de minister verzoeken om hun collectieve arbeidsovereenkomst (cao) algemeen verbindend te verklaren (avv). Hierdoor geldt de cao niet alleen voor de aangesloten partijen, maar voor alle werkgevers en werknemers in de cao-sector. Een belangrijke voorwaarde voor het algemeen verbindend verklaren van cao's is het meerderheidsvereiste. Dat vereiste houdt in dat de cao is afgesloten door cao-partijen die een belangrijke meerderheid van de in de cao-sector werkzame personen vertegenwoordigen.

Cao-partijen voeren representativiteitsonderzoek uit in het kader van het avv-verzoek. Het doel van dit onderzoek is om het representativiteitscijfer te berekenen: het aandeel werknemers en werkgevers dat de cao-partijen vertegenwoordigen in de totale cao-sector. Er is onder cao-partijen behoefte aan alternatieve gegevensbronnen voor het uitvoeren van representativiteitsonderzoek. Dat geldt voornamelijk voor cao-sectoren met beperkte toegang tot registratiegegevens op sectorniveau (zoals pensioenfondsbestanden).

Eerder evaluatieonderzoek van SEO Economisch Onderzoek stelt dat CBS Microdata in potentie een geschikte bron zijn voor het uitvoeren van representativiteitsonderzoek (zie Van Kesteren & Heyma, 2021). In de CBS Microdata is informatie beschikbaar over de cao-code die de werkgever invult bij de loonopgave van de werknemer. De cao-code geeft weer onder welke bedrijfstak-cao de werknemer valt en is voor alle bedrijfstakcao's beschikbaar. Dit onderzoek gaat nader in op de geschiktheid van CBS Microdata voor het berekenen van het representativiteitscijfer.

## Eigenschappen CBS Microdata

CBS Microdata zijn een geschikte bron voor het uitvoeren van representativiteitsonderzoek. Ten eerste zijn de CBS Microdata representatief voor de hele cao-sector, omdat werkgevers voor alle werknemers een cao-code moeten registreren als zij vallen onder een bedrijfstak-cao. Ten tweede is het mogelijk om het representativiteitscijfer op interne consistente wijze te berekenen, waarbij het aantal direct gebonden werkgevers en werknemers op vergelijkbare wijze wordt vastgesteld als het aantal indirect gebonden werkgevers en werknemers. Ten derde is het gebruik van CBS Microdata valide in de zin dat de operationalisering van de cao-code in lijn is met de werkings sfeer van de cao. Tot slot zijn de gegevens uit de CBS Microdata voldoende actueel doordat ze met één kwartaal vertraging beschikbaar komen.

Het gebruik van CBS Microdata heeft ook beperkingen. Zo moeten cao-partijen een onderzoeksbureau inschakelen met toegang tot de CBS Microdata om het representativiteitscijfer te berekenen, en moeten de resultaten van deze berekeningen openbaar worden gemaakt, terwijl de achterliggende gegevens niet mogen worden gedeeld (met beperkingen voor reproduceerbaarheid als gevolg).

## Betrouwbaarheid van CBS Microdata

Empirische analyses suggereren dat de CBS Microdata een betrouwbare bron zijn voor het uitvoeren van representativiteitsonderzoek. De berekeningen van het representativiteitscijfer op basis van CBS Microdata komen

grotendeels overeen met de oorspronkelijke representativiteitsonderzoeken van cao-partijen die zijn aangeduid als 'best practices'. Vooral na correctie voor registratiefouten zijn de uitkomsten vergelijkbaar met die van de oorspronkelijke onderzoeken. Registratiefouten lijken vooral op te treden als er uitzonderingen zijn gemaakt voor bepaalde werkgevers en werknemers binnen de werkingssfeer van de cao. Hoe beter de uitgezonderde groepen werkgevers en werknemers zijn te identificeren in de CBS Microdata, hoe betrouwbaarder de uitkomsten van de CBS Microdata zijn.

### **Reikwijdte van CBS Microdata**

Er is een aantal voorwaarden waaronder het gebruik van CBS Microdata tot betrouwbare resultaten leidt. Alleen voor de cao-sectoren die voldoen aan deze voorwaarden is het gebruik van CBS Microdata een waardevol alternatief voor de huidige berekeningswijze. Zo zijn de CBS Microdata alleen bruikbaar als de cao van de cao-sector al eerder algemeen verbindend is verklaard en er een complete en betrouwbare KvK-registratie van direct aan de cao gebonden werkgevers beschikbaar is. Idealiter zijn ook de werknemers- en werkgeversdefinities uit de cao onveranderd ten opzichte van de vorige cao en zijn de in de cao uitgezonderde categorieën werkgevers en werknemers te identificeren in de CBS Microdata. Als één van deze voorwaarden niet wordt nageleefd, zal dit de betrouwbaarheid van de uitkomsten aantasten.

# Inhoudsopgave

Samenvatting		i	
Inhoudsopgave		3	
1	Inleiding	1	
	1.1	Aanleiding	1
	1.2	Doelstellingen	2
	1.3	Onderzoeksvragen	2
	1.4	Onderzoeksaanpak	3
	1.5	Leeswijzer	3
2	Eigenschappen van CBS Microdata	4	
	2.1	Gunstige eigenschappen van de CBS Microdata	4
	2.2	Beperkingen van de CBS Microdata	5
	2.3	Tussenconclusie	5
3	Betrouwbaarheid van de CBS Microdata	6	
	3.1	Pilotfase	6
	3.2	Externe validiteit	8
	3.3	Tussenconclusie	9
4	Reikwijdte van de CBS Microdata	10	
Referenties		12	

# 1 Inleiding

Dit onderzoek toetst de bruikbaarheid van cao-gegevens uit de CBS Microdata voor het uitvoeren van representativiteitsonderzoek. Daarvoor zijn door cao-partijen uitgevoerde onderzoeken gereproduceerd op basis van CBS Microdata. Verder is een praktische handleiding bijgevoegd als een apart document voor cao-partijen die hun representativiteitsonderzoek willen laten uitvoeren op basis van CBS Microdata.

## 1.1 Aanleiding

Werkgevers- en werknemersorganisaties (vanaf nu: cao-partijen) kunnen de minister van Sociale Zaken en Werkgelegenheid (SZW) verzoeken om een algemeenverbindendverklaring (avv) van hun collectieve arbeidsovereenkomsten (cao's). De cao geldt dan niet alleen voor de aangesloten werkgevers en hun werknemers, maar voor alle werkgevers en werknemers die werkzaam zijn in de betreffende cao-sector. Hieronder vallen ook werkgevers die niet vertegenwoordigd zijn bij de cao-onderhandelingen, omdat zij geen lid zijn van de betrokken werkgeversorganisatie.

In het kader van het avv-verzoek stellen cao-partijen een representativiteitsopgave op. Hierin onderbouwen cao-partijen dat zij voldoen aan het meerderheidsvereiste. Dat vereiste houdt in dat de cao is afgesloten door werkgevers- en werknemersorganisaties die een belangrijke meerderheid van de in de cao-sector werkzame personen vertegenwoordigen. Om aan te tonen dat cao-partijen voldoen aan het meerderheidsvereiste, berekenen zij de mate waarin zij de cao-sector vertegenwoordigen (zie Box 1.1 voor een nadere toelichting).

SEO Economisch Onderzoek heeft onderzoek uitgevoerd naar de kwaliteit van de representativiteitsopgaven die zijn aangeleverd voor avv-verzoeken in de periode 2017 tot en met 2020 (zie Van Kesteren & Heyma, 2021). Het onderzoek biedt inzicht in tekortkomingen van deze representativiteitsopgaven, maar reikt ook handvatten aan voor het gebruik van alternatieve gegevensbronnen voor het uitvoeren van een representativiteitsonderzoek. De behoefte aan alternatieve bronnen is voornamelijk groot in cao-sectoren met beperkte toegang tot registratiegegevens op sectorniveau (bijv. pensioenfondsbestanden).

Uit het onderzoek blijkt dat de CBS Microdata een mogelijke geschikte alternatieve bron zijn voor het uitvoeren van representativiteitsonderzoek. In de CBS Microdata is informatie beschikbaar over de cao-code die de werkgever invult bij de loonopgave van de werknemer. De cao-code geeft weer onder welke bedrijfstak-cao de werknemer valt en is voor alle bedrijfstakcao's beschikbaar. Het gebruik van de cao-code in de CBS Microdata is in potentie een betrouwbaar alternatief voor het berekenen van het representativiteitscijfer.

Er zijn bij cao-partijen echter twijfels over de betrouwbaarheid van de ingevulde cao-code. Doordat werkgevers zelf de cao-code registreren en omdat vanuit UWV en CBS geen controle plaatsvindt op de juistheid hiervan, is er een reëel risico dat er in de loonopgave een verkeerde cao-code wordt ingevuld. Het is daarom van belang de betrouwbaarheid van de ingevulde cao-code nader te onderzoeken.

## Box 1.1 Achtergrond avv-verzoek

Een belangrijke voorwaarde die de Wet AVV stelt is het meerderheidsvereiste. Dat vereiste houdt in dat de cao waarop het avv-verzoek betrekking heeft is afgesloten door werkgevers- en werknemersorganisaties die een belangrijke meerderheid van de in de bedrijfstak werkzame personen vertegenwoordigen. Om te laten zien dat aan het meerderheidsvereiste wordt voldaan, dienen de cao-partijen de representativiteit als volgt te berekenen:

$$\frac{\text{Het aantal personen werkzaam bij werkgevers gebonden door de cao, die naar aard van hun functie respectievelijk werkzaamheden binnen de werkingssfeer van de cao vallen}}{\text{Totaal aantal personen die binnen de werkingssfeer van de cao vallen, indien deze algemeen verbindend wordt verklaard}} \times 100$$

De 'teller' van de berekening bevat het aantal georganiseerde werkgevers en werknemers. De 'noemer' bevat alle werkgevers en werknemers die, georganiseerd of ongeorganiseerd, via een algemeen verbindend verklaring van de cao, onder de desbetreffende cao vallen.

Om te bepalen of is voldaan aan het vereiste van een belangrijke meerderheid, worden de volgende uitgangspunten gehanteerd:

- Een meerderheid van 60 procent of meer wordt in ieder geval als 'een belangrijke meerderheid' gekwalificeerd;
- Een meerderheid tussen 55 procent en 60 procent wordt ook nog als een belangrijke meerderheid gekwalificeerd, tenzij het draagvlak voor de cao binnen het werkingssfeergebied gering is of er een zeer scheve verdeling van de meerderheid binnen het werkingssfeergebied bestaat (omdat een cao-sector bijvoorbeeld gedomineerd wordt door enkele zeer grote werkgevers);
- Bij een meerderheid tussen 50 en 55 procent vindt avv niet plaats, tenzij er naar het oordeel van de minister sprake is van bijzondere omstandigheden.

## 1.2 Doelstellingen

Het primaire doel van het onderzoek is om inzicht te geven in de mate waarin de cao-code in de CBS Microdata gebruikt kan worden voor het representativiteitsonderzoek. Hiervoor is het allereerst van belang dat de cao-code zelf *betrouwbaar* en *valide* is. Dat is het geval als werkgevers de cao-code uitsluitend invullen bij werknemers die daadwerkelijk onder de werkingssfeer van de cao vallen. Ten tweede is de *reikwijdte* van de CBS Microdata van belang. Er is een aantal voorwaarden waaronder het gebruik van CBS Microdata tot betrouwbare resultaten leidt. Alleen voor de cao-sectoren die voldoen aan deze voorwaarden is het gebruik van CBS Microdata een waardevol alternatief voor de huidige berekeningswijze. Tot slot speelt de *praktische uitvoerbaarheid* een rol in de bruikbaarheid van de CBS-gegevens. Hierbij gaat het om de kosten die cao-partijen moeten maken, de doorlooptijd van het onderzoek en de expertise die nodig is om de analyse uit te voeren. Dit onderzoek biedt inzicht in al deze afzonderlijke elementen.

## 1.3 Onderzoeksvragen

De hoofdvraag van dit onderzoek kan als volgt worden geformuleerd:

*"In hoeverre is de CBS Microdata bruikbaar voor de verplichte representativiteitsonderzoeken die cao-partijen voor het indienen van een avv-verzoek moeten uitvoeren?"*

Deze hoofdvraag bestaat uit een aantal afzonderlijke deelvragen:

1. In hoeverre is de cao-code die een werkgever invult bij de loonaangifte een betrouwbare en valide indicator in relatie tot de werkingssfeer van de cao en het bepalen van het representativiteitscijfer?
2. Kunnen cao-sectoren te allen tijde hun representativiteitsopgaven voor avv-verzoeken baseren op CBS-Microdata of zijn er ook restricties (reikwijdte)?
3. Hoe ziet een representativiteitsonderzoek op basis van CBS Microdata er praktisch gezien voor cao-partijen uit (praktische uitvoerbaarheid)?

## 1.4 Onderzoeksaanpak

Voor de beantwoording van de onderzoeksvragen zijn drie stappen doorlopen: de pilotfase, de fase gericht op bepaling van de externe validiteit en het opstellen van een praktische handleiding.

### Pilotfase

Voor de pilotfase zijn in samenspraak met SZW **vijf representativiteitsopgaven** geselecteerd die in Van Kesteren & Heyma (2021) zijn beoordeeld als *best practices*. Vervolgens is het originele representativiteitsonderzoek dat ten grondslag ligt aan de representativiteitsopgave gereproduceerd op basis van CBS Microdata. Hiervoor hebben de cao-sectoren hun ledenlijsten aangeleverd, die zijn gekoppeld aan de CBS-bestanden. Daarna zijn de uitkomsten van de analyse op basis van CBS Microdata en de uitkomsten van het originele onderzoek met elkaar vergeleken. Eventuele verschillen zijn nader verklaard op basis van documentanalyses en interviews met cao-partijen.

### Bepalen externe validiteit

Voor het bepalen van de externe validiteit zijn in samenspraak met SZW **dertien representativiteitsopgaven** geselecteerd. Voor deze selectie is gekeken of de omvang van de cao-sector zoals weergegeven in de representativiteitsopgaven, overeenkomt met de omvang van de cao-sector zoals berekend op basis van CBS Microdata.<sup>1</sup> Eventuele verschillen zijn nader verklaard op basis van documentanalyses.

### Opstellen praktische handleiding

Tot slot is een praktische handleiding opgesteld voor het gebruik van de CBS Microdata bij representativiteitsonderzoek. Hiervoor zijn de ervaringen van SEO in het uitvoeren van het onderliggende onderzoek gedeeld. Daarbij zijn praktische tips gegeven voor het uitvoeren van een analyse met behulp van de CBS Microdata, het doorvoeren van correcties en het aanleveren van output voor evaluatieonderzoek. De handleiding is voornamelijk bedoeld voor onderzoeksbureaus die de analyse gaan uitvoeren in opdracht van cao-partijen, zodat het representativiteitscijfer consistent berekend kan worden op basis van de CBS Microdata.

## 1.5 Leeswijzer

De indeling van het rapport is als volgt. Hoofdstuk 2 bevat een overzicht van voordelen en beperkingen van het gebruik van CBS Microdata in het kader van representativiteitsonderzoek. Dit overzicht is opgesteld op basis van documentanalyses. Hoofdstuk 3 biedt vervolgens zicht op de betrouwbaarheid op basis van empirische analyses. Hoofdstuk 4 gaat daarna in op de reikwijdte van de CBS Microdata. De praktische handleiding is bijgevoegd als een apart document, omdat de handleiding vooral relevant is voor onderzoeksbureaus die het representativiteitsonderzoek gaan uitvoeren.

---

<sup>1</sup> Het gaat hier specifiek om de 'noemer' van de representativiteitsberekening.

## 2 Eigenschappen van CBS Microdata

De cao-gegevens uit de CBS Microdata zijn representatief, interne consistent, actueel en valide. Beperkingen zijn de selectieve toegang tot de CBS Microdata, verplichte openbaarmaking en beperkte reproduceerbaarheid.

Dit hoofdstuk bevat een weergave van een aantal voordelen en beperkingen van de CBS Microdata in het kader van het avv-verzoek. Het overzicht is opgesteld op basis van documentanalyses. In het volgende hoofdstuk zijn deze kenmerken verder getoetst met empirische analyses.

### 2.1 Gunstige eigenschappen van de CBS Microdata

In de CBS Microdata is sinds 2014 informatie beschikbaar over de bedrijfstak-cao-code die de werkgever invult bij de loonopgave van de werknemer.<sup>2</sup> Deze cao-code geeft weer onder welke bedrijfstak-cao de werknemer valt en is voor *alle* bedrijfstakcao's beschikbaar.<sup>3</sup>

Het gebruik van de cao-code in de CBS Microdata kan een geschikt alternatief zijn voor het berekenen van het representativiteitscijfer. Dit komt omdat de cao-code aan bepaalde criteria voldoet die belangrijk zijn voor het uitvoeren van representativiteitsonderzoek:

- **Representativiteit**  
De cao-gegevens zijn representatief voor de hele sector omdat werkgevers voor alle werknemers een cao-code moeten registreren als zij vallen onder een bedrijfstak-cao;
- **Validiteit**  
De operationalisering van de cao-code is in lijn met de werkingssfeer van de cao omdat de werkgever de cao-code alleen mag invullen als de werknemer verplicht onder de desbetreffende cao valt;<sup>4</sup>
- **Interne consistentie**  
Door KvK-registraties van aangesloten werkgevers te koppelen aan de CBS Microdata kunnen in hetzelfde databestand direct aan de cao gebonden werkgevers en werknemers worden onderscheiden van indirect, via avv, aan de cao gebonden werkgevers en werknemers. Dit maakt het mogelijk om het representativiteitscijfer op consistente wijze te berekenen;
- **Actualiteit**  
De gegevens waarop het representativiteitscijfer is gebaseerd mogen niet ouder zijn dan één jaar, wat mogelijk is omdat de cao-gegevens uit de CBS Microdata doorgaans met één kwartaal vertraging beschikbaar komen.

<sup>2</sup> Het betreft hier specifiek de variabele "bedrijfstakcao" uit de SPOLISBUS-bestanden. Deze variabele is gemaakt op basis van de SCAOCODE: de cao-code die werkgevers moeten opgeven in de loonopgave. De bedrijfstakcao's die bij het ministerie van Sociale Zaken en Werkgelegenheid (SZW) worden aangemeld, krijgen een cao-code tussen nummer 1 en 7999. Deze code wordt afgegeven door SZW. Overige cao's - zoals ondernemingscao's - krijgen door het CBS een code van boven de 8000. In de CBS Microdata zijn ondernemingscao's geanonimiseerd omdat anders sprake zou zijn van onthulling. Indien geen collectieve regeling van toepassing is, of een bepaalde regeling op vrijwillige basis wordt gevolgd, geldt de code 9999.

<sup>3</sup> De cao-codes zijn [hier](#) te vinden.

<sup>4</sup> De werknemer valt alleen onder de cao als (i) de werkgever voldoet aan de werkgeversdefinitie uit de cao en (ii) de werknemer voldoet aan de werknemersdefinitie uit de cao. Als de cao op vrijwillige basis wordt ingevoerd, mag de cao-code niet worden gebruikt. Door deze voorwaarden vullen in theorie alleen werkgevers de cao-code in die onder de werkingssfeer van de cao vallen, en doen ze dit alleen voor werknemers die voldoen aan de werknemersdefinitie van de cao.



## 2.2 Beperkingen van de CBS Microdata

Er zijn echter ook beperkingen aan het gebruik van de CBS Microdata voor representativiteitsonderzoek:

- **Vertrouwelijkheid**

Het is verplicht om de uitkomsten van analyses met CBS Microdata openbaar te maken.<sup>5</sup> Dit kan een belemmering vormen voor cao-partijen in het kader van de vertrouwelijkheid van de uitkomsten. Het is echter niet nodig om de cao-sector te noemen in de openbaarmaking, waardoor de vertrouwelijkheid toch gewaarborgd kan blijven;

- **Reproduceerbaarheid**

Het is niet mogelijk om de gebruikte gegevens en bewerkingen te delen met een derde partij, wat voor beperkingen zorgt in de reproduceerbaarheid van het representativiteitscijfer. Wel is het mogelijk voor een derde partij met toegang tot de CBS Microdata om de berekeningen in de beveiligde CBS-omgeving te reproduceren;

- **Beschikbaarheid**

Een cao-partij kan zelf doorgaans geen gebruikmaken van de CBS Microdata. De gegevens worden vrijwel alleen ter beschikking gesteld aan wetenschappelijke en onderzoeksinstituten, die hiervoor een procedure moeten doorlopen. De cao-partij kan wel een onderzoeksbureau met toegang tot CBS Microdata inschakelen om het representativiteitscijfer te laten berekenen.

## 2.3 Tussenconclusie

De CBS Microdata bevatten gunstige eigenschappen voor het uitvoeren van representativiteitsonderzoek, zoals representativiteit voor de hele cao-sector, interne consistentie, actualiteit en validiteit. De betrouwbaarheid van de CBS Microdata in de praktijk is in een later hoofdstuk nader onderzocht. Het gebruik van de CBS Microdata heeft echter ook beperkingen. Zo moeten cao-partijen een onderzoeksbureau inschakelen met toegang tot de CBS Microdata om het representativiteitscijfer te berekenen, en moeten de resultaten van deze berekeningen openbaar gemaakt worden, terwijl de achterliggende gegevens niet mogen worden gedeeld (met beperkingen voor de reproduceerbaarheid als gevolg).

---

<sup>5</sup> Dit omdat door het CBS de Microdata in het kader van statistisch onderzoek van overheidswege worden verzameld. Daarom is bepaald dat alle resultaten van statistisch en/of wetenschappelijk onderzoek op basis van die data openbaar gemaakt moeten worden.

## 3 Betrouwbaarheid van de CBS Microdata

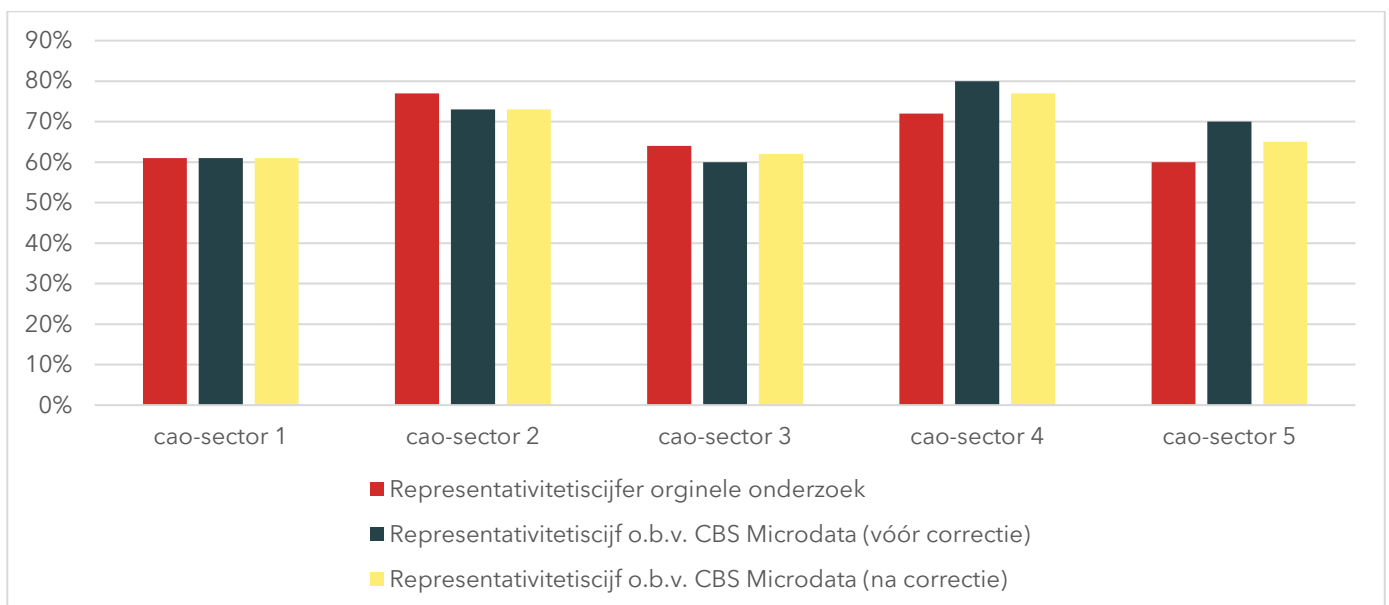
Empirische analyses suggereren dat de CBS Microdata een betrouwbare bron vormen voor het uitvoeren van representativiteitsonderzoek. Vooral na correctie voor registratiefouten zijn de uitkomsten veelal vergelijkbaar met die van *best practices*.

Dit hoofdstuk biedt inzicht in de betrouwbaarheid van de cao-code uit de CBS Microdata. Specifiek gaat het om de resultaten van de pilotfase (Paragraaf 3.1) en de fase waarin de externe validiteit is bepaald (Paragraaf 3.2). De synthese (Paragraaf 3.3) bevat een algemene conclusie over de betrouwbaarheid van de CBS Microdata.

### 3.1 Pilotfase

Figuur 3.1 geeft de resultaten weer uit de pilotfase. In deze fase is voor vijf cao-sectoren het originele representativiteitsonderzoek geproduceerd op basis van CBS Microdata. Per cao-sector zijn de uitkomsten van het originele onderzoek (rode balk), de uitkomsten op basis van CBS Microdata vóór correctie (groene balk) en de uitkomsten op basis van CBS Microdata na correctie (gele balk) weergegeven. De cao-sectoren zelf zijn niet genoemd in verband met de vertrouwelijkheid van de gegevens. De aanpak is beschreven in Box 3.1.

**Figuur 3.1** De uitkomsten op basis van CBS Microdata komen in veel gevallen overeen met de resultaten uit het originele onderzoek



Bron: CBS Microdata, bewerking SEO Economisch Onderzoek (2022) hierboven 3x representativiteitscijfer

De resultaten vóór correctie verschillen voor een aantal sectoren van de resultaten van het originele onderzoek. Vooral in cao-sector 4 en 5 zijn de afwijkingen relatief groot. Dit komt voor een belangrijk deel door registratiefouten door werkgevers (zie Tabel 3.1). In een aantal gevallen vullen werkgevers de cao-code in bij werknemers die niet onder de cao vallen. Het gaat dan met name om stagiaires en DGA's, die in de regel uitgesloten zijn van de cao. Als

die registratiefouten vaker (minder vaak) voorkomen bij direct dan bij indirect aan de cao gebonden werkgevers, dan leidt dat tot vertekeningen in het representativiteitscijfer.

Tabel 3.1 Registratiefouten komen voor in de CBS Microdata

	<b>Registratiefouten direct gebonden werkgevers</b> Aandeel foutief ingevoerde cao-codes van totaal aantal ingevoerde cao-codes	<b>Registratiefouten indirect gebonden werkgevers</b> Aandeel foutief ingevoerde cao-codes van totaal aantal ingevoerde cao-codes
Cao-sector 1	5%	7%
Cao-sector 2	1%	2%
Cao-sector 3	5%	9%
Cao-sector 4	4%	2%
Cao-sector 5	9%	17%

Bron: CBS Microdata, bewerking SEO Economisch Onderzoek (2022)

De resultaten ná correctie voor registratiefouten zijn in de meeste gevallen beter vergelijkbaar met de originele onderzoeksresultaten. Dit suggereert dat de verschillen tussen het originele onderzoek en het onderzoek gebaseerd op CBS Microdata deels zijn toe te schrijven aan registratiefouten. Tegelijkertijd blijft er zeker in cao-sector 4 en 5 nog een aanmerkelijk verschil bestaan tussen de resultaten van beide onderzoeken. Uit aanvullende analyses en interviews blijkt dat hier twee verklaringen voor zijn:

- **Niet-waargenomen registratiefouten** Er zijn mogelijk werkgevers die onder de werkingssfeer van de cao vallen, maar die de cao-code niet invoeren voor hun werknemers. Vooral indirect aan de cao gebonden werkgevers zijn vermoedelijk minder goed op de hoogte van de cao-bepalingen van de werknemers, en voeren daarom een verkeerde cao-code op voor hun werknemers. Hierdoor wordt de totale omvang van de cao-sector onderschat (de 'noemer'), waardoor het representativiteitscijfer hoger uitvalt dan het daadwerkelijke cijfer. Dat kan een verklaring zijn voor de hogere representativiteitscijfers in cao-sector 4 en 5;
- **Verschillen in registratiewijze van origineel onderzoek** De originele onderzoeken kennen hun beperkingen. Veel cao-partijen maken bijvoorbeeld gebruik van pensioenfondsbestanden, omdat de verplichtstelling van pensioenfondsen vaak overeenkomt met de werkingssfeer van de cao. Er zijn echter verschillen tussen het domein van de verplichtstelling en de werkingssfeer van de cao, waar niet altijd volledig voor kan worden gecorrigeerd.

### Box 3.1 Aanpak pilotfase

De cao-partijen hebben de KvK-registratie van hun ledenlijsten aangeleverd. Vervolgens heeft CBS de KvK-nummers versleuteld en heeft SEO de koppeling gemaakt met de CBS-microdatabestanden. Doordat de werkgevers (middels hun KvK-nummers) zijn gekoppeld aan versleutelde loonheffingsnummers, die op hun beurt gekoppeld kunnen worden aan Polisadministratiegegevens afkomstig van de Belastingdienst en UWV, is het mogelijk voor iedere werknemer met een bepaalde cao-code te identificeren of hij/zij:

- Werkzaam is voor een bedrijf dat lid is van de werkgeversorganisatie;
- Werkzaam is voor een bedrijf dat geen lid is van de werkgeversorganisatie.

Vervolgens is op basis van de gekoppelde Microdatabestanden het representativiteitscijfer berekend. Daarbij is eenzelfde peildatum gehanteerd omwille van de vergelijkbaarheid tussen het oorspronkelijke representativiteitsonderzoek en het onderzoek op basis van CBS Microdata. Doordat de cao-code iedere maand wordt ingevuld door de werkgever, is het mogelijk om een vergelijkbare peildatum te selecteren.

## 3.2 Externe validiteit

Tabel 3.2 laat de uitkomsten zien van de fase waarin de externe validiteit van de pilotfase is bepaald. Hiervoor is voor 13 cao-sectoren de omvang van de cao-sector (de 'noemer') gereproduceerd en vergeleken met het originele onderzoek. De eerste kolom bevat de uitkomsten van het originele onderzoek, de tweede kolom de resultaten op basis van CBS Microdata vóór correctie en de derde kolom de resultaten na correctie. De cao-sectoren zelf zijn niet genoemd in verband met de vertrouwelijkheid van de gegevens.

**Tabel 3.2** De schattingen van de omvang van de cao-sector zijn beter vergelijkbaar na correctie voor registratiefouten

Cao-sector	Originele onderzoek	CBS Microdata (vóór correctie)	CBS Microdata (na correctie)
6	111.000	151.000	112.000
7	8.000	6.000	6.000
8	11.000	10.000	10.000
9	40.000	37.000	35.000
10	6.000	7.000	7.000
11	4.000	4.000	4.000
12	12.000	11.000	11.000
13	4.000	3.000	3.000
14	4.000	7.000	5.000
15	100	200	200
16	23.000	26.000	24.000
17	4.000	4.000	4.000
18	12.000	9.000	9.000

Bron: CBS Microdata, bewerking SEO Economisch Onderzoek (2022)

Noot: Afgerond op 1000-tallen (cao-sector 1-13, 16-18) en 100-tallen (sector 15).

De resultaten vóór correctie verschillen voor een aantal sectoren van de resultaten van het originele onderzoek. Dat geldt met name voor de grotere sectoren, zoals cao-sector 6, 9, 16 en 18. In deze cao-sectoren zijn relatief veel groepen werknemers uitgezonderd van de cao. Het gaat dan bijvoorbeeld om werknemers boven een bepaalde inkomensgrens of leeftijdsgrens, scholieren/studenten, vakantiewerkers en specifieke functies. Werkgevers lijken zich niet altijd bewust van deze uitzonderingen, en registreren daarom ten onrechte de cao-code bij de uitgezonderde groepen werknemers. Ook zijn uitzonderingen in de cao mogelijk doorgevoerd nadat de werkgever de cao-code heeft ingevuld. Hierdoor wordt de omvang van de cao-sector overschat.

Doordat de CBS Microdata op detailniveau informatie bieden over kenmerken van werknemers in de cao-sector, is het mogelijk om voor registratiefouten van werkgevers te corrigeren in de berekeningen. Tabel 3.3 geeft een overzicht van de mogelijkheden om in de werkingsfeerbepalingen uitgezonderde groepen werknemers en werkgevers te onderscheiden. Zo is het mogelijk per werknemer de leeftijd, de contractvorm, het inkomen, de

regio/woonplaats en de arbeidsvorm in kaart te brengen. Dat maakt het mogelijk om op basis van deze kenmerken te corrigeren voor registratiefouten van werkgevers.

Tegelijkertijd is het niet mogelijk om voor alle uitgezonderde groepen registratiefouten te identificeren (en dus ook niet om hiervoor te corrigeren in de berekeningen). Bijvoorbeeld in het geval van uitgezonderde functies of directieleden, omdat iemand zijn/haar beroep niet waar te nemen is in de CBS Microdata. Bovendien is het lastig om voor registratiefouten met betrekking tot de werkgeversdefinitie te corrigeren. Bijvoorbeeld: een werkgever die wél onder de cao valt, maar de cao-code niet voert voor haar werknemers. Of: een werkgever die niet onder de cao valt, maar de cao-code wel voert voor haar werknemers. Dit omdat de werkgeversdefinitie zich vaak richt op bedrijfsactiviteiten, die niet direct waar te nemen zijn in de CBS Microdata. Het is wel mogelijk om SBI-codes en bedrijfsgrootte van werkgevers te observeren, maar dit maakt het maar beperkt mogelijk om werkgevers die onder cao vallen te identificeren. SBI-codes dekken immers in veel gevallen niet de (gehele) werkingssfeer van de cao.

Tabel 3.3 Mogelijkheid om te corrigeren voor registratiefouten van werkgevers binnen de CBS Microdata

Beschikbare informatie voor verdere afbakening/correctie werkingssfeer	Niet beschikbaar informatie voor verdere afbakening/correctie werkingssfeer
Leeftijd	Directieleden
Contractvorm (regulier, oproep, uitzend, stagiair, DGA)	Specifieke beroepen
Student/scholier	Bedrijfsactiviteiten
Inkomen	Vrijwillig aangesloten/gedispenseerde werkgevers
Sector / SBI-code	
Bedrijfsgrootte	
Regio/woonplaats	

Tabel 3.2 laat zien dat de resultaten na correctie voor registratie beter vergelijkbaar zijn met de oorspronkelijke onderzoeken. Wel valt op dat het aantal registratiefouten in sommige gevallen groot is (bijv. in sector 6). Het is daarom van belang om in de analyses met CBS Microdata te corrigeren voor eventuele registratiefouten. Verschillen die nog bestaan zijn te verklaren vanuit niet-waargenomen registratiefouten (bijv. door uitgezonderde specifieke functies of bedrijfsactiviteiten) of beperkingen in het originele onderzoek.

### 3.3 Tussenconclusie

De resultaten suggereren dat de CBS Microdata een betrouwbare bron zijn voor het uitvoeren van representativiteitsonderzoek. De resultaten komen grotendeels overeen met de originele onderzoeken, die zijn aangeduid als 'best practices'. Vooral na correctie voor registratiefouten zijn de uitkomsten vergelijkbaar met die van de originele onderzoeken. Registratiefouten lijken vooral op te treden als er uitzonderingen zijn gemaakt voor bepaalde werkgevers en werknemers binnen de werkingssfeer van de cao. Hoe beter de uitgezonderde groepen werkgevers en werknemers zijn te identificeren in de CBS Microdata, hoe betrouwbaarder de uitkomsten van de CBS Microdata zijn.

## 4 Reikwijdte van de CBS Microdata

De CBS Microdata kunnen alleen worden gebruikt als de cao al eerder algemeen verbindend is verklaard én als de KvK-registraties van betrokken werkgeversorganisaties betrouwbaar en volledig zijn. De betrouwbaarheid is groter als de van de cao uitgezonderde groepen zijn te identificeren in de CBS Microdata.

Dit hoofdstuk bevat een overzicht van voorwaarden waaronder het gebruik van CBS Microdata leidt tot betrouwbare uitkomsten van het representativiteitsonderzoek. Onderstaande voorwaarden 1 en 2 zijn essentieel om het representativiteitsonderzoek uit te kunnen voeren. Als aan één van deze voorwaarden niet wordt voldaan, is het onmogelijk om op basis van CBS Microdata het representativiteitscijfer te bepalen. Voorwaarden 3 en 4 verhogen de betrouwbaarheid van het representativiteitsonderzoek. Als aan deze voorwaarden niet wordt voldaan, zal dit ten koste gaan van de betrouwbaarheid van de uitkomsten.

### 1. De voorgaande cao van de cao-sector is algemeen verbindend verklaard

De CBS Microdata zijn alleen bruikbaar als de cao in het verleden ook algemeen verbindend is verklaard. Werkgevers mogen immers alleen de cao-code invullen als zij op het moment van invullen verplicht onder de werkingssfeer van de cao vallen. Als de cao van de cao-sector nog nooit eerder algemeen verbindend is verklaard, dan mogen dus alleen direct aan de cao gebonden werkgevers de cao-code registreren. In dat geval is het alleen mogelijk om de CBS Microdata te gebruiken voor de berekening van het aantal direct gebonden werknemers en werkgevers, maar niet om het aantal indirect gebonden werkgevers en werknemers mee te berekenen. Dit maakt het onmogelijk om op basis van alleen CBS Microdata het representativiteitscijfer te bepalen.

### 2. De KvK-registratie van direct gebonden werkgevers is volledig en betrouwbaar

De ledenlijsten van direct aan de cao gebonden werkgevers kunnen alleen op basis van KvK-nummers worden gekoppeld aan de CBS-microdatabestanden. Daarom is het cruciaal dat de KvK-registratie van direct gebonden werkgevers volledig en betrouwbaar is. Is dat niet het geval, dan is er vermoedelijk veel koppelverlies bij het koppelen van het ledenbestand aan de CBS-microdatabestanden. Hierdoor kunnen niet alle direct gebonden werknemers en werkgevers worden geïdentificeerd, wat zorgt voor een onderschatting van het representativiteitscijfer. Bij een klein koppelverlies kan men nog werken met veronderstellingen over het aantal werknemers dat werkt bij direct gebonden werkgevers, bij een groot koppelverlies leidt de analyse op basis van CBS Microdata tot onbetrouwbare uitkomsten.

### 3. De werknemers- en werkgeversdefinities zijn onveranderd ten opzichte van de vorige cao

De cao-code die indirect gebonden werkgevers invullen heeft in principe betrekking op de cao die op dat moment algemeen verbindend is verklaard. Dit betekent dat indirect gebonden werkgevers niet de werkingssfeer van de nieuwe cao, waarvoor een avv-verzoek wordt ingediend, aanhouden bij het bepalen van de juiste cao-code. Als er veranderingen hebben plaatsgevonden in de werkgevers- en werknemersdefinities van de cao, dan dekt de cao-code van indirect gebonden werkgevers niet volledig de werkingssfeer. In dat geval hangt de bruikbaarheid van de CBS Microdata af van de vraag of het mogelijk is om te corrigeren voor de verschillen in werkgevers- en werknemersdefinities in de Microdata. Dit is bijvoorbeeld mogelijk als de veranderingen in definities plaatsvinden in contracttypen, inkomensgroepen of werkgevers met specifieke SBI-codes (zie ook Tabel 3.3).

#### 4. De in de cao uitgezonderde categorieën werkgevers en werknemers zijn te identificeren in de CBS Microdata

Het is van belang om te bepalen in hoeverre de in de cao uitgezonderde groepen werkgevers en werknemers zijn te identificeren in de CBS Microdata. Als de uitgezonderde groepen goed zijn te identificeren in de CBS Microdata, dan is het mogelijk om in de berekeningen te corrigeren voor registratiefouten van werkgevers die de cao-code (onterecht) hebben ingevuld bij uitgezonderde werkgevers en werknemers. Hoewel dit geen essentiële voorwaarde is voor het gebruik van CBS Microdata voor het representativiteitsonderzoek, zijn de uitkomsten van de CBS Microdata wel betrouwbaarder als aan deze voorwaarde wordt voldaan.

# Referenties

Van Kesteren, J., & Heyma, A. (2021). *Kwaliteit representativiteitsopgaven 2017-2020*. Amsterdam: SEO Economisch Onderzoek.





# “De wetenschap dat het goed is.”

SEO Economisch Onderzoek doet onafhankelijk toegepast onderzoek in opdracht van overheid en bedrijfsleven. Ons onderzoek helpt onze opdrachtgevers bij het nemen van beslissingen. SEO Economisch Onderzoek is gelieerd aan de Universiteit van Amsterdam. Dat geeft ons zicht op de nieuwste wetenschappelijke methoden. We hebben geen winstoogmerk en investeren continu in het intellectueel kapitaal van de medewerkers via promotietrajecten, het uitbrengen van wetenschappelijke publicaties, kennisnetwerken en congresbezoek.

**SEO-rapport 2023-08**

**ISBN 978-90-5220-256-3**

## **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 magazine 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](mailto:secretariaat@seo.nl).

Roetersstraat 29  
1018 WB Amsterdam

**+31 20 525 1630**  
[secretariaat@seo.nl](mailto:secretariaat@seo.nl)  
[www.seo.nl](http://www.seo.nl)