

## ARTIKEL

# Online panelonderzoek als panacee voor dalende respons in surveyonderzoek?\*

*Annemieke Luiten\*\**

## 1 Inleiding

In oktober 2021 organiseerde de Vereniging voor Statistiek en Onderzoek (VSO) een discussiemiddag voor haar leden met de prikkelende en enigszins onheilspellende titel 'Wie doet nog mee? Uitdagingen aan surveyonderzoek'. Verschillende gemeenten die surveyonderzoek doen onder hun bevolking, illustreerden dat het steeds moeilijker en duurder wordt om mensen te bewegen om mee te doen met hun onderzoeken en lieten inspirerende voorbeelden zien van hun inspanningen om het tij misschien te keren, of in ieder geval enigszins tegen te houden. Verschillende deelnemers spraken de overweging uit om over te stappen op een bewoners(internet)panel als alternatief. Een panel zou de kosten van veldwerk beperken. Anderzijds maakten de aanwezigen zich zorgen over de representativiteit van dergelijke panels.

Het afnemen van de respons en het toenemen van de kosten van traditioneel surveyonderzoek is voor velen een reden om te zoeken naar alternatieve bronnen. In medisch, sociologisch, psychologisch en marktonderzoek wordt inmiddels voornamelijk gebruikgemaakt van online panels of ook big data bronnen, zoals sociale media en online data bases. Zie Little et al. (2020) voor een lange lijst onderzoeken op basis van deze methoden. Het verschil met de traditionele survey-methode is dat de gegevens afkomstig zijn van panels of databronnen waarvan de kans dat mensen erin zitten onbekend is en dat sommigen helemaal geen kans hebben om erin te komen.

Ik was namens het CBS aanwezig om te spreken over de respons bij CBS-onderzoeken en de maatregelen die het CBS in de loop der jaren heeft genomen om de respons te stimuleren. Ook het CBS heeft een aantal jaren geleden onderzocht of een panel een oplossing zou zijn. Gedurende meer dan een jaar is diepgaand onderzoek verricht naar de mogelijkheden, de kosten, representativiteit en statistieken op basis van het panel. In deze bijdrage geef ik een samenvatting van de opzet van dit pilot-panel, de bevindingen en de overwegingen die hebben geleid tot de beslissing om géén CBS-panel op te richten.

\* De in dit rapport weergegeven opvattingen zijn die van de auteur en komen niet noodzakelijk overeen met het beleid van het Centraal Bureau voor de Statistiek. Met dank aan Joep Burger en Ger Snijkers (beiden CBS) en twee anonieme reviewers.

\*\* Annemieke Luiten is werkzaam bij het CBS als methodoloog op het gebied van primaire waarneming en is (co-)auteur van verschillende artikelen op dit gebied. Contact: a.luiten@cbs.nl.

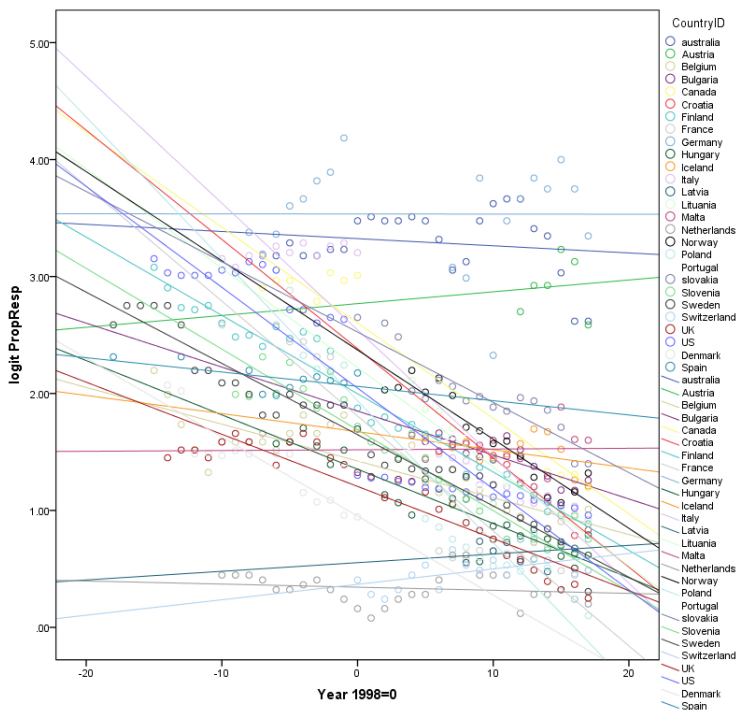
Annemieke Luiten

Eerst geef ik een kort overzicht van de responsontwikkelingen in traditioneel surveyonderzoek. Vervolgens beschrijf ik de soorten panels die onderscheiden worden en de voor- en nadelen van het werken met panels. Speciale aandacht zal daarbij worden gegeven aan representativiteit van de verschillende panels en wat we daarvoor moeten verstaan. Paragraaf 4 geeft een samenvatting van de opzet en bevindingen van de proef van het CBS met het opzetten van een panel. In paragraaf 5 wordt kort ingegaan op het wegen van surveyresultaten. Paragraaf 6 ten slotte bevat een korte samenvatting en aanbevelingen.

## 2 Responsontwikkelingen

De al jaren gehoorde zorg over de kwaliteit van surveyonderzoek klinkt de laatste jaren steeds luider. De Leeuw, Luiten en Stoop (2020) constateren in hun inleiding tot een 'special issue' over non-respons in het *Journal of Official Statistics* dat responscijfers zowel bij overheids-, markt- en academisch onderzoek onder druk staan. Lage responscijfers maken de kans dat de cijfers structureel vertekend zijn groot. Er zijn zorgen over bevolkingsgroepen die nauwelijks vertegenwoordigd zijn in onderzoek. Pogingen om de respons op peil te houden resulteren in oplopende kosten, terwijl budgetten eerder teruglopen. Verschillende factoren zijn van invloed op de responsontwikkelingen, zoals de terugloop van de mogelijkheid tot telefonisch onderzoek (telefoonnummers zijn steeds moeilijker te vinden), de introductie van online en mixed mode vragenlijsten (die, ondanks de vaak uitgesproken hoop op het tegendeel, over het algemeen geen responsstijging tot gevolg hebben, zie bijvoorbeeld Schouten et al., 2021), het toenemende gebruik van smartphones (waardoor mensen de vragenlijst op de smartphone willen invullen, met grotere kans op afbreuk tijdens het invullen), en de grote hoeveelheid verzoeken om allerlei soorten vragenlijsten in te vullen, met 'enquête-moeheid' tot gevolg.

De Leeuw, Hox en Luiten (2018) tonen in een analyse van de responscijfers van 36 jaar *Labour Force Surveys* (LFS) aan dat de responscijfers in Europa, de Verenigde Staten, Groot-Brittannië en Australië gestadig en lineair afnemen, gemiddeld over alle landen met 0.73 procentpunten per jaar. Dit wordt veroorzaakt door een toename van het aantal non-contacten (0.42 procentpunten per jaar) en een toename van het aantal weigeringen (0.40 procentpunten per jaar). Figuur 1 laat zien dat de gemiddelde achteruitgang in de responscijfers wordt geflatteerd door een aantal landen dat niet of nauwelijks kampt met responsdaling.

**Figuur 1** Responstrends in de Labour Force Survey, 1980-2015 in 27 landen

Noot: De responscijfers zijn logaritmisches getransformeerd. In 2015 is de hoogste respons gevonden in Duitsland (97%), Australië (93%), Portugal (89%) en de VS (87). In deze landen is deelname verplicht en neemt de respons ook nauwelijks af. De laagste respons in 2015 wordt gevonden in Polen (53%), Nederland (55%), het VK (56%) en Zweden (58%). In Nederland is de respons in de meetperiode tamelijk stabiel gebleven, maar in de andere landen is de respons respectievelijk 19, 23, en 25 procentpunten achteruitgegaan.

Luiten, Hox en De Leeuw (2020) tonen aan dat deze trends ook gelden voor andere surveys dan de (vaak verplichte) LFS. In dit onderzoek werd geanalyseerd welke (veldwerk)omstandigheden het verschil maken tussen landen die wel en niet met teruggang te kampen hebben. De deelnemende landen werden bevraagd over alle veldwerk-omstandigheden en design-aspecten waarvan de auteurs een mogelijke relatie met respons veronderstelden, zoals inzet van andere modes, incentives, aantal afgelegde bezoeken of telefonische contactpogingen, inzet van speciale brieven voor weigeringsconversie, toestaan van proxy-interviews, et cetera. Al deze factoren samen bepaalden echter maar een klein deel van de variantie tussen landen. Opvallend was dat er regelmatig een contra-intuïtieve relatie werd gevonden tussen maatregelen en respons. Zo was het inzetten van incentives positief gerelateerd aan het aantal weigeringen en het berekenen van maten voor representativiteit was negatief gerelateerd aan respons. De auteurs concludeerden dat vooral landen met een lage respons maatregelen zijn gaan inzetten om de respons om-

Annemieke Luiten

hoog te krijgen, en dat vooral landen met een lage respons de representativiteit van die respons willen bestuderen.

Een andere trend die Luiten et al. (2020) constateerden, was de toename van de kosten voor veldwerk, hoewel kosten door slechts zes landen werden gerapporteerd. Kosten van veldwerk hangen vooral samen met non-contact: het terugdringen van non-contact is kostbaar, waarschijnlijk als gevolg van het hoger aantal benodigde bezoeken. Kosten hingen niet significant samen met de overall respons en met het percentage weigeringen.

De responscijfers van het CBS zijn, tegen de trend in, redelijk stabiel binnen de onderzochte periode. Een van de redenen daarvoor is mogelijk dat de respons in Nederland al vanaf het begin van de meetperiode (1980) het laagst was van alle gemeten landen (de onderste lijn in figuur 1 is Nederland). De responscijfers voor de LFS, of de Enquête Beroepsbevolking zoals deze in Nederland heet, bewegen zich tussen de 50 en 60%. Dit responscijfer werd in het begin van de reeks behaald met een face-to-face waarneming, inmiddels via een mixed-mode design met web, telefonische en face-to-face waarneming. Zoals figuur 1 ook toont, zijn er inmiddels verschillende landen die op hetzelfde responsniveau zijn aangeland. Een volgende editie van dit internationaal responsonderzoek moet aantonen of dit een plateau blijkt waarop de respons van officiële statistieken zal blijven hangen. Een andere reden voor de relatieve stabiliteit in Nederland is dat het CBS in de loop der jaren heel veel onderzoek heeft gedaan naar responsbeïnvloeding, zowel in onderzoek met interviewers, als onderzoek via webvragenlijsten.

In de volgende paragraaf wordt beschreven in hoeverre de wens om de responscijfers en de kosten in de hand te houden kan worden ingewilligd door het opzetten van een (online) panel.

### 3 Soorten panels

Welk soort panel de voorkeur heeft, wordt onder meer bepaald door het doel waarvoor het gebruikt wordt (Pffor & Schröder, 2016). Het belangrijkste verschil betreft panels op basis van kanssteekproeven en panels die dat principe loslaten (non-probability panels<sup>1</sup>). Kanssteekproeven maken het in principe mogelijk om de resultaten te generaliseren naar een populatie (heel Nederland in het geval van het CBS, een hele gemeente of provincie in het geval van meer lokale steekproeven), mits er sterke hulpvariabelen aanwezig zijn om te wegen en de respons niet te laag of te selectief is. Bij non-probability panels kan dat over het algemeen niet en gelden de bevindingen alleen voor het panel, hoewel ook hier uitzonderingen mogelijk zijn indien sterke hulpvariabelen aanwezig zijn uit registers of een aanvullend survey op basis van een kanssteekproef.

Er zijn verschillende manieren om panels te werven, met verschillende methodologische implicaties. De American Association for Public Opinion Research (AAPOR)

1 Hier wordt de Engelse term gebruikt bij gebrek aan een goed ingevoerd Nederlands alternatief. Ook op andere plaatsen worden vaak Engelse termen gebruikt; de literatuur over surveymethodologie is over het algemeen Engelstalig; vertalingen van sommige termen zijn eerder verwarrend dan verhelderend.

publiceerde in 2011 een rapport over het destijds opkomende fenomeen ‘online panels’. Zij onderscheidde drie typen online panels:

- Probability (kanssteekproef-)panels. Aan de basis van een probability panel ligt een steekproef. Personen of huishoudens in de steekproef worden intensief benaderd in verschillende modi (telefonisch, persoonlijk, web) om deel uit te gaan maken van het online panel. Typerend is dat mensen die geen toegang hebben tot internet gefaciliteerd worden om wel toegang te krijgen, door het verstrekken van laptops of tablets en/of een internetaansluiting. Voorbeelden zijn het LISS-panel in Nederland, het Duitse Internet Panel en de Franse *Longitudinal Study by Internet for the Social Sciences* (het ELIPSS-panel) (Blom et al., 2016). Het eveneens Duitse GESIS-panel faciliteert mensen die geen internet hebben om mee te doen via schriftelijke vragenlijsten. Een iets minder elegante, maar ook minder kostbare, manier is om paneeldeelnemers te werven uit respondenten van een ander onderzoek met een kanssteekproef als basis. Eventuele vertekening in het eerste onderzoek kan worden gecorrigeerd door selectieve bijwerving van de groepen die ondervertegenwoordigd zijn. Dit is de benadering die door het CBS is gekozen in het onderzoek naar het opzetten van een CBS-panel, zie paragraaf 4. In beide gevallen kan de insluitkans in het panel worden uitgerekend, en is het mogelijk onzekerheidsmarges te berekenen. Resultaten kunnen worden beschreven in termen van populatieschattingen. Gemeenten die een kanssteekproef trekken uit hun gemeentelijke basisadministratie, kunnen op deze manier generaliseren naar alle bewoners van hun gemeente. In paragraaf 3.1 wordt een voorbeeld van een online probability panel, het LISS-panel, nader besproken.
- De tegenhanger van probability panels zijn de non-probability panels. Hierbij worden twee typen onderscheiden:
  - Vrijwilligerspanels. Deelnemers in vrijwilligerspanels worden specifiek geworven om deel te gaan nemen aan een panel. Dat kan op allerlei manieren, via advertenties, websites, tijdschriften, tv-programma’s. Ook kunnen e-mails van mensen die voor de een of andere reden worden geworven, worden gedeeld met derden. Deze panels zijn ook bekend onder andere namen, bijvoorbeeld opt-in panels en access panels. Deelnemers vullen eerst een vragenlijst in met demografische achtergrondkenmerken en eventueel ‘webographic’ of ‘psychographic’ variabelen die leefstijl en internetgedrag meten. Op basis van hun profiel worden zij vervolgens geworven voor vervolgonderzoek.

De belangrijkste reden voor de meeste mensen om aan een dergelijk panel mee te doen is de incentive die verdiend kan worden door deelname (Poynter & Comley, 2003 in AAPOR, 2011). Het risico hiervan is dat er vertekening optreedt als gevolg van economische zelfselectie, waardoor relatief te veel werklozen, mensen met pensioen, en studenten in een dergelijk panel zitten (Lehdonvirta et al., 2020).
- Panels op basis van zogenaamd ‘river sampling’. River sampling is een methode waarbij respondenten worden geworven wanneer zij online zijn. Bij river sampling wordt een uitnodiging voor een enquête, bijvoorbeeld via een link, voorgelegd aan een websitebezoeker terwijl deze bezig is met een

Annemieke Luiten

andere online activiteit. Respondenten selecteren zichzelf op basis van hun interesse in de inhoud (*topical self selection*). De achtergrond van de bezoekers wordt niet eerst geanalyseerd. Hierdoor zal het onderwerp van onderzoek vaker voorkomen in een river sample dan in meer neutrale steekproeven (Lehdonvirta et al., 2020).

In paragraaf 3.2 wordt nader ingegaan op non-probability panels.

### 3.1 *Het LISS-panel: voorbeeld van een probability panel*

Het onderzoeksinstituut CentERdata, onderdeel van de Universiteit van Tilburg, heeft in 2008 het online LISS-panel (Longitudinal Internet Studies for the Social Sciences) opgezet. Het bestaat uit bijna 8000 personen van 16 jaar en ouder in 5000 huishoudens, die elke maand een vragenlijst invullen. Daarvoor krijgen zij een incentive van € 15 per uur interviewtijd. Zij worden per e-mail uitgenodigd om de vragenlijst in te vullen. Eén keer per jaar wordt een basisvragenlijst afgenomen met steeds dezelfde vragen.

Het panel is opgebouwd door een aselechte steekproef van huishoudens te trekken uit het bevolkingsregister van Nederland. De geselecteerde huishoudens werden telefonisch of face-to-face benaderd. Bovendien werden bereidwillige huishoudens zonder internettoegang voorzien van apparatuur waarmee zij toegang tot internet kregen. Om longitudinale analyses mogelijk te maken wordt geïnvesteerd in het zo lang mogelijk in het panel houden van de deelnemers. Van de benaderde adressen heeft uiteindelijk 41% in het panel plaatsgenomen. Elke twee jaar wordt er een verversingssteekproef getrokken door het CBS om uitval op te vangen. Per jaar valt ongeveer 10% van de huishoudens uit. Elke maand respondeert tussen de 66% en 87% van de panelleden op een uitnodiging om een vragenlijst in te vullen.

Er zijn verschillende experimenten gedaan om te achterhalen hoe mensen het best geïnteresseerd gehouden kunnen worden. Succesvolle ingrepen zijn: mensen bellen die meer dan twee maanden niet hebben gerespondeerd (32% van de gebelde 'slapers' werd weer actief). Het tevens aanbieden van een incentive van € 10 om weer te beginnen was nog succesvoller (43% van de 'slapers' werd weer actief na een telefoontje en de beloofde incentive). Alleen een incentive bieden was minder succesvol dan de combinatie met het bellen (31-40%, afhankelijk van het bedrag). Niet succesvol bleek het geven van feedback over de verzamelde resultaten (Scherpenzeel, 2014).

Verschillende auteurs hebben onderzoek gedaan naar de representativiteit van het LISS-panel (o.a. Hoogendoorn & Daalmans, 2008; Van der Laan, 2009; Scherpenzeel & Bethlehem, 2011; Scherpenzeel & Schouten, 2011). Scherpenzeel en Bethlehem (2011) vergelijken de samenstelling van het LISS-panel met de Nederlandse populatie, met de verdeling in een traditioneel CBS-survey (het Nederlands Kiezersonderzoek van 2006), met een gemiddelde van 19 Nederlandse zelfselectiepanels (de 19 panels die participeerden in het Nederlands Online Panel VergelijkingsOnderzoek), waarbij de panels tegelijkertijd een zelfde soort steekproef trokken uit hun panel en deze dezelfde vragenlijst voorlegden (NOPVO; Van Ossenbruggen et al., 2006) en met een zelfselectie-survey (21minuten.nl). De verdeling van een aantal demografische kenmerken werd bestudeerd in al deze bron-



nen (70+ers, singles, en deelnemers met een niet-westerse migratieachtergrond). Ook werd de verdeling van een aantal webographic variabelen bestudeerd: stemmen in de laatste landelijke verkiezingen, stemmen voor specifieke partijen (SP en CDA) en internetgebruik.

De vergelijking met de populatietotalen toonde aan dat het LISS-panel een ondervertegenwoordiging had van mensen van 70 jaar en ouder, van mensen die alleen wonen en van mensen met een niet-westerse migratieachtergrond. Er was een oververtegenwoordiging van mensen die hadden gestemd in de laatste landelijke verkiezingen. Een vergelijking met het NKO toonde aan dat de verdeling van het LISS erg dicht lag tegen de verdeling bij het face-to-face onderzoek, met uitzondering van de verdeling van ouderen en niet-internetbezitters. Een vergelijking van het LISS met het 21minuten.nl-onderzoek toonde aan dat het LISS een betere verdeling had van de ouderen en niet-internetgebruikers, niet-westerse migranten en CDA-stemmers. Andere groepen, zoals SP-stemmers en eenpersoonshuishoudens waren net zo goed of zelfs iets beter vertegenwoordigd in het 21minuten.nl-onderzoek. Een vergelijking van het LISS-panel met steekproeven uit de 19 NOPVO-panels toonde aan dat de NOPVO-panels meer van de populatie verschilden dan het LISS-panel op vier van de zes kenmerken die werden bestudeerd. Maar er waren ook kenmerken waarop de NOPVO-panels minder van de populatie verschilden dan het LISS-panel, met name het aandeel eenpersoonshuishoudens.

De conclusie van Scherpenzeel en Bethlehem (2011) was dat (internet)panels dicht in de buurt kunnen komen van een traditioneel survey, mits een correct sample design wordt gebruikt. De ondervertegenwoordiging van ouderen en de niet-internetgebruikers werd door deze aanpak niet helemaal tenietgedaan. Een aanvulling met papieren vragenlijsten zoals gedaan wordt in het GESIS-panel (Blom et al., 2016), zou dit probleem mogelijk verhelpen, maar heeft uiteraard aanzienlijke praktische en financiële consequenties.

### 3.2 *Non-probability online panels*

Verschillende vooraanstaande surveymethodologen hebben bijgedragen aan het overzicht dat de AAPOR in 2011 publiceerde over de kwaliteit van (non-probability) online panels (Baker et al., 2013). De schrijvers vatten samen dat er een grote kans op fouten (*survey errors*) is bij non-probability panels door een aantal redenen. In 2011 waren er nog aanzienlijke dekkingproblemen: 25% van de Amerikanen, waar het literatuuroverzicht voornamelijk op gebaseerd was, had nog geen internet. Dit is inmiddels voor de Nederlandse situatie veel minder een probleem: in 2019 maakte 97% van de Nederlanders van 12 jaar en ouder gebruik van internet. De Nederlanders die in 2019 nog nooit hadden geïnternet, waren over het algemeen vrouwen van 75 jaar of ouder (CBS: 'Hoeveel procent van de Nederlanders beschikt thuis over internet?'). Probability panels die hun leden aansluiten op internet en/of een tablet of laptop ter beschikking stellen, hebben dit probleem minder.

Een tweede reden die de auteurs noemen, is dat de respons erg laag is: vaak respondeert nog geen 10% van de panelleden op een uitnodiging om een vragenlijst in te vullen. Ook is er een grote kans op vertekening door zelfselectie, doordat vooral mensen ingeschreven zijn met een specifieke belangstelling voor surveys in zijn

Annemieke Luiten

algemeenheid, surveys over een bepaald onderwerp, of voor de beloningen die ze kunnen krijgen. Anderzijds zijn grote groepen mensen nauwelijks aanwezig in de panels (jongeren, 75+ers, mensen met een migratieachtergrond).

Afgezien van de samenstelling van de panels spraken de auteurs ook zorgen uit over de kwaliteit van de antwoorden. Er kan sprake zijn van alarmerende niveaus van 'satisficing'. Hiermee wordt het gedrag bedoeld waarmee mensen met zo min mogelijk (geestelijke) moeite door een vragenlijst gaan en bijvoorbeeld structureel de midden-categorie van een schaal kiezen, niet differentiëren tussen antwoorden op een schaal, bovenmatig de 'weet niet'-optie kiezen, of filtervragen zo invullen dat vervolgvragen worden vermeden.

Een apart gevaar van met name zelfselectie-surveys of river panels, zeker op het gebied van beleidsonderzoek, is dat het moeilijk is om te controleren wie er meedoet; lobbyisten kunnen groepen mensen activeren om mee te doen, en ook is het mogelijk om een vragenlijst meer dan één keer in te vullen.

Sinds 2011 zijn er verschillende aanvullende onderzoeken gedaan in andere culturen en omstandigheden. Brügggen et al. (2016) vergeleken de kwaliteit van 18 vrijwilligerspanels in Nederland. De benchmark voor de demografische verdeling was de Basis Registratie Personen. Voor de verdeling van werksituatie en opleiding was de benchmark de (face-to-face afgenomen) CBS Enquête Beroepsbevolking en voor sociale kenmerken was de benchmark het LISS-panel. De 18 online studies namen deel aan NOPVO-onderzoek, zie paragraaf 3.1, en vulden voor het onderzoek van Brügggen et al. een vragenlijst in over kenmerken van hun onderzoek. Een van de 19 panelhouders vulde de vragenlijst niet in en werd niet opgenomen in het onderzoek.

Het bleek dat de resultaten van de verschillende non-probability panels onderling heel verschillend waren, soms op een bepaald soort variabele, maar soms ook op bijna alle variabelen. Ook verschilden de panels in de relatie tussen variabelen en soms zelfs in de richting van die relatie: waar in het ene panel een positieve correlatie werd gevonden, werd in het andere panel een negatieve gevonden. Het LISS-panel presteerde beter dan de 18 opt-in panels. Pogingen om de resultaten te wegen verminderden de vertekening niet, noch qua aandeel mensen in subgroepen, noch qua relatie tussen variabelen.

Lehdonvirta et al. (2020) beschrijven onderzoek waarbij twee river samples, een vrijwilligerspanel en een probability sample in Finland worden vergeleken. In elk van de vier onderzoeken werden dezelfde vragen gesteld. Doel was te onderzoeken hoe goed de non-probability river en panel samples de populatie beschreven in termen van demografische verdeling en niet-demografische subgroepen. Het vrijwilligerspanel was een commercieel online panel. De eerste river sample werd geworven via een link in een nieuwsartikel op de website van de Finse televisie over het onderwerp van de vragenlijst. De tweede river sample werd geworven via Facebook-advertenties. De benchmark was een telefonisch onderzoek van het Finse statistiekbureau. Doelvariabelen waren antwoorden op twee vragen: 'Heb je ooit op internet hatelijke of vernederende taal gezien die bepaalde groepen mensen of individuen aanvalt?' en 'Ben je het doelwit geweest van online pesten of intimidatie?'.



Demografische verdeling werd bepaald op basis van leeftijd, geslacht, Facebook-gebruik (ja/nee) en voltijd student zijn. De resultaten toonden dat de demografische verdeling van het vrijwilligerspaneel heel dicht tegen de benchmark aan lag, maar dat de river samples tot 27 procentpunten afweken voor het aandeel mannen, en 29 procentpunten voor het aandeel studenten. Voor de doelvariabelen werd bij beide river samples 15.3 tot 38.9 procentpunten verschil gevonden. Het vrijwilligerspaneel deed het beter, met afwijkingen van 0.7 en 2.5 procentpunten op de demografische variabelen, en 2.3 tot 14.4 procentpunten op de doelvariabelen. In alle gevallen werd geprobeerd een weging toe te passen die echter geen effect had op de resultaten.

Lehdonvirta et al. (2020) concluderen dat river samples geen bruikbare manier vormen om de aanwezigheid van een verschijnsel in de algemene bevolking te meten. Het vrijwilligerspaneel toonde een nauwkeurigere afspiegeling van de algemene bevolking. Desondanks suggereren zij dat er wel gebieden zijn waarop river samples gebruikt kunnen worden, zolang er voorzichtige conclusies worden getrokken. Voorbeelden zijn exploratie van een thema bij specifieke groepen, methodologische studies, of (voorlopig) onderzoek naar verschillen tussen groepen.

#### 4 Ervaringen met een CBS-internetpaneel

In 2013 besloot het CBS tot het onderzoeken van de mogelijkheden van een internetpaneel in verband met de verwachte kortere doorlooptijd en lagere kosten. Het paneel moest een probability paneel zijn: het was de bedoeling om op basis van het paneel te generaliseren naar de hele populatie.

Gedurende een jaar werd een paneel gebouwd en onderhouden, bestaande uit respondenten van een regulier CBS-survey (het Onderzoek Verplaatsingsgedrag in Nederland), die aangaven deel te willen uitmaken van het paneel. OViN was in 2013 een mixed-mode onderzoek waarbij mensen zowel via internet, telefoon als face-to-face mee konden doen. Panedeelnemers zouden een jaar lang elke maand een van de reguliere CBS-vragenlijsten voorgelegd krijgen, om te zien of de resultaten (na weging) overeenkwamen met parallel uitgezette reguliere onderzoeken. De werving was relatief laag intensief (Snijkers & Luiten, 2013): er werden geen incentives ingezet, noch bij de werving, noch voor het maandelijks invullen van de vragenlijst, er waren geen speciale wervingsmaterialen, er werd niet geprobeerd om de niet-internetpopulatie te faciliteren, er werden geen pogingen gedaan om mensen die niet bereidwillig waren alsnog over te halen, allemaal factoren die het LISS-paneel zo succesvol maakten.

Ongeveer 60% van de OViN-respondenten toonde zich in principe bereid tot deelname aan het paneel. Niet iedereen deed uiteindelijk ook mee, echter: 55% van de bereidwilligen die in OViN ook via internet hadden gedaan, en maar 17% van de bereidwilligen die via telefoon of face-to-face hadden gerespondeerd. Dat kwam neer op respectievelijk 30, 10 en 13% van de OViN-respondenten en een uiterst klein percentage van de oorspronkelijke steekproef. Op het moment dat het paneel echt aan de slag was, was de maandelijks respons ongeveer 75%, vergelijkbaar met de maandelijks respons in het LISS-paneel.

Annemieke Luiten

Om te zien of de samenstelling van het panel overeenkwam met de populatie werd gekeken naar een aantal kenmerken: leeftijd, geslacht, huishoudinkomen, herkomst, samenstelling van het huishouden, stedelijkheid, eigendom van de woning, burgerlijke staat en sociaaleconomische klasse, kenmerken die ook bekend waren van de steekproef. Er waren een aantal groepen waarvan de samenstelling aanzienlijk afweek. Zo waren jongeren (onder de 25 jaar) en ouderen (75+) aanzienlijk ondervertegenwoordigd, en was de groep 55-64-jarigen aanzienlijk oververtegenwoordigd. Ook gehuwden, mensen met een Nederlandse achtergrond, paren zonder kinderen, en mensen met een hoog huishoudinkomen waren oververtegenwoordigd. Ter vergelijking: in de steekproef had 25% van de mensen een hoog inkomen, in de respons van het wervingsonderzoek was dat 27%, maar bij de uiteindelijke paneldeelnemers was dat 40%.

In de analyses van de antwoorden op de vragenlijsten die aan het panel werden voorgelegd, is geprobeerd deze scheefheid zo te wegen dat deze overeen zouden komen met de respons op de wervingsvragenlijst. De te schatten variabelen waren opleiding en werk. Dit bleek echter niet mogelijk. Correctie door weging was vaak te veel of te weinig, of zelfs niet in de goede richting: het percentage gepensioneerden in het panel was te hoog, en na weging nog hoger. Ook bij onderzoeken die in latere maanden werden uitgezet bij het panel, bleek dat er in een groot aantal vergelijkingen significante verschillen waren tussen de resultaten uit het panel en die van het reguliere onderzoek, die met weging niet (volledig) gecorrigeerd konden worden. Het was echter niet helemaal duidelijk in hoeverre deze verschillen het gevolg waren van de samenstelling van het panel (een selectie-effect), of van het feit dat in het panel alle respons uit een internetvragenlijst afkomstig was, terwijl het reguliere onderzoek een mixed-mode design had met zowel internet, telefonische en face-to-face waarneming (een meet-effect).

De eerste bevindingen toonden dus dat de respons van het panel laag was, de samenstelling scheef, dat de inhoudelijke bevindingen afweken van de reguliere bevindingen en niet afdoende konden worden gewogen. Een vierde bevinding gaf de uiteindelijke doorslag om niet op de ingeslagen weg verder te gaan, en dat waren de kosten. Het werven van een representatief panel waarbij mensen die geen internet hebben worden gefaciliteerd om mee te doen, kost erg veel geld, zeker als er in andere modes dan alleen via internet geworven wordt. Vervolgens moet er minimaal jaarlijks worden bijgeworven: mensen beëindigen hun deelname, en het panel wordt steeds ouder, zodat er jongeren bij moeten komen. Het CBS voorzag bovendien dat eens in de drie jaar het hele panel ververst zou moeten worden, om te voorkomen dat mensen te zeer op routine zouden gaan antwoorden (*panel conditioning*). Op basis van de ervaringen van het LISS-panel werd de jaarlijkse uitval geschat op 20%. Daarnaast zou 27% verversing nodig zijn (80% verdeeld over drie jaar), en zou jaarlijks 2% jongeren toegevoegd moeten worden. Samen is dat 49% verversing en aanvulling per jaar (Visschers, 2013). Ook moet er een systeem worden gebouwd om panelbeheer, contactregistratie, communicatie met panelleden en het beheer van financiële incentives uit te voeren. Deze combinatie van factoren leidde ertoe dat het CBS de businesscase voor het bouwen, beheren en onderhouden van een eigen panel negatief beoordeelde.

## 5 Wegen

Op verschillende plaatsen in deze bijdrage wordt gewag gemaakt van pogingen om resultaten te wegen of te kalibreren om de verdeling van het panel meer in overeenstemming te brengen met de verdeling van de populatie. In deze pogingen werd gebruikgemaakt van technieken die waren ontwikkeld voor kanssteekproeven, hoewel ook opkomende alternatieven werden uitgetoetst. Steeds bleek dat deze weegmethodes niet in staat waren om de vertekening volledig te corrigeren. De mate waarin met succes vertekening kan worden gecorrigeerd, hangt af van de mate waarin de beschikbare hulpvariabelen het data-generatieproces verklaren (Baker et al., 2013; Bethlehem & Biffignandi, 2011).

Hulpvariabelen zijn variabelen die zijn gemeten in de survey en waarvan de verdeling in de populatie bekend is. Door de verdeling in de populatie te vergelijken met de verdeling in de survey kan worden berekend of deze representatief is, in ieder geval met betrekking tot de hulpvariabelen. Door te wegen kan een gebrek aan representativiteit worden opgeheven, wederom met betrekking tot de hulpvariabelen. De hoop is dat de respons daardoor ook representatief wordt voor de doelvariabelen. De keus en de meting van de hulpvariabelen zijn cruciaal om succesvol te kunnen wegen. De vertekening wordt alleen minder als gevolg van wegen als de hulpvariabelen zowel responsgedrag verklaren als sterk zijn gerelateerd aan de onderzoeksvariabelen (Bethlehem & Biffignandi, 2011).

Baker et al. (2013) schrijven dat probability sampling niet de enige manier is waarop valide statistische conclusies kunnen worden getrokken over doelpopulaties en dat in andere wetenschapsgebieden geavanceerde statistische methoden worden gebruikt voor data uit non-probability bronnen. Tot op de dag van vandaag proberen surveymethodologen en statistici om dergelijke geavanceerde methoden te ontwikkelen voor non-probability surveys en andere data, zie bij voorbeeld Valliant (2019), Rao (2020), Little et al. (2020). De gouden graal lijkt echter nog niet gevonden.

## 6 Conclusie

De aanleiding voor het schrijven van dit artikel was het samenvallen van de overweging van enkele VSO-leden om over te stappen op een bewonerspanel, en de ervaringen van het CBS die ertoe leidden een dergelijk panel niet op te richten. Gedurende het schrijven van het artikel kwam ik tot de overtuiging dat die twee ervaringen misschien niet zo veel met elkaar te maken hebben als op het eerste oog leek. Het CBS heeft met het panel getest of het mogelijk was om de schattingen van populatietotalen uit surveyonderzoek te reproduceren, iets dat niet voldoende mogelijk bleek. Maar methodologisch onderzoek zou met een dergelijk panel uitstekend kunnen. Het hangt er dus volledig van af wat het precieze doel is van een panel: moet het de populatie (zoals alle personen van 16 jaar en ouder in een bepaalde gemeente) beschrijven, of moet het de gelegenheid bieden snel inzicht te krijgen in de mening van geëngageerde bewoners die zich hebben aangemeld voor een panel over gemeenteaangelegenheden.

Annemieke Luiten

Het doel bepaalt dus het soort onderzoek en het soort panel dat gebruikt kan worden. Populatieschattingen kunnen nog steeds het best worden gemaakt via kanssteekproeven, die ook de vorm kunnen aannemen van een panel op basis van een kanssteekproef, zoals bij het LISS-panel het geval is. Aan een dergelijk panel zijn echter aanzienlijke kosten en inspanningen verbonden om het panel representatief en geïnteresseerd te houden. Het CBS berekende dat voor zijn doelen het panel niet kostenbesparend zou zijn. Vrijwilligerspanels kunnen aanzienlijk goedkoper zijn, maar hebben het nadeel dat er waarschijnlijk een sterk selectie-effect aanwezig is van overmatig geïnteresseerde deelnemers, plus een onderdekking van belangrijke groepen in de samenleving. Het resultaat is systematische vertekening als gevolg van correlaties tussen het onderwerp van de studie, de kans dat iemand (geen) internet heeft, en de kans dat iemand aan een specifiek panel mee wil doen. Hoewel onderzoek aantoont dat ook probability samples zelden of nooit volledig representatief zijn omdat de non-respons over het algemeen niet random is (zie bijv. Groves & Couper, 1989), was de conclusie van Brick (2011: 884) in een blik op de geschiedenis en toekomst van surveys dat '(...) a well-conducted probability sample with a low response rate is likely to be of higher quality (lower average bias for a broad range of estimates) than a sample of volunteers'. Misschien is dat een geruststelling voor VSO-leden die zich zorgen maken over de lage respons van hun surveys.

## Literatuur

- AAPOR. (2011). AAPOR report on online panels: Research synthesis. *Public Opinion Quarterly*, 74, 711-781.
- Baker, R., Brick, J.M., Bates, N.A., Battaglia, M., Couper, M.P., Dever, J.A., . . . Tourangeau, R. (2013). Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology*, 1, 90-143.
- Bethlehem, J., & Biffignandi, S. (2011). *Handbook of web surveys*. Hoboken, NJ: Wiley & Sons.
- Blom, A.G., Bosnjak, M., Cornilleau, A., Cousteaux, A.S., Das, M., Douhou, S., & Krieger, U. (2016). A comparison of four probability-based online and mixed-mode panels in Europe. *Social Science Computer Review*, 34, 8-25. doi:10.1177/0894439315574825
- Brick, M. (2011). The future of survey sampling. *Public Opinion Quarterly*, 73, 872-888.
- Brüggen, E., Van den Brakel, J., & Krosnick, J. (2016). *Establishing the accuracy of online panels for survey research*. CBS discussion paper, toegankelijk via <https://www.cbs.nl/en-gb/background/2016/15/establishing-the-accuracy-of-online-panels-for-survey-research>
- CBS. (2020). *Hoeveel procent van de Nederlanders beschikt thuis over internet?* Te raadplegen via <https://www.cbs.nl/nl-nl/cijfers/detail/83429NED?dl=27A20>
- De Leeuw, E., Hox, J., & Luiten, A. (2018). International nonresponse trends across countries and years: An analysis of 36 years of Labour Force Survey data. *Survey Insights: Methods from the Field*. Retrieved from <https://surveyinsights.org/?p=10452>
- De Leeuw, E., Luiten, A., & Stoop, I. (2020). Preface to the special issue on nonresponse. *Journal of Official Statistics*, 36(3), 463-468. <https://doi.org/10.2478/jos-2020-0024>
- Groves, R., & Couper, M. (1998). *Nonresponse in household interview surveys*. New York: Wiley.

- Hoogendoorn, A., & Daalmans, J. (2008). *Nonresponse in the recruitment of an internet panel based on a probability sample*. CBS Discussion paper 08007.
- Lehdonvirta, V., Oksanen, A., Räsänen, P., & Blank, G. (2020). Social media, web, and panel surveys: Using non-probability samples in social and policy research. *Policy and Internet*, 13, 134-155.
- Little, R., West, B., Boonstra, P., & Hu, J. (2020). Measures of the degree of departure from ignorable sample selection. *Journal of Survey Statistics and Methodology*, 8, 932-964. <https://doi.org/10.1093/jssam/smz023>
- Luiten, A., Hox, J., & De Leeuw, E. (2020). Survey nonresponse trends and fieldwork effort in the 21st century: Results of an international study across countries and surveys. *Journal of Official Statistics*, 36, 469-487. <http://dx.doi.org/10.2478/JOS-2020-0025>
- Pfarr, K., & Schröder, J. (2016). Why panel surveys? *GESIS Survey Guidelines*. Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences. Doi:10.15465/gesis-sg\_en\_008
- Rao, J. (2020). On making valid inferences by integrating data from surveys and other sources. *Sankhya B*, 1-31.
- Scherpenzeel, A. (2014). Survey participation in a probability-based internet panel in the Netherlands. In U. Engel, B. Jann, P. Lynn, A. Scherpenzeel, & P. Sturgis (Eds.), *Improving survey methods: Lessons from recent research* (p. 223-236). New York: Routledge.
- Scherpenzeel, A., & Bethlehem, J. (2011). How representative are online panels? Problems of coverage and selection and possible solutions. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the internet: Advances in applied methods and research strategies* (p. 105-132). New York: Routledge.
- Scherpenzeel, A., & Schouten, B. (2011). *LISS panel R-indicator: Representativity in different stages of recruitment and participation of an internet panel*. Paper presented at the 22nd International Workshop on Household Survey Nonresponse, 5th-7th September 2011, Bilbao.
- Schouten, B., Van den Brakel, J., Buelens, B., Giesen, D., Luiten, A., & Meertens, V. (2021). *Mixed-mode official surveys: Design and analysis*. Chapman and Hall/CRC. <https://doi.org/10.1201/9780429461156>
- Snijkers, G., & Luiten, A. (2013). *Voorstellen voor de opzet van een CBS-webpanel*. CBS-rapport.
- Valliant, R. (2019). Comparing alternatives for estimation from nonprobability samples. *Journal of Survey Statistics and Methodology*, 8, 231-263. <https://doi.org/10.1093/jssam/smz003>
- Van der Laan, J. (2009). *Representativity of the LISS panel*. CBS discussion paper 09041.
- Van Ossenbruggen, R., Vonk, T., & Willems, P. (2006). *Results Dutch Online Panel Comparison Study (NOPVO)*. Verkrijgbaar via Results\_NOPVO\_English.pdf (websm.org)
- Visschers, J. (2013). *Business case CBS web panel fase 2*. CBS-rapport.
- West, B., Little, R., Andridge, R., Boonstra, P., & Alvarado-Leiton, F. (2021). *New measures for assessing non-ignorable selection bias in non-probability samples and low response rate probability samples*. Social Statistics Seminar, University of Manchester, December 7th, 2021.