

Een terugblik op de ‘Proeve van onderzoek naar de kwaliteit van gegevens bekomen via het survey-interview’

Geert Loosveldt*¹

Samenvatting

Het oorspronkelijke artikel ‘Proeve van onderzoek naar de kwaliteit van gegevens bekomen via het survey-interview’ van Geert Loosveldt is terug te vinden op sociologos.be/terugblik.

De abstract van het oorspronkelijke artikel luidde:

“In dit artikel wordt nagegaan in welke mate binnen een bepaald onderzoek de doelstelling van het survey-interview, namelijk het op betrouwbare wijze bekomen van geldige informatie, gerealiseerd werd. Uit veronderstellingen omtrent de aard van een interview werden hiervoor drie diagnostische momenten afgeleid: interviewvariantie, analyse van de interviewduur en de rangorde van de interviews. De afhankelijke variabelen die in de analyses werden gebruikt zijn: het aantal sporten dat men het laatste jaar beoefend heeft, het aantal ooit beoefende sporten en het aantal sporten waarvoor men ooit belangstelling heeft gehad.

Uit de resultaten bleek dat de interviewer op zichzelf en kenmerken van interviewers zoals ervaring, snelheid en eigen rapportering van actuele en vroegere sportbeoefening en belangstelling een zekere verklaringskracht hebben ten aanzien van deze drie basisvariabelen.

De vaagheid van de vraagstelling – met daaraan verbonden de grotere mogelijkheid voor interpretatief gedrag in functie van eigen rapportering van de interviewer – en een negatieve kosten-batenanalyse konden daarmee in verband gebracht worden”.

Inleiding

In het derde nummer van de derde jaargang (1982) van het *Tijdschrift voor Sociologie* werd een artikel gepubliceerd waarin een aantal aanzetten werden gegeven om de datakwaliteit van survey-interviews te evalueren (Loosveldt, 1982). Daarbij werden gegevens geanalyseerd die verzameld werden door 67 interviewers in het kader van

* geert.loosveldt@soc.kuleuven.be

1 KU Leuven, Centrum voor Sociologisch Onderzoek

een surveyonderzoek naar sport en vrijetijdsbesteding in Vlaanderen 1980 (n = 1502). De probleemstelling van het artikel kaderde in een algemene kritiek op het surveyonderzoek, zoals die onder andere werd geformuleerd in boeken van Derek Phillips in de jaren 70 met de niet mis te verstane titels *Knowledge from what?* en *Abandoning Method* (Phillips, 1971, 1973). Kort en ruw samengevat kwam de kritiek erop neer dat sociologen veel te afhankelijk zijn van interviews en vragenlijsten en dat deze methoden van dataverzameling gebruikt worden in omstandigheden die daarvoor niet geschikt zijn. Verder kunnen survey-interviews waarbij interviewers toevallig geselecteerde respondenten op een gestandaardiseerde wijze bevragen aan de hand van een gestandaardiseerde vragenlijst omwille van een aantal vertekende factoren (o.a. sociale wenselijkheid) niet beschouwd worden als een betrouwbare en bruikbare methode van dataverzameling in sociaalwetenschappelijk onderzoek (Phillips, 1971, p. 11). In het artikel van 1982 wordt ingezoomd op één van deze vertekende factoren, met name 'intervieweffecten'. Algemeen gedefinieerd gaat het om de impact die interviewers hebben op de verkregen data. Alvorens verder in te gaan op deze specifieke factor, formuleren we kort enkele algemene bedenkingen over het gebruik van survey-interviews in sociaalwetenschappelijk onderzoek en de ontwikkeling van de surveymethodologie tijdens de afgelopen 35 jaar.

Ondanks de kritiek in de jaren 70 worden survey-interviews nog steeds frequent gebruikt in sociaal wetenschappelijk onderzoek. Dit blijkt onder andere uit een aantal recurrente grootschalige cross-nationale surveys die momenteel worden georganiseerd: European Value Study (EVS); European Social Survey (ESS), Survey of Health, Ageing and Retirement in Europe (SHARE), Programme for the International Assessment of Adult Competencies (PIAAC) en Labor Force Survey (LFS). Ook in de Vlaanderen en België zijn er terugkerende surveys, zoals de Nationale Gezondheidsenquête die in 2013 voor de vijfde keer werd georganiseerd en de survey naar 'Sociaal-culturele verschuivingen in Vlaanderen' (SCV-survey) die reeds vanaf 1996 jaarlijks wordt georganiseerd. Deze illustratieve lijst van face-to-face surveys laat niet toe om precieze uitspraken te doen over de reële omvang van het actuele gebruik van survey-interviews. Wel maakt de lijst duidelijk dat de kritiek die al in de jaren 70 werd geuit er niet voor gezorgd heeft dat survey-interviews moeten of kunnen worden beschouwd als een 'Abandoning Method', zijnde een data verzamelingsmethode waarvan we beter geen gebruik meer maken. Men zou zelfs terecht kunnen opmerken dat een dergelijke kritiek het surveymethodologisch onderzoek heeft gestimuleerd en heeft bijgedragen tot de verdere ontwikkeling van surveymethodologie als een volwaardige discipline. In die zin was de kritiek wel terecht maar niet de daaraan gekoppelde conclusie. Bij de ontwikkeling van de surveymethodologie tijdens de afgelopen 35 jaar speelt het concept 'Total Survey Error' (TSE) een belangrijke rol. Dit blijkt onder andere uit de publicatie in 2010 van een themanummer van *Public Opinion Quarterly* over dit concept (Biemer & Lyberg, 2010). De oorsprong van de TSE-benadering kan gesitueerd worden in een in 1979 gepubliceerd boek van Anderson, Kasper en Frankel met de titel *Total Survey Error*. In het eerste hoofdstuk van dit boek presenteren de auteurs een decompositie van de totale surveyfout op basis van het onderscheid tussen (1) variabele (*variance*) en systematische

fouten (*bias*), (2) steekproeffouten (*sampling errors*) en fouten die betrekking hebben op de representativiteit maar niet aan de steekproef gerelateerd zijn (*nonsampling errors*) en (3) fouten die men maakt omwille van het observeren (*errors of observation*) en fouten die optreden omdat men niet observeert (*nonobservational errors*). De combinatie van deze drie criteria die gebruikt worden om fouten te onderscheiden resulteert in een systematische classificatie van fouten die tijdens de verschillende fases van een surveyonderzoek kunnen optreden. De classificatie maakt duidelijk dat er tijdens een surveyonderzoek meer dan alleen maar steekproeffouten kunnen optreden. De TSE-benadering hoeft zich niet te beperken tot een afzonderlijke cross-sectionele survey maar kan ook worden toegepast op cross-culturele of cross-nationale surveys (Smith, 2011). Bij een dergelijke toepassing kan de TSE-benadering worden uitgebreid met fouten die veroorzaakt worden door een gebrek aan onderlinge vergelijkbaarheid van de surveys. Wanneer verschillen tussen landen worden beschouwd als reële verschillen maar in feite veroorzaakt worden door een verschillende impact van diverse fouten in verschillende landen, dan heeft men te maken met een vergelijkingsfout (*comparison error*). De laatste uitbreiding van de TSE-benadering naar cross-nationale surveys illustreert dat dit conceptuele kader zich nog steeds verder ontwikkelt en in de toekomst een belangrijke rol zal blijven spelen binnen de surveymethodologie.

In wat volgt gaan we na hoe de probleemstelling over de kwaliteitsevaluatie van data die verzameld worden door middel van survey-interviews, zoals die werd uitgewerkt in het artikel van 1982, kan gesitueerd worden in de TSE-benadering.

Situering van de interviewereffecten in het 'Total Survey Error'-kader

Zoals reeds werd opgemerkt in de inleiding, staat de evaluatie van interviewereffecten centraal in de evaluatie van de datakwaliteit die in het artikel van 1982 werd gerapporteerd. Wanneer we interviewereffecten situeren in het TSE-kader, kunnen we interviewereffecten algemeen omschrijven als niet steekproef gerelateerde fouten die kunnen optreden omdat interviewers een belangrijke rol spelen tijdens de dataverzamelingsfase van een surveyonderzoek. In surveyonderzoek zijn de interviewers de go-between tussen de onderzoeker die informatie wil verkrijgen bij een representatieve steekproef en de respondenten die deel uitmaken van die steekproef. Bij de uitvoering van hun taak kunnen interviewers verantwoordelijk zijn voor systematische en variabele fouten. Er treden systematische fouten op wanneer alle interviewers een gelijkaardige impact hebben op de verzamelde data, zoals het ontlokken van sociaal wenselijke antwoorden. Variabele fouten treden bijvoorbeeld op wanneer interviewers een vraag niet altijd op dezelfde wijze correct verduidelijken. Interviewereffecten kunnen in termen van het TSE-kader dus gedefinieerd worden als niet steekproef gerelateerde observationele fouten die de schatters kunnen vertekenen en die ook kunnen bijdragen tot de variabiliteit in de schatters (ruis in de data). Interviewereffecten worden doorgaans beschouwd als een oorzaak van meetfouten. Men beperkt zich dan tot de impact die interviewers kunnen hebben tijdens de interactie met de respondent op de geregistreerde antwoorden.

De taak van de interviewer beperkt zich echter niet tot het face-to-face interview met de respondent. De interviewer moet de respondent ook contacteren en overtuigen om mee te werken aan het interview. Ook daarbij kunnen er interviewereffecten optreden. Zo is het best mogelijk dat bepaalde interviewers er beter in slagen om bepaalde groepen van respondenten te overtuigen om mee te werken aan het interview dan andere interviewers (West & Olson, 2010). Indien dit het geval is, dan zijn interviewers niet alleen mede verantwoordelijk voor meetfouten maar ook voor non-respons fouten. Dit laatste illustreert dat de verschillende bronnen van fouten die in de TSE-benadering onderscheiden worden onderling kunnen interageren. Het onderzoek naar hoe verschillende fouten elkaar kunnen beïnvloeden is vrij recent en wordt beschouwd als de volgende logische stap in de verdere ontwikkeling van de TSE-benadering (Groves & Lyberg, 2010).

Analyse van interviewereffecten als oorzaak van meetfouten

Om interviewereffecten als bron van meetfouten in surveyonderzoek te evalueren, kan men onder meer nagaan wat de bijdrage is van de interviewers aan de variantie van de variabelen die in een surveyproject worden gemeten. Dit kan door middel van een interviewer-variantieanalyse waarbij de interviewer als onafhankelijk variabele in de analyse wordt gespecificeerd. De proportie variantie in de afhankelijke variabele die door de interviewer verklaard kan worden kan beschouwd worden als een numerieke schatting van de impact van de interviewer op de geregistreerde antwoorden. Deze proportie is de intra-interviewer correlatiecoëfficiënt (IIC) en drukt het homogeniserende effect uit dat 'veroorzaakt' wordt door de interviewers. Het is duidelijk dat de IIC's best zo laag mogelijk zijn.

Een kwaliteitsevaluatie op basis van IIC's kent reeds een lange traditie en de resultaten ervan hebben de ontwikkeling van de surveypraktijk mede beïnvloed (Schaeffer, Dykema & Maynard, 2010). In 1989 rapporteerde Groves resultaten van 7 onderzoeken die gepubliceerd werden tussen 1962 en 1985 en die gebruik maakten van face-to-face interviews. De vermelde gemiddelde intra-interviewer correlaties varieerden van 0,005 tot 0,102 (Groves, 1989). De IIC's die voor de drie basisvariabelen in het artikel van 1982 werden vermeld zijn opmerkelijk hoger. Voor het aantal beoefende sporten bedraagt de proportie verklaarde variantie 0,10; voor het aantal ooit beoefende sporten is dit 0,17 en voor het aantal sporten waarvoor men belangstelling heeft is dit 0,20.

Het evalueren van interviewereffecten door middel van het berekenen van IIC's is nog altijd een courante praktijk bij de evaluatie van de kwaliteit van surveydata. Deze benadering is in de loop der jaren echter conceptueel verfijnd en analytisch vernieuwd.

Conceptuele uitwerking

Door de ontwikkeling van het TSE-kader is het duidelijker geworden dat de IIC's slechts een gedeelte van de mogelijke interviewereffecten uitdrukken. Met name gaat het om

het verschil tussen interviewers wat betreft de systematische impact die elke interviewer heeft op de antwoorden van de respondenten. De interviewervariantie meet dus de variabiliteit tussen interviewers wat betreft de systematisch vertekening van elk van de interviewers afzonderlijk (Biemer & Lyberg, 2003). Zuivere vertekening, zijnde een dominant en systematisch effect van alle interviewers, en de ruis in de data die interviewers veroorzaken (variabele fouten) worden door de IIC niet gecaptreed (Loosveldt, 2008). Dit impliceert dus dat wanneer de IIC's laag zijn, dit niet noodzakelijk betekent dat er geen interviewereffecten in de data aanwezig zijn. Om die andere types van interviewereffecten op het spoor te komen, kan men andere kwaliteitsevaluatieprocedures toepassen. Zo kan men een geluidsopname maken van het interview en op basis daarvan de feitelijke interactie tussen interviewer en respondent evalueren. Op basis van de resultaten van een dergelijke analyse kan men andere aspecten van interviewereffecten opsporen en krijgt men doorgaans ook een beter inzicht in de factoren die de IIC's mede bepalen (bv. verschillende interpretaties bij de verduidelijking van een bepaalde vraag). Publicaties over de resultaten van interactieanalyse zijn echter minder talrijk dan publicaties waarin de interviewer-variantie benadering wordt toegepast. Het feit dat het uitvoeren van een dergelijke interactieanalyse behoorlijk arbeidsintensief is (opnemen, beluisteren en coderen), is daar wellicht voor verantwoordelijk. Toch merken sommigen op dat het opnemen van interviews gemakkelijker is geworden door het gebruik van laptops en tablets en dat er in de toekomst meer publicaties over interactieanalyse mogen worden verwacht (Schaeffer, Dykema & Maynard, 2010).

Een andere conceptuele aangelegenheid waaraan heel wat aandacht is besteed, is het onderscheid tussen interviewereffecten die worden beschouwd als meeteffecten en regionale effecten die optreden omwille van clustering in de steekproef. Omdat interviewers om de verplaatsingskost te drukken doorgaans in een beperkt geografisch gebied actief zijn, is het niet evident om interviewereffecten te onderscheiden van de mogelijke effecten die optreden omwille van de regionale clustering van de respondenten. Dit probleem werd in het artikel van 1982 ook reeds opgemerkt. Er werd toen een rudimentaire werkwijze toegepast om te controleren voor de regioeffecten. De interviewervariantie voor de drie basisvariabelen werd geschat voor de respondenten in één homogene regio, met name de 'industriële steden'. Er werd vastgesteld dat de interviewervariantie voor deze drie variabelen in deze homogene regio nagenoeg niet verschilde van de schattingen van de interviewervariantie voor de volledige steekproef. Dit werd beschouwd als een duidelijke indicatie dat de impact van de regioeffecten veel geringer was dan die van de vastgestelde interviewereffecten. Deze conclusie werd achteraf bevestigd door ander onderzoek, waarin werd vastgesteld dat interviewers voor een groter deel verantwoordelijk zijn voor homogeniseringseffecten dan de ruimtelijke clustering. Dit betekent ook dat ruimtelijk gerelateerde variantie gedeeltelijk kan en moet worden toegeschreven aan de impact van de interviewers (O'Muircheartaigh & Campanelli, 1998; Schnell & Kreuter, 2005; West, Kreuter, & Jaenichen, 2013). Wanneer men bij de evaluatie van interviewereffecten geen rekening houdt met mogelijke regionale effecten, dan bevatten de interviewereffecten wellicht ook gedeeltelijk deze effecten en worden de interviewereffecten overschat. Bemerkt

dat zowel interviewereffecten als regio-effecten ervoor zorgen dat de steekproefvariantie van een gemiddelde proportie of percentage groter wordt dan bij een zuivere toevalsteekproef. Hiermee moet tijdens een substantiële analyse van de data rekening worden gehouden.

Analytische vernieuwing

De analytische vernieuwing bij de evaluatie van interviewereffecten is ongetwijfeld het gebruik van multi-levelmodellen. Andere benamingen die hiervoor soms gebruikt worden zijn multi-levelregressiemodellen of hiërarchische lineaire modellen. Multi-levelanalyse is bijzonder geschikt om hiërarchische datasets te analyseren. In een hiërarchische dataset zijn de eenheden op een lager niveau genest binnen de eenheden van een hoger niveau. Een typisch voorbeeld van een dergelijke hiërarchische structuur zijn de respondenten die genest zijn binnen de interviewers die gebruikt werden om de respondenten te interviewen. De respondenten zijn het eerste of het laagste niveau en de interviewers zijn het tweede of het hoogste niveau. In een artikel dat in 1994 gepubliceerd werd in *Sociological Methods & Research* werd opgemerkt dat de toepassingen van hiërarchische regressiemodellen om interviewereffecten te evalueren op dat moment nog vrij zeldzaam waren (Hox, 1994, p. 304). Het is pas in de loop van de jaren negentig dat multi-levelanalyse zich ontwikkeld heeft als het standaard analysekader voor het evalueren van interviewereffecten. Met behulp van multi-levelanalyse kunnen immers vrij gevarieerde en complexe vragen over interviewereffecten worden beantwoord. Zo kan men bijvoorbeeld nagaan of het effect van de leeftijd van een respondent op zijn houding tegenover migranten verschilt al naargelang wie interviewt en of deze verschillen tussen de interviewers kunnen verklaard worden door de leeftijd van de interviewer.

Het eenvoudigste multi-levelmodel ziet er als volgt uit:

$$Y_{ij} = \beta_{0j} + \varepsilon_{ij}$$

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

We kunnen beide uitdrukkingen integreren in een vergelijking:

$$Y_{ij} = \gamma_{00} + \mu_{0j} + \varepsilon_{ij}$$

In dit model is Y_{ij} de waarde van de afhankelijke variabele Y die geregistreerd werd bij de i -de respondent die geïnterviewd werd door de j -de interviewer, β_{0j} is het intercept voor interviewer j en ε_{ij} is de residuele waarde voor de i -de respondent die door de j -de interviewer werd geïnterviewd. Dit model bevat geen onafhankelijke variabelen (nulmodel). Het intercept voor een interviewer (β_{0j}) kan vervolgens worden opgesplitst in een algemeen intercept γ_{00} en een interviewerspecifieke afwijking μ_{0j} . De varianties op het niveau van de respondenten en van de interviewers worden respectievelijk uitgedrukt door σ_e^2 en σ_u^2 . Wanneer de variantie van de interviewerspecifieke afwijkingen van het globaal intercept (σ_u^2) significant van nul verschilt, dan concludeert men dat er

significante verschillen zijn tussen de interviewers. Voor elke variabele kan het interviewereffect geschat worden door middel van intra-interviewer correlatie. Het is de proportie variantie in de afhankelijke variabele die door de interviewers kan verklaard worden:

$$\rho_{\text{int}} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

Dit eenvoudig basismodel kan verder worden uitgebreid met onafhankelijke variabelen die worden gemeten bij de respondenten en die relevant zijn om de afhankelijke variabele te verklaren. Een dergelijke uitbreiding van het model kan onder andere als gevolg hebben dat de variantie van de interviewerspecifieke afwijkingen van het globaal intercept kleiner wordt. De proportionele reductie van deze variantie drukt dan uit hoeveel de opgenomen onafhankelijke variabelen op het niveau van de respondenten kunnen verklaren van de variantie op het interviewerniveau. Verder kan men niet alleen het intercept laten variëren over de interviewers, maar kan men ook de regressiecoëfficiënten van de onafhankelijke variabelen laten variëren. Dit betekent dat het effect van een onafhankelijke variabele die men heeft gemeten bij een respondent niet voor elke interviewer gelijk hoeft te zijn. Verder kan men het model nog uitbreiden met interviewerkenmerken die gebruikt worden om de variabiliteit voor het intercept en de slopes op het interviewerniveau te verklaren. In de volgende paragraaf komen een aantal interviewerkenmerken aan bod die daarvoor in aanmerking komen.

De evaluatie van interviewereffecten hoeft zich niet te beperken tot een analyse waarin slechts één afhankelijke variabele is betrokken. Interviewereffecten kunnen ook mee in rekening worden genomen bij de evaluatie van een meetmodel van een latente variabele waarin meerdere indicatoren worden gebruikt. Daarbij worden de indicatoren beschouwd als herhaaldelijke metingen die genest zijn binnen de respondenten. Dit resulteert in een multi-levelmodel met drie niveaus: de metingen van de verschillende indicatoren situeren zich op het eerste niveau, deze metingen zijn genest binnen de respondenten (tweede niveau) die op hun beurt genest zijn binnen de interviewers (Beullens & Loosveldt, 2014). Dit model laat toe om een variantiecomponent te bepalen op het niveau van de interviewers die de impact uitdrukt van de interviewers op de variabiliteit van de herhaaldelijke metingen.

Verklaring van interviewereffecten

Kenmerken van de vraag

De eerder vermelde IIC's van de basisvariabelen van het sportonderzoek die in het artikel van 1982 werden gerapporteerd zijn niet alleen vrij hoog maar ook erg verschillend en zij illustreren dat niet alle vragen even gevoelig zijn voor interviewereffecten. Een mogelijke verklaring die in het artikel wordt vermeld, is de vaagheid van de vraagstelling. Wanneer de vaagheid van de vraagstelling toeneemt (bv. beoefenen

versus belangstelling) neemt de ‘verklaringskracht’ van de interviewer toe. Vagere vragen bieden meer ruimte tot interpretatie en dit geeft aanleiding tot een uitgebreidere en soms minder vlot lopende interactie tussen interviewer en respondent (bv. vragen om verduidelijking). Tijdens dergelijke complexe interacties neemt de kans op systematische beïnvloeding van de interviewer toe. Ander onderzoek bevestigde achteraf dat vraagkenmerken mede verantwoordelijk kunnen zijn voor de interviewereffecten. Interviewereffecten zijn groter bij attitudevragen, bij gevoelige vragen en bij open vragen (Schnell & Kreuter, 2005). Vraagkenmerken zoals de moeilijkheidsgraad en de daaraan gerelateerde complexiteit van de interviewertaak zijn ook bepalend voor de interviewereffecten bij *item nonresponse* (Pickery & Loosveldt, 2001). Het effect van de vraagkenmerken op de interviewereffecten is dus indirect. Vraagkenmerken hebben een impact op de wijze waarop de interactie tussen interviewer en respondent verloopt en de wijze waarop de interactie verloopt kan bijdragen tot het optreden van interviewereffecten (Schaeffer, Dykema, Maynard, 2010).

Kenmerken van de interviewer

Om interviewereffecten te verklaren is ook onderzoek gedaan naar de impact van interviewerkenmerken. Daarbij kan een onderscheid gemaakt worden tussen voor de respondent direct waarneembare kenmerken van de interviewer zoals geslacht, leeftijd en ras en niet direct waarneembare kenmerken zoals eigen opvatting over het onderwerp van het interview, ervaring met interviewen en verwachtingen over hoe respondenten zullen reageren en hoe het interview zal verlopen.

In de analyse die in het artikel van 1982 wordt gerapporteerd worden er geen *waarneembare interviewerkenmerken* gebruikt. Toch is er een traditie om waarneembare interviewerkenmerken te gebruiken in de evaluatie van interviewereffecten. Een klassiek voorbeeld is het onderzoek van Schuman en Converse (1971) naar de invloed van het ras van de interviewer op de antwoorden van zwarte respondenten. Ze stelden vast dat zwarte respondenten zich op vragen over raciale onderwerpen minder vijandig en militant opstelden wanneer ze geïnterviewd werden door blanke interviewers dan tegenover zwarte interviewers. De mate waarin anti-blanke gevoelens worden geregistreerd bij een steekproef van zwarte respondenten wordt dus mede bepaald door de raciale samenstelling van de groep van interviewers (Schuman & Converse, 1971). Algemeen wordt aangenomen dat waarneembare interviewerkenmerken de antwoorden van de respondent kunnen beïnvloeden wanneer de kenmerken gerelateerd kunnen worden aan de inhoud van de vragen of het onderwerp van het onderzoek. Een minder klassiek maar mooi voorbeeld daarvan is de impact van de ‘Body Mass Index’ (BMI) van de interviewer in een onderzoek naar eetgewoonten. De onderzoekers stelden vast dat interviewers met een hoge BMI eetgewoonten registreerden waarbij men minder probeerde te eten. Dit effect was sterker bij respondenten die een dieet moesten volgen. Een mogelijke verklaring daarvoor was dat een gemakkelijk waarneembaar kenmerk, met name de zwaarlijvigheid van de interviewer, de noodzaak of de wenselijkheid

activeert om minder te eten en dit in het bijzonder bij respondenten die een dieet moesten volgen (Eisinga *et al.*, 2011).

De eigen opinie, opvatting of het feitelijke gedrag van de interviewer over de onderwerpen die tijdens het interview aan bod komen zijn voor de respondent *niet direct waarneembare kenmerken* van de interviewer die gebruikt kunnen worden en relevant zijn om de interviewereffecten te verklaren. Zo werd in 1982 vastgesteld dat interviewers die zelf veel belangstelling hadden voor sport ook veel belangstelling registreerden bij de respondenten die ze interviewden. Om een dergelijke koppeling van interviewerkenmerken en geregistreerde antwoorden te kunnen maken, moet men voor elke afzonderlijke interviewer over de nodige informatie beschikken. Die informatie kan men verkrijgen door de interviewers zelf de vragenlijst te laten invullen. Dit vereist een bijkomende dataverzameling tijdens het opstarten van het veldwerk. De koppeling is ook om privacy redenen niet evident. Het is wellicht om die redenen dat dergelijke kenmerken van de interviewer weinig of niet worden gebruikt in het onderzoek naar interviewereffecten en dat dit aspect van de analyse die in het artikel van 1982 werd gerapporteerd, weinig navolging heeft gekend.

Naast de eigen opinies en opvattingen van de interviewer over het onderwerp kan ook de ervaring van de interviewer beschouwd worden als een relevant niet waarneembaar kenmerk om interviewereffecten te verklaren. Ervaring is een taakgerelateerd kenmerk en wordt normaal beschouwd als een pluspunt. Uit onderzoek blijkt echter dat dit bij interviewerwerk niet altijd het geval is. Ervaren interviewers registreren bijvoorbeeld antwoorden die gekenmerkt worden door een grotere mate van volgzzaamheid (Olson & Bilgen, 2011). Een ander voorbeeld vinden we in het onderzoek naar persoonlijke netwerken. In dit onderzoek werden grote interviewereffecten vastgesteld (IIC = 0,214). Interviewers met de minste ervaring registreerden de omvangrijkste persoonlijke netwerken. Dit kan beschouwd worden als een indicatie van betere datakwaliteit. Een mogelijke verklaring is dat minder ervaren interviewers meer open staan voor en meer aandacht besteden aan de specifieke instructies die bij de vragen over de omvang en de samenstelling van het netwerk moeten gevolgd worden (van Tilburg, 1998, p. 315). De associatie tussen ervaring en het minder nauwgezet opvolgen van instructies en een zekere nonchalance bij het interviewen komt ook aan bod in het artikel van 1982. Uit de toen gepubliceerde resultaten blijkt onder meer dat er grote verschillen zijn tussen interviewers in interviewduur en dat de gemiddelde duur van het interview bij de ervaren interviewers iets lager ligt dan bij de niet ervaren interviewers. Wanneer de rangorde van het interview mee in rekening werd gebracht, bleek dat de gemiddelde duur bij de ervaren interviewers sneller daalt. Olson en Peytchev (2007) rapporteren een gelijkaardig verband tussen duur van het interview, rangorde van het interview en interviewerervaring. Dit lijkt erop te wijzen dat de interactie tussen de interviewer en de respondenten verandert naarmate een interviewer in de loop van een project meer interviews doet en dat dit veranderingsproces verschillend is voor ervaren en onervaren interviewers. Welke aspecten van de interactie er juist wijzigen (bv. minder doorvragen, meer suggestieve vragen stellen, minder antwoorden

registreren die vervolgvragen impliceren) en wat de impact is op de geregistreerde antwoorden is echter nog niet onderzocht.

Het reeds in 1982 vastgestelde interviewereffect op de duurtijd van een interview werd onder andere ook vastgesteld bij een analyse van de data van ronde 4 en ronde 5 van de European Social Survey (EES). In beide rondes van ESS werden in alle landen hoge waarden voor de IIC voor de duurtijd vastgesteld. Opvallend daarbij zijn ook de aanzienlijke verschillen tussen de landen. Zo bedraagt de IIC voor de duurtijd in Tsjechië in ronde 5 0,15 en in Israël 0,62 (Loosveldt & Beullens, 2013a). Met de data van ronde 5 van ESS werd ook de impact van de interviewer en de respondent op de snelheid waarmee het interview verloopt (= aantal gestelde vragen per minuut = snelheid interview) geëvalueerd. Uit de resultaten van deze analyse blijkt dat in 8 van de 12 landen de impact van de interviewer op de snelheid van het interview groter is dan de impact van de respondenten. Het is dus de interviewer die in sterkere mate dan de respondent bepaalt aan welke snelheid het interview verloopt. Bemerkt dat de snelheid waarmee het interview verloopt niet gelijk is aan de snelheid van interviews (Loosveldt & Beullens, 2013b).

Wanneer we de duur van het interview en de snelheid van het interview beschouwen als een indirecte indicator van de wijze waarop de interviewer het interview afneemt (Olson & Peytchev, 2007), dan zijn de sterke interviewereffecten die op deze kenmerken van het interview (duurtijd en snelheid) werden vastgesteld, een duidelijke indicatie dat interviewers hun taak tijdens het interview op een erg verschillende wijze uitvoeren. Dit is zeker niet wat men verwacht wanneer men als uitgangspunt neemt dat interviewers de basisprincipes van het gestandaardiseerd interviewen moeten toepassen bij het afnemen van de interviews. Deze vaststellingen wijzen erop dat de wijze waarop interviewers tijdens briefings en trainingen worden voorbereid op hun taak als interviewer voor verbetering vatbaar is.

Conclusie

Wanneer we afsluitend de vraag stellen wat er anno 2014 overeenkomt van de 'Proeve van onderzoek naar de kwaliteit van gegevens bekomen via het survey interview' (Loosveldt, 1982) die in 1982 werd gepubliceerd, dan is de eerste reactie dat bepaalde aspecten van de kwaliteitsevaluatie en van de gerapporteerde resultaten momenteel zeker niet als gedemodeerd, verouderd of achterhaald kunnen worden beschouwd. Integendeel! De evaluatie van interviewereffecten waarop het artikel focust is nog steeds een courante procedure om de kwaliteit van surveydata te evalueren en er worden nog regelmatig artikelen gepubliceerd met resultaten van dergelijke analyses. Sommige resultaten van 1982 werden trouwens in recentere publicaties bevestigd (bv. interviewereffecten op interviewduur). De analysetechnieken hebben zich wel verder ontwikkeld. De analytische tools die momenteel gebruikt worden (multi-levelanalyse) zijn adequater en laten toe om meer complexe vragen te beantwoorden. Die meer complexe vragen worden in sterke mate geïnspireerd door de ontwikkelingen in het

TSE-kader. Dit kader maakt ook duidelijk dat de kwaliteitsbenadering in het artikel van 1982 zeer partieel is en dat vele andere bronnen van fouten die de kwaliteit van een surveyonderzoek kunnen beïnvloeden niet aan bod komen in de toen gerapporteerde analyse. Zo is het bijvoorbeeld opmerkelijk dat in het artikel niet wordt verwezen naar nonrespons fouten en dat zelfs de responsgraad van het onderzoek waarmee gewerkt wordt niet wordt vermeld. Het onderzoek naar interviewereffecten zal zich in de toekomst wellicht richten op de vraag of en in welke mate interviewers mede verantwoordelijk zijn voor andere foutenbronnen die binnen het TSE-kader worden gedefinieerd (bv. interviewereffecten op antwoordtendensen).

Nieuwe technologisch ontwikkelingen hebben de afgelopen jaren de dataverzameling in sociaalwetenschappelijk onderzoek sterk beïnvloed (bv. websurveys en het gebruik van smartphones) en zullen er wellicht voor zorgen dat er minder face-to-face surveys worden georganiseerd en dat dit type onderzoek zich meer en meer zal beperken tot academisch onderzoek en onderzoek voor de overheid. Interviewers zullen daarin een cruciale rol blijven spelen. Gegeven het belang van dit onderzoek, blijft het een uitdaging om de impact van de interviewers op de geregistreerde antwoorden te beperken. De selectie en de training van de interviewers enerzijds en de opvolging van de interviewers tijdens het veldwerk anderzijds zijn daarbij essentieel. Het is trouwens belangrijk dat er voldoende en degelijke interviewercapaciteit beschikbaar blijft. Om dit te realiseren zal een permanente investering noodzakelijk zijn van alle onderzoeksgroepen die gebruik maken van face-to-face survey-interviews. Samenwerking tussen de verschillende betrokken partijen (onderzoeksgroepen en veldwerkorganisaties) is daarbij meer dan wenselijk.

Bibliografie

- Anderson, R., Kasper J. & Frankel, M. (1979). *Total Survey Error: Application to Improve Health Surveys*. San Francisco, CA: Jossey-Bass.
- Beullens, K. & Loosveldt, G. (2014). Interviewer Effects on Latent Constructs in Survey Research. *Journal of Survey Statistics and Methodology*, 2(4), 433-58.
- Biemer, P. & Lyberg, L. (2003). *Introduction to Survey Quality*. Hoboken, NJ: John Wiley & Sons.
- Biemer, P. & Lyberg, L. (Eds.) (2010). Special Issue: Total Survey Error. *Public Opinion Quarterly*, 74(5). Oxford: Oxford University Press.
- Eisinga, R., te Grotenhuis, M., Larsen, J., Pelzer, B. & van Strien, T. (2011). BMI of Interviewer Effects. *International Journal of Public Opinion Research*, 23(4), 530-43.
- Groves, B. (1989). *Survey Errors and Survey Costs*. New York: John Wiley & Sons.
- Groves, B. & Lyberg, L. (2010). Total Survey Error, Past, Present, and Future. *Public Opinion Quarterly*, 74(5), 849-79.
- Hox, J. (1994). Hierarchical Regression Models for Interviewer and Respondent Effects. *Sociological Methods & Research*, 22(3), 300-18.
- Loosveldt, G. (1982). Proeve van onderzoek naar de kwaliteit van gegevens bekomen via het survey-interview. *Tijdschrift voor Sociologie*, 3(3), 215-28.

- Loosveldt, G. (2008). Face-To-Face Interviews. In E. D. De Leeuw, J. J. Hox & D. A. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 201-20). New York, NY: Taylor & Francis Group.
- Loosveldt, G. & Beullens, K. (2011). Evaluation of Interviewer Effects in Round Four of the European Social Survey. *The European Social Survey – Data for a Changing Europe, deliverable 12.2*.
- Loosveldt, G. & Beullens, K. (2013a). ‘How Long Will it Take?’ An Analysis of Interview Length in the Fifth Round of the European Social Survey. *Survey Research Methods*, 7(2), 69-78.
- Loosveldt, G. & Beullens, K. (2013b). The Impact of Respondents and Interviewers on Interview Speed in Face-to-face Interviews. *Social Science Research*, 42(6), 1422-30.
- Olson, K. & Bilgen, I. (2011). The Role of Interviewer Experience on Acquiescence. *Public Opinion Quarterly*, 75(1), 99-114.
- Olson, K. & Peytchev, A. (2007). Effect of Interviewer Experience on Interviewer Pace and Interviewer Attitudes. *Public Opinion Quarterly*, 71(2), 273-86.
- O’Muircheartaigh, C. & Campanelli, P. (1998). The Relative Impact of Interviewer Effects and Sample Design Effects on Survey Precision. *Journal of the Royal Statistical Society. Series A*, 161, 63-77.
- Phillips, D. (1971). *Knowledge from what? Theories and Methods in Social Research*. Chicago: Rand McNally and Company.
- Phillips, D. (1973). *Abandoning Method: Sociological Studies in Methodology*. San Francisco: Jossey-Bass.
- Pickery, J. & Loosveldt, G. (2001). An Exploration of Question Characteristics that Mediate Interviewer Effects on Item Nonresponse. *Journal of Official Statistics*, 17(3), 337-50.
- Schaeffer, N., Dykema J. & Maynard, D. (2010). Interviewers and Interviewing. In P. Marsden & J. Wright (Eds.), *Handbook of Survey Research (second edition)* (pp. 437-70). Bingley, UK: Emerald Group Publishing Limited.
- Schnell, R. & Kreuter, F. (2005). Separating Interviewer and Sampling-points Effects. *Journal of Official Statistics*, 21(3), 389-410.
- Schuman, H. & Converse, J. (1971). The Effects of Black and White Interviewers on Black Responses in 1968. *Public Opinion Quarterly*, 35(1), 44-68.
- Smith, T. (2011). Refining the Total Survey Error Perspective. *International Journal of Public Opinion Research*, 23(4), 464-84.
- van Tilburg, T. (1998). Interviewer Effects in the Measurement of Personal Network Size: A Nonexperimental Study. *Sociological Methods & Research*, 26(3), 300-28.
- West, B., Kreuter, F. & Jaenichen, U. (2013). “Interviewer” Effects in Face-to-face Surveys: A Function of Sampling, Measurement Error or Nonresponse? *Journal of Official Statistics*, 29(2), 277-97.
- West, B. & Olson, K. (2010). How much of Interviewer Variance is really Nonresponse Error Variance? *Public Opinion Quarterly*, 74(5), 1004-26.