

Het Zwitserlevengevoel. Selectieve uitval door onderdekking en non-respons in onderzoek naar ouderen: Een overzicht

H. van Goor^a

Selective loss of participants as a consequence of under-coverage and non-response in research among the aged: An overview

This article presents an overview of biases in studies among the aged as a consequence of under-coverage and non-response. The question studied is whether participants and non-participants in research differed in their socioeconomic status (education; income; SES of neighbourhood), social integration (marital status; living arrangements; social participation; ethnicity) and health (general health situation; mortality in the period after study; lifestyle and risk factors, namely obesity and alcohol abuse; contact with physicians). In all, 65 studies were reviewed. The results showed that research participants had a higher socioeconomic status, were more integrated socially and in better health than research non-participants. The differences (effect sizes) were small to moderate. On the basis of these results, we can conclude that univariate distributions and prevalence estimates as well as relationships between variables in published research will frequently be biased.

Key words: Under-coverage, non-response, bias, aged, survey research
Tijdschr Gerontol Geriatr 2009; 40: 2-16

Samenvatting

In dit artikel wordt een overzicht gegeven van de selectieve uitval in onderzoek naar ouderen die het gevolg is van onderdekking en non-

respons. Nagegaan is of deelnemers en niet-deelnemers aan onderzoek verschillen in sociaaleconomische status (opleiding; inkomen; SES woonbuurt), maatschappelijke integratie (burgerlijke staat; woonsituatie; maatschappelijke participatie; etniciteit) en gezondheid (algehele gezondheid; mortaliteit in de periode na onderzoek; leefstijl en risicofactoren, te weten overgewicht en ernstige alcoholproblemen; contacten met artsen). In totaal zijn 65 studies onderzocht. Uit het onderzoek blijkt dat deelnemers aan onderzoek een ho-

^a Vakgroep Sociologie, Faculteit Gedrags- en Maatschappijwetenschappen, Rijksuniversiteit Groningen, Groningen
Correspondentie: Dr. H. van Goor, Vakgroep Sociologie Rijksuniversiteit Groningen, Grote Rozenstraat 31, 9712 TG Groningen. T: 050-5415796. E: henk.van.goor@rug.nl

gere sociaaleconomische status hebben, maatschappelijk meer geïntegreerd zijn en een betere gezondheid hebben dan niet-deelnemers. De verschillen (effectgroottes) zijn klein tot matig groot. Op basis van deze uitkomsten kan worden geconcludeerd dat zowel univariate verdelingen en prevalentieschattingen, als verbanden tussen variabelen in onderzoek regelmatig vertekend zullen zijn.

Trefwoorden: onderdekking; non-respons; selectieve vertekeningen; ouderen; survey-onderzoek

Inleiding

Uitval van potentiële onderzoeksdeelnemers is een van de ernstige problemen waarmee onderzoekers worden geconfronteerd. Deelname aan onderzoek neemt al decennia lang gestaag af, zowel in sociaalwetenschappelijke, als in epidemiologische studies.¹⁻⁵ Het uitvalprobleem is zo mogelijk nog groter in onderzoek naar ouderen, aangezien deelname aan onderzoek samenhangt met leeftijd: met het klimmen der jaren neemt de participatie af.⁶⁻⁸ Niet alleen neemt het aantal personen af dat beschikbaar is voor analyse, maar nog belangrijker is dat de uitval meestal selectief is.^{9,10} Deelnemers en niet-deelnemers verschillen in allerlei opzichten van elkaar, waardoor onderzoeksuitkomsten vertekend raken. Zo wordt de prevalentie van depressie onder ouderen systematisch onderschat als gevolg van methodologische tekortkomingen in onderzoek op dit terrein.¹¹ Ook het voorkomen van demantie onder ouderen wordt in survey-onderzoek onderschat.¹² Het kleiner worden van het verschil in prevalentie van hoge bloeddruk tussen blanken en zwarten in de VS, dat in de afgelopen decennia in onderzoek is vastgesteld, is tot slot eveneens in belangrijke mate een methodologisch artefact: een derde tot de helft van de afname is toe te schrijven aan dalende responsniveaus.¹³ Gezonde en vitale ouderen lijken in onderzoek dus te zijn oververtegenwoordigd. Dit wekt associaties met het beeld van het Zwitserlevengevoel uit de reclame – van ouderen als kapitaalkrachtige, sociaalactieve en gezonde levensgenieters. Het Zwitserlevengevoel biedt echter geen representatief beeld van de situatie waarin ouderen in doorsnee verkeren: veel ouderen hebben juist een laag inkomen, verkeren in een geïsoleerde positie en kampen met lichamelijke en geestelijke gebreken en stoornissen.^{14,15} De vraag rijst in hoeverre onderzoeksuitval leidt tot een systematisch vertekend beeld van de leefomstandigheden en gezondheid van ouderen. Deze vraag staat in dit artikel centraal.

Factoren die een rol spelen bij vertekeningen in onderzoek, zijn onder meer de uitsluiting van

een deel van de onderzoekspopulatie, hier verder aangeduid als onderdekking, en het niet deelnemen aan onderzoek van personen die daartoe worden uitgenodigd, in het vervolg non-respons genoemd. In dit artikel wordt een overzicht gegeven van een aantal factoren dat samenhangt met de uitval van potentiële deelnemers aan onderzoek naar ouderen, zowel door onderdekking als non-respons. Nagegaan wordt of participanten en niet-participanten systematisch van elkaar verschillen in kenmerken die betrekking hebben op twee centrale thema's in het leven van ouderen: maatschappelijke positie en gezondheid. Hebben deelnemers een hogere maatschappelijke positie, zijn zij maatschappelijk meer geïntegreerd en hebben zij een betere gezondheid dan niet-deelnemers? En zo ja: hoe groot zijn deze verschillen en wat zijn de consequenties hiervan voor onderzoek?

Van beide foutenbronnen die in dit artikel worden onderzocht, heeft onderdekking duidelijk minder aandacht gekregen dan non-respons. Ten onrechte, want als gevolg van onderdekking kan een aanzienlijk deel van de onderzoekspopulatie al bij voorbaat van onderzoek zijn uitgesloten.¹² Onderdekking kan op twee manieren ontstaan: door het steekproefkader dat wordt gebruikt om personen voor onderzoek te selecteren en door de verwijdering van bepaalde categorieën door de onderzoekers zelf. Om te beginnen dekt een steekproefkader zelden de gehele populatie. Marginale en mobiele bevolkingscategorieën, evenals mensen in inrichtingen en tehuizen zijn vaak ondervertegenwoordigd.¹⁶ Daar komt nog bij dat steekproefkaders sterk verschillen in dekkinggraad en in selectiviteit: zo leveren een bevolkingsregister of een (mini-)volkstelling in het algemeen een meer representatief beeld van een populatie dan een telefoonboek of een kiesregister.^{8,17,18} Daarnaast verwijderen onderzoekers, veelal om praktische redenen, zelf vaak bepaalde categorieën personen uit hun onderzoek. Zo worden personen die lastig te onderzoeken zijn of die niet in staat geacht worden om mee te doen (hoewel ze wel tot de onderzochte populatie behoren), met grote regelmatig weggelaten, ofwel door hen reeds vooraf uit te sluiten, ofwel door hen achteraf uit de getrokken steekproef te verwijderen. Deze werkwijze leidt er toe dat de onderdekking toeneemt: bepaalde subcategorieën uit de populatie worden op deze manier immers van onderzoek uitgesloten. Problemen inherent aan het steekproefkader en verwijdering van subcategorieën door onderzoekers worden daarom beide tot de onderdekking gerekend. Uitsluiting van bepaalde personen zal soms onvermijdelijk zijn, maar het probleem is dat onderzoekers zelden of niet ingaan op de consequenties die dit heeft voor de representativiteit van het onderzoek. Het probleem is namelijk dat het bijna altijd om dezelfde bevolkingscategorieën gaat, de uitval is dus selectief.¹⁶ In dit artikel gaat de aandacht vooral uit naar de

vraag welke categorieën ouderen aldus van onderzoek worden uitgesloten.

Naar non-respons wordt regelmatig onderzoek gedaan. Non-respons kan verschillende vormen aannemen. Om te beginnen slaagt men er niet altijd in om met iedereen contact te leggen (non-contact of niet-bereikbaar). Vervolgens wenst niet iedereen met wie contact wordt gelegd, aan het onderzoek mee te werken (weigering). Tot slot zijn er personen die niet in staat zijn tot deelname, bij voorbeeld omdat zij in de tussentijd zijn overleden, ernstig ziek zijn, verhuisd of langdurig afwezig. Het onderscheid is van belang, omdat de factoren die met niet-bereikbaarheid, weigering en niet-in-staat-zijn samenhangen, deels verschillen.¹⁹ Een probleem is echter dat de verschillende typen non-respons in de praktijk soms moeilijk zijn te onderscheiden. Ook presenteren onderzoekers lang niet altijd gegevens over de verschillende typen afzonderlijk. Om deze redenen worden in dit verslag alle vormen van non-respons samengevoegd tot één categorie.

In studies naar non-responsvertekeningen in survey-onderzoek is herhaaldelijk gewezen op het bestaan van een 'middle class bias'.^{8,20} Vooral mensen uit het 'brede maatschappelijke midden' zouden aan onderzoek meedoen. Als gevolg hiervan zouden zowel de maatschappelijk toplaag, als de maatschappelijke onderlaag ondervertegenwoordigd zijn. Nagegaan zal worden in hoeverre ook in onderzoek naar ouderen sprake is van een 'middenklassenvertekening'. Het begrip middenklasse kan op twee manieren worden opgevat, in termen van sociaaleconomische status en van maatschappelijke integratie. Sociaaleconomische status (SES) verwijst naar de lagere of hogere positie die mensen in de maatschappij innemen en wordt vaak gemeten aan de hand van opleiding, beroep en inkomen. Onderzoek toont aan dat mensen uit lagere sociale milieus minder vaak meedoen aan onderzoek.^{8,20-23} Vermoedelijk geldt dit ook voor de maatschappelijke bovenlaag, maar dit is in algemeen populatieonderzoek lastig aantoonbaar. De nadruk zal hier dan ook worden gelegd op de vraag of ouderen met een lagere SES ondervertegenwoordigd zijn. Maatschappelijke integratie heeft betrekking op binding en deelname aan centrale instituties in de maatschappij, als huwelijk, arbeid, politiek en (officiële) cultuur. Uit onderzoek blijkt dat personen die maatschappelijk een marginale positie innemen, minder vaak deelnemen aan onderzoek dan mensen die maatschappelijk wel zijn geïntegreerd.^{8,20,22-24} Onderzocht zal worden of ook voor ouderen geldt dat sterker geïntegreerde personen vaker in onderzoek participeren.

Naast maatschappelijke positie is gezondheid het tweede thema waarvan hier wordt nagaan of het samenhangt met uitval. Gezondheid – hier opgevat als lichamenlijk en geestelijk welzijn – kan om te beginnen van invloed zijn op de mogelijkheid tot deelname. Mensen met een

zwakke gezondheid zullen minder goed in staat zijn om aan onderzoek deel te nemen dan gezonde mensen. Uit onderzoek naar de samenhang tussen gezondheid en non-respons onder de volwassen bevolking blijkt inderdaad dat respondenten lichamenlijk en geestelijk gezonder zijn en meer tevreden over medische zorg dan non-respondenten.²⁴⁻²⁶ Nagegaan zal worden in hoeverre ook voor ouderen geldt dat hun algehele gezondheidssituatie van invloed is op onderzoeksdeelname. Daarnaast is het mogelijk dat leefstijl en risicofactoren van invloed zijn op non-respons. Aannemelijk is dat mensen met riskante, maatschappelijk afgekeurde gedragingen en kenmerken minder participeren in onderzoek vanwege gevoelens van sociale bedreiging of schaamte. Zo heeft de vrees om negatief beoordeeld te worden een negatieve invloed op de bereidheid om deel te nemen aan onderzoek.²⁷ In dit verband zal worden onderzocht in hoeverre overgewicht en ernstige alcoholproblemen leiden tot non-respons.

Tot dusverre is steeds aangenomen dat gezondheidsproblemen een negatief effect hebben op onderzoeksdeelname. Er zijn echter studies waaruit blijkt dat respondenten meer gezondheidsklachten hebben dan non-respondenten.²⁸ Dit zou er op kunnen wijzen dat mensen met gezondheidsproblemen juist meer geïnteresseerd zijn in onderzoek en daarom vaker meedoen. Bekend is dat belangstelling voor het onderwerp van onderzoek een positief effect heeft op deelname aan surveys.^{20,29-30} Directe indicatoren voor belangstelling voor het onderwerp van studie zijn echter moeilijk te vinden. Vandaar dat gebruik is gemaakt van variabelen die betrekking hebben op het gebruik van medische voorzieningen, meer in het bijzonder contacten met artsen. Aangenomen wordt dat mensen die frequent contact hebben met een arts, vaker meedoen vanwege hun belangstelling voor het onderwerp van studie. Een complicatie is uiteraard dat een hoge medische consumptie ook kan wijzen op een slechte gezondheid. Bij de bespreking van de resultaten zullen we hier nader op ingaan.

Data en methoden

Dit onderzoek richt zich op onderdekking en non-responsvertekeningen in survey-onderzoeken naar ouderen, waarbij personen van 50 jaar en ouder als ouderen zijn beschouwd (zie verder de selectiecriteria). Zowel studies op sociaal-wetenschappelijk terrein, als op (sociaal-) medisch en epidemiologisch gebied zijn opgenomen. Onderzoeken naar ouderen waarin van andere onderzoeksopzetten gebruik is gemaakt (zoals interventiestudies, 'clinical trials', experimenten en dergelijke), zijn buiten beschouwing gelaten.

Gezocht is naar studies over non-respons in surveys onder ouderen. Van de opgespoorde stu-

dies is tevens nagegaan of bepaalde categorieën van deelname zijn uitgesloten (onderdekking). Voor het vinden van studies is gebruik gemaakt van elektronische databases op het gebied van de medische en sociale wetenschappen: Web of Science, Medline, AMED, CINAHL, EMBASE, PsycINFO en SocINDEX. Gezocht is met een combinatie van zoektermen over non-respons (als nonresponse, non-response, refusal) en ouderen (old, elderly, aged). De literatuurverzameling is in december 2007 afgesloten. Er is alleen gezocht naar publicaties in het Engels, Duits, Frans en Nederlands. Op basis van titel en (indien aanwezig) samenvatting is een eerste selectie gemaakt. De geselecteerde publicaties zijn opgezocht en aan de hand van de selectiecriteria (zie onder) is een nadere schifting aangebracht. De overgebleven publicaties zijn in detail bestudeerd. Voorts is aanvullend verder gezocht aan de hand van de literatuurlijsten van de geselecteerde publicaties.

Bij de selectie zijn de volgende criteria gehanteerd.

1. Als ouderen zijn personen van 50 jaar en ouder beschouwd. Nu verschillen studies onderling sterk in de leeftijdscategorie die men onderzoekt. Bovendien worden in studies naar ouderen geregeld ook personen jonger dan 50 jaar opgenomen. Dat heeft er uiteindelijk toe geleid dat studies zijn geselecteerd met daarin personen van 50 jaar of ouder en waarin geen personen jonger dan 30 jaar zijn opgenomen. Voorts zijn enkele studies opgenomen zonder leeftijdsgrenzen, die overwegend op ouderen betrekking hebben. Tenslotte is gebruik gemaakt van enkele studies over de gehele volwassen bevolking, waarin (ook) gegevens worden gepresenteerd over non-responsvertekeningen in oudere leeftijdscategorieën.

2. Het onderzoek is beperkt tot studies waarin gegevens worden verschaft over de respondenten en non-respondenten uit de oorspronkelijke steekproef of populatie, waarmee het onderzoek is gestart. Zowel eenmalige, cross-sectionele onderzoeken, als longitudinale ('panel') studies zijn opgenomen; bij longitudinale onderzoeken zijn alleen de respondenten en non-respondenten uit de eerste afnameronde ('baseline') in de analyse betrokken.

3. Er zijn alleen metingen van variabelen opgenomen waarover (bijna) volledige informatie (85 procent of meer) beschikbaar is voor respondenten en non-respondenten. Een probleem van veel non-responsonderzoek is dat vaak voor slechts een beperkt deel van de non-respondenten gegevens voorhanden zijn. In deze gevallen is niet zeker dat de non-respondenten over wie, bij voorbeeld uit een aanvullende enquête, informatie is verkregen, representatief zijn voor alle non-respondenten. Vandaar dat in dit overzicht alleen studies zijn opgenomen waarin over (nagenoeg) alle (non-)respondenten informatie beschikbaar is. Het aantal studies dat voor opname in aanmerking komt, is daardoor beperkt. Daar

staat tegenover dat de geselecteerde studies van hoge kwaliteit zijn.

Eenheid van analyse (*studie*) is elk onderzoek dat betrekking heeft op een specifiek databestand. Publicaties die op eenzelfde databestand betrekking hebben, zijn als één studie geteld. Omgekeerd is een publicatie waarin over twee (of meer) afzonderlijk verzamelde databestanden wordt gerapporteerd, als twee (of meer) studies beschouwd.

Van elke in dit onderzoek opgenomen variabele uit een studie zijn – voor zover aanwezig – alle gegevens over de *meting* genoteerd, zoals gemiddelde en standaarddeviatie, frequentieverdeling, statistische significantietoets. Vervolgens zijn de gegevens op twee manieren geanalyseerd. Om te beginnen is gekeken naar de *richting* van de verschillen tussen respondenten en non-respondenten, ongeacht of deze verschillen statistisch significant waren of niet. De variabelen zijn steeds zo ingedeeld dat een hogere waarde verwijst naar een hogere sociaaleconomische status, een sterkere maatschappelijke integratie, een betere gezondheid, minder riskante gedragingen en kenmerken, meer contact met artsen. Vervolgens is gekeken naar de *grootte* van de verschillen tussen respondenten en non-respondenten: de effectgrootte ('effect size', ES). Aangezien de variabelen, bij voorbeeld opleidingsniveau of algehele gezondheidssituatie, in de bestudeerde onderzoeken meestal op verschillende wijze zijn gemeten, is gebruik gemaakt van een gestandaardiseerde maat voor de bepaling van de effectgrootte: het gestandaardiseerde verschil in gemiddelde ('standardized mean difference', ES_{sm}) tussen respondenten en non-respondenten. ES_{sm} wordt uitgedrukt in standaarddeviatie-eenheden. In formule:

$$ES_{sm} = \frac{\bar{X}_R - \bar{X}_{NR}}{s_{pooled}}$$

waarin \bar{X} = gemiddelde (van respondenten, respectievelijk non-respondenten) en s_{pooled} = standaarddeviatie voor respondenten en non-respondenten samen.³¹ Bij de berekening van de effectgroottes zijn de oorspronkelijke, ongecorrigeerde verdelingen van de variabelen gebruikt. Indien mogelijk, is gebruik gemaakt van gemiddelden of frequentieverdelingen om ES_{sm} te berekenen. In een aantal gevallen was het noodzakelijk om terug te vallen op alternatieve maten, die met ES_{sm} vergelijkbare uitkomsten opleveren.³¹ Een bijkomend probleem is dat de gegevens uit de onderzoeksverslagen niet altijd toereikend zijn om de effectgrootte te berekenen. Dit betekent dat ES_{sm} soms niet kon worden berekend; in andere gevallen was het alleen mogelijk de grootte van ES_{sm} te schatten. Ook kwalitatieve uitspraken in de studies over het ontbreken van een verschil tussen respondenten en non-respondenten zijn opgenomen; in deze gevallen is ES_{sm} op o gesteld. Vanwege deze problemen en

Tabel 1 Overzicht van gebruikte studies naar onderzochte populaties en leeftijdscategorieën				
Populatie	Leeftijdscategorieën			Totaal
	30-65	≥65	zowel <65 als ≥65 ^a	
Algehele populatie of cohort ^b	10	22	24	56
Specifieke populatie: medisch ^c	0	2	5	7
Specifieke populatie: niet-medisch ^d	1	0	1	2
Totaal	11	24	30	65

In alle studies zijn personen opgenomen van 50 jaar of ouder. ^a Ouder dan 30 jaar; ook opgenomen zijn enkele studies zonder leeftijdsgrenzen die voornamelijk op ouderen betrekking hebben. ^b Alle personen behorend tot een bepaald leeftijdscohort (zelfstandig wonend en in verzorgings- of verpleeghuis, of alleen zelfstandig wonende personen; inclusief studies onder alleen mannen of vrouwen). ^c Populatie geselecteerd op basis van een bepaalde medische aandoening (bij voorbeeld kanker) of medische behandeling (bij voorbeeld ziekenhuisopname) (inclusief studies onder alleen mannen of vrouwen). ^d Mannelijke werknemers van één bedrijf; gepensioneerde vrouwelijke werknemers van één bedrijf.

vanwege het beperkte aantal metingen per variabele waarover wij kunnen beschikken, is hier volstaan met de presentatie – per variabele – van de kleinste en de grootste gevonden effectgrootte en van de mediane effectgrootte. Op deze wijze wordt een beeld verkregen van de effectgroottes die men in onderzoek kan verwachten.

Resultaten

Studies

In totaal waren er op grond van bovenstaande criteria 65 studies beschikbaar, waarin wordt gerapporteerd over ten minste één voor dit onderzoek relevante variabele. Een overzicht van de gebruikte publicaties is verkrijgbaar bij de auteur. De omvang van de onderzochte steekproeven loopt uiteen van zo'n 150 tot meer dan 200.000 personen. De studies zijn verschenen in de periode 1959-2007 en hebben betrekking op 14 geïndustrialiseerde landen: USA (23), UK (12), Zweden (9), Nederland (6), Duitsland (3), Italië (3), Finland (2), Australië (1), Canada (1), Denemarken (1), Israël (1), Japan (1), Noorwegen (1) en Zwitserland (1). De meeste studies (56) hebben betrekking op ouderen in het algemeen, negen hebben betrekking op specifieke subpopulaties, waarvan de meeste zijn afgebakend aan de hand van medische criteria (ziekte of behandeling). Er zijn maar weinig studies geselecteerd die uitsluitend op 'jongere ouderen' (65-minners) betrekking hebben; de meeste hebben betrekking op 65-plussers of op jongere en oudere ouderen samen (tabel 1).

Onderdekking: Uitsluiting van categorieën ouderen
Onderdekking veroorzaakt door het steekproefkader alleen blijft hier buiten beschouwing, daar in de onderzochte studies zelden of nooit wordt

ingegaan op de uitval die daarvan het gevolg is. De aandacht richt zich op de categorieën ouderen die onderzoekers zelf – vooraf of achteraf – van onderzoek uitsluiten. Hoe vaak worden personen verwijderd en welke typen personen verdwijnen op deze wijze uit onderzoek? Vanwege het afwijkende karakter van de steekproeven van de specifieke populaties is de analyse beperkt tot de onderzoeken naar ouderen in het algemeen (N=56). Tabel 2 verschaft de gegevens.

In 36 van de 56 onderzoeken (64 procent) worden door de onderzoekers *expliciet* personen uitgesloten van onderzoek, personen die niet worden meegenomen in de non-responsanalyse. In hoeverre in de resterende 20 studies bij voorbaat al mensen zijn afgefallen is niet duidelijk: van een aanzienlijk deel van deze studies is aanneemelijk dat de non-responsanalyse op de gehele oorspronkelijke ('bruto') steekproef betrekking heeft, maar niet in alle gevallen bestaat hierover zekerheid. De omvang van de gerapporteerde uitval varieert sterk, van minder dan één procent tot ruim 40 procent. Wat in tabel 2 opvalt, is dat uitval het vaakst voorkomt in de oudste leeftijdscategorie (65+) en het minst in de jongste leeftijdscategorie (65-). Tevens is de omvang van de uitval het grootst in de 65+ categorie. Er lijkt dus een samenhang te bestaan met leeftijd: naarmate de onderzochten ouder zijn, wordt een groter deel van de onderzoekspopulatie bij voorbaat reeds van onderzoek uitgesloten.

Bekijken we waarom onderzoekers personen uit onderzoek verwijderen, dan blijkt dat uitsluiting vooral te maken heeft met het *niet in staat zijn* tot deelname aan onderzoek: vooral medische omstandigheden (in ruime zin) en voorts verhuizing en taalproblemen. Om te beginnen zien we dat in 30 studies (54 procent) personen worden verwijderd vanwege overlijden (in de periode tussen steekproeftrekking en onder-

Tabel 2 Uitsluiting van deelname aan onderzoek naar leeftijdscategorie (alleen algemene populaties, N=56)				
Redenen van uitsluiting ^a (<i>expliciet vermeld</i>)	Leeftijdscategorie			Totaal
	65-	65+	65-/65+	
(N=)	(10)	(22)	(24)	(56)
Overlijden ^b	1	12	8	21
Medische redenen ^c	2	7	3	12
In verzorgings- of verpleeghuis	1	7	4	12
Een of meer van voorgaande redenen ^d	3	15	12	30
Verhuisd/afwezig/onvindbaar	1	10	7	18
Taalproblemen/minderheidsgroepering	2	6	1	9
Overige	0	7	8	15
Totaal studies met expliciete uitsluiting ^e	4	18	14	36

^a Meer dan één reden per onderzoek mogelijk. ^b Tussen steekproeftrekking en uitvoering van het veldwerk. ^c Ernstige of terminale ziekte, zwakke gezondheid, dementie of andere geestelijke aandoening, cognitief gebrekkig functioneren. ^d Overlijden en/of medische oorzaken en/of verblijf in verzorgings- of verpleeghuis. ^e Een of meer van de zes onderscheiden redenen van uitsluiting genoemd.

zoek), medische omstandigheden en/of verblijf in een verzorgings- of verpleeghuis. Aannemelijk is dat deze mensen in het algemeen een (veel) zwakkere gezondheid hadden dan degenen die wel in de analyse zijn betrokken.

In 18 onderzoeken (32 procent) worden meer mobiele mensen (verhuisd/niet te vinden) van onderzoek uitgesloten. Mobiliteit kan zowel wijzen op een slechte gezondheid die mensen noopt om te verhuizen, bij voorbeeld naar een verzorgings- of verpleeghuis, als juist op een grotere vitaliteit. Vermoedelijk doen beide situaties zich voor. Daarnaast worden in 9 studies (16 procent) mensen uitgesloten die de officiële voertaal van het land onvoldoende beheersen. Dit zijn mensen die maatschappelijk minder sterk zijn geïntegreerd. Ten slotte worden in 15 studies (27 procent) ‘overige’ redenen genoemd, waaronder ‘administratieve redenen’, uitsluiting van niet-bereikbaren (waaronder mensen zonder telefoon), of van weigeraars bij een ‘eligibility screening’.

We kunnen concluderen dat door de – overigens vaak noodgedwongen – gehanteerde uitsluitingsgronden reeds een eerste, soms aanzienlijke selectie van onderzoekssubjecten plaats vindt. De netto-steekproef die veelal het uitgangspunt is voor de analyse van de non-respons in een onderzoek, is daardoor al duidelijk veel minder representatief voor de onderzochte po-

pulatie dan de bruto-steekproef. Ouderen in de netto-steekproef zijn vermoedelijk (veel) gezonder en maatschappelijk meer geïntegreerd dan de uitvallers. Wat eerder al bleek in onderzoek naar dementie,¹² blijkt dus veel algemener te gelden voor onderzoek onder ouderen.

Non-respons: Verschillen tussen respondenten en non-respondenten

In het navolgende overzicht van non-responsvertekeningen wordt eerst aan de hand van de *richting* van de verschillen tussen respondenten en non-respondenten nagegaan of er een systematisch patroon in de non-respons valt te ontdekken. Is een dergelijk systematisch patroon aanwezig, dan is het vervolgens interessant om na te gaan hoe groot deze verschillen zijn. Dit laatste wordt vastgesteld aan de hand van de boven geïntroduceerde maat voor effectgrootte ES_{sm} .

Bij de bespreking van de verschillen tussen respondenten en non-respondenten wordt een aantal keren aandacht besteed aan ‘afwijkende’ gevallen. Gelet op de kleine aantallen studies hebben deze aanvullende analyses een speculatief, hypothesevormend karakter. Ze dienen vooral een heuristisch doel: het zoeken naar factoren die ons kunnen helpen om verschillen in non-responspatroon tussen studies te verklaren.

Hoe staat het met de respons in de onderzochte studies? De gerapporteerde responsniveaus variëren tussen 34 en 92 procent, met een mediaan van 74 procent. Slechts 20 procent van de onderzoekers meldt een respons lager dan 60 procent, terwijl bijna een derde een responspercentage van meer dan 80 noemt. De respons is dus in het algemeen hoog. De cijfers dienen wel met enige voorzichtigheid te worden geïnterpreteerd: de werkelijke respons ligt in een aantal gevallen vermoedelijk lager. Herberekeningen laten zien dat de overschatting van het responsniveau overigens meestal beperkt is (minder dan vijf procent).

*Richting van non-responsverschillen:
Sociaaleconomische status (SES)*

Voor sociaaleconomische status zijn gegevens beschikbaar over opleidingsniveau, inkomen en de sociaaleconomische status van de woonbuurt, in totaal 20 metingen uit 19 verschillende studies (tabel 3). De resultaten van alle de analyses zijn zeer eenduidig: in 18 van de 20 metingen (90 procent) hebben respondenten een hogere SES dan non-respondenten; eenmaal scoren de non-respondenten hoger en eenmaal is er geen verschil. Respondenten hebben gemiddeld een hoger opleidingsniveau, een hoger inkomen en wonen vaker in een buurt met een hoge SES.

Non-respondenten scoren eenmaal hoger, te weten bij opleiding. Bekijken we deze 'afwijkende' score nader, dan blijkt deze score zich voor te doen in een studie met een zeer specifieke en van de andere studies sterk afwijkende

populatie, namelijk personen met longkanker. Tachtig procent van de non-respondenten in deze studie blijkt te zijn overleden.³² Door de overheersende betekenis van sterfte als reden van uitval, lijkt de betekenis van andere factoren naar de achtergrond te verdwijnen. De 'controls' uit dit 'case-control' onderzoek, die hier als een afzonderlijke studie zijn beschouwd, vertonen wel het bekende patroon: respondenten zijn gemiddeld hoger opgeleid dan non-respondenten.³² Deze uitkomsten suggereren dat in geval van ernstige gezondheidsproblemen andere factoren die non-respons kunnen veroorzaken naar de achtergrond worden gedrongen. Nader onderzoek zal moeten uitwijzen of dit inderdaad het geval is.

*Richting van non-responsverschillen:
Maatschappelijke integratie*

Voor maatschappelijke integratie zijn er gegevens over burgerlijke staat, woonsituatie, maatschappelijke participatie en etniciteit. Met uitzondering van woonsituatie zijn de resultaten ook hier zeer consistent en in overeenstemming met de verwachtingen (tabel 4). Respondenten zijn vaker gehuwd, hun maatschappelijke participatie is hoger en ze zijn vaker afkomstig uit de maatschappelijke meerderheidsgroepering: in 35 (92 procent) van 38 metingen (exclusief woonsituatie).

Een opvallende afwijking van het algemene patroon vinden we bij woonsituatie: in de ene helft van de studies wonen de respondenten vaker samen met anderen, in de andere helft

Tabel 3 Verschillen in sociaaleconomische status (SES) tussen respondenten en non-respondenten in onderzoek onder ouderen					
Variabele	Aantal studies	Aantal metingen	Hogere score voor		
			R	NR	Geen verschil
Opleidingsniveau (laag-hoog)	6	6	5	1	0
Inkomen (laag-hoog) ^a	5	5	5	0	0
SES van woonbuurt (laag-hoog)	9	9	8	0	1 ^b
Totaal SES	19 ^c	20	18	1	1

In de tabel wordt per variabele achtereenvolgens het aantal studies genoemd waarin deze variabele is onderzocht, het aantal metingen van deze variabele in de betreffende studies (soms is in een studie meer dan één indicator gebruikt voor een variabele) en tot slot wordt aangegeven of respondenten (R) dan wel non-respondenten (NR) een hogere score hebben op de onderzochte metingen voor de betreffende variabele, of dat respondenten en non-respondenten even hoog scoren (geen verschil). 'Geen verschil' betekent dat het verschil 0 is, dat wil zeggen kleiner is dan $\pm 0,1$ procent. Leesvoorbeeld: de variabele opleidingsniveau is in zes studies onderzocht, wat zes metingen oplevert (in elke studie eenmaal gemeten); de respondenten scoren vijf maal hoger dan de non-respondenten (dat wil zeggen: de respondenten hebben in deze vijf studies een hoger opleidingsniveau dan de non-respondenten) en eenmaal scoren de non-respondenten hoger (in dit ene onderzoek hebben de non-respondenten een hoger opleidingsniveau dan de respondenten). ^a Inclusief een studie waarin gevraagd is of het inkomen (on)voldoende was. ^b Kwalitatieve uitspraak in tekst ("no variation in response rate by socioeconomic status"), geen cijfers gegeven. ^c Exclusief dubbeltellingen.

Tabel 4 Verschillen in maatschappelijke integratie tussen respondenten en non-respondenten in onderzoek onder ouderen					
Variabele	Aantal studies	Aantal metingen	Hogere score voor		
			R	NR	Geen verschil
Burgerlijke staat (niet gehuwd – gehuwd)	16	16	14	2	0
Woonsituatie (alleen – met anderen)	8	8	4	4	0
Maatschappelijke participatie (laag – hoog) ^a	4	8	8	0	0
Etniciteit (minderheid – meerderheid)	14	14	13	0	1 ^b
Totaal maatschappelijke integratie	33 ^c	46	39	6	1

Zie voor algemene toelichting en leesvoorbeeld tabel 3. ^a Hiertoe gerekend zijn (alle één keer): contacten met vrienden; 'job deprivation'; 'job commitment'; buurtsamenstelling (% werkloosheid en % alleenstaanden); lidmaatschap politieke partij; aantal petitie ondertekend; geregistreerd als kiezer; thuis andere taal sprekend dan officiële voertaal in land. ^b Kwalitatieve uitspraak in tekst ("no difference"), geen cijfers gepresenteerd. ^c Exclusief dubbeltellingen.

juist de non-respondenten. Dit is verrassend, aangezien burgerlijke staat wel het verwachte patroon laat zien. Een mogelijke verklaring is dat het bij de non-respondenten die met anderen samenwonen, gaat om personen met een zwakke gezondheid die door hun directe sociale omgeving worden afgeschermd voor onderzoek.^{6,33} Als deze verklaring juist is, dan zouden samenwonende ouderen met een goede gezondheid vaker meedoen aan onderzoek, terwijl samenwonende ouderen met een zwakke gezondheid juist minder vaak zouden deelnemen! Integratie in sociale verbanden zou dan – afhankelijk van iemands gezondheidssituatie – tegengestelde effecten hebben op onderzoeksdeelname. Nader onderzoek zal moeten uitwijzen of de sociale omgeving van ouderen inderdaad dit tegengesteld effect heeft op deelname aan onderzoek.

Richting van non-responsverschillen: Gezondheid
Over gezondheid zijn de meest gegevens beschikbaar: 72 metingen uit 42 studies (tabel 5). Kijken we eerst naar de algehele gezondheidssituatie, die op zeer uiteenlopende wijze is gemeten, dan zien we dat respondenten in het algemeen een betere gezondheid hebben (14 van 18 metingen, ofwel 78 procent). De vier 'afwijkende' scores komen uit twee studies. Drie metingen zijn afkomstig uit een studie onder 85-jarigen.³⁴ Het is mogelijk dat de hoge leeftijd een rol speelt, maar hierover valt op basis van onze gegevens weinig met zekerheid te zeggen. De vierde meting is afkomstig uit een studie

naar oudere ziekenhuispatiënten (60+) waarin de meest ernstig zieke personen reeds vooraf van onderzoek zijn uitgesloten: een onderdekking van bijna 60 procent.³⁵ Interessant is dat er in een andere studie onder ziekenhuispatiënten (opgenomen voor een chirurgische ingreep vanwege prostaatproblemen) wel een positief verband is tussen gezondheid en respons (vier indicatoren).³⁶ Het meest opvallende verschil tussen beide studies is dat in de laatste studie weinig onderdekking optreedt, minder dan vijf procent, tegen bijna 60 procent in de eerste. Het lijkt er dus op dat de (omvang van de) onderdekking van invloed is geweest op de uitkomsten van de non-responsanalyses.

Naast gegevens over de algehele gezondheid zijn er gegevens beschikbaar over de mortaliteit onder respondenten en non-respondenten in de periode na onderzoek. Reeds Gordon et al. hebben gewezen op de mogelijkheid om de mortaliteit in de periode na onderzoek te gebruiken als indicator voor (verschillen in) de gezondheidssituatie van respondenten en non-respondenten: verwacht mag worden dat onder personen met een zwakkere gezondheid ten tijde van het onderzoek meer (en eerder) sterfte optreedt in de periode na onderzoek.³⁷ Sindsdien zijn mortaliteitscijfers regelmatig voor dit doel gebruikt. Ook hier zijn de resultaten eenduidig: de sterfte is in de jaren na onderzoek veel hoger onder non-respondenten dan onder respondenten (in 24 van 27 studies, oftewel 89 procent). Het verschil in mortaliteit is meestal het grootst in

Tabel 5 Verschillen in algemene gezondheidssituatie, risicofactoren en contacten met artsen tussen respondenten en non-respondenten in onderzoek onder ouderen					
Variabele	Aantal studies	Aantal metingen	Hogere score voor		
			R	NR	Geen verschil
<i>Algemene gezondheidssituatie en mortaliteit</i>					
Algehele gezondheid (slecht-goed) ^a	8	18	14	4	0
Mortaliteit (na survey) (hoog-laag) ^b	27	27	24	3	0
<i>Risicofactoren</i>					
Overgewicht (BMI) (ja-nee)	2	2	2	0	0
Ernstige alcoholproblemen (ja-nee)	6	9	9	0	0
<i>Contacten met artsen</i>					
Contact met huisarts (frequentie) ^c	6	6	4	1	1
Bezoek aan arts (frequentie) ^d	3	5	1	4	0
Bezoek van (huis)arts (frequentie) ^e	2	5	0	5	0
Totaal gezondheid	42 ^f	72	54	17	1

Zie voor algemene toelichting en leesvoorbeeld tabel 3. ^a Hiertoe gerekend zijn (aantal): perceptie eigen gezondheid (2), objectieve vaststelling gezondheid (3); schatting van dagelijks functioneren door verpleegkundige (1); moeilijkheden met dagelijks functioneren (ADL) (1); ontvangen van thuiszorg (1); jaarlijkse frequentie van ziekteverzuim en aantal ziektedagen (2); arbeidsongeschiktheidsuitkering (1); opname in verpleeghuis in jaar van onderzoek: frequentie, duur en kosten (3); opname in verpleeghuis in 3 jaar na onderzoek (1); aantal dagen in ziekenhuis voor operatie (1); reden ziekenhuisopname (urgent versus routine) (1); voorspelling chirurg over uitslag operatie voor patiënt (1). ^b De onderzochte periode na het onderzoek varieert van één jaar tot 14 jaar. ^c Bezoek aan of contact met huisarts/huisartsenpraktijk; inclusief algemeen gezondheidsonderzoek (1). ^d Bezoek aan arts (zonder meer) (2); bezoek aan oogarts (2); bezoek aan psychiater (1). ^e Bezoek van arts thuis, in verzorgingshuis, in verpleeghuis, op eerste hulp (alle afkomstig uit één onderzoek); contact op initiatief van huisarts. ^f Exclusief dubbeltellingen.

de eerste jaren na onderzoek, waarna de verschillen langzaam afnemen. In drie studies is de sterfte juist onder de respondenten hoger, zij het dat in alle drie gevallen de verschillen tussen respondenten en non-respondenten gering zijn. Eén studie heeft betrekking op werknemers van één bedrijf: de sterfte is relatief laag en de verschillen tussen respondenten en non-respondenten zijn miniem.³⁸ In deze studie is sprake van onderdekking waarbij zieke werknemers en personen overleden na steekproef-trekking zijn uitgesloten, wat de lagere sterfte

onder non-respondenten zou kunnen verklaren. De twee overige studies hebben betrekking op 65-plussers, waaronder een studie onder 85-jarigen. Opvallend in deze laatste studie is dat directe respondenten en non-respondenten nauwelijks verschillen in mortaliteit, maar dat met name de additionele respondenten die alsnog over de streep zijn getrokken om aan het onderzoek mee te doen, een veel hogere mortaliteit kennen.³⁴ In de derde studie valt niet uit te sluiten dat het bij voorbaat uitsluiten van personen, onder andere wegens overlijden, verblijf in een

Tabel 6	Verschillen tussen respondenten en non-respondenten in sociaaleconomische status, maatschappelijke integratie en gezondheid: Effectgroottes		
	Variabelen (N metingen)	Effectgrootte	
	Minimum	Maximum	Mediaan
<i>Sociaaleconomische status</i>			
Opleiding (5)	0,16	0,36	0,19
Inkomen (5)	0,03	0,65	0,18
SES woonbuurt (7)	0,00	0,40	0,19
Totaal SES (17)	0,00	0,65	0,19
<i>Maatschappelijke integratie</i>			
Burgerlijke staat (15)	-0,06	0,71	0,05
Woonsituatie (8)	-0,14	0,19	-0,03
Maatschappelijke participatie (7)	0,09	1,17	0,16
Etniciteit (13)	0,00	0,43	0,19
Totaal maatschappelijke integratie (exclusief woonsituatie) (35)	-0,06	1,17	0,17
<i>Gezondheid</i>			
Algehele gezondheid (14)	-0,48	0,38	0,14
Mortaliteit in eerste of in eerste twee jaar na onderzoek (10) ^a	0,04	0,28	0,13
Mortaliteit in 5 jaar na onderzoek (11)	-0,05	0,45	0,11
Overgewicht (2)	0,10	0,23	(0,17)
Ernstige alcoholproblemen (9)	0,04	0,76	0,19
Contact met huisarts (6)	-0,06	0,27	0,07
Bezoek aan arts (5)	-0,05	0,12	-0,01
Bezoek van (huis)arts (5)	-0,27	-0,06	-0,11
Totaal gezondheid (exclusief bezoek aan/van arts) (52)	-0,48	0,76	0,13

Berekend is het gestandaardiseerde verschil tussen de gemiddelden van respondenten en non-respondenten (bij voorbeeld het verschil in opleidingsniveau of in mortaliteit). De verschillen worden uitgedrukt in standaarddeviatie-eenheden. Voor elke variabele worden de kleinste en de grootste gevonden effectgrootte (minimum en maximum) vermeld, evenals de mediaan. Bij kwalitatieve uitspraken van geen verschil is de effectgrootte op 0,00 gezet. Tussen haakjes is steeds het aantal metingen vermeld waarvoor het gestandaardiseerde verschil in gemiddelden kon worden berekend. ^a Voor zes studies was het mogelijk om de verschillen in mortaliteit tussen respondenten en non-respondenten vast te stellen voor het eerste jaar na onderzoek; hieraan zijn vier andere studies toegevoegd waarin de mortaliteit kon worden berekend over de eerste twee jaar na onderzoek.

verpleeghuis en verhuizing, van invloed is geweest op de mortaliteitsverschillen tussen respondenten en non-respondenten.³⁹

Een tweede categorie factoren heeft betrekking op leefstijl en risicofactoren. De uitkomsten zijn zeer eenduidig. Alle elf metingen wijzen in dezelfde richting: overgewicht en ernstige alcoholproblemen leiden tot een geringere deelname aan onderzoek. Deze uitkomsten vormen een ondersteuning voor de theorie dat deelname aan onderzoek mede wordt bepaald door gevoelens van bedreiging, schaamte en de vrees om negatief beoordeeld te worden.

De laatste categorie variabelen heeft betrekking op contact met artsen. In tegenstelling tot de voorgaande analyses komt hier een gemengd beeld naar voren. De contacten met artsen zijn uitgesplitst in drie deelcategorieën: contacten met de eigen huisarts; contacten met andere artsen (hoewel niet steeds even duidelijk, gaat het vermoedelijk meestal om contacten met een specialist); en bezoek aan (of contact met) de patiënt door (of op initiatief van) een (huis)arts. Respondenten hebben vaker contact met de huisarts (vier uit zes metingen), terwijl non-respondenten vaker contact hebben met andere artsen (vier uit vijf metingen) en/of bezoek krijgen van een arts (alle vijf metingen). Het ligt niet voor de hand dat dit patroon van uitkomsten wordt bepaald door verschillen in belangstelling voor het onderwerp van studie. In dat geval zouden de respondenten immers steeds over meer contacten met artsen dienen te beschikken dan de non-respondenten. Aangezien non-respondenten vaker contact hebben met een specialist en vaker bezoek krijgen van een arts in een verzorgingshuis, verpleeghuis, of op de eerste hulp, lijkt het aannemelijker en meer in overeenstemming met de voorgaande resultaten dat non-respondenten een slechtere gezondheid hebben dan respondenten. De gegevens uit de beschikbare studies staan niet toe om deze samenhangen verder te onderzoeken. Het is wenselijk om in toekomstig onderzoek contacten met artsen te analyseren in combinatie met gezondheidsfactoren ten einde na te gaan in hoeverre contacten, naast gezondheid, een zelfstandig effect hebben op non-respons.

Vatten we de resultaten over de factor gezondheid samen, dan is duidelijk dat er in de bestudeerde onderzoeken sprake is van een 'healthy participant effect'. Alle onderzochte indicatoren wijzen er op dat ouderen met een goede gezondheid vaker aan onderzoek meedoen dan ouderen met een zwakke gezondheid. Dat geldt voor de algehele gezondheidssituatie, de mortaliteit in de periode na onderzoek, leefstijl en risicofactoren en vermoedelijk ook voor contacten met artsen. Gezondheid is dus een factor die leidt tot een sterke systematische differentiatie tussen respondenten en non-respondenten.

Non-respons: Effectgroottes

We hebben in het voorgaande gezien dat er inderdaad sprake is van een duidelijk patroon van uitkomsten: respondenten en non-respondenten verschillen systematisch van elkaar in sociaal-economische status, maatschappelijke integratie en gezondheid. De vraag is nu hoe groot deze verschillen zijn. Tabel 6 verschaft ons de gegevens.

Om te beginnen zien we dat er tussen studies een aanzienlijke spreiding is in effectgroottes: van betrekkelijk kleine effecten (kleiner dan $\pm 0,25$ standaarddeviatie) tot in sommige gevallen sterke effecten (groter dan $\pm 0,75$ standaarddeviatie). Nemen we de mediane effectgrootte als richtsnoer, dan zien we bij alle drie onderzochte SES-variabelen ongeveer even grote effecten van 0,18 tot 0,19. Bij maatschappelijke integratie is de mediane effectgrootte zowel bij burgerlijke staat als bij woonsituatie klein. De mediane effectgrootte bij maatschappelijke participatie en etniciteit liggen daarentegen weer op een niveau vergelijkbaar met dat van de SES-variabelen. De effectgroottes laten bij de variabelen die op de gezondheid betrekking hebben een wisselend beeld zien. Ook hier zijn de mediane effectgroottes betrekkelijk klein. In het voorgaande hebben we al enkele keren aangegeven dat onderdekking als gevolg van uitsluiting van minder gezonde personen uit de steekproef vermoedelijk effect heeft gehad op de non-responsuitkomsten. Het lijkt er op dat onderdekking een drukkend effect heeft op de grootte van de verschillen tussen respondenten en non-respondenten. Het is daarom aannemelijk dat de gevonden effectgroottes de ware omvang van de vertekeningen in gezondheid onderschatten. Bij gebrek aan gegevens is het echter onmogelijk om hierover verdergaande uitspraken te doen.

Discussie

Vatten we de uitkomsten van de voorgaande analyse van 65 studies samen, dan zien we dat respondenten en non-respondenten systematisch van elkaar verschillen. Ouderen die aan onderzoek meedoen, hebben een hogere sociaal-economische status en zijn in het algemeen meer geïntegreerd in de maatschappij dan ouderen die niet meedoen. Ook is de gezondheid van de respondenten duidelijk beter dan die van de non-respondenten: er is sprake van een 'healthy participant effect'. Maar niet alleen de non-respons is selectief. Ook in de stap die daaraan vooraf gaat – de uitsluiting van bepaalde bevolkingscategorieën van onderzoek – treedt selectieve uitval op: minder gezonde mensen worden relatief vaak van onderzoek uitgesloten, evenals mensen die maatschappelijk minder zijn geïntegreerd. Onderdekking en non-respons lijken een elkaar versterkend, cumulatief effect te hebben op de vertekeningen in maatschappelijke positie en gezondheid.

Voegen we de uitkomsten van de analyses van onderdekking en non-respons samen, dan kunnen we constateren dat vooral op het punt van de gezondheid sprake is van aanzienlijke selectie-effecten, waardoor de meest gezonde personen in onderzoek sterk oververtegenwoordigd raken. De selectieve uitval leidt tot een vertekend beeld van de leefomstandigheden en gezondheid van ouderen: ouderen worden als welvarender, maatschappelijk meer geïntegreerd, gezonder en vitaler voorgesteld dan zij in feite zijn. Het beeld van het Zwitserlevengevoel treffen we dus ook in onderzoek aan. Deze uitkomsten bevestigen de conclusies uit eerder onderzoek over de onderschatting van medische aandoeningen in survey-onderzoek.^{11,12}

In hoeverre zijn deze uitkomsten generaliseerbaar naar survey-onderzoek onder ouderen in het algemeen (externe validiteit)? Zoals ook elders, is deze vraag niet met volledige zekerheid te beantwoorden, maar er valt wel iets over te zeggen. Bij externe validiteit gaat het om de mate waarin uitkomsten van onderzoek consistent (repliceerbaar) en robuust zijn (geldig voor verschillende (deel)populaties en onder uiteenlopende omstandigheden). Een derde en laatste aspect heeft betrekking op de grenzen van de generaliseerbaarheid, dat wil zeggen de condities waaronder gevonden verbanden wel en niet opgaan.^{40,41} De onderzochte non-response-effecten blijken in de meeste gevallen consistent te zijn. Weliswaar varieert de grootte van de effecten van studie tot studie, de richting daarentegen is zeer consistent: het zijn meestal dezelfde subgroepen die vaker aan onderzoek meedoen. Tevens zijn de non-responsvertekeningen robuust: de onderzochte studies zijn zeer heterogeen, onder meer naar bestudeerde leeftijds categorie, onderwerp van studie, plaats en tijd van onderzoek, steekproefkader, meetmethode, wijze van uitvoering van het onderzoek en responspercentage. Onder al deze uiteenlopende condities en omstandigheden vinden we non-response-effecten die steeds weer in dezelfde richting wijzen. Op basis van consistentie en robuustheid van uitkomsten kan worden geconcludeerd dat in survey-onderzoek onder ouderen de kans op vertekeningen op de onderzochte drie dimensies groot geacht moet worden. Dit is in overeenstemming met uitkomsten van non-responsstudies onder andere bevolkingscategorieën.^{8,20-26} Onderzoek onder ouderen lijkt hierop geen uitzondering te vormen. Over het derde aspect van externe validiteit (condities en grenzen) kan helaas weinig worden gezegd. Daartoe is het nodig om de verschillende condities waaronder onderzoek wordt verricht, systematisch te analyseren. Dit is een nog grotendeels onontgonnen terrein van studie. Meer aandacht hiervoor is echter dringend gewenst.⁴² Om de grenzen van de generaliseerbaarheid vast te stellen zijn multivariate analyses noodzakelijk (zie onder voor een verdere uiteenzetting hierover).

Welke betekenis dienen we nu te hechten aan de gevonden verschillen tussen respondenten en non-respondenten? Er zijn systematische verschillen, maar de analyse van de effectgroottes wijst uit dat de verschillen hooguit matig groot zijn. Houdt dit in dat wij deze vertekeningen kunnen negeren, of dienen we er wel degelijk rekening mee te houden? Om de gevolgen van non-responsverschillen te beoordelen moeten we zowel rekening houden met het responspercentage als met de verschillen tussen respondenten en non-respondenten (effectgrootte). Naarmate het responspercentage hoger is en de verschillen tussen respondenten en non-respondenten kleiner zijn, zijn de gevolgen geringer. Aan de absolute grootte van effecten kunnen dus niet zonder meer conclusies worden verbonden; we moeten rekening houden met de responshoogte. In dit verband is een onderzoek van Jones voor ons van belang.⁴³ Jones toont aan dat reeds bij relatief kleine verschillen tussen respondenten en non-respondenten ($ES = 0,25$ standaarddeviatie) een respons van 90 procent nodig is voor een betrouwbare meting van gemiddelden en dus ook voor een betrouwbare schatting van de prevalentie van allerlei medische aandoeningen in een populatie. Dit betekent dat de hier gevonden effectgroottes bij de non-respons, gevoegd bij de (onbekende) vertekeningen als gevolg van onderdekking, doen vermoeden dat onnauwkeurige prevalentieschattingen eerder regel dan uitzondering zijn. Door de al decennia lange daling van responsniveaus is dat probleem alleen maar groter geworden. Dalende responsniveaus maken het bovendien moeilijk om trends in de tijd vast te stellen, omdat de kans groot is dat recent onderzoek sterker is vertekend dan ouder onderzoek. Hierdoor ontstaat een onjuist beeld van ontwikkelingen in de tijd, zoals wordt bevestigd door onderzoeken naar veranderingen in de prevalentie van hoge bloeddruk en naar ontwikkelingen in politieke belangstelling.^{13,44}

In onderzoek naar non-respons wordt meestal alleen gekeken naar vertekeningen in afzonderlijke variabelen (univariaat). Dat geldt ook voor de hier bestudeerde studies. Onderzoek naar het effect van meerdere variabelen op non-respons tegelijk (multivariaat) is evenwel dringend gewenst. Om te beginnen kan men zo het 'netto non-response-effect' van afzonderlijke factoren vaststellen (door introductie van controlevariabelen of co-varianten). Overigens is het wel zaak om bij het gebruik van co-varianten rekening te houden met de relaties tussen de betrokken variabelen. Zeker bij een factor als SES is dit van belang.⁴⁵ Daarnaast kan men de gezamenlijke werking van variabelen analyseren (bij voorbeeld interactie-effecten). In de aanvullende, heuristische analyses van 'afwijkende' gevallen is reeds enkele malen gewezen op een mogelijke gezamenlijke werking van factoren, bij voorbeeld woonsituatie en gezondheid. Aanvullend wil ik op één factor nog iets verder ingaan, namelijk

leeftijd. Het is denkbaar dat het non-responspatroon onder de alleroudsten verschilt van dat van jongere ouderen als gevolg van homogenisering van oudere leeftijdscategorieën op de onderzochte dimensies, bij voorbeeld SES. Omdat mensen met een lagere SES een hogere sterftekans hebben dan mensen met een hogere SES, is de kans groot dat bij het ouder worden homogenisering op SES plaatsvindt.^{46,47} Hierdoor zou bij het ouder worden de betekenis van SES in de non-respons kunnen afnemen (en gezondheid bij voorbeeld belangrijker kunnen worden). Helaas worden dit soort samenhangen zelden bestudeerd. Het voorgaande maakt echter genoegzaam duidelijk dat multivariate analyses ons inzicht in non-responsvertekeningen (en de grenzen en condities waaronder ze optreden, zie boven: het derde aspect van externe validiteit) aanmerkelijk kunnen vergroten.

Niet alleen univariate of multivariate verdelingen en prevalentieschattingen kunnen vertekend zijn, door de selectieve uitval kunnen verbanden tussen variabelen eveneens vertekend raken. Dit klemt te meer daar in onderzoek naar ouderen SES, maatschappelijke integratie en gezondheid regelmatig met elkaar in verband worden gebracht. En zoals we in het voorgaande hebben kunnen zien, doen zich bij alle drie factoren vertekeningen voor. Er wordt nogal eens aangenomen dat verbanden tussen variabelen 'robuust' zijn, ook als er sprake is van selectieve uitval. Er is betrekkelijk weinig onderzoek gedaan naar de effecten van non-respons op verbanden, maar duidelijk is dat verbanden wel degelijk ernstig vertekend kunnen raken.^{27,32,48-51} Zo leidt een ondervertegenwoordiging van personen met een lage SES onder onderzoeksdeelnemers tot een onderschatting van SES-verschillen in gezondheid en mortaliteit in survey-onderzoek.^{32,50,51} Niet alleen afzonderlijke variabelen, maar ook verbanden tussen variabelen blijken dus gevoelig te zijn voor selectieve uitval.

Een oplossing voor de vertekeningen als gevolg van non-respons is niet eenvoudig te geven. Zelfs bij veldwerk waarbij gebruik wordt gemaakt van 'best practices' is tegenwoordig de uitval aanzienlijk.¹⁹ Verlies van onderzoekssubjecten is daarmee welhaast onvermijdelijk. Het is daarom van groot belang om – naast goed en zorgvuldig uitgevoerd veldwerk – als vast onderdeel van het onderzoek uit andere bronnen aanvullende gegevens te verzamelen over alle onderzoekseenheden uit de steekproef en om de gevolgen van de uitval door non-respons nauwkeurig te analyseren.

Een apart probleem is tot slot de uitsluiting van bepaalde subcategorieën ouderen door onderzoekers. Vaak zal dat onvermijdelijk zijn, maar het is niet juist, zoals vaak gebeurt, om deze personen vervolgens geheel te negeren, alsof ze geen deel uitmaken van de onderzochte populatie. Personen die na het trekken van de steekproef zijn overleden, te zwak, of te dement om aan onderzoek mee te doen, maken ook deel uit van de onderzochte populatie. Uitsluiting leidt tot vertekeningen, reeds voordat het eigenlijke onderzoek is begonnen. Onderzoekers zullen zich meer rekenschap dienen te geven van de consequenties van uitsluiting voor de representativiteit van hun onderzoek. Aangezien aan deze uitval tot dusverre betrekkelijk weinig aandacht is besteed, is meer onderzoek hiernaar dringend gewenst. Ook zal onderzocht dienen te worden of en hoe selectie-effecten van onderdekking en non-respons elkaar versterken. Pas als we meer weten over beide uitvalbronnen, zijn we in staat om de gevolgen ervan volledig op hun waarde te beoordelen.

Dankbetuiging

Gaarne dank ik Martha Bijlholt, Geert Braam, Joop Ellemers en de anonieme beoordelaars van dit tijdschrift voor hun commentaar op een eerdere versie van dit artikel.

Literatuur

- 1 Baruch Y. Response rate in academic studies: A comparative analysis. *Human Relations* 1999; 52: 421-438.
- 2 Leeuw E de, Heer W de. Trends in household survey nonresponse. In: Groves R.M. et al red. *Survey Nonresponse*. New York: Wiley, 2002: 41-54.
- 3 Curtin R, Presser S, Singer E. Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Q* 2005; 69: 87-98.
- 4 Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: A survey of practice. *Am J Epidemiol* 2006; 163: 197-203.
- 5 Galea S, Tracy M. Participation rates in epidemiologic studies. *Ann Epidemiol* 2007; 17: 643-653.
- 6 Herzog AR, Rodgers WL. Age and response rates to interview sample surveys. *J Gerontol: Soc Sci* 1988; 43: S200-S205.
- 7 Kaldenberg DO, Koenig HF, Becker BW. Mail survey response rate patterns in a population of the elderly: Does response deteriorate with age? *Public Opinion Q* 1994; 58: 68-76.
- 8 Goor H van, Rispens S. A middle class image of society: A study of undercoverage and non-response bias in a telephone survey. *Quality & Quantity* 2004; 38: 35-49.
- 9 Groves RM. Nonresponse rates and non-

- response bias in household surveys. *Public Opinion Q* 2006; 70: 646-675.
- 10 Groves RM, Peytcheva E. The impact of non-respons rates on nonresponse bias: A meta-analysis. *Public Opinion Q* 2008; 72: 167-189.
 - 11 Thompson MG, Heller K, Rody CA. Recruitment challenges in studying late-life depression: Do community samples adequately represent depressed older adults? *Psychol Aging* 1994; 9: 121-125.
 - 12 Riedel-Heller SG, Busse A, Angermeyer MC. Are cognitively impaired individuals adequately represented in community surveys? Recruitment challenges and strategies to facilitate participation in community surveys of older adults: A review. *Eur J Epidemiol* 2001; 16: 827-835.
 - 13 Gorey KM, Trevisan M. Secular trends in the United States black/white hypertension prevalence ratio: Potential impact of diminishing response rates. *Am J Epidemiol* 1998; 147: 95-99.
 - 14 Braam GPA. Zwitserleven: Niet voor alle ouderen; Over definities van armoede en de situatie onder ouderen. *Tijdschr. Gerontol. Geriatr.* 1999; 30: 102-104.
 - 15 Boer A de (red) Rapportage ouderen 2006: Veranderingen in de leefsituatie en levensloop. Den Haag: Sociaal en Cultureel Planbureau, 2006.
 - 16 Schnell R. Wer ist das Volk? Zur faktischen Grundgesamtheit bei "allgemeinen Bevölkerungsumfragen": Undercoverage, Schwererreichbare und Nichtbefragbare. *Kölner Zeitschr Soziologie Sozialpsychologie* 1991; 43: 106-137.
 - 17 Smith W, Mitchell P, Attebo K, Leeder S. Selection bias from sampling frames: Telephone directory and electoral roll compared with door-to-door population census; Results from the Blue Mountains Eye Study. *Aust N Z J Public Health* 1997; 21: 127-133.
 - 18 Keeter S. The impact of cell phone noncoverage bias on polling in the 2004 presidential election. *Public Opinion Q* 2006; 70: 88-98.
 - 19 Stoop IAL. The Hunt for the Last Respondent: Nonresponse in Sample Surveys. The Hague: Social and Cultural Planning Office of the Netherlands, 2005.
 - 20 Goyder J. The Silent Minority: Nonrespondents on Sample Surveys. Cambridge: Polity Press, 1987.
 - 21 Goyder J, Warriner K & Miller S. Evaluating socio-economic status (SES) bias in survey non-response. *J Official Statistics* 2002; 18: 1-11.
 - 22 Green KE. Sociodemographic factors and mail survey response. *Psychol & Marketing* 1996; 13: 171-184.
 - 23 Chretien J-P, Chu LK, Smith TC, Smith B, Ryan MAK & the Millennium Cohort Study Team. Demographic and occupational predictors of early response to a mailed invitation to enroll in a longitudinal health study. *BMC Med Res Methodol* 2007; 7: 6. <http://www.biomedcentral.com/147-2288/7/6>.
 - 24 Goldberg M, Chastang JF, Leclerc A et al. Socioeconomic, demographic, occupational, and health factors associated with participation in a long-term epidemiological survey: A prospective study of the French GAZEL cohort and its target population. *Am J Epidemiol* 2001; 154: 373-384.
 - 25 Hansen V, Jacobsen BK & Arnesen E. Prevalence of serious psychiatric morbidity in attenders and nonattenders to a health survey of a general population: the Tromsø Health Study. *Am J Epidemiol* 2001; 154: 891-894.
 - 26 Mazor KM, Clauser BE, Field T, Yood RA, Gurwitz JH. A demonstration of the impact of response bias on the results of patient satisfaction surveys. *HSR: Health Serv Res* 2002; 37: 1403-1417.
 - 27 Rosenthal R, Rosnow RL. *The Volunteer Subject*. New York: Wiley, 1975.
 - 28 Monsó E, Muñoz-Rino F, Izquierdo J et al. Occupational asthma in the community: Risk factors in a Western Mediterranean population. *Arch Environ Health* 1998; 53: 93-98.
 - 29 Van Kenhove P, Wijnen K, De Wulf K. The influence of topic involvement on mail-survey response behavior. *Psychol & Marketing* 2002; 19: 293-301.
 - 30 Groves RM, Presser S, Dipko S. The role of topic interest in survey participation decisions. *Public Opinion Q* 2004; 68: 2-31.
 - 31 Lipsey MW, Wilson DB. *Practical Meta-Analysis*. Thousand Oaks, CA: Sage, 2001.
 - 32 Richiardi L, Boffetta P, Merletti F. Analysis of nonresponse bias in a population-based case-control study on lung cancer. *J Clin Epidemiol* 2002; 55: 1033-1040.
 - 33 McNeely EA, Clements SD. Recruitment and retention of the older adult into research studies. *J Neurosci Nurs* 1994; 26: 57-61.
 - 34 Bootsma-Van der Wiel A, Exel E van, Craen AJM de et al. A high response rate is not essential to prevent selection bias: Results from the Leiden 85-plus study. *J Clin Epidemiol* 2002; 55: 1119-1125.
 - 35 Schneider G, Driesch G, Kruse A, Nehen HG, Heuft G. Veränderung einer Stichprobe ≥ 60 -jähriger Krankenhauspatienten nach 5 Jahren: Ergebnisse der ELDERMEN-Study. *Z Gerontol Geriat* 2004; 37: 136-144.
 - 36 Emberton M, Black N. Impact of non-response and of late-response by patients in a multi-centre surgical outcome audit. *Int J Qual Health Care* 1995; 7: 47-55.
 - 37 Gordon T, Moore FE, Shurtleff D, Dawber, TR. Some methodological problems in the long-term study of cardiovascular disease: Observations on the Framingham study. *J Chron Dis* 1959; 10: 186-206.
 - 38 Paul O, Lepper MH, Phelan WH et al. A longitudinal study of coronary heart disease. *Circulation* 1963; 28: 20-31.
 - 39 Ganguli M, Mendelsohn A, Lytle M, Dodge H. A follow-up comparison of study participants and refusers within a rural elderly population. *J Gerontol: Med Sci* 1998; 53: M455-470.
 - 40 McGrath JE, Brinberg D. External validity and

- the research process: A comment on the Calder/Lynch dialogue. *J Consumer Res* 1983; 10: 115-124.
- 41 Green LW, Glasgow RE. Evaluating the relevance, generalization, and applicability of research: Issues in external validation and translation methodology. *Eval Health Prof* 2006; 29: 126-153.
- 42 Steckler A, McLeroy KR. The importance of external validity. *Am J Public Health* 2008; 98: 9-10.
- 43 Jones J. The effects of non-response on statistical inference. *J Health & Social Policy* 1996; 8: 49-62.
- 44 Visscher G. Kiezersonderzoek op een dwaalspoor: De in politiek geïnteresseerde burger als selffulfilling prophecy. Den Haag: Sdu, 1995.
- 45 Kaufman JS, Kaufman S. Assessment of structured socioeconomic effects on health. *Epidemiol* 2001; 12: 157-167.
- 46 Alwin DF, Wray LA. A life-span developmental perspective on social status and health. *J Gerontol* 2005; 60B (Special Issue II): 7-14.
- 47 Willson AE, Shuey KM, Elder GH. Cumulative advantage processes as mechanisms of inequality in life course health. *Am J Sociol* 2007; 112: 1886-1924.
- 48 Goor H van, Stuiver B. A wave-analysis of distributional bias, substantive bias and data quality in a mail survey among Dutch municipalities. *Acta Politica* 1998; 33: 179-196.
- 49 Voogt RJJ, Kempen H van. Nonresponse bias and stimulus effects in the Dutch National Election Study. *Quality & Quantity* 2002; 36: 325-345.
- 50 Martikainen P, Valkonen T. Bias related to the exclusion of the economically inactive in studies on social class differences in mortality. *Int J Epidemiol* 1999; 28: 899-904.
- 51 Lorant V, Demarest S, Miermans P-J & Van Oyen H. Survey error in measuring socio-economic risk factors of health status: A comparison of a survey and a census. *Int J Epidemiol* 2007; 36: 1292-1299.