

De gevolgen van nonrespons en retrospectieve vertekeningen voor de uitkomsten van een onderzoek naar ziekteverzuim

H. van Goor en A.L. Verhage¹

1 Inleiding

Het terugdringen van ziekteverzuim en arbeidsongeschiktheid staat in Nederland al een aantal jaren hoog genoteerd op de politieke agenda. Vanaf het begin van de jaren negentig is er met dat doel een reeks van maatregelen getroffen, zoals onder andere de Wet Terugdringing Arbeidsongeschiktheidsvolume (1992) en de Wet Terugdringing Ziekteverzuim (1994). Om vast te stellen hoe deze maatregelen worden uitgevoerd en welke – bedoelde en onbedoelde – gevolgen ze hebben voor de betrokken werknemers, is evaluatieonderzoek noodzakelijk. Onderzoekers kunnen daarbij gebruikmaken van verschillende databronnen. Zo wordt wel gebruikgemaakt van administratieve gegevens die door bedrijven en instellingen zelf worden geregistreerd. Een belangrijke beperking van deze databron is echter het geringe aantal variabelen waarover gewoonlijk gegevens beschikbaar zijn. Enquêtes zijn dan een aantrekkelijk alternatief, omdat zij de mogelijkheid bieden de betrokkenen zelf rechtstreeks te vragen naar hun opvattingen en gedragingen en naar hun ervaringen met allerlei beleidsmaatregelen om ziekteverzuim en arbeidsongeschiktheid terug te dringen.

Tegenover de voordelen van enquêtes staan echter ook nadelen. Niet iedereen is bereid om aan onderzoek mee te doen, waardoor nonresponseeffecten op kunnen treden. Maar ook onder de respondenten bestaat het gevaar van vertekening, bijvoorbeeld omdat in de antwoorden op vragen naar gedrag, zoals naar frequentie en duur van het ziekteverzuim, retrospectieve vertekeningen kunnen voorkomen. Enquêtes zijn dus gevoelig voor verschillende foutenbronnen. In het algemeen worden deze foutenbronnen onafhankelijk van elkaar bestudeerd, maar er zijn aanwijzingen dat verschillende typen fouten met elkaar samenhangen. Zo wordt bijvoorbeeld vaak aangenomen dat pogingen om de respons te verhogen, leiden tot een verslechtering van de kwaliteit van de verkregen data (Groves 1987: S168).

In deze studie willen wij het effect van twee foutenbronnen – nonrespons en retrospectieve vertekeningen – op de verdeling van en verbanden tussen variabelen onderzoeken in een enquête onder de werknemers van Koninklijke Wegenbouw Stevin (KWS) Regio Noord. Wij zijn vooral geïnteresseerd in de samenhang van bei-

de foutenbronnen met het onderwerp van studie: ziekteverzuim. In welk opzicht verschillen respondenten en nonrespondenten van elkaar in frequentie en duur van hun ziekteverzuim? In hoeverre treden vervolgens onder respondenten retrospectieve vertekeningen op in de rapportage van hun ziekteverzuim? En welke gevolgen hebben ten slotte eventuele vertekeningen in het ziekteverzuim voor de verbanden tussen variabelen? Daarnaast zullen we ook aandacht besteden aan mogelijke vertekeningen in een aantal sociaal-demografische kenmerken.

Een gezamenlijke analyse van beide foutenbronnen maakt het niet alleen mogelijk om de *totale vertekening* in de verzuimgegevens vast te stellen, maar verschaft ons ook inzicht in de *relatieve betekenis* van beide foutenbronnen. Voorts kunnen we op deze wijze nagaan in hoeverre beide foutenbronnen eenzelfde of juist een verschillende werking hebben. Versterken ze elkaar, zodat een *cumulatie* van vertekeningen in één richting optreedt? Of treedt juist een *neutralisatie* van de vertekeningen op, omdat beide foutenbronnen in tegengestelde richting werken?

2 Literatuuroverzicht

Nonrespons. Het onderzoek naar vertekeningen ten gevolge van nonrespons is vooral gericht op sociaal-demografische kenmerken, zoals geslacht, leeftijd, opleiding en burgerlijke staat. Dat is begrijpelijk, want andere gegevens zijn vaak niet voorhanden. Tot dusverre zijn er echter weinig duidelijke en systematische verbanden tussen sociaal-demografische kenmerken en nonrespons vastgesteld. Hiervoor kan men verschillende redenen aanvoeren (vergelijk Goyder 1987: 80 e.v.). Eén reden, waarop wij hier de aandacht willen vestigen, betreft de *motivatie* om deel te nemen aan een enquête. Het is aannemelijk dat de bereidheid om aan onderzoek mee te doen niet primair een gevolg is van sociaal-demografische kenmerken, maar juist wordt beïnvloed door het *onderwerp* van studie.

Wat is er bekend over de samenhang tussen onderwerp van studie en nonrespons (zie ook Van Goor en Stuiver 1995: 390-391)? Om te beginnen is veel onderzoek gedaan naar het effect van de *belangstelling* voor het onderwerp van studie op de bereidheid om aan onderzoek mee te doen. Het blijkt dat personen die geïnteresseerd zijn in het onderwerp, vaker meedoen aan een enquête dan personen die niet geïnteresseerd zijn (Goyder 1987; Dillman 1991; Martin 1994).

Respondenten en nonrespondenten verschillen voorts in allerlei opzichten van elkaar in *houdingen en gedrag*. Zo zien we dat met name vroege respondenten meer uitgesproken en positieve opvattingen hebben dan late respondenten of nonrespondenten (Stinchcombe et al. 1981; Pearl en Fairley 1985; Green 1991; Dalecki et al. 1993).

Uit het beschikbare onderzoek komt naar voren dat degenen die niet aan onderzoek meedoen, vaak minder goede prestaties leveren dan degenen die wel meedoen. Dit is onder andere aangetoond voor schoolsucces van middelbare scholieren en stu-

denten (Ellis et al. 1970; Pavalko en Lutterman 1973), omzet van verkopers (Kirchner en Mousley 1963), contacten van boeren met bedrijfsvoorlichters (Fluttert 1980), effectiviteit van werknemers (Gannon et al. 1971), beleid inzake woonwagendstandplaatsen van gemeenten (Van Goor en Stuiver 1995). Van bijzonder belang voor ons is een onderzoek van Dijkstra (1981) naar ziekteverzuim in een drietal bedrijven: een scheepswerf, een grafisch bedrijf en een reinigingsdienst. Dijkstra vond dat nonrespondenten gemiddeld vaker verzuimden dan respondenten en dat het gemiddeld aantal verzuimde dagen per verzuimgeval en het gemiddeld aantal verzuimdagen per jaar onder nonrespondenten groter was dan onder respondenten. Uit al deze studies blijkt dat respondenten beter presteren en succesvoller zijn met betrekking tot het onderwerp van onderzoek dan nonrespondenten. De verklaring voor deze verschillen in deelname moet vermoedelijk worden gezocht in het feit dat degenen die slechter presteren of weinig succesvol zijn, het onderzoek bedreigend vinden en daardoor minder vaak meedoen (O'Neill 1979: 229; Van Goor en Stuiver 1995: 390-391).

Tot zover hebben we aandacht besteed aan de vraag in hoeverre nonrespons leidt tot vertekeningen in de verdeling van de 'inhoudelijke' ofwel 'afhankelijke' variabelen in een onderzoek. Minstens zo belangrijk is echter de vraag in hoeverre nonrespons leidt tot vertekeningen in de verbanden tussen variabelen. Helaas zijn studies op dit terrein zeer schaars. Niet altijd blijken vertekeningen op te treden (Kivlin 1965), maar de meeste onderzoeken wijzen uit dat nonrespons inderdaad tot vertekeningen leidt in de onderzochte verbanden (Hawkins 1975; Goudy 1976 en 1978; Van Goor en Stuiver 1995). Hoewel de effecten niet altijd even groot zijn, leiden ze vaak wel tot inhoudelijk andere conclusies. Dit betekent dus dat verschillen in uitkomsten tussen studies veroorzaakt kunnen zijn door verschillen in nonrespons (zie ook Van Goor 1996).

Retrospectieve vertekeningen. Onderzoek heeft aangetoond dat de nauwkeurigheid waarmee respondenten rapporteren over omstandigheden, gebeurtenissen en gedragingen in een meer of minder recent verleden, in het algemeen veel te wensen overlaat (Sudman en Bradburn 1974; Raadschelders en Van der Zouwen 1976; Bernard et al. 1984; Marquis et al. 1986; Pearson et al. 1992; Dex 1995). Bernard et al. (1984: 503) stellen zelfs dat door de bank genomen waarschijnlijk ongeveer de helft van de verschaft informatie onjuist is. Een uitzondering vormen vragen naar feitelijke gegevens, zoals persoonlijke bezittingen en personalia, die redelijk accuraat worden beantwoord. De meeste enquêtevragen blijken echter aanmerkelijk moeilijker te beantwoorden. Respondenten vermelden soms meer activiteiten dan gerechtvaardigd is (overrapportage), zoals bij stemgedrag (Cahalan 1968; Silver et al. 1986) of het lezen van boeken (Stefanowska 1979). In andere gevallen rapporteren ze juist minder gebeurtenissen (onderrapportage), bijvoorbeeld bij ontslag en werkeloosheid (Evans en Leighton 1995) of gezondheidsproblemen (Cannell et al. 1979).

Verschillende factoren zijn van invloed op de accuratesse van de informatie die

respondenten verschaffen (Burton en Blair 1991: 53-54). Twee daarvan zijn hier voor ons vooral van belang: 1. geheugeneffecten en 2. problemen van zelfpresentatie (zie uitvoeriger hierover Van Goor en Verhage MS).

1. Het beantwoorden van retrospectieve vragen doet een beroep op het geheugen van de respondent. Het zich herinneren van bepaalde gebeurtenissen kost vaak veel inspanning en moeite. Niet alleen moet de respondent in zijn geheugen zoeken naar relevante informatie, maar ook moet hij vervolgens op basis van de gevonden informatie trachten het correcte antwoord te bepalen. Individuen maken hierbij gebruik van verschillende cognitieve strategieën. Welke strategie wordt gehanteerd, is mede afhankelijk van het type informatie dat men zoekt (Bradburn et al. 1987; Blair en Burton 1987; Burton en Blair 1991; Pearson et al. 1992).

Eén zo'n cognitieve strategie – die vermoedelijk door respondenten ook wordt toegepast bij vragen naar ziekteverzuim – is het 'episode enumeration' of 'episode retrieval' model van Sudman en Bradburn (1974). Volgens dit model kunnen twee typen geheugenfouten optreden, namelijk omissies als gevolg van het geheel vergeten van relevante gebeurtenissen en dateringsfouten ('telescoping'). Omissies leiden tot onderrapportage; dateringsfouten kunnen tot onder- en overrapportage leiden. Onderzoekers zijn meestal niet in staat beide foutenbronnen te onderscheiden, zodat de door de respondenten gerapporteerde frequenties het netto-effect van beide foutenbronnen samen weergeven. Netto treedt gewoonlijk vooral onderrapportage op. Bij vragen over de gezondheid is dit zelfs zeer nadrukkelijk het geval (Cannell et al. 1979; Marquis et al. 1986; Loftus et al. 1992).

2. Bij zelfpresentatie gaat het om vertekeningen die de respondent bewust of onbewust aanbrengt in de verschaft informatie, vanwege de manier waarop hij zichzelf wenst te presenteren tegenover de onderzoeker. Bij sociaal gewenst gedrag leidt dit tot overrapportage. Indien wordt gevraagd naar sociaal ongewenst (afwijkend) gedrag, of naar gedragingen of gebeurtenissen die bij de respondent gevoelens van schaamte, verlegenheid of bedreiging oproepen, treedt daarentegen juist onderrapportage op. Aangezien vragen naar ziekteverzuim vermoedelijk eveneens met gevoelens van onzekerheid en bedreiging gepaard gaan (vergelijk Dijkstra 1981: 271-272), kan ook hier onderrapportage worden verwacht.

Naar de effecten van retrospectieve vertekeningen op verbanden tussen variabelen is voorzover ons bekend amper systematisch onderzoek gedaan. Sommige auteurs veronderstellen dat foutieve antwoorden vooral een verzwakkend effect op verbanden hebben (Weaver en Swanson 1974: 78-79; O'Muirheartaigh 1976: 110-111), maar Weiss (1968: 627-628) en De Graaf et al. (1996: 10-12) laten zien dat de associatie tussen variabelen zowel sterker als zwakker kan worden (zie voorts ook Marquis et al. 1986).

3 Hypothesen

Aangezien wij de retrospectieve vertekeningen in de antwoorden op de verzuimvragen elders reeds uitvoerig hebben geanalyseerd (Van Goor en Verhage MS), zullen wij hier alleen hypothesen formuleren over de effecten van nonrespons op de verdeling van de verzuimvariabelen. Daarnaast formuleren we hypothesen over het gecombineerde effect van beide foutenbronnen op verdelingen en verbanden.

In het voorgaande hebben we gezien dat belangstelling voor het onderwerp van studie en gevoelens van bedreiging die dat onderwerp oproept, van invloed zijn op nonrespons. Onze eerste twee hypothesen hebben hierop betrekking. Wij verwachten dat werknemers als gevolg van gevoelens van bedreiging minder vaak zullen meewerken aan een enquête, naarmate zij vaker en langduriger verzuimen. Eveneens verwachten wij dat degenen die niet met ziekteverzuim zijn geweest, niet erg geïnteresseerd zullen zijn in onze enquête en daardoor eveneens minder vaak zullen meedoen.

H1: *Bedreiging*. Nonrespondenten verzuimen gemiddeld vaker en langduriger dan respondenten.

H2: *Belangstelling*. Werknemers die gedurende een bepaalde periode niet verzuimd hebben, zullen minder vaak aan een enquête over ziekteverzuim deelnemen dan werknemers die wel hebben verzuimd.

Uit het literatuuroverzicht blijkt dat bij nonrespons en retrospectieve vertekeningen, althans gedeeltelijk, dezelfde factoren werkzaam zijn. Een onderwerp als ziekteverzuim kan bij degenen die bij deze onderwerpen 'negatief' uit de bus dreigen te komen, zowel leiden tot nonrespons, als tot onderrapportage. De onderrapportage wordt soms nog versterkt door geheugeneffecten. Het is dus aannemelijk dat nonrespons en retrospectieve vertekeningen een cumulatief effect hebben op de vertekeningen die optreden in verdelingen.

H3: *Cumulatie van vertekeningen in verdelingen*. Nonrespons en retrospectieve vertekeningen werken in dezelfde richting (cumulatief) en versterken dus elkaars effect: personen die vaker of langduriger verzuimen, doen ofwel niet mee aan onderzoek, ofwel geven op minder verzuimd te hebben dan werkelijk het geval is geweest.

Tot slot formuleren wij nog een hypothese over de gevolgen van nonrespons en retrospectieve vertekeningen voor de *verbanden* tussen variabelen. De mogelijkheden om beide effecten in combinatie te onderzoeken zijn helaas beperkt door het beperkte aantal variabelen in het administratieve bestand. Wij zullen de gevolgen van beide typen vertekeningen analyseren in de relaties tussen enkele demografische

kenmerken van de werknemers en ziekteverzuim en tussen de ziekteverzuimvariabelen onderling. Aangezien demografische kenmerken in het licht van theorieën over ziekteverzuim niet als de meest interessante variabelen beschouwd kunnen worden, hebben we daarnaast ook nog een analyse gemaakt van de gevolgen van retrospectieve vertekeningen voor de relaties tussen enkele aspecten van de arbeidsomstandigheden en ziekteverzuim. Wij volstaan met een algemene formulering van de hypothese, aangezien richting en grootte van de effecten van nonrespons en retrospectieve vertekeningen op de relaties tussen variabelen zich moeilijk nauwkeurig laten voorspellen.

H4: Effecten op verbanden. Nonrespons en retrospectieve vertekeningen leiden tot vertekeningen in de verbanden tussen variabelen.

4 Onderzoeksgegevens

Het onderzoek. De gegevens voor het hier gerapporteerde onderzoek zijn ontleend aan een evaluatieonderzoek van het ziekteverzuimbeheersingsplan (ZVP) dat in juli 1994 door de wegenbouwmaatschappij KWS Regio Noord is ingevoerd (Verhage 1996). Als onderdeel van het onderzoek is een schriftelijke enquête onder het personeel gehouden in september/oktober 1995, ruim een jaar na invoering van het plan. Omdat het onderzoek pas in 1995 van start is gegaan, had de enquête noodgedwongen een retrospectieve opzet.

Alle 205 personeelsleden die ten tijde van het afnemen van de enquête in dienst waren van KWS Regio Noord, zijn in het onderzoek betrokken. Eind september 1995 is het enquêteformulier in een enveloppe van het bedrijf verzonden naar het huisadres van de werknemers. Tevens werd een portvrije antwoortenveloppe meegezonden, geadresseerd aan het huisadres van de onderzoekster. Twee weken later werd een herinneringsbrief gestuurd aan alle werknemers die nog niet hadden gereageerd.

De enquêteformulieren waren voorzien van een identificatienummer. Dit maakte het mogelijk om gegevens van enquête en administratie te koppelen zonder dat het bedrijf inzage kreeg in de individuele gegevens die de werknemers aan de onderzoekster hadden verstrekt, zodat hun anonimiteit gegarandeerd bleef. Het gebruik van een identificatienummer was voorts ingegeven door de wens om een zo hoog mogelijke respons te bereiken bij maximaal één herinneringsbrief. Uit een onderzoek van McKee (1992) blijkt namelijk dat het gebruik van een identificatienummer, gekoppeld aan de aankondiging van een herinneringsoproep in de begeleidingsbrief, tot een duidelijke toename in respons leidt zonder dat de datakwaliteit daar noemenswaardig onder te lijden heeft.

Na de eerste toezending reageerden 136 werknemers (66,3%), na de herinneringsbrief nog eens 21, waarmee de totale respons op 157 werknemers, ofwel 76,6%, uitkwam.

Variabelen. Uit zowel administratief bestand als enquête beschikken we over gegevens over vijf sociaal-demografische kenmerken: geslacht, leeftijd (geboortejaar), aantal dienstjaren (jaar van indiensttreding), functiecategorie (CAO, POP, Overige)² en district waarin men werkzaam is (Groningen, Friesland, Drenthe, Regiokantoor).

Voor ziekteverzuim beschikken we uit beide bronnen over vier verzuimmaten:

- a. *verzuimfrequentie*: aantal ziekmeldingen in de periode voor, respectievelijk na invoering van het ziekteverzuimbeheersingsplan (ZVP);
 - b. *verzuimduur van de laatste verzuimperiode*: aantal verzuimdagen gedurende de laatste ziekmelding in de periode voor, respectievelijk na invoering van het ZVP.
- Daarnaast beschikken we uit de administratie nog over gegevens over: de *totale verzuimduur*: aantal verzuimdagen in de periode voor, respectievelijk na invoering van het ZVP. Bij de administratieve gegevens hebben we ten slotte ook: de *verzuimfrequentie* en de *totale verzuimduur* voor beide perioden samen (voor en na het ZVP) onderzocht. De verzuimmetingen hebben betrekking op de periode 1 januari 1992 - 30 juni 1994 (voor ZVP) en de periode 1 juli 1994 - 30 september 1995 (na ZVP).

Bij de analyse van de nonrespons is uitsluitend van administratieve gegevens gebruikgemaakt. Bij de analyse van retrospectieve vertekeningen zijn de gegevens over de respondenten uit de administratie en uit de enquête met elkaar vergeleken.

5 Resultaten

Nonrespons en verdelingen. Zoals reeds eerder is opgemerkt, gaat onze aandacht vooral uit naar de gevolgen van nonrespons voor de verdeling van de centrale inhoudelijke ('afhankelijke') variabelen: frequentie en duur van het ziekteverzuim. We hebben echter ook gekeken in hoeverre nonrespons effect heeft op de verdeling van een vijftal sociaal-demografische ('onafhankelijke') kenmerken. Over deze demografische kenmerken kunnen we kort zijn (tabel 1). Het blijkt namelijk dat het al dan niet meedoen aan het onderzoek met geen van deze vijf kenmerken samenhangt.

Bij ziekteverzuim treffen we een geheel ander beeld aan. Zoals tabel 2 laat zien, wordt hypothese 1 bevestigd door de uitkomsten. Voor alle acht metingen geldt dat nonrespondenten gemiddeld meer verzuimen dan respondenten. Hierbij dient echter wel te worden aangetekend dat de verschillen bij de verzuimfrequentie na invoering van het verzuimbeheersingsplan slechts miniem zijn.³ Onze resultaten stemmen overeen met de uitkomsten van het onderzoek van Dijkstra (1981), die eveneens vond dat nonrespondenten gemiddeld vaker en langduriger verzuimen dan respondenten. Het feit dat KWS Regio Noord een veel lager ziekteverzuim kent dan de bedrijven die Dijkstra indertijd onderzocht, maakt interessant genoeg geen verschil voor het al dan niet optreden van nonrespons.

Tabel 1: Percentage respons naar sociaal-demografische kenmerken

	<i>Geslacht</i>				χ^2	<i>p</i>
	<i>man</i>	<i>vrouw</i>				
% respons (<i>N</i>)	76,7 (193)	75,0 (12)			0,018	0,89
<i>Leeftijd (jaren)</i>						
	≤ 30	<i>31-40</i>	<i>41-50</i>	≥ 51		
% respons (<i>N</i>)	75,5 (53)	83,6 (55)	70,8 (65)	78,1 (32)	2,830	0,42
<i>Diensttijd (jaren)</i>						
	≤ 5	<i>6-10</i>	<i>11-20</i>	≥ 21		
% respons (<i>N</i>)	69,4 (49)	79,2 (53)	80,4 (56)	76,6 (47)	2,069	0,56
<i>Functiecategorie</i>						
	<i>CAO</i>	<i>POP</i>	<i>Overige</i>			
% respons (<i>N</i>)	75,8 (124)	78,7 (75)	66,7 (6)		0,552	0,76
<i>District</i>						
	<i>Groningen</i>	<i>Friesland</i>	<i>Drenthe</i>	<i>Regiokantoor</i>		
% respons (<i>N</i>)	76,1 (46)	77,1 (48)	75,0 (92)	84,2 (19)	0,758	0,86

De meest uitgesproken verschillen doen zich voor bij de totale verzuimduur: nonrespondenten verzuimen gemiddeld bijna twee keer zo lang als respondenten. Voorts valt op dat de verschillen in verzuim tussen respondenten en nonrespondenten het geringst zijn in de periode na invoering van het plan (verzuimfrequentie en verzuimduur van de laatste ziekteperiode). Het lijkt er dus veel op dat niet het recentste verzuim, maar vooral de totale verzuimduur over een langere periode effect heeft op de beslissing om al dan niet aan onderzoek mee te doen. Dit zou er op kunnen wijzen dat juist diegenen van wie de algehele gezondheidssituatie al geruime tijd minder goed is, zich sterker bedreigd voelen door een enquête en daarom minder vaak meedoen.

Onze tweede hypothese veronderstelt dat degenen die niet ziek zijn geweest, minder geïnteresseerd zijn in de enquête en daarom minder vaak aan onderzoek meedoen. Dit blijkt niet het geval te zijn. Degenen die niet hebben verzuimd, blijken in alle acht gevallen de enquête juist vaker te retourneren dan degenen die wel hebben verzuimd. Hypothese 2 dient dus verworpen te worden.

Tabel 2: Gemiddeld ziekteverzuim onder respondenten en nonrespondenten

<i>Ziekteverzuim voor, respectievelijk na invoering plan</i>	<i>Resp. N = 157</i>	<i>Nonresp. N = 48</i>	<i>t-test</i>	<i>p (eenz.)</i>
1 ziekmeldingsfrequentie voor ^{ab}	0,63	1,05	2,20	0,015
2 ziekmeldingsfrequentie na	0,69	0,71	0,13	0,448
3 ziekmeldingsfrequentie voor en na ^a	0,54	0,69	1,47	0,072
4 totale verzuimduur voor ^{ab}	7,79	13,57	1,89	0,030
5 totale verzuimduur na	5,24	10,10	1,80	0,037
6 totale verzuimduur voor en na ^a	5,78	10,51	2,34	0,010
7 duur verzuim laatste ziekmelding voor ^{bc}	1,41	1,74	1,52	0,066
8 duur verzuim laatste ziekmelding na ^c	1,31	1,54	0,87	0,192

Toelichting: Frequentie is gemeten in aantal ziekmeldingen, duur in werkdagen. 'Voor' = januari 1992 t/m juni 1994; 'na' = juli 1994 t/m september 1995; 'voor en na' = beide perioden samen.

a. Bij deze maten zijn gewogen gemiddelden berekend, waarbij is gecorrigeerd voor het aantal dienstjaren.

b. $N = 141$, resp. 43.

c. Bij 'duur laatste verzuim' is gebruik gemaakt van gecategoriseerde gegevens.

Factoren als belangstelling voor het onderwerp van studie en gevoelens van bedreiging kunnen uiteraard tegelijkertijd werkzaam zijn. Interessant is nu dat er aanwijzingen zijn dat belangstelling en bedreiging bij onderwerpen waarbij de 'prestaties' van de betrokkenen worden gemeten, een tegengesteld effect op nonrespons hebben (Van Goor en Stuiver 1995).

Tabel 3: Percentage nonrespons naar verzuimfrequentie en totale verzuimduur voor beide perioden (voor en na invoering van het plan) gezamenlijk

<i>Verzuimfrequentie voor en na ZVP</i>			
	<i>0</i>	<i>1-2</i>	<i>≥ 3</i>
% nonrespons	21,7	20,8	27,9
(N)	(60)	(77)	(68)
<i>Totale verzuimduur voor en na ZVP (in werkdagen)</i>			
	<i>0</i>	<i>1-18</i>	<i>≥ 19</i>
% nonrespons	21,7	16,7	31,5
(N)	(60)	(72)	(73)

In dit onderzoek zou dit betekenen dat nonrespons het vaakst voorkomt onder degenen die zich in het geheel niet ziek hebben gemeld (geen verzuim) of die juist vaak en/of lang hebben verzuimd (ernstig verzuim), terwijl degenen die relatief weinig hebben verzuimd (licht verzuim), het vaakst de enquête terug zouden sturen. Aan gezien hypothese 2 verworpen moest worden, hoeft het geen verbazing te wekken dat we dit 'combinatie-effect' niet algemeen aantreffen. Slechts twee verzuimmaten – verzuimfrequentie voor en na ZVP gezamenlijk en totale verzuimduur voor en na ZVP gezamenlijk – beantwoorden gedeeltelijk aan het verwachte combinatiepatroon (tabel 3). We zien hier een J-vormig patroon: de nonrespons is iets hoger onder de categorie geen verzuim dan onder de categorie licht verzuim; de nonrespons is duidelijk het hoogst onder de categorie ernstig verzuim. Deze uitkomsten zijn interessant, omdat hieruit blijkt dat degenen die nooit hebben verzuimd, gedurende de gehele onderzochte periode van drie jaar en negen maanden, relatief wat minder vaak meedoen aan het onderzoek. Van hen mag worden aangenomen dat ze een goede gezondheid hebben en dat het onderwerp van studie voor hen dus inderdaad weinig interessant is. Voorts wijzen de uitkomsten er op dat de factor bedreiging van meer betekenis is dan de factor interesse: de nonrespons is verreweg het hoogst onder degenen die frequent en/of langdurig verzuimen.

Retrospectieve vertekeningen en verdelingen. Voor het vaststellen van retrospectieve vertekeningen in de antwoorden van de respondenten gebruiken wij de administratieve gegevens, zoals die door KWS Regio Noord zijn geregistreerd, als maatstaf. Om de mate van overeenstemming tussen gegevens uit administratief bestand en enquête te bepalen hebben we gebruikgemaakt van het percentage overeenstemming tussen beide databronnen, ofwel het *percentage correcte antwoorden* en de overeenstemmingsmaat Cohen's *Kappa*. Kappa is aanvankelijk ontwikkeld om de mate van overeenstemming vast te stellen voor nominale data; latere uitbreidingen maken het echter ook mogelijk om gegevens op ordinaal meetniveau te analyseren (Popping 1992). In dit onderzoek is gebruikgemaakt van de partieel lineair gewogen Kappa, waarbij de berekeningen zijn uitgevoerd met behulp van het computerprogramma Agree 6.0 (Popping 1995). Als $Kappa \geq 0,81$, dan is de mate van overeenstemming dusdanig hoog dat enquêtegegevens en administratieve gegevens uitwisselbaar worden geacht. Bij een Kappa tussen 0,41 en 0,80 spreekt men van matige tot aanzienlijke overeenstemming; beneden 0,40 is de overeenstemming (zeer) zwak (vergelijk Popping 1984: 27).⁴

Bij de vijf door ons onderzochte sociaal-demografische kenmerken mogen we verwachten dat respondenten deze accuraat weergeven (vergelijk Sudman en Bradburn 1974; Raadschelders en Van der Zouwen 1976; Dex 1995). Zoals tabel 4 laat zien, is de overeenstemming tussen administratie en enquête bij de variabelen geslacht en leeftijd perfect.⁵ Bij de variabelen aantal dienstjaren en functiecategorie is de mate van overeenstemming iets minder, maar nog steeds heel hoog. De overeen-

stemming is bij het kenmerk district het laagst. De oorzaak van deze relatief lage overeenstemming moet vermoedelijk voor een groot deel worden gezocht in definitieverschillen tussen administratie en werknemers en niet zozeer in fouten in de registratie.⁶ Wij kunnen dus concluderen dat in de vragen naar sociaal-demografische kenmerken geen of slechts geringe vertekeningen optreden. Omgekeerd kunnen deze uitkomsten natuurlijk ook worden beschouwd als een aanwijzing dat de administratieve gegevens accuraat zijn! Hoewel hiermee niet direct iets is gezegd over de validiteit van de administratieve verzuimgegevens, geeft het toch een zeker vertrouwen dat ook deze gegevens valide zijn.

Tabel 4: Mate van overeenstemming tussen administratieve gegevens en enquête voor vijf demografische kenmerken

Kenmerk	Kappa	Percentage	N
		overeenstemming	
geslacht	1,00	100	156
leeftijd	1,00	100	151
functiecategorie	0,96	98	157
dienstjaren	0,94	95	155
district	-*	86	156

Toelichting: N is het aantal gevallen waarover overeenstemming is berekend. $N_{\max} = 157$. De oorzaak van een kleinere N is gelegen in ontbrekende antwoorden in de enquête.

* Indien de te vergelijken variabelen verschillen in antwoordcategorieën kan Kappa niet worden berekend (zie verder de tekst).

Bezien we nu de mate van overeenstemming bij de vier verzuimmaten waarvoor administratie en enquête kunnen worden vergeleken, dan blijkt deze slechts matig te zijn (tabel 5: kolommen 1 en 2). Deze bevindingen bevestigen het algemene beeld uit de onderzoeksliteratuur van een matige validiteit van retrospectieve enquêtegegevens over gezondheidsproblemen.

Voorts blijkt uit tabel 5 (kolommen 3 en 4) dat onderrapportage vaker voorkomt dan overrapportage: respondenten zijn eerder geneigd een geringere verzuimfrequentie of een kortere laatste verzuimperiode te rapporteren dan een hogere frequentie, respectievelijk langere laatste verzuimperiode. Er treedt hierbij overigens wel een verschil op tussen de perioden voor en na invoering van het ZVP. In de oudste periode treedt veel vaker overrapportage op dan in de recentste periode. Zoals wij elders hebben betoogd, is de grotere overrapportage in de periode voor het ZVP vermoedelijk het gevolg van dateringsfouten (Van Goor en Verhage MS).

Tabel 5: Mate van overeenstemming tussen administratieve gegevens en enquête voor vier verzuimmaten

Verzuimmaat	Kappa	Overeen- stemming (%)	Over- rapportage (%)	Onder- rapportage (%)	Overige verschillen (%)
	(1)	(2)	(3)	(4)	(5)
ZMFREQ _{voor ZVP} (N = 156)	0,61	60	12	25	3
ZMFREQ _{na ZVP} (N = 157)	0,58	78	4	18	-
Duur laatste VP _{voor ZVP} (N = 156)	0,58	53	17	27	3
Duur laatste VP _{na ZVP} (N = 149)	0,51	68	6	26	-

Toelichting: Bij het vaststellen van de mate van overeenstemming, respectievelijk van de mate van onder- en overrapportage zijn de administratieve gegevens als uitgangspunt gehanteerd. De kolommen (2), (3), (4) en (5) zijn samen steeds 100%. In de kolom 'Overige verschillen' zijn die gevallen opgenomen waarin '0' in plaats van 'nvt' wordt opgegeven, of 'nvt' in plaats van '0'. In deze gevallen kan niet van onder- of overrapportage worden gesproken.

Het gecombineerd effect van nonrespons en retrospectieve vertekeningen op de verdeling van de verzuimvariabelen. Op basis van de tot zover gepresenteerde gegevens kunnen we nu ook een uitspraak doen over hypothese 3. Nonrespons en retrospectieve vertekeningen blijken inderdaad een cumulatief effect te hebben op de verdeling van de verzuimvariabelen (tabellen 2 en 5). Werknemers die vaker of langer verzuimen, werken minder vaak aan onderzoek mee. Doen ze wel mee, dan rapporteren ze minder of korter verzuim dan juist is.⁷ Zo neemt bijvoorbeeld het percentage personen dat niet heeft verzuimd, voor het ZVP toe van 31% via 34% naar 37% en na het ZVP van 55% via 56% naar 66%. Alles samengenomen blijken nonrespons en retrospectieve vertekeningen elkaar dus te versterken. Gaan we ten slotte na hoe groot de bijdrage van elk van beide foutenbronnen afzonderlijk aan de vertekeningen is, dan kunnen we concluderen dat retrospectieve vertekeningen een groter effect hebben op de verdeling van onze verzuimvariabelen dan de nonrespons.

De effecten van nonrespons en retrospectieve vertekeningen op de verbanden (correlatiecoëfficiënten). Om vast te stellen in hoeverre nonrespons en retrospectieve vertekeningen van invloed zijn op de relaties tussen onafhankelijke en afhankelijke variabelen, zijn correlatiecoëfficiënten berekend voor de verschillende databestanden. Tabel 6 bevat de resultaten.

Tabel 6: Productmoment-correlaties tussen sociaal-demografische kenmerken, arbeidsomstandigheden en ziekteverzuim, berekend voor alle werknemers/administratieve gegevens (A/A), respondenten/administratieve gegevens (R/A) en respondenten/enquêtegegevens (R/E)

	ZMF _v			ZMF _n			DZM _v			DZM _n		
	A/A	R/A	R/E	A/A	R/A	R/E	A/A	R/A	R/E	A/A	R/A	R/E
leeftijd	.118	.075	.173*	.016	.032	.085	.100	.133	.038	-.009	.028	.090
dienstjaren	.092	.049	.112	.038	.059	.069	.067	.098	-.017	.014	.055	.031
functie	-.071	-.060	-.126	-.143*	-.136*	.011	-.169*	-.158*	-.081	-.230*	-.219*	-.033
zwaar ^a	-	-.142*	-.064	-	-.054	-.168*	-	-.129	-.189*	-	.185*	-.167*
aankunnen ^a	-	-.011	-.110	-	-.083	-.112	-	.035	-.109	-	-.167*	-.133
spanning ^a	-	-.188*	-.161*	-	-.184*	-.208*	-	-.202*	-.286*	-	-.137*	-.267*
ZMF _n	.331*	.286*	.509*	-	-	-	-	-	-	-	-	-
DZM _v	.668*	.707*	.692*	.188*	.138	.294*	-	-	-	-	-	-
DZM _n	.239*	.209*	.491*	.859*	.856*	.847*	.188*	.124	.393*	-	-	-

Toelichting: Leeftijd = geboortjaar in jaren; dienstjaren = jaar van indiensttreding in jaren; functiecategorie = CAO - POP; zwaar = werk zwaar?; aankunnen = werk niet aankunnen?; spanning = spanningen met collega's?; ZMF_v = ziekmeldingsfrequentie voor ZVP; ZMF_n = idem na ZVP; DZM_v = duur laatste ziekmelding voor ZVP; DZM_n = idem na ZVP.

^a Bij werkomstandigheden beschikken wij alleen over informatie over de respondenten.

* $p \leq 0,05$ (eenzijdig).

Zoals uit de cijfers duidelijk wordt, vertonen de correlatiecoëfficiënten meer of minder grote fluctuaties, als we de verschillende berekeningen voor alle werknemers, respectievelijk alleen de respondenten met elkaar vergelijken. Eén manier om de verschillen te analyseren is door ons af te vragen wat de consequenties van deze fluctuaties zijn voor conventionele nulhypothese toetsing ($\alpha \leq 0,05$).

Van de in totaal dertig geanalyseerde verbanden leidt significantietoetsing in tien gevallen tot tegenstrijdige uitspraken. In deze gevallen blijkt de 'keuze' van het te analyseren databestand inderdaad van invloed te zijn op de uitkomsten. Acht van de tien foutieve uitkomsten treffen we aan bij de correlaties die zijn berekend op basis van de enquêtegegevens (R/E). De correlaties die op basis van de enquête zijn berekend, leveren dus het vaakst onjuiste resultaten op. Wat hierbij opvalt, is dat verbanden onder invloed van retrospectieve vertekeningen niet altijd zwakker worden, zoals sommige auteurs wel hebben gesteld, maar soms ook sterker. Interessant in dit verband is, dat met name de samenhang tussen de verzuimmaten voor en na het ZVP in de enquête veel groter is dan in het administratief bestand. Deze uitkomst wijst er ons inziens op dat respondenten bij het beantwoorden van de verzuimvragen hun

antwoorden tot op zekere hoogte 'uniformeren', waardoor de samenhang tussen het verzuim in beide perioden ten onrechte toeneemt.

Nu is conventionele nulhypothese-toetsing al zeer lang een omstreden zaak (zie voor een recente kritiek Cohen 1994). Ook voor ons onderzoek is het de vraag of significantietoetsing wel de meest aangewezen analysetechniek is om de verschillen tussen de uitkomsten van de verschillende analyses te vergelijken. Bij significantietoetsing nemen we immers een dichotome beslissing: de uitkomsten liggen boven, dan wel onder het gekozen significantieniveau, maar een dergelijke benadering zegt weinig over de grootte van de verschillen tussen de correlatiecoëfficiënten. Desondanks hebben wij de bovenstaande significantietoetsing uitgevoerd, omdat aldus duidelijk wordt welke gevaren conventionele hypothese-toetsing met zich meebrengt, als men zich onvoldoende rekenschap geeft van de kwaliteit van de gebruikte databestanden.

Om de mate van *overeenstemming* tussen de correlatiecoëfficiënten uit de verschillende berekeningen vast te stellen, is een andere benadering nodig. We kunnen de mate van overeenstemming systematisch onderzoeken door de *relatieve vertekening* (*RV*) ten gevolge van nonrespons en retrospectieve vertekeningen te analyseren. *RV* kunnen we vaststellen, door de *r*-waarden om te zetten in *Z*-scores (zie Van Goor en Stuiver MS).

$$RV = \frac{|Z_0 - Z_v|}{s}$$

Z_0 = de *Z*-score van de waargenomen correlatiecoëfficiënt;

Z_v = de *Z*-score van de correlatiecoëfficiënt die voor valideringsdoeleinden wordt gebruikt;

s = de standaarddeviatie van de correlatiecoëfficiënt die voor validering wordt gebruikt.

$$s = \frac{1}{\sqrt{N-3}}$$

RV geeft de afwijking weer in standaarddeviatie-eenheden ten opzichte van de correlatiecoëfficiënt die als (validerings)criterium wordt gebruikt.⁹ We kunnen *RV* op drie verschillende manieren berekenen: voor nonrespons (RV_n), voor retrospectieve vertekeningen (RV_r) en voor het totaleffect: nonrespons en retrospectieve vertekeningen samen (RV_{n+r}). In het eerste en het derde geval (RV_n en RV_{n+r}) nemen we de correlaties, berekend voor alle werknemers op basis van de administratieve gegevens, als valideringscriterium en in het tweede geval (RV_r) de correlaties berekend voor alle respondenten op basis van de administratieve gegevens.

RV kan zowel positieve als negatieve waarden aannemen, maar wij zullen hier verder geen aandacht besteden aan de richting van de vertekeningen. *RV* is een aan-

trekkelijke maat, aangezien de uitkomsten geïnterpreteerd kunnen worden in termen van betrouwbaarheidsintervallen. Wel dienen we er ons rekenschap van te geven dat *RV* door ons hier als maat voor *overeenstemming* (tussen waargenomen en 'ware' correlatiecoëfficiënt) wordt gebruikt.

Bij de interpretatie van *RV* is de cruciale vraag tussen welke grenzen *RV* mag variëren om (nog) van aanvaardbare, dat wil zeggen niet-vertekende, correlatiecoëfficiënten te kunnen spreken. Deze vraag kan ten minste op twee verschillende manieren worden beantwoord (Van Goor en Stuiver MS). In de eerste plaats kunnen we zelf vooraf de grenzen aangeven van het gebied waarbinnen aanvaardbare (niet-vertekende) uitkomsten liggen. In de tweede plaats kunnen we de normaalverdeling als uitgangspunt nemen en nagaan of de waargenomen verdeling van *RV* overeenkomt met de verwachte (normaal)verdeling. Wij zullen beide benaderingen achtereenvolgens behandelen.

- 1 De keuze van een bepaald gebied waarbinnen *RV* dient te vallen, is afhankelijk van de steekproefomvang en van de (verwachte) sterkte van de te onderzoeken verbanden. Nu zijn relaties in sociologisch onderzoek vaak betrekkelijk zwak of hooguit matig sterk. Vandaar dat wij de grenzen niet te breed dienen te trekken om geen onjuiste conclusies te trekken. In dit onderzoek zullen wij de volgende indeling gebruiken:

<i>RV</i> (in <i>s</i>)	betrouwbaarheids-interval	mate van overeenstemming
$\leq 0,675$	$< 50\%$	hoog
1,000 - 0,676	68 - 50%	aanzienlijk
1,650 - 1,001	90 - 68%	matig
1,960 - 1,651	95 - 90%	laag
$\geq 1,961$	$> 95\%$	zeer laag

Indien $RV \leq 0,675s$, zullen wij spreken van een *aanvaardbaar* niveau van afwijkingen, dat wil zeggen dat we de gevonden correlatiecoëfficiënten als *niet-vertekend* zullen beschouwen. Indien we bereid zijn een wat grotere fluctuatie in waarden te accepteren, dan zouden we bijvoorbeeld een interval van 1,000s kunnen nemen. Een breder interval lijkt ons onwenselijk, vanwege het gevaar om onjuiste conclusies te trekken.

Nonresponseeffecten en het gecombineerd effect van nonrespons en retrospectieve vertekeningen kunnen we alleen onderzoeken voor de 18 relaties tussen de sociaal-demografische kenmerken en de ziekteverzuimvariabelen; bij de retrospectieve vertekeningen kunnen we ook de 12 relaties tussen arbeidsomstandigheden en ziekteverzuim in de analyse betrekken.

Uit tabel 7 blijkt, dat 14 van de 18 waarden van RV_n binnen het 50% betrouwbaarheidsinterval liggen; alle 18 liggen binnen het 68%-interval. Dat betekent dat de non-respons (in dit onderzoek 23%) niet of slechts in geringe mate leidt tot vertekende uitkomsten. RV_r laat een geheel ander beeld zien: slechts negen van de 30 waarden zijn $\leq 0,675s$; 17 zijn $\leq 1.000s$. Zeven vallen zelfs buiten het 90% betrouwbaarheidsinterval. De onnauwkeurigheid in de antwoorden op de enquêtevragen over ziekteverzuim leiden dus in een aantal gevallen tot onaanvaardbaar grote vertekeningen. Bekijken we ten slotte het gecombineerd effect van nonrespons en retrospectieve vertekeningen (RV_{n+r}), dan is het totaalresultaat nog iets ongunstiger. We kunnen op basis van deze uitkomsten concluderen, dat in dit onderzoek de gevolgen van retrospectieve vertekeningen beduidend groter zijn dan die van nonrespons. De grootste afwijkingen doen zich voor bij de relaties tussen functie en beide verzuimaten na het ZVP, tussen spanningen met collega's en de duur van de laatste verzuimperiode na het ZVP en bij de vier relaties tussen het verzuim voor en na het ZVP. Het zijn dus met name relaties met de verzuimaten *in de periode na het ZVP* die sterk worden vertekend.

Tabel 7: Relatieve vertekening (RV) ten gevolge van nonrespons, retrospectieve vertekeningen en beide samen

	Afwijkingen van de criteriumscores, in standaard-deviatie eenheden				
	> 1,96	1,96-1,65	1,65-1,00	1,00-0,675	< 0,675
1. Nonrespons (RV_n)	0	0	0	4	14
2. Retrospectieve vertekeningen (RV_r)	4	3	6	8	9
3. Totaaleffect: nonrespons en retrospectieve vertekeningen samen (RV_{n+r})	5	0	4	4	5

2 Als gevolg van allerlei toevallige meetfouten zullen er altijd verschillen optreden tussen correlatiecoëfficiënten berekend voor verschillende databestanden. Als de meetfouten inderdaad toevallig zijn, mogen we aannemen dat de verschillen tussen de correlatiecoëfficiënten normaal verdeeld zijn. We kunnen dus de normaalverdeling als uitgangspunt nemen en nagaan of de waargenomen verdeling van RV , zoals weergegeven in tabel 7, overeenkomt met de verwachte (normaal)verdeling. Gelet op het aantal gevallen is deze berekening alleen zinvol uit te voeren op de verschillen tussen correlatiecoëfficiënten ten gevolge van retrospectieve vertekeningen. De feitelijke verdeling blijkt hier inderdaad significant te verschillen van de verwachte normaalverdeling ($\chi^2 = 9,03$; $df = 3$ [de >95% en de 90-95%

intervallen zijn vanwege de verwachte kleine aantallen samengenomen]; $p < 0,05$). We kunnen dus concluderen dat de retrospectieve vertekeningen in de enquête ook leiden tot systematische vertekeningen in de verbanden tussen de variabelen.

Tabel 8: Een vergelijking van multiële regressieberekeningen voor administratieve gegevens respectievelijk enquête bij de verklaring van ziekteverzuim: ongestandaardiseerde effecten en tussen haakjes gestandaardiseerde coëfficiënten (alleen respondenten)

	<i>Administratieve gegevens</i>		<i>Enquête</i>	
	B	β	B	β
1. Afhankelijke variabele: Verzuimfrequentie na ZVP				
Leeftijd	0,005	(0,086)	0,003	(0,059)
Functie	-0,238	(-0,198)*	0,026	(0,026)
Werk zwaar?	0,064	(0,064)	-0,082	(-0,099)
Werk niet aankunnen?	-0,032	(-0,042)	0,003	(0,005)
Spanningen met collega's?	-0,086	(-0,107)	-0,032	(-0,047)
ZMF	0,170	(0,251)**	0,328	(0,495)***
(constante)	1,531**		0,891*	
r^2 (adj.)	0,093		0,248	
2. Afhankelijke variabele: Duur laatste verzuimperiode na ZVP				
Leeftijd	0,011	(0,089)	0,009	(0,092)
Functie	-0,669	(-0,274)**	-0,032	(-0,015)
Werk zwaar?	-0,142	(-0,072)	-0,121	(-0,072)
Werk niet aankunnen?	-0,213	(-0,139)	-0,039	(-0,029)
Spanningen met collega's?	-0,102	(-0,063)	-0,181	(-0,137)
DLM	0,036	(0,037)	0,283	(0,341)***
(constante)	3,876***		1,708	
r^2 (adj.)	0,080		0,162	

Toelichting: De variabelen zijn steeds ontleend aan de vermelde databestanden, uitgezonderd de variabelen met betrekking tot de arbeidsomstandigheden, die uit de enquête afkomstig zijn. *N* varieert tussen 134 en 140.

* $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$.

Retrospectieve vertekeningen en de verklaring van ziekteverzuim (regressieanalyse). Tot zover hebben we alleen aandacht besteed aan de bivariate relaties met ziekteverzuim. Stel echter dat we vooral geïnteresseerd zijn in de vraag hoe het ziekteverzuim in de periode na het ZVP het best kan worden verklaard uit de sociaal-demografische

kenmerken, de arbeidsomstandigheden en het ziekteverzuim in de periode voor het ZVP gezamenlijk. Welke consequenties heeft nu het gebruik van de ziekteverzuimgegevens uit het administratieve bestand dan wel uit de enquête voor de uitkomsten? Om te bepalen welke gevolgen deze 'keuze' heeft, hebben we gebruikgemaakt van multiële regressieanalyse. We zullen ons hier beperken tot een analyse van de retrospectieve vertekeningen, niet alleen omdat de nonresponseffecten gering zijn, maar ook omdat we alleen over gegevens over de arbeidsomstandigheden beschikken uit de enquête.

Zoals blijkt uit tabel 8, worden de uitkomsten van de analyses inderdaad beïnvloed door de keuze van het bestand. Gebruik van de verzuimgegevens uit de enquête leidt om te beginnen tot een veel hogere verklaarde variantie (r^2) dan gebruik van die uit het administratieve bestand! Maar ook de factoren die als de belangrijkste onafhankelijke variabelen naar voren komen, verschillen gedeeltelijk of geheel, afhankelijk van het gebruikte bestand. Zoals viel te verwachten, oefenen vooral de variabelen waarbij in de voorgaande analyse van RV de grootste verschillen in de relatie met het verzuim na het ZVP optraden – met name functie en ziekteverzuim voor het ZVP –, ook grote invloed uit op de uitkomsten van onze regressieanalyse. De retrospectieve vertekeningen in de verzuimvariabelen beïnvloeden dus in hoge mate de uitkomsten van onze analyses.

6 Samenvatting en discussie

In dit artikel hebben wij de gevolgen onderzocht van twee foutenbronnen in een enquête naar ziekteverzuim: nonrespons en retrospectieve vertekeningen. Welke effecten hebben deze foutenbronnen op de verdeling van en verbanden tussen variabelen? Onze aandacht ging hierbij vooral uit naar de samenhang van beide foutenbronnen met het onderwerp van studie: ziekteverzuim. Drie vragen stonden centraal. In welk opzicht verschillen respondenten en nonrespondenten van elkaar in frequentie en duur van hun ziekteverzuim? In hoeverre treden vervolgens onder respondenten retrospectieve vertekeningen op in de rapportage van hun ziekteverzuim? En welke gevolgen hebben ten slotte eventuele vertekeningen in het ziekteverzuim voor de verdelingen van en verbanden tussen variabelen?

Kort gezegd komen de uitkomsten er op neer dat inderdaad vertekeningen optreden, zowel in de verdeling van de ziekteverzuimvariabelen, als in de verbanden tussen een aantal variabelen en ziekteverzuim. Nonrespons en een retrospectieve vraagstelling dragen beide aan deze vertekeningen bij, waarbij het aandeel van de retrospectieve vertekeningen groter blijkt te zijn dan die van de nonrespons.

Zelfs bij een hoge respons, zoals in dit onderzoek (77%), blijkt nonrespons nog samen te hangen met ziekteverzuim. Degenen die vaker en/of langer verzuimen, sturen de vragenlijst minder vaak terug. Hypothese 1 is hiermee bevestigd. De verzuimduur blijkt hiërarchisch van grotere betekenis te zijn dan de verzuimfrequentie. Inte-

ressant is dat vooral degenen die over een langere periode gezien (in ons onderzoek drie jaar en negen maanden) een slechtere gezondheid hebben, minder vaak meedoen. Dit betekent dat als een vraag naar verzuim over een korte periode (een maand, een half jaar, of zelfs een jaar) niet of slechts weinig vertekend is, toch een selectie-effect aanwezig kan zijn, omdat met name degenen die over een langere periode gezien minder gezond zijn, minder vaak aan het onderzoek zullen meedoen. Er kan dus met andere woorden een 'verborgen' selectie-effect optreden.

Hypothese 2, waarin wij de verwachting uitspreken dat degenen die niet verzuimen, niet in het onderwerp geïnteresseerd zijn en daarom de vragenlijst minder vaak terugsturen, blijkt niet op te gaan. Wel blijkt uit een analyse van het verzuim over de hele onderzochte periode (voor en na het ZVP samen) dat degenen die in het geheel niet hebben verzuimd, inderdaad iets minder vaak meedoen dan degenen die slechts weinig hebben verzuimd; degenen die relatief veel hebben verzuimd doen evenwel verreweg het minst vaak mee.

Van groot belang voor de praktijk is de bevinding dat geen nonresponseeffecten optreden in de onderzochte vijf sociaal-demografische (onafhankelijke) variabelen, maar wel in de afhankelijke variabelen (ziekteverzuim). Deze uitkomsten ondersteunen nogmaals dat controle op nonresponsvertekeningen in onafhankelijke variabelen weinig zegt. Het gevaar is niet denkbeeldig dat de onderzoeker zich door een dergelijke controle ten onrechte gerustgesteld voelt. Immers, zoals ons onderzoek overduidelijk aantoonde, nonrespons blijkt vooral samen te hangen met het *onderwerp* van studie.

Ons onderzoek bevestigt de uitkomsten van het nonresponsonderzoek naar ziekteverzuim van Dijkstra (1981). Hoewel beide onderzoeken zijn uitgevoerd in perioden die in allerlei opzichten verschillen, zowel in verzuimniveau, maatschappelijke en politieke belangstelling voor het onderwerp als in de bestrijding van ziekteverzuim en arbeidsongeschiktheid, wijzen de uitkomsten toen en nu in dezelfde richting: personen die vaker of langer verzuimen, voelen zich bedreigd door een enquête en verlenen daarom minder vaak hun medewerking aan een dergelijk onderzoek.

In de vragen naar ziekteverzuim blijken retrospectieve vertekeningen zeer frequent voor te komen. Onze uitkomsten stemmen in dit opzicht overeen met die van Gray (1955) onder Britse ambtenaren en van O'Muircheartaigh (1976) onder de werknemers van een Iers industrieel bedrijf. De overeenstemming tussen administratieve gegevens en enquête is matig. Beide databronnen zijn zeker niet uitwisselbaar. De vertekeningen nemen vooral het karakter aan van onderrapportage: er worden minder ziekmeldingen gerapporteerd en er wordt een kortere duur van het verzuim opgegeven dan werkelijk het geval is. Daarentegen blijkt dat de respondenten wel accuraat rapporteren over hun sociaal-demografische kenmerken.

De vertekeningen ten gevolge van nonrespons en retrospectieve vertekeningen hebben inderdaad een cumulatief effect op de verdeling van de verzuimvariabelen, zoals in hypothese 3 werd verondersteld: samen leiden ze tot een aanzienlijke onder-

schatting van het werkelijke verzuim. De verdeling van de sociaal-demografische kenmerken wordt (uitgezonderd bij district) daarentegen niet of nauwelijks beïnvloed.

De vertekeningen in de verzuimvariabelen werken door in de verbanden tussen variabelen. Op basis van de analyse van de correlatiecoëfficiënten (conventionele nulhypothese-significantietoetsing en *RV*) en van de multiële regressie kunnen we concluderen dat in dit onderzoek nonrespons alleen slechts een betrekkelijk gering effect heeft op de verbanden tussen de onderzochte variabelen; het effect van retrospectieve vertekeningen is veel groter. Hypothese 4 gaat dus niet op voor de nonrespons, maar wel voor de retrospectieve vertekeningen. Het gevolg is dat de uitkomsten van de enquête in een (te) groot aantal gevallen niet valide blijken te zijn. Retrospectieve vertekeningen (soms versterkt door de vertekeningen ten gevolge van nonrespons) dragen dus in belangrijke mate bij aan de onzekerheid over en ambiguïteit in de empirische resultaten.

De uitkomsten van dit onderzoek roepen een weinig bemoedigend beeld op van de vertekeningen die nonrespons en retrospectieve vertekeningen samen in een enquête kunnen veroorzaken. Ons onderzoek staat in dit opzicht overigens niet alleen, het bevestigt het beeld dat uit de onderzoeksliteratuur naar voren komt. Een belangrijke vraag is uiteraard hoe deze fouten kunnen worden teruggedrongen. Dit is geen eenvoudige opgave, gelet op de praktische omstandigheden waaronder veel enquêtes worden afgenomen. Toch is over deze vraag wel iets te zeggen.

Om de vertekeningen ten gevolge van nonrespons terug te dringen staan in principe twee wegen open. Men kan er ofwel naar streven om de respons te maximaliseren, ofwel kan men achteraf, nadat de gegevens zijn verzameld, proberen om voor mogelijke fouten te corrigeren, bijvoorbeeld door herweging. De tweede weg blijkt moeilijk begaanbaar te zijn. Zonder informatie over de inhoudelijke variabelen uit een onderzoek (houdingen en gedrag van de respondenten) blijkt weging weinig zin te hebben (vergelijk Van Goor en Stuiver 1995). Aangezien men gewoonlijk niet over deze informatie beschikt, is deze oplossingsroute in het algemeen van weinig betekenis. Belangrijker is het om vooraf te streven naar een zo hoog mogelijke respons. Met een goed doordachte veldwerkstrategie blijkt men inderdaad heel ver te kunnen komen, maar een zekere uitval is bijna onvermijdelijk. In hoeverre de maximaal bereikbare responsniveaus in alle gevallen toereikend zullen zijn om vertekeningen voldoende tegen te gaan, is overigens nog altijd een open vraag (vergelijk Van Goor en Stuiver MS).

Met betrekking tot retrospectieve vertekeningen is op dit moment veel minder duidelijk hoe vertekeningen het best kunnen worden tegengegaan. Dit heeft naar onze mening te maken met het beperkte aantal echte valideringsonderzoeken dat tot nu toe is uitgevoerd (vergelijk Dex 1995). Hierdoor is nog veel onbekend over de precieze werking van de verschillende mechanismen die bij dit type vertekeningen een rol spelen. Weliswaar wordt er vanaf het begin van de jaren tachtig ook vanuit de

cognitieve psychologie in toenemende mate aandacht besteed aan geheugeneffecten in surveys (zie onder andere Tanur 1992; Schwarz en Sudman 1996), maar deze studies hebben tot dusverre nog maar weinig praktisch toepasbare inzichten opgeleverd. Daar komt nog bij dat in onderzoek waarin verbeteringen worden uitgetest om dit type vertekeningen tegen te gaan, te vaak gebruik wordt gemaakt van gebrekkig of niet gevalideerde gegevens.

Ons onderzoek maakt duidelijk dat het van groot belang is om gelijktijdig aandacht te besteden aan verschillende foutenbronnen in surveys. Nonrespons en retrospectieve vertekeningen blijken elkaars effecten immers te versterken, juist in variabelen die centraal staan in het onderzoek. Met een hoge respons alleen zijn we er dan ook nog niet. Retrospectieve vertekeningen blijken eveneens een grote rol te spelen. Zolang wij er niet in slagen deze, maar ook andere foutenbronnen op bevredigende wijze uit te schakelen, blijft er grote onzekerheid bestaan over de validiteit van de uitkomsten van empirisch onderzoek. Dit bemoeilijkt niet alleen de toetsing van wetenschappelijke theorieën, maar doet ook ernstig afbreuk aan het nut van sociologisch onderzoek voor praktische doeleinden.

Noten

1. De heer J. Bos, hoofd Personeel en Organisatie van Koninklijke Wegenbouw Stevin (KWS) Regio Noord, zijn wij erkentelijk voor het beschikbaar stellen van gegevens, op basis waarvan wij dit onderzoek konden uitvoeren. Roel Popping was ons behulpzaam bij het berekenen van Kappa aan de hand van het computerprogramma Agree 6.0. Gaarne danken wij hen beiden en voorts Martha M.C. Bijlholt, Geert P.A. Braam en Joop E. Ellemers voor hun commentaar op een eerdere versie van dit artikel.

2. Het onderscheid tussen CAO en POP bij functiecategorie valt ruwweg samen met het onderscheid naar productiepersoneel (CAO) en kantoorpersoneel en leidinggevend personeel in het veld, zoals uitvoerders (POP); onder Overig personeel valt tijdelijk personeel, zoals praticanten en stagiaires.

3. Het is mogelijk dat dit geringe verschil samenhangt met de geringe variatie in verzuimfrequentie na invoering van het ziekteverzuimbeheersingsplan: 55% van het personeel heeft zich in deze periode niet ziek gemeld, 41% 1 of 2 keer en slechts 9% heeft zich 3 of 4 keer ziek gemeld.

4. Voor de berekening van Kappa zijn bij de sociaal-demografische kenmerken leeftijd en aantal dienstjaren als volgt gehercodeerd: *leeftijd*: ≤ 30 ; 31-40, 41-50, ≥ 51 jaar; *aantal dienstjaren*: <2, 2-5, 6-10, 11-20, ≥ 21 jaar.

Bij de ziekteverzuimvariabelen zijn de antwoordcategorieën uit de enquête gebruikt: (1) *frequentie ziekteverzuim voor invoering ZVP*: n.v.t. (niet in dienst van KWS), 0, 1-2, 3-5, ≥ 6 ziekmeldingen; (2) *frequentie ziekteverzuim na invoering ZVP*: 0, 1-2, 3-5 ziekmeldingen; (3) *duur laatste verzuimperiode voor invoering ZVP*: n.v.t., 0, 1-2, 3-7, 8-42, ≥ 43 werkdagen; (4) *duur laatste verzuimperiode na invoering ZVP*: 0, 1-2, 3-7, 8-42, ≥ 43 werkdagen (de categorie 'momenteel ziek' is buiten beschouwing gelaten). Bij de berekening van Kappa is de categorie n.v.t. als nominaal beschouwd en 0 en hoger als ordinaal.

5. Aangezien de mate van overeenstemming wordt beïnvloed door het aantal en de breedte van de gehanteerde categorieën, hebben we bij leeftijd ook nog eens de mate van overeenstem-

ming vastgesteld op basis van het geboortjaar. Ook dan blijkt de overeenstemming 100% te zijn. Dit is opvallend hoog, daar in enkele andere studies voor leeftijd overeenstemmingspercentages van 83, 92 en 92 worden genoemd (Cahalan 1968: 610; Weaver en Swanson 1974: 73).

6. In totaal geven 16 respondenten, van wie er volgens de administratieve gegevens negen werkzaam zijn in één van de drie districten en zeven op het regiokantoor, spontaan als antwoord dat men werkzaam is in meer dan één district. Deze discrepantie tussen administratie en enquête is vermoedelijk het gevolg van een verschil tussen formele aanstelling en feitelijk werkterrein. In de praktijk blijken werknemers door de districten namelijk tijdelijk aan elkaar te worden uitgeleend. Ook zijn sommige werknemers in meer regio's werkzaam.

7. Bovendien neemt de overrapportage af en neemt de onderrapportage toe, naarmate men frequenter heeft verzuimd en naarmate de duur van de laatste verzuimperiode langer is (Van Goor en Verhage MS).

8. Uit de onderzoeksliteratuur (vergelijk Verhage 1996: 12-13) is bekend dat leeftijd (gecodeerd als geboortjaar) samenhangt met verzuim: jongere werknemers verzuimen vaker (verwachting: positieve correlaties), maar ouderen verzuimen langer, waardoor ouderen in totaal meer verzuimen (verwachting: negatieve correlaties). Omdat het aantal dienstjaren (gecodeerd als jaar van indiensttreding) sterk met leeftijd samenhangt ($r = 0,636$; $N = 205$), verwachten we dat het aantal dienstjaren op dezelfde wijze als leeftijd met verzuim samenhangt. Aangezien productiepersoneel veelal meer verzuimt dan kantoorpersoneel, nemen we aan dat het CAO-personeel meer verzuimt dan de POP-pers (verwachting: negatieve correlaties). Voorts nemen we aan dat de ziekteverzuimvariabelen onderling positief correleren. Ten slotte verwachten we dat bedreigende arbeidsomstandigheden leiden tot frequenter en langduriger verzuim (verwachting: negatieve correlaties).

Deze verwachtingen worden door ons eenzijdig getoetst. Waar de richting van de verbanden tegengesteld blijkt te zijn aan onze verwachtingen, toetsen wij tweezijdig. Wij wijzen er hier, wellicht ten overvloede, op dat N van invloed is op het verkrijgen van significante uitkomsten. Dit betekent dat bij alle werknemers eerder (dat wil zeggen bij lagere waarden van r) statistische significantie wordt bereikt dan bij de respondenten.

9. Voor de omzetting van r -waarden in Z -scores hebben we gebruik gemaakt van tabel G in De Jonge en Wielenga (1969: 306-307). Zie voor de formules van s Blalock (1979: 418-421).

10. Het aantal dienstjaren is buiten beschouwing gelaten, omdat deze variabele sterk samenhangt met leeftijd (zie noot 8). Bovendien blijkt uit de correlatieberekeningen dat het aantal dienstjaren niet of nauwelijks samenhangt met ziekteverzuim. Aangezien verzuimfrequentie en duur van de laatste verzuimperiode voor het ZVP sterk met elkaar samenhangen, is bij de voorspelling van de verzuimfrequentie na het ZVP alleen de verzuimfrequentie voor het ZVP als voorspeller meegenomen en bij de voorspelling van de duur van de laatste verzuimperiode na het ZVP alleen de duur van de laatste verzuimperiode ervoor.

Geraadpleegde literatuur

- Bernard, H.R. et al. (1984) The problem of informant accuracy: The validity of retrospective data. *Annual Review of Anthropology* 13: 495-517.
- Blair, E. en S. Burton (1987) Cognitive processes used by survey respondents to answer behavioral frequency questions. *Journal of Consumer Research* 14: 280-288.
- Blalock, H.M. (1979) *Social statistics* (Revised Second Edition). New York: McGraw-Hill.
- Bradburn, N.M., L.J. Rips en S.K. Shevell (1987) Answering autobiographical questions: The impact of memory and inference on surveys. *Science* 263 (4798, 10 april): 157-161.
- Burton, S. en E. Blair (1991) Task conditions, response formulation processes, and response

- accuracy for behavioral frequency questions in surveys. *Public Opinion Quarterly* 55: 50-79.
- Cahalan, D. (1968) Correlates of respondent accuracy in the Denver Validity Survey. *Public Opinion Quarterly* 32: 607-621.
- Cannell, Ch.F., L. Oksenberg en J.M. Converse (red.) (1979) *Experiments in Interviewing techniques: Field experiments in health reporting, 1971-1977*. Michigan: Survey Research Center.
- Cohen, J. (1994) The earth is round. *American Psychologist* 49: 997-1003.
- Dalecki, M.G., J.C. Whitehead en G.C. Blomquist (1993) Sample nonresponse bias and aggregate benefits in contingent valuation: An examination of early, late and nonrespondents. *Journal of Environmental Management* 38: 133-143.
- Dex, S. (1995) The reliability of recall data: A literature review. *Bulletin de Méthodologie Sociologique* 49: 58-89.
- Dijkstra, A. (1981) Ziekteverzuim en nonresponse: Representativiteit van deelname aan enquête-onderzoek in arbeidsorganisaties ten aanzien van een te verklaren variabele. *Gezondheid en Samenleving* 2: 266-273.
- Dillman, D.A. (1991) The design and administration of mail surveys. *Annual Review of Sociology* 17: 225-249.
- Ellis, R.A., C.M. Endo en J.M. Armer (1970) The use of potential nonrespondents for studying nonresponse bias. *Pacific Sociological Review* 13: 103-109.
- Evans, D.S. en L.S. Leighton (1995) Retrospective bias in the displaced worker surveys. *The Journal of Human Resources* 30: 386-396.
- Fluttert, P.H.M. (1980) Wat is een respons van 80% waard?: Een onderzoeksnotitie. *Sociologische Gids* 27: 160-165.
- Gannon, M.J., J.C. Notheman en S.J. Carroll, Jr. (1971) Characteristics of nonrespondents among workers. *Journal of Applied Psychology* 55: 586-588.
- Goor, H. van (1996) Het lijkt uit de kast: Nonrespons in het Nationaal Kiezersonderzoek. *Sociologische Gids* 43: 166-170.
- Goor, H. van en B. Stuiver (1995) Succes en falen van beleid en nonrespons: Een empirisch onderzoek naar het terugzenden van schriftelijke enquêtes door Nederlandse gemeenten. *Sociologische Gids* 42: 388-406.
- Goor, H. van en B. Stuiver (MS) Are high response rates necessary? A wave-analysis of distributional bias, substantive bias and data quality in a mail survey among Dutch municipalities. Manuscript.
- Goor, H. van en A.L. Verhage (MS) Retrospectieve vertekeningen in een onderzoek naar ziekteverzuim: Nauwkeurigheid van verzuimrapportage, geheugeneffecten en gevoelens van bedreiging. Manuscript.
- Goudy, W.J. (1976) Nonresponse effects on relationships between variables. *Public Opinion Quarterly* 40: 360-369.
- Goudy, W.J. (1978) Interim response to a mail questionnaire: Impacts on variable relationships. *Sociological Quarterly* 19: 253-265.
- Goyder, J. (1987) *The silent minority: nonrespondents on sample surveys*. Oxford: Polity Press.
- Graaf, P.M. de, A. Poortman en W.C. Ultee (1996) De kwaliteit van retrospectieve beroepsgegevens: Een onderzoek op basis van huwelijksaktes. *Sociale Wetenschappen* 39(3) 1-15.
- Green, K.E. (1991) Reluctant respondents: Differences between early, late, and nonresponders to a mail survey. *Journal of Experimental Education* 59: 268-276.
- Groves, R.M. (1987) Research on survey data quality. *Public Opinion Quarterly* 51: S156-S172.

- Hawkins, D.F. (1975) Estimation of nonresponse bias. *Sociological Methods and Research* 3: 461-488.
- Jonge, H. de en G. Wielenga (1969) *Statistische methoden voor psychologen en sociologen* (derde druk). Groningen: Wolters-Noordhoff.
- Kirchner, W.K. en N.B. Mousley (1963) A note on job performance: Differences between respondent and nonrespondent salesmen to an attitude survey. *Journal of Applied Psychology* 47: 223-224.
- Kivlin, J.E. (1965) Contributions to the study of mail-back bias. *Rural Sociology* 30: 322-326.
- Loftus, E.F. et al. (1992) Memory and mismemory for health events. In: J.M. Tanur (red.) *Questions about questions: Inquiries into the cognitive bases of surveys*. New York: Russell Sage Foundation, p. 102-137.
- McKee, D.O. (1992) The effect of using a questionnaire identification code and message about nonresponse follow-up plans on mail survey response characteristics. *Journal of the Market Research Society* 34: 179-191.
- Marquis, K.H., M.S. Marquis en J.M. Polich (1986) Response bias and reliability in sensitive topic surveys. *Journal of the American Statistical Association* 81: 381-389.
- Martin, Ch.L. (1994) The impact of topic interest on mail survey response behaviour. *Journal of the Market Research Society* 36: 327-338.
- O'Muircheartaigh, C.A. (1976) Response errors in an attitudinal sample survey. *Quality and Quantity* 10: 97-115.
- O'Neill, M.J. (1979) Estimating the nonresponse bias due to refusals in telephone surveys. *Public Opinion Quarterly* 43: 218-232.
- Pavalko, R.M. en K.G. Lutterman (1973) Characteristics of willing and reluctant respondents. *Pacific Sociological Review* 16: 463-476.
- Pearl, D.K. en D. Fairley (1985) Testing for the potential for nonresponse bias in sample surveys. *Public Opinion Quarterly* 49: 553-560.
- Pearson, R.W. et al. (1992) Personal recall and the limits of retrospective questions in surveys. In: J.M. Tanur (red.) *Questions about questions: Inquiries into the cognitive bases of surveys*. New York: Russell Sage Foundation, p. 65-94.
- Popping, R. (1984) *Overeenstemmingsmaten voor nominale data*. Meppel: Krips Repro (dissertatie Rijksuniversiteit Groningen).
- Popping, R. (1992) *Taxonomy on Nominal Scale Agreement 1945-1990*. Groningen: iec ProGAMMA.
- Popping, R. (1995) *Computing Agreement on Nominal Data: The Computer Program AGREE 6.0*. Groningen: iec ProGAMMA.
- Raadschelders, J. en J. van der Zouwen (1976) *De juistheid van interviewgegevens: Een analyse van gegevens omtrent de overeenstemming tussen de antwoorden op interviewvragen en externe gegevens*. Amsterdam: Vakgroep Methoden en Technieken VU.
- Schwarz, N. en S. Sudman (red.) (1996) *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Fransisco: Jossey-Bass.
- Silver, B.D., B.A. Anderson en P.R. Abramson (1986) Who overreports voting? *American Political Science Review* 80: 613-624.
- Stefanowska, M. (1979) Feelings of 'cultural inadequacy' and validity of respondents answers in surveys on book reading. *Quality and Quantity* 13: 327-338.
- Stinchcombe, A.L., C. Jones en P. Sheatsley (1981) Nonresponse bias for attitude questions. *Public Opinion Quarterly* 45: 359-375.
- Sudman, S. en N.M. Bradburn (1974) *Response effects in surveys: A review and synthesis*. Chicago: Aldine.

- Tanur, J.M. (red.) (1992) *Questions about questions: Inquiries into the cognitive bases of surveys*. New York: Russell Sage Foundation.
- Verhage, A.L. (1996) *Op weg naar een lager ziekteverzuim: Een evaluatie van het ziekteverzuimbeheersingsplan KWS Regio Noord*. Groningen: KWS Regio Noord/Vakgroep Sociologie Rijksuniversiteit Groningen (doctoraalscriptie).
- Weaver, Ch.N. en C.L. Swanson (1974) Validity of reported date of birth, salary, and seniority. *Public Opinion Quarterly* 38: 69-80.
- Weiss, C.H. (1968) Validity of welfare mothers' interview responses. *Public Opinion Quarterly* 32: 622-633.