



## Het meten van meetfouten

**Auteur(s):**

Bochove, C. van  
Thijssen, J.

*Directeur Sociaal-economische statistieken, resp. hoofd Sociaal-economische persoonsenquêtes bij het CBS.*

**Verschenen in:**

ESB, 83e jaargang, nr. 4136, pagina 60, 23 januari 1998

**Rubriek:**

Discussie

**Trefwoord(en):**

werkloosheid

Dit artikel is een naschrift n.a.v.: G. Visscher, [Werkloosheidscijfers op basis van de Enquête Beroepsbevolking](#), ESB, 23 januari 1998, blz. 58-60.

**Volgens Visscher leidt selectieve non-respons in persoonsenquêtes van het CBS tot overschatting van het opleidingspeil van de bevolking en daardoor tot onderschatting van de werkloosheid en overschatting van de werkgelegenheid. Kennelijk zijn wij er in meerdere contacten niet in geslaagd hem de essentie van onze methodiek en onderzoeksresultaten duidelijk te maken. In dit commentaar proberen we het nogmaals, want bij misverstanden over sociaal-economische kerncijfers is niemand gebaat.**

Het CBS past diverse methoden toe om voor de effecten van non-respons te corrigeren en om te analyseren in hoeverre dat succesvol is. De stand van zaken is als volgt.

» Al in 1991 hebben we gepubliceerd dat de Enquête beroepsbevolking (EBB) het aantal mensen met een lage opleiding iets onderschat. Onlangs is aangetoond dat die onderschatting maximaal 200.000 bedraagt op een totaal van 4,3 mln. Dat is minder dan de helft van het aantal dat Visscher noemt.

» Door de methoden die het CBS hanteert maakt het voor de cijfers over de geregistreerde werkloosheid niets uit dat laag opgeleiden minder responderen dan hoog opgeleiden.

» De werkgelegenheid wordt met zekerheid niet overschat in de Enquête beroepsbevolking.

### Non-respons: hoe gaan we er mee om?

De hoge non-respons die in Nederland bij persoonsenquêtes optreedt is een grote zorg. Visscher signaleert dat terecht. Op zichzelf hoeft non-respons niets te betekenen voor de uitkomsten, zolang de non-respons maar niet selectief is naar de doelvariabelen. Maar bij een hoge non-respons moet je je daar zorgen over maken. Daarom heeft het CBS sinds de non-respons in Nederland vanaf medio jaren zeventig begon op te lopen, honderden mensjaren geïnvesteerd in het ontwikkelen en toepassen van methoden om de effecten van non-respons vast te stellen en er voor te corrigeren. En er duizenden pagina's over gepubliceerd.

Het vaststellen doen we met een veelheid aan methoden. De belangrijkste is het vergelijken van enquête-informatie met andere bronnen. Er zijn twee groepen van andere bronnen. De eerste bestaat uit registers die voor niet statistische doeleinden worden bijgehouden: de bevolkingsadministraties, de bestanden van uitkeringsinstanties, arbeidsbureaus, studiefinanciering, werknemersverzekeringen, de belastingdienst, enzovoorts. De tweede groep van bronnen bestaat uit onze eigen enquêtes bij bedrijven, vooral over werknemers, en bij onderwijsinstellingen. De twee belangrijkste methoden die we tot dusverre hebben gehanteerd om te corrigeren voor non-respons zijn weging van de uitkomsten en de bouw van rekeningenstelsels. Bij de weging van de uitkomsten gebruiken we totalen die we kennen uit andere bronnen. Tot dusverre hebben we daarbij in de arbeidsstatistieken vooral twee bronnen gebruikt: bevolkingsgegevens en, voor de werkloosheidsstatistieken, de bestanden van de arbeidsbureaus. De respons onder allochtone jonge mannen in Amsterdam is laag. Door de weging bereiken we dat het aantal allochtone jonge mannen in Amsterdam klopt. Zolang de respondenten uit deze groep representatief zijn voor de betrokken groep, kloppen dan ook de uitkomsten voor andere variabelen. Of dat zo is, moet weer door verdere analyse worden vastgesteld.

Bij weging kan niet met alle informatie die er is rekening worden gehouden: er zijn al snel veel meer wegingscellen dan respondenten. Bovendien hoeven de registers niet goed te zijn. Alle registraties kennen vervuiling. Alleen door alle informatie uit alle bronnen te analyseren, beoordelen en prudent te benutten kan een redelijk betrouwbaar beeld van de werkelijkheid worden gegeven. Waar we dan vervolgens opnieuw wantrouwig naar kijken om het verder te kunnen verbeteren. Dit gebeurt in rekeningenstelsels, zoals de arbeidsrekeningen. Uit de gecombineerde informatie van de enquêtes bij bedrijven, de Enquête beroepsbevolking, de bestanden van de werknemersverzekeringen proberen we daarin, in een 'ware jacht op fouten', bijvoorbeeld een zo juist mogelijk aantal banen vast te stellen.

### Hoe goed lukt dat?

## Opleidingsniveau

Het opleidingsniveau is een lastige variabele om greep op te krijgen. Om twee redenen. Er bestaat geen integraal register waarin het opleidingsniveau van de bevolking vast ligt. De andere reden dat het lastig is om het totale opleidingsniveau van de bevolking te bepalen, is dat het juist vaststellen van het opleidingsniveau van een individu niet simpel is. Daarvoor moet een wirwar van in de loop van de tijd veranderende opleidingen juist worden getypeerd. En post-initiële opleidingen niet worden vergeten. Keer op keer blijkt bij vergelijkingen, onder meer op individueel niveau, dat door vertekeningen in vrijwel alle registraties en in administraties van bedrijven het werkelijk behaalde opleidingsniveau van grote groepen individuen systematisch wordt onderschat.

Toch zijn er wel subtiele methoden om een indruk krijgen of de uitkomsten van de Enquête beroepsbevolking (EBB) voor het opleidingsniveau juist zijn. En dan blijkt Visscher wel een punt te hebben waarop het CBS overigens al in 1991 heeft gewezen<sup>1</sup>: het is aannemelijk dat er een zekere onderschatting zit in de EBB-uitkomsten voor laag-opgeleiden en dus een overschatting van het aantal hoger opgeleiden. Maar lang niet in de omvang die Visscher noemt.

Recent zijn de over- en onderschattingen gekwantificeerd<sup>2</sup>. Daaruit blijkt dat het aantal mensen van 15-64 jaar met alleen basisonderwijs met maximaal 110.000 wordt onderschat en het aantal mensen met vbo/mavo met 90.000. In totaal wordt het aantal laagopgeleiden met maximaal 200.000 onderschat (op een gemeten totaal van 4,3 mln, het aantal mensen met een middelbare of hogere opleiding wordt dus met 200.000 overschat (op een totaal van 6,3 mln). De resultaten zijn verkregen door voor de bij een arbeidsbureau ingeschreven mensen de verdeling naar opleidingsniveau volgens de EBB te vergelijken met die uit de 'koppelings-steekproef' voor de Statistiek geregistreerde werkloosheid. Deze laatste steekproef omvat EBB-respondenten die in de registratie van de arbeidsbureaus zijn teruggevonden. Bij de weging van deze koppelingssteekproef is gecorrigeerd voor opleidingsniveau, bij de weging van de EBB-steekproef niet.

De door het CBS als maximum vastgestelde onderschatting van het aantal laagopgeleiden is dus minder dan de helft van het door Visscher genoemde aantal (500.000). Hij heeft de onderschatting berekend door non-respons percentages te gebruiken die zijn ontleend aan een door een particulier onderzoeksbureau uitgevoerde telefonische enquête in één regio bij mensen die ingeschreven staan bij een arbeidsbureau dan wel recent zijn uitgeschreven. De EBB is een enquête aan huis bij huishoudens uitgevoerd door het CBS in geheel Nederland. Wellicht zit hier een oorzaak van de verschillen. Er zijn overigens nog meer bewijzen dat de afwijkingen in de EBB bij lange na niet de omvang kunnen hebben die Visscher aangeeft. Het voert echter te ver om die hier allemaal te gaan noemen.

## Werkenden

De overschatting bij hoogopgeleiden en de onderschatting bij laagopgeleiden in combinatie met de verschillen in participatiegraad tussen deze groepen leiden volgens Visscher tot een overschatting van de werkende beroepsbevolking (mensen met een baan van 12 uur of meer per week) met 200.000. In de ook al genoemde Arbeidsrekeningen<sup>3</sup> blijkt voor 1994 het aantal banen volgens de EBB 228.000 hoger uit te komen dan volgens de bedrijven-statistieken. Dit verschil is echter vrijwel volledig geconcentreerd bij deeltijdbanen (de 228.000 banen betreffen 65.000 arbeidsjaren). Het gaat om banen die in bedrijven-enquêtes expliciet niet worden meegenomen omdat het om niet-ziectewet verzekerde arbeid gaat, en om arbeid die ten onrechte wordt onttrokken aan de verplichting tot registratie bij de fiscus of de sociale verzekeringsorganen. Uit enquêtes naar zwarte arbeid begin jaren tachtig bleek al dat zeker een aantal niet geregistreerde, vooral kleine, banen in deze orde van grootte bestond. Van een overschatting van het aantal werkenden in de EBB is dus in ieder geval geen sprake. Dat wordt nog eens bevestigd als we het aantal werknemers van een aantal zeer grote individuele bedrijven (waar geen niet geregistreerde banen zijn) volgens de EBB en de opgaven van deze bedrijven zelf naast elkaar leggen. Dat spoort telkens op één of twee duizend na.

## Werkloosheidscijfers

Het aantal mensen dat tot de werkloze beroepsbevolking behoort, is volgens Visscher weinig gevoelig voor de over- en onderschatting van hoog- en laagopgeleiden. Wel wordt de verdeling van werklozen naar opleidingsniveau hier sterk door beïnvloed. Omdat de onder- en overschatting van hoog- en laagopgeleiden aanzienlijk kleiner is dan Visscher veronderstelt, zal de werkelijke verdeling veel minder afwijken van de gepubliceerde cijfers dan hij suggereert. Ook het aantal geregistreerde werklozen is niet gevoelig voor de selectieve non-respons naar opleidingsniveau. Dat komt alleen al doordat het CBS in de statistiek geregistreerde werkloosheid voor alle zekerheid weegt naar de opleidingsaantallen uit de respons en de arbeidsbureaus. Voor de bepaling van het totale aantal geregistreerde werklozen is dat echter niet eens nodig. In 1996 scheelde deze weging namelijk bijna niets: het totale aantal geregistreerde werklozen komt voor 1996 zonder weging naar opleidingsniveau uit op 439.000 en met weging op 440.000.

Visscher heeft nog een andere redenering waarom de werkloosheidscijfers niet zouden kloppen. Die heeft betrekking op de verschillen in ontwikkeling tussen de aantallen werkloosheidsuitkeringen, ingeschreven mensen bij een arbeidsbureau enerzijds en de werkloze beroepsbevolking en geregistreerde werkloosheid anderzijds<sup>4</sup>. De verschillen zijn voor het grootste deel goed te verklaren. Zo is het aantal bij de arbeidsbureaus ingeschreven mensen met een grote afstand tot de arbeidsmarkt de laatste jaren gegroeid. In 1988 gaven bijna 150.000 van alle ingeschreven mensen op niet beschikbaar te zijn voor de arbeidsmarkt, in 1996 was dit aantal gestegen tot 220.000. Ook het aantal mensen met een baan van 12 uur of meer onder de ingeschrevenen nam, mede door de flexibilisering van de arbeidsmarkt, aanzienlijk toe. Hun aantal steeg van bijna 280.000 in 1988 tot ruim 400.000 in 1996. Logisch dat dan de ontwikkeling van het aantal bij een arbeidsbureau ingeschreven mensen niet spoort met die van het aantal geregistreerde werklozen.

Ook de cijfers over uitkeringen in de EBB deugen volgens Visscher niet. Ze zijn zo'n 20 procent lager dan volgens de uitkeringsinstanties. Dit klopt, blijkt eerdere CBS publicaties<sup>5</sup> waarin wordt gerapporteerd over onderzoek waarbij de bestanden zijn gekoppeld. De effecten van selectieve non-respons blijken bij werkloosheidsuitkeringen echter niet het hoofdprobleem te zijn. Driekwart van de onderschatting ontstaat door het opgeven van een andere uitkering of het niet opgeven van een uitkering. Slechts een kwart van de onderschatting is toe te schrijven aan selectieve non-respons. Heel recent heeft het CBS een koppelingsonderzoek naar arbeidsongeschikten uitgevoerd<sup>6</sup>. Bij deze groep blijkt selectieve non-respons verantwoordelijk te zijn voor een groter deel, namelijk de helft, van het gat. Niet zo raar, omdat veel arbeidsongeschiktheid wordt veroorzaakt door of gepaard gaat met psychische problemen.

Visscher zit er dus bij het aanwijzen van de oorzaak van de onderschatting de helft tot driekwart naast.

---

**1** S.W.H.C. Loozen en G.J. Post, Het onderwijsniveau van de Nederlandse bevolking: uitkomsten van de Enquête beroepsbevolking 1990, Supplement bij de *Sociaal-economische maandstatistiek*, 1991, nr. 4.

**2** I.H.A. Knops en C.M. Kracht, Telt het CBS te weinig laag opgeleiden?, *Sociaal-economische maandstatistiek*, oktober 1997.

**3** CBS, *Arbeidsrekeningen 1993-1995*, Voorburg/Heerlen, 1996. Over de verschillen tussen de uitkomsten van de EBB en bedrijvenstatistieken werd overigens al gerapporteerd in de Arbeidsrekeningen 1988.

**4** Voor het waarom van het bestaan van verschillende cijfers over werkloosheid zie J. Thijssen, Kijk op werkloosheid, ESB, 28 mei 1997 en C. Van Bochove, Werkloosheid is voor statistiek een te rekbaar begrip, *de Volkskrant*, 28 november 1997.

**5** Zie onder meer C.H. Arts e.a., *Vergelijkingsonderzoek werkloosheidsuitkeringen. Een vergelijking van gegevens van mensen met een bijstandsuitkering*, CBS, Voorburg, 1995.

**6** J.D. Kartopawiro en S.A. Ploeger, Vergelijkingsonderzoek arbeidsongeschiktheidsuitkeringen, 1995, *Sociaal-economische maandstatistiek*, januari 1998.