

De effecten van interviewer- en respondentkenmerken op antwoordgedrag in survey-onderzoek: een multilevel-benadering*

Johannes H. Smit, Pieter van den Eeden,
Dorly J.H. Deeg en Aartjan T.F. Beekman

1. Inleiding

Wanneer informatie wordt verzameld met behulp van mondelinge interviews, is het niet onwaarschijnlijk dat er verschillen ontstaan tussen de werkelijke attributen van respondenten ('true values') waarnaar men vraagt en de antwoorden door respondenten gegeven. Deze verschillen worden aangeduid met 'meetfouten'. De laatste jaren is er een hernieuwde aandacht voor het ontstaan en de gevolgen van meetfouten bij onderzoeksdesigns waarin van mondelinge interviews gebruik wordt gemaakt (Groves 1989; Fuller 1992; Biemer e.a. 1991).

In de literatuur worden vier mogelijke oorzaken voor het ontstaan van deze meetfouten onderscheiden: fouten ontstaan door het instrument (als gevolg van bijvoorbeeld te vage termen in vragen, antwoordalternatieven die niet goed aansluiten bij de vragen); fouten ontstaan door de respondenten (zoals veroorzaakt door hun gebrek aan motivatie, sociaal wenselijk antwoordgedrag); fouten ontstaan door de context waarin de dataverzameling plaatsvindt (bij de respondenten thuis of op het werk, de aanwezigheid van derden tijdens het interview) en fouten ontstaan door de interviewers.

Om meetfouten van het laatste type (fouten ontstaan door interviewers) op te sporen is het belangrijk de antwoordvariatie die kan worden toegeschreven aan de interviewer vast te stellen. Vervolgens kan men proberen bij de interviewer factoren te identificeren die mogelijk aanleiding zijn voor het ontstaan van deze variatie. Ten slotte kunnen deze inzichten worden gebruikt om de wijze waarop het instrument wordt afgenomen te verbeteren.

* Deze publikatie is gebaseerd op gegevens verzameld in het kader van de Longitudinal Aging Study Amsterdam (LASA). LASA wordt mogelijk gemaakt door de directie Ouderenbeleid, ministerie van Volksgezondheid, Welzijn en Sport, en wordt uitgevoerd in de vakgroep Psychiatrie en de vakgroep Sociologie en Sociale Gerontologie aan de Vrije Universiteit te Amsterdam. Een eerdere versie van dit paper is gepresenteerd op de Sociaal-wetenschappelijke studiedagen 1994, gehouden op 7 en 8 april 1994 te Amsterdam. Correspondentie: J. H. Smit, vakgroep Sociologie en Sociale Gerontologie, Prins Hendriklaan 27-29, 1075 AZ Amsterdam.

Meetfouten die te wijten zijn aan interviewers worden 'interviewereffecten' genoemd. Het onderzoek naar interviewereffecten richt zich enerzijds op de kenmerken (opleiding, geslacht, leeftijd) van interviewers en anderzijds op het gedrag van de interviewer (zoals: niet correct voorlezen van de vraag, suggereren van een antwoord, vrijelijk interpreteren van een antwoord). Zowel interviewerkenmerken als interviewergedrag kunnen de wijze waarop het antwoord wordt gegeven op ongewenste wijze beïnvloeden.

Het onderzoek naar interviewereffecten heeft zich tot voor kort voornamelijk gericht op het vaststellen van varianties van een enkele responsvariabele, welke deels wordt toegeschreven aan de interviewer (bijvoorbeeld Groves & Fultz 1985). Vervolgens probeert men deze variantie te verklaren in termen van achtergrondkenmerken van de interviewers (zoals hun opleiding en geslacht). Bij deze benadering wordt voorbijgegaan aan het gegeven dat respondenten 'genest' zijn onder interviewers. Respondenten die geïnterviewd worden door dezelfde interviewer hebben gemeen dat ze onderhevig zijn aan dezelfde interviewerkenmerken. Recent is belangstelling ontstaan voor een benadering waarin meer aandacht besteed wordt aan het hiërarchische karakter van het design. Het uitgangspunt is dat antwoorden van respondenten geïnterviewd door interviewer A onderling meer met elkaar gemeenschappelijk hebben dan de antwoorden van respondenten geïnterviewd door interviewer B. Het ligt voor de hand dit uitgangspunt uit te werken in modellen die de door interviewers veroorzaakte antwoordvertekening beschrijven.

Dit uitgangspunt hangt nauw samen met de opvatting dat interviewers mede de sociale situatie creëren waarin het interview wordt afgenomen. Zij creëren een specifieke sociale omgeving, die onder meer wordt gekenmerkt door een sterke structurering en gericht is op het verschaffen van informatie door de respondent aan de interviewer. Maar zij kunnen dit voor een gedeelte op hun eigen wijze doen.

Ter verklaring van de verschillen in interviewermeetfouten gaan we er voorts van uit dat de interviewers verschillen in de wijze waarop ze respondenten ondervragen. Deze verschillen ontstaan door aan de respondenten in verschillende mate mentale hulpmiddelen ('resources') ter beschikking te stellen die nauw samenhangen met de bovengenoemde achtergrondkenmerken en gedragingen van de interviewer. Een voorbeeld van een dergelijk hulpmiddel is de zorgvuldigheid die een interviewer vertoont en die een respondent zal stimuleren zo zorgvuldig mogelijk een antwoord op de gestelde vraag te formuleren. De achtergrond- en gedragskenmerken van de interviewer vormen daarmee kenmerken van de interviewsituatie. Verschillen in de mate waarin de respondenten gebruik maken van deze 'resources' voor het geven van juiste antwoorden leiden dan tot verschillen in het antwoordgedrag van respondenten (zie voor een uitwerking van de Responding Environment Theory: Van den Eeden & Bruinsma 1992).

Een model waarmee deze theorie over het effect van interviewerkenmerken op

het beantwoordingsproces bestudeerd kan worden is het random coëfficiënt model voor multilevel-analyse.

In deze bijdrage zullen we dit model voor de bestudering van interviewereffecten uitwerken met behulp van een depressiescore bij oudere respondenten. Onderzocht zal worden of met behulp van dit model kan worden vastgesteld of interviewers een bron van meetfouten zijn bij het verzamelen van deze gegevens en, indien dat inderdaad het geval is, of met behulp van het random coëfficiënt model interviewerkenmerken geïdentificeerd kunnen worden ter verklaring van deze meetfouten.

2. Veldwerk en instrumenten

Gegevens voor deze studie zijn verzameld in het kader van de 'Longitudinal Aging Study Amsterdam' (LASA). LASA is een longitudinaal, multidisciplinair onderzoeksproject over veranderingen in de autonomie en de kwaliteit van leven van ouderen en naar de predictoren en consequenties van die veranderingen (Deeg, Knipscheer & Van Tilburg 1993). Het steekproefkader wordt gevormd door gemeentelijke bevolkingsbestanden uit drie cultureel verschillende geografische gebieden in het westen, het oosten en het zuiden van Nederland. De steekproef is zodanig opgebouwd dat deze de landelijke verdeling naar urbanisatiegraad weerspiegelt. Elke regio bestaat uit een middelgrote tot grote stad en twee of meer landelijke gemeenten die in de nabijheid van de stad liggen. LASA volgt op de NESTOR-studie naar leefvormen en sociale netwerken (NESTOR-LSN) (Knipscheer e.a., nog te verschijnen). In deze studie, waarin het veldwerk van januari 1992 tot juni 1992 plaatsvond, werden ouderen van 55 tot 89 jaar ondervraagd. Het responspercentage bedroeg 61,7. In het kader van LASA werden de in de NESTOR-LSN geïnterviewde personen opnieuw benaderd met uitzondering van de groep 85- tot 89-jarige respondenten (gecorrigeerd responspercentage 62,3).

De interviewers werden geworven via advertenties in de lokale media en via advertenties op de prikborden van universiteiten. Belangrijke selectiecriteria waren eerdere ervaring met interviewen, een duidelijk handschrift en enige ervaring met computers. In totaal werden 43 interviewers (41 vrouwen en twee mannen) geselecteerd. De interviewtraining bestond uit vijf sessies van elk zes uur. Deze was gebaseerd op Emans (1985), een training ontwikkeld door De Bie & Dijkstra (1989) en een video voor interviewerinstruatie van het National Institute on Aging (1988). Door middel van videovoorbeelden en door het houden van rollenspellen werden de basisinterviewvaardigheden aangeleerd. Elke interviewer deed een proefinterview dat op cassette werd opgenomen en vervolgens werd doorgesproken met de onderzoeksleiding.

Tijdens het veldwerk werden aanvullende trainingssessies georganiseerd om interviewers bij te sturen. Alle interviewers vulden een vragenlijst in om interviewerkenmerken vast te stellen. Het veldwerk werd gehouden van september 1992

tot oktober 1993. Van de oorspronkelijke groep ondervraagden ($I = 3805$) bleken 3107 personen in staat en bereid opnieuw aan een interview deel te nemen (81,6 procent). Alle interviews werden na toestemming van de respondent op cassette opgenomen. Binnen de drie regio's werden de respondenten aselekt aan de interviewers toegewezen. Het interview werd bij respondenten 'thuis' (woning, verzorgingstehuis, enzovoort) afgenomen, waarbij gebruik gemaakt werd van draagbare computers.

De scores van de depressieschaal werden vastgesteld met de CES-D-schaal: Center for Epidemiologic Depression Scale; voor een beschrijving zie Radloff 1977, Beekman e.a. 1994. Daarop werden in eerder onderzoek interviewereffecten aangetoond (Choi & Comstock 1975; Handlin e.a. 1974). Deze schaal is als volgt geoperationaliseerd. Binnen deze schaal (bestaande uit 20 uitspraken) worden vier factoren onderscheiden (Radloff 1977; Hertzog e.a. 1990): 'depressief gevoel', 'somatische symptomen en remmingen', 'positief gevoel' en 'interpersoonlijke relaties' (zie voor de formulering van de uitspraken Bijlage I). Van de 20 uitspraken zij er vier verwoord in een positieve richting en de overige niet, om tendenties waarbij de respondent het antwoord afleidt van de richting van de gestelde vraag (response set) te voorkomen en het al dan niet aanwezig zijn van de eerder genoemde factor 'positief gevoel' te meten. Elke uitspraak wordt gescoord van nul tot drie op een frequentieschaal. Voor dit onderzoek werden op respondentniveau leeftijd, sekse en absolute subjectieve gezondheid (evaluatie van de eigen gezondheid) en relatieve subjectieve gezondheid (evaluatie van de eigen gezondheid in vergelijking met leeftijdgenoten) als – in de literatuur – bekend staande aan depressie gerelateerde variabelen geselecteerd (Kennedy & Kelman 1989).

Als kenmerken van de interviewers werden leeftijd, opleiding, enkele persoonlijkheidskenmerken (extraversie, vriendelijkheid en nauwkeurigheid) en twee maten van assertiviteit (zich zeker voelen en omgaan met ongewenste situaties) geselecteerd. De keuze voor deze variabelen op interviewerniveau was ontleend aan het feit dat deze variabelen reeds eerder in verband waren gebracht met de kwaliteit van gegevens van het survey-onderzoek (Hox, De Leeuw & Kreft 1991). In Bijlage 2 wordt een (statistische) beschrijving van de genoemde variabelen gegeven.

Respondenten en interviewers met een ontbrekend antwoord op één van de uitspraken in de genoemde lijsten werden voor aanvang van de analyses verwijderd. De analyses werden uiteindelijk uitgevoerd met 35 interviewers en 2383 respondenten. Er is geen aanwijzing dat de uitval systematisch is.

3. Methode

Het random coëfficiënt model kan worden beschreven in twee vergelijkingen. De eerste vergelijking (niveau 1: respondenten) heeft betrekking op binnen-intervie-

wer regressie, waarbij de afhankelijke variabele verklaard wordt uit respondentkenmerken en de meetfouten (waaronder de effecten van interviewerkenmerken) weergegeven worden in de storingsterm (E_{ij}). Er geldt voor respondent i geïnterviewd door interviewer j :

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \dots + \beta_{Kj}X_{Kij} + E_{ij} \quad (1)$$

waarin:

Y_{ij} : de afhankelijke variabele

i : de respondent ($i = 1, \dots, I_j$), in totaal $I = I_1 + \dots + I_j$ respondenten

j : de interviewer ($j = 1, \dots, J$)

β_{0j} : het intercept behorende bij de regressievergelijking van Y_{ij}

β_{kj} : de regressiecoëfficiënt van de afhankelijke variabele Y op de onafhankelijke respondentvariabele k

X_{ki} : onafhankelijke variabelen ($k = 1, \dots, K$) voor respondent i

E_{ij} : de storingsterm behorende bij Y_{ij} met variantie s_i^2 en gemiddelde 0

In de tweede vergelijking (niveau 2: interviewers) worden de interviewerkenmerken ingebracht. Voor elke coëfficiënt wordt daartoe een nieuwe regressievergelijking opgesteld. Het intercept β_{0j} en de hellingen β_{kj} ($k = 1, \dots, K$) worden met de interviewerkenmerken in verband gebracht.

Er geldt:

$$\beta_{kj} = \gamma_{k0} + \gamma_{k1}Z_{1j} + \dots + \gamma_{kn}Z_{nj} + e_{kj} \quad (2)$$

waarin:

Z_{nj} : onafhankelijke variabelen ($n = 1 \dots N$) voor interviewer j

γ_{k0} : het intercept van de regressie van β_{kj} op Z_1, \dots, Z_n

γ_{kn} : de regressiecoëfficiënt van de afhankelijke variabele β_{kj} op de onafhankelijke interviewervariabele Z

e_{kj} : de storingsterm behorende bij β_{kj} met variantie t_k^2 en gemiddelde 0 en de covariantie $t_{k_m k_n}$ van k_m en k_n ($m \neq n$)

Substitutie van vergelijking (2) in vergelijking (1) levert het uiteindelijke model op.

Het random coëfficiënt regressiemodel voor multilevel-analyse, waarin de binnen-interviewer regressies op het niveau van de respondent dienst doen als afhankelijke variabelen op het niveau van de interviewer, is betrekkelijk nieuw voor de bestudering van dit type relaties. Ter introductie wordt dit model uitgewerkt met behulp van een voorbeelduitspraak: 'De afgelopen week voelde ik me *evenveel waard* als andere mensen'. De score op deze uitspraak wordt dan in een eerste stap

uitgedrukt in de scores van de respondentvariabelen 'leeftijd', 'seks', absolute subjectieve gezondheid (*absgez*) en relatieve subjectieve gezondheid (*relgez*) volgens regressievergelijking (1):

$$\text{evenveel waard}_{ij} = \beta_{0j} + \beta_{1j} \text{leeftijd}_{ij} + \beta_{2j} \text{seks}_{ij} + \beta_{3j} \text{absgez}_{ij} + \beta_{4j} \text{relgez}_{ij} + E_{ij}$$

waarin i de respondent en j de interviewer aangeeft. β_{0j} is een constante die geldt voor elke respondent die is geïnterviewd door interviewer j ; β_{1j} , β_{2j} , β_{3j} en β_{4j} zijn de interviewer-specifieke regressiecoëfficiënten van achtereenvolgens leeftijd, seks, absolute subjectieve gezondheid en relatieve subjectieve gezondheid. E_{ij} is de storingsterm met variantie s^2 .

Vervolgens worden in de tweede stap de binnen-interviewer regressies β_{0j} , β_{1j} , β_{2j} , β_{3j} en β_{4j} (in het algemeen β_{kj} - met $k = 0, \dots, 4$) in verband gebracht met de interviewervariabelen leeftijd (*intleeft*), opleiding (*intpleid*), extraversie (*intextrav*), instemming (*intinstem*), zorgvuldigheid (*intzorgv*), zekerheid (*intzeker*), en omgaan met ongewenste situaties (*intongew*). Dit gebeurt volgens de regressievergelijking (2):

$$\beta_{kj} = \gamma_{k0} + \gamma_{k1} \text{intleeft}_j + \gamma_{k2} \text{intpleid}_j + \gamma_{k3} \text{intextrav}_j + \gamma_{k4} \text{intinstem}_j + \gamma_{k5} \text{intzorgv}_j + \gamma_{k6} \text{intzeker}_j + \gamma_{k7} \text{intongew}_j + e_{kj}$$

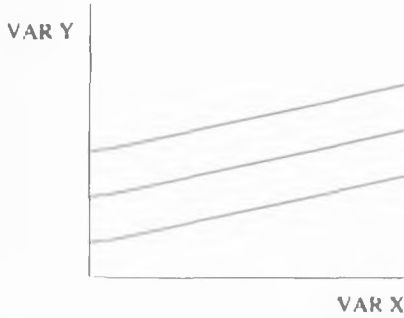
waarin de storingstermen e_{kj} de variantie t_k^2 en de covariantie $t_{k_m k_n}$ (met $k_m, k_n = 0, \dots, 4$ en $m \neq n$) hebben. Voor nadere toelichting zie Bryk en Raudenbush (1992) en Longford (1993).

De coëfficiënten γ_{kj} (met $j: 1, \dots, 7$) vormen het 'verklarende' deel van de vergelijkingen. De coëfficiënten β_{kj} waarin een storingsterm voorkomt vormen het 'toevals' deel daarvan; ze kunnen worden afgelezen na substitutie.

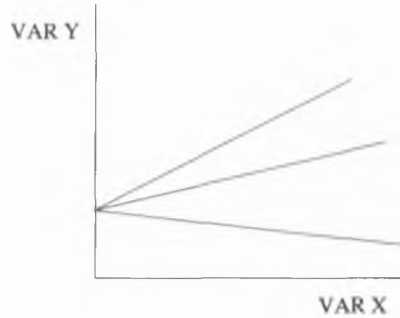
Laten we twee extreme gevallen beschrijven. Situatie A: de regressiecoëfficiënten β_{kj} zijn gelijk voor iedere interviewer j . De verschillen in het intercept γ_{k0} tussen de interviewers kunnen worden weergegeven door evenwijdige regressielijnen die de Y-as, die de afhankelijke variabele 'evenveel waard' weergeeft, snijden. In dit geval zijn de effecten van de respondentvariabelen (zoals 'leeftijd') op de uitspraak 'evenveel waard' onafhankelijk van de interviewer. De startpositie van waaruit respondenten vertrekken bij een gegeven interviewer en die veroorzaakt wordt door de algemene presentatie, indruk, enzovoort, van de interviewer, verschilt van interviewer tot interviewer. Dit wordt uitgedrukt in de verschillen tussen de intercepten. Elke lijn drukt dan de interviewer-specifieke regressie uit (zie figuur 1).

Situatie B: de intercepten γ_{k0} verschillen niet tussen interviewers. De verschillen in de coëfficiënten β_{kj} ($j = 1, \dots, J$) tussen de interviewers kunnen worden weergegeven door j convergerende lijnen met één snijpunt op de Y-as (zie figuur

2). In dit geval is de startpositie van alle respondenten gelijk, hetgeen inhoudt dat de 'algemene' presentatie, indruk, enzovoort, die interviewers op respondenten maken onderling niet verschillen. Wel zijn er verschillen in de effecten van de afgebeelde respondentvariabele. In het voorbeeld van leeftijd kan een oudere interviewer naarmate de respondenten ouder zijn (= X) sterkere gevoelens van evenveel waard te zijn (= Y) ontlokken (dus hogere scores te zien geven) dan een jongere interviewer. In de meeste gevallen zal zich een combinatie voordoen van beide situaties.



Figuur 1



Figuur 2

Omdat er niet of nauwelijks een specifieke theorie bestaat over interviewereffecten op beantwoordingsprocessen bij respondenten, wordt in het volgende een procedure gebruikt die uitgaat van een model waarin alleen de totale variantie van de afhankelijke variabele op de twee genoemde niveaus (respondent, interviewer) wordt gesplitst. Daarna wordt het regressiemodel eerst op het niveau van de respondent berekend, en vervolgens wordt gekeken of dit regressiemodel van interviewer tot interviewer verschilt. Is dit laatste het geval, dan wordt onderzocht of die verschillen samenhangen met verschillen op de interviewervariabelen volgens de laatstgenoemde vergelijking.

Gebruik is vooral gemaakt van het programma VARCL3 (Longford 1986). Bij de opeenvolgende stappen in de analyse zijn niet-significante coëfficiënten telkens op nul gesteld, en is gecontroleerd of een volgend model beter bij de gegevens past dan het vorige (op basis van de verschillen in de log-likelihood waarde, onderweg van het verschil in vrijheidsgraden).

4. Resultaten

In eerste instantie werd een exploratieve multilevel-analyse uitgevoerd om vast te stellen of interviewereffecten konden worden aangetoond op de CES-D totaalscore. De variantie tussen de interviewers bleek drie procent te zijn. Met andere woorden, beantwoording van de CES-D was niet alleen afhankelijk van bepaalde res-

pondentkenmerken, maar was ook afhankelijk van de interviewer die de vragen stelde.

Vervolgens werd de factorstructuur bestudeerd. Door middel van een orthogonale factoranalyse (met een nonrestricted varimax rotatie) werden vier factoren vastgesteld (in overeenstemming met Radloff 1977). De ladingen staan vermeld in Bijlage 3. De eerste factor (depressief gevoel) neemt 30,9 procent van de variantie voor zijn rekening, de tweede factor (somatische symptomen en remmingen) 6,9 procent, de derde factor (positief gevoel) 6,1 procent, en de vierde factor (interpersoonlijke relaties) 5,6 procent. Het percentage verklaarde variantie bedraagt 49,5 procent. Bij een inspectie van de tussen-interviewer variantie bij de factoren bleek dat achtereenvolgens 3,5 procent, 0 procent, 7,7 procent en 0,4 procent tussen-interviewer variantie kon worden vastgesteld. Ter illustratie van het model beperken we ons in deze studie tot de factor 'positief gevoel', zijnde de factor met het hoogste percentage tussen-interviewer variantie.

'Positief gevoel' wordt vastgesteld met behulp van een viertal vragen, die als volgt zijn geformuleerd:

Wilt u me voor elk van de volgende uitspraken zeggen of u zich de afgelopen week zo voelde. U kunt steeds antwoorden met behulp van de antwoordmogelijkheden: 0. zelden of nooit, 1. soms, 2. vaak, 3. altijd of bijna altijd.

'De afgelopen week voelde ik me evenveel waard als andere mensen.'

'De afgelopen week was ik hoopvol gestemd over de toekomst.'

'De afgelopen week was ik gelukkig.'

'De afgelopen week had ik plezier in het leven.'

Onze analyse richt zich speciaal op de interviewereffecten die optreden bij deze uitspraken.

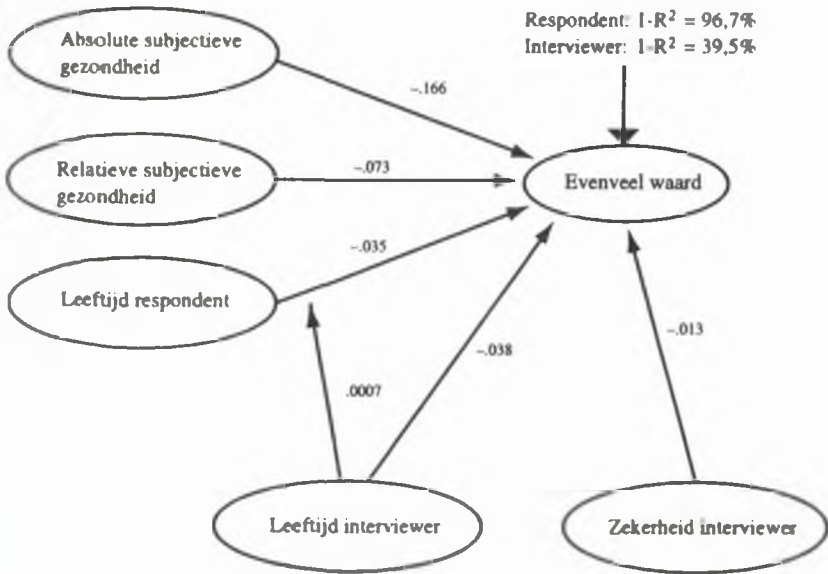
De uitkomsten van de analyse worden gegeven in ongestandaardiseerde coëfficiënten. In tabel 1 staan de gemiddelden en de standaarddeviaties van deze vier uitspraken.

Tabel 1: Gemiddelden en standaarddeviaties van uitspraken met betrekking tot de dimensie 'positief gevoel' (I = 2383)

Uitspraken	Gemiddelde	Standaarddeviatie
'Evenveel waard'	2,57	0,88
'Hoopvol gestemd'	1,67	1,21
'Gelukkig'	2,32	0,97
'Plezier in het leven'	2,31	0,98

De uitkomsten van de multilevel-analyse. Voor elk van de vier uitspraken werd eerst nagegaan in hoeverre de kenmerken van respondenten samenhangen met de

uitspraak. Vervolgens werd vastgesteld in hoeverre er tussen-interviewer variantie bestond op de intercepten en regressiecoëfficiënten van de relatie tussen de respondentkenmerken en de uitspraak. Ten slotte werd de hoeveelheid verklaarde variantie berekend.



Figuur 3: Weergave resultaten 'evenveel waard'

Een uitwerking voor de uitspraak 'evenveel waard' wordt gegeven. De aanvankelijke tussen-interviewer variantie bedraagt 3,3 procent. De multilevel-analyse met betrekking tot deze uitspraak laat de volgende resultaten zien:

- a. De leeftijd, en zowel de absolute als de relatieve subjectieve gezondheid van de respondent hangen negatief samen met deze uitspraak. De coëfficiënten zijn achtereenvolgens $-0,035$, $-0,166$ en $-0,073$. Met andere woorden: naarmate respondenten ouder ($-0,035$) zijn en de relatieve subjectieve gezondheid ($-0,166$) en absolute subjectieve gezondheid ($-0,073$) minder goed geëvalueerd worden, voelen ze zich minder vaak 'evenveel waard als andere mensen'. Er zijn geen verschillen tussen mannen en vrouwen.
- b. De gemiddelde score verkregen op de uitspraak 'evenveel waard' van respondenten hangt direct samen met de leeftijd en de mate waarin de interviewer zich zeker voelt (de coëfficiënten zijn achtereenvolgens $-0,038$ en $-0,013$). Hoe ouder de interviewer en hoe zekerder, hoe lager de gemiddelde score van de respondenten op de uitspraak 'evenveel waard'.

- c. Er bestaan verschillen in de effecten van leeftijd op het antwoord van de uitspraak 'evenveel waard' tussen de interviewers. De sterkte van het effect van het respondentkenmerk leeftijd op 'evenveel waard' (de coëfficiënt is 0,0007) hangt af van de leeftijd van de interviewer. Hoe ouder een interviewer, des te sterker de leeftijd van de respondent met de score op de uitspraak 'evenveel waard' samenhangt.
- d. De totale verklaarde variantie op het niveau van de respondent bedraagt 3,3 procent; de verklaarde variantie op het niveau van de interviewer (de reductie van de aanvankelijke tussen-interviewer variantie van 3,3 procent) bedraagt 61,5 procent.

Figuur 3 geeft een overzicht van deze uitkomsten.

Tabel 2: Uitkomsten van de multilevel-analyse per uitspraak die behoort bij positief gevoel: 'evenveel waard', 'hoopvol gestemd', 'was gelukkig' en 'plezier in het leven'; niet-significante coëfficiënten (Prob. > 0,05) zijn weggelaten

		'evenveel waard'	'hoopvol gestemd'	'was gelukkig'	'plezier in het leven'*
verklarend deel					
<i>respondent level</i>					
leeftijd		-0,035	-0,015		
seks			-0,207	-0,207	-0,207
absolute subjectieve gezondheid		-0,166	-0,218	-0,218	-0,218
relatieve subjectieve gezondheid		-0,073	-0,063		
<i>verklaring van interceptverschillen tussen de interviewers</i>					
door leeftijd interviewer		-0,038		0,015	0,015
zekerheid interviewer		-0,013			
<i>verklaring van hellingverschillen tussen de interviewers</i>					
van leeftijd respondent					
door leeftijd interviewer		0,0007	0,016		
toevalsdeel					
<i>respondent level</i>	s_i^2	0,697	1,306	0,842	0,842
<i>interviewer level</i>					
gemiddelde	t_{10}	0,010	0,047	0,019	0,019
leeftijd	t_{11}	0,0001			
seks	t_{12}	0,029			
absolute subjectieve gezondheid	t_{13}	0,007			
relatieve subjectieve gezondheid	t_{14}	0,013			

* De verdelingen van de uitspraken 'was gelukkig' en 'plezier in het leven' zijn vrijwel gelijk, met het gevolg dat ook hun coëfficiënten gelijk zijn

Een samenvatting van de analyses van de vier uitspraken staat vermeld in tabel 2. Met name leeftijd van de interviewer heeft gevolgen voor de verkregen scores van de uitspraken behorende bij positief gevoel.

De verklaarde varianties van het interviewerniveau en het respondentniveau met betrekking tot de vier uitspraken worden weergegeven in tabel 3.

Tabel 3: Percentage tussen-interviewer variantie en percentage verklaarde variantie op respondent- en interviewerniveau per uitspraak

	Percentage tussen-interviewer variantie	Percentage verklaarde variantie	
		Respondent	Interviewer
'evenveel waard'	3,3	3,3	2,1
'hoopvol gestemd'	6,1	7,3	3,0
'gelukkig'	4,4	6,1	2,5
'plezier in het leven'	4,8	6,2	2,7

5. Interpretatie en discussie

Door het random coëfficiënt regressiemodel in een hiërarchisch design te hanteren konden we op microniveau naar de processen kijken die leiden tot interviewereffecten tijdens het vaststellen van een depressiescore. Daarbij kwamen aspecten aan het licht welke enig inzicht verschaffen in het beantwoordingsproces van respondenten. Daarbij was het uitgangspunt dat interviewers voor een belangrijk deel de interviewsituaties creëren waarbinnen het beantwoordingsproces plaatsvindt. De achtergrond- en gedragskenmerken van interviewers fungeren daarbij als hulpbronnen voor het geven van de (juiste) antwoorden. Het gevolg van deze opvatting is dat de interviewsituatie en de respondent als verschillende niveaus kunnen worden beschouwd en het effect van de interviewerkenmerken kan worden bestudeerd met behulp van het random coëfficiënt model van multilevel-analyse.

Uit de analyse bleek dat de beantwoording door de respondenten vooral wordt beïnvloed door de leeftijd van de interviewer. Persoonlijkheidskenmerken en assertiviteitsmaten spelen vrijwel geen rol. Positief geformuleerde uitspraken zijn meer onderhevig aan interviewervariantie dan negatief of neutraal geformuleerde items. Volgens Molenaar (1988) is er een effect waar te nemen van de wijze waarop een vraagobject is omschreven. Naarmate het vraagobject positiever is omschreven is het verdelingsgemiddelde hoger. Als verklaring wordt aangevoerd dat positieve formuleringen op de respondent een emotionele of psychosociale druk uitoefenen om een antwoord te kiezen in de richting waarin de omschrijving verte-

kend is. De mate waarin respondenten deze druk ervaren kan afhankelijk zijn van verschillen in 'resources' die de interviewer aanbiedt. De leeftijd van de interviewer kan voor respondenten een aanleiding zijn om het antwoord aan te passen. Verder valt op dat het leeftijdseffect van interviewers bij de uitspraak 'evenveel waard' negatief is terwijl dat effect bij de uitspraken 'gelukkig' en 'plezier in het leven' positief is (zie tabel 2). Tijdens de factoranalyse bleek eveneens dat de uitspraak 'evenveel waard' minder goed op deze factor laadde. Een en ander zou aanleiding kunnen zijn om de betekenis van de uitspraak 'evenveel waard' voor de factor 'positief gevoel' en voor de totale schaal opnieuw te evalueren.

Deze bevindingen over het optreden van meetfouten die kunnen worden toegeschreven aan de interviewer, kunnen ook worden gebruikt voor een aantal praktische implicaties. Uitspraken over de relaties tussen de factor 'positief gevoel' en andere variabelen zullen met de nodige voorzichtigheid gedaan moeten worden. Een andere implicatie is dat tijdens het trainen van interviewers ten behoeve van de afname van de CES-D-schaal voor welbevinden rekening kan worden gehouden met de opgedane kennis, en men bijvoorbeeld tijdens proefinterviews nauwkeuriger de interacties tussen interviewer en respondent nagaat.

Het onderzoek naar interviewereffecten als type meetfout is van belang voor de onderkenning van problematische situaties van ouderen in zoverre deze zijn gebaseerd op de genoemde schaal voor welbevinden. De CES-D-schaal wordt ook wel gebruikt als indicatie voor potentiële depressieve klachten van een wat ernstiger aard. Een score van 16 wordt daarbij meestal als criterium genoemd. Wanneer er bij het afnemen van de uitsprakenlijst fouten optreden, betekent dit dat het criterium waarop de indicatie gebaseerd is onbetrouwbaarder wordt. Dit houdt in dat het aantal mensen toeneemt dat ten onrechte niet de aandacht krijgt van zorgverleners, terwijl anderen deze aandacht ten onrechte wel krijgen.

Geraadpleegde literatuur

- Beekman, A. T. F., J. van Limbeek, D. J. H. Deeg, L. Wouters & W. van Tilburg (1994), Screeninginstrumenten voor depressie bij ouderen in de algemene bevolking: De bruikbaarheid van de Center For Epidemiologic Studies Depression Scale (CES-D). *Tijdschrift voor Gerontologie en Geriatrie*, 25 (3), 95-104.
- Bie, S. E. de, & W. Dijkstra (1989), *Interviewen Cursusboek*. Amsterdam: Siswo, Vereniging van Onderzoeksinstituten.
- Biemer, P. P., R. M. Groves, L. E. Lyberg, N. A. Mathiowetz & S. Sudman (red.) (1991), *Measurement Errors in Surveys*. New York: John Wiley & Sons.
- Bryk, A. S., & S. W. Raudenbush (1992), *Hierarchical Linear Models: Applications and Data Analysis Methods*. Londen: Sage Publications.
- Choi, I. C., & G. W. Comstock (1975), Interviewer effect on responses to a questionnaire relating to mood. *American Journal of Epidemiology*, 101, 4-92.
- Deeg, D. J. H., C. P. M. Knipscheer & W. van Tilburg (1993), *Autonomy and Well-being in the Aging Population: Concepts and Design of the Longitudinal Aging Study Amsterdam*.

- NIG-overzichtstudie 7. Bunnik: Nederlands Instituut voor Gerontologie.
- Eeden, P. van den, & C. Bruinsma (1992), *Responding Environment Theory*. Ongepubliceerd manuscript. Amsterdam: Vakgroep Methoden en Technieken. Vrije Universiteit.
- Emans, B. (1985), *Interviewen: Theorie, Techniek en Training*. Groningen: Wolters-Noordhoff.
- Fowler, F.J., & Th.W. Mangione (1990), *Standardized Survey Interviewing; Minimizing Interviewer-Related Error*. Newbury Park, CA: Sage.
- Fuller, W.A. (1987), *Measurement Error Models*. New York: Wiley.
- Groves, R.M. (1989), *Survey Error and Survey Cost*. New York: John Wiley & Sons.
- Groves, R.M., & N.H. Fultz (1985), Gender effects among telephone interviewers in a survey of economic attitudes. *Sociological Methods & Research*, 14, 31-52.
- Handlin, V., D. Klassen, R. Hornstra & A. Roth (1974), *Interviewer Effects in a Community Mental Health Survey*. Technical Report. The greater Kansas City Mental Health Foundation. Contract PH 43-66-1324, National Institute of Mental Health.
- Hertzog, C., J. van Alstine, P.D. Usala, D.F. Hultsch & R. Dixon (1990), Measurement Proportions of The CES-D Scale in older populations. *Psychological Assessment*, 2, 64-72.
- Hox, J.J., E.D. de Leeuw & I.G.G. Kreft (1991), The effect of interviewer and respondent characteristics on the quality of survey data: A multilevel model. In: P.P. Biemer, R.M. Groves, L.E. Lyburg, N.A. Mathiowetz & S. Sudman (red.), *Measurement Errors in Surveys*. New York: Wiley.
- Kennedy, G.J., & H.R. Kelman (1989), Hierarchy of characteristics associated with depressive symptoms in an urban elderly sample. *American Journal of Psychiatry*, 146, 220-225.
- Knipscheer, C.P.M., J. de Jong Gierveld, T.G. van Tilburg & P.A. Dijkstra (forthcoming), *Living Arrangement and Social Network of Older Adults: First Results*.
- Longford, N.T. (1986), *VARCL-Interactive Software for Variance Component Analysis: Applications for Survey Data*. Lancaster: Lancaster University. Centre for Applied Statistics.
- Longford, N.T. (1993), *Random Coefficient Models*. Oxford: Clarendon Press.
- Molenaar, N. (1988), *Formuleringseffecten in Interviews*. Amsterdam: Vrije Universiteit.
- NIA (1988), *Assessing Physical Performance in the Home*. Part 1 and 2. Bethesda, MD: National Institute on Aging.
- Radloff, L.S. (1977), The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 3, 385-401.

Bijlage 1: De CES-D-vragenlijst

Wilt u me voor elk van de volgende uitspraken zeggen of u zich de afgelopen week zo voelde. U kunt steeds antwoorden met behulp van de antwoordmogelijkheden.

0. zelden of nooit 1. soms 2. vaak 3. altijd of bijna altijd

1. De afgelopen week maakte ik me zorgen om dingen waar ik me anders geen zorgen over maak.
2. De afgelopen week had ik geen zin in eten, was mijn eetlust slecht.
3. De afgelopen week kon ik een neerslachtige stemming niet van me afschudden, zelfs niet met hulp van mijn familie en vrienden.
4. De afgelopen week voelde ik me evenveel waard als andere mensen.
5. De afgelopen week had ik moeite mijn gedachten te houden bij wat ik aan het doen was.
6. De afgelopen week voelde ik me depressief.
7. De afgelopen week had ik het gevoel dat alles wat ik deed me moeite kostte.
8. De afgelopen week was ik hoopvol gestemd over de toekomst.
9. De afgelopen week vond ik mijn leven een mislukking.
10. De afgelopen week voelde ik me angstig.
11. De afgelopen week had ik een onrustige slaap.
12. De afgelopen week was ik gelukkig.
13. De afgelopen week praatte ik minder dan gewoonlijk.
14. De afgelopen week voelde ik me eenzaam.
15. De afgelopen week waren de mensen onvriendelijk.
16. De afgelopen week had ik plezier in het leven.
17. De afgelopen week moest ik soms huilen.
18. De afgelopen week voelde ik me bedroefd.
19. De afgelopen week had ik het gevoel dat de mensen me niet aardig vonden.
20. De afgelopen week kon ik maar niet goed op gang komen.

Bijlage 2: Voorbeelden van uitspraken

- absgez1:** Hoe is over het algemeen uw gezondheidstoestand? (zeer goed → zeer slecht)
- relgez2:** Hoe is uw gezondheidstoestand in vergelijking met andere mensen van uw leeftijd? (veel beter → veel slechter)
- intopleid:** Wat is de hoogste opleiding waar u een diploma voor heeft behaald? (1 = lagere school → 9 = academisch (oude stijl))
- intextrav** (14 uitspraken): Spraakzaam. Praat veel, tegen iedereen. (1 = absoluut niet van toepassing → 7 = goed van toepassing)
- intvriend** (14 uitspraken): Heeft weinig vertrouwen in zijn/haar medemens. Cynisch. (1 = absoluut niet van toepassing → 7 = goed van toepassing)
- intzorgv** (14 uitspraken): Nonchalant, doet de dingen wanneer het hem/haar zo uitkomt. Vergeet nogal eens een afspraak. (1 = absoluut niet van toepassing → 7 = goed van toepassing)

- intzekeer** (10 uitspraken): Positieve of negatieve reacties van anderen hebben niet zoveel invloed op wat ik doe. (1 = absoluut niet van toepassing → 7 = goed van toepassing)
- intongew** (10 uitspraken): Als iemand mij onderbreekt als ik iets tegen een ander zeg, dan vraag ik hem/haar te wachten tot ik uitgesproken ben. (1 = absoluut niet van toepassing → 7 = goed van toepassing)

* De oorspronkelijke uitspraken worden hier gepresenteerd. In verband met de neiging van mensen om steeds hetzelfde antwoord te geven ('response set') is een aantal uitspraken 'negatief' geformuleerd

Bijlage 3

Resultaat van de factor analysis (orthogonale varimax rotatie). De factoren zijn: 'depressief gevoel', 'somatische symptomen en remmingen', 'positief gevoel' en 'interpersoonlijke relaties'. Ladingen kleiner dan 0.5 zijn niet vermeld.

	Depressief gevoel	Somatische symptomen en remmingen	Positief gevoel	Interpersoonlijke relaties
bedroefd	0.799			
soms huilen	0.795			
neerslachtig	0.662			
depressief	0.588			
zorgen	0.557			
eenzaam	0.509			
angstig				
kost moeite		0.711		
niet op gang		0.677		
geen eetlust		0.597		
gedachten er niet bij				
minder praten				
onrustige slaap				
mislukking				
gelukkig			0.749	
plezier in het leven			0.736	
hoopvol gestemd			0.705	
evenveel waard				
onvriendelijk				0.803
niet aardig				0.791
Eigenwaarde	6.179	1.374	1.221	1.229

Bijlage 4

Beschrijving relevante variabelen

Aantal respondenten = 2383; aantal interviewers = 35.

Variabele	Gemid	StdD	Min	Max	Label
Leeftijd	69,56	8,65	55	86	respondent leeftijd (jaren)
sekse	1,51	0,50	1	2	(man = 1)
Absolute gezh	2,38	0,92	1	5	absolute subjectieve gezondheid
relat gezh	2,45	0,90	1	5	relatieve subjectieve gezondheid
intleeft	42,65	9,64	20	58	interviewer leeftijd (jaren)
intopleid	5,85	1,38	4	8	interviewer opleiding (laag → hoog)
intextrav	64,87	10,11	38	89	interviewer extraversie (laag → hoog)
intvriend	81,05	8,05	64	96	interviewer vriendelijk-heid (laag → hoog)
intzorgv	71,00	9,52	53	94	interviewer zorgvuldig-heid (laag → hoog)
intzeker	39,09	5,26	28	50	interviewer zekerheid (laag → hoog)
intongew	34,85	5,05	29	46	interviewer omgaan ongewenst gedrag (laag → hoog)