

## Enkele problemen rond de nonresponse\*

D. Kruyt, *Rijksuniversiteit Utrecht*

Veel gedragswetenschappelijk onderzoek en met name survey-onderzoek maakt gebruik van steekproeven van waaruit gegeneraliseerd moet worden naar achterliggende populaties. Een daarbij nogal vergeten vraagstuk is dat van de nonresponse.

Hier worden een aantal problemen rond de nonresponse ter sprake gebracht. Daartoe wordt eerst een aantal begrippen ingevoerd ter verduidelijking van de begrippen „bias” en „representativiteit”. Vervolgens worden enige voorwaarden voor generalisatie uit steekproeven naar voren gebracht. Daarna wordt met een voorbeeld inzichtelijk gemaakt hoe eventuele bias generalisatieuitspraken kan invalideren. Tenslotte komen een aantal redenen voor nonresponse ter sprake en worden een vijftal reductie- en schattingsmethoden voor bias tengevolge van nonresponse besproken.

### *Bias en representativiteit*

*Elementen* uit een populatie zijn die eenheden voor wie informatie wordt verzameld. De *populatie* is de verzameling elementen over welke in een onderzoek conclusies worden getrokken.

Elk onderzoek definieert de populatie in termen van (a) de definitie der elementen, (b) (eventueel) de definitie van grotere eenheden waarbinnen de elementen zich bevinden, (c) ruimte, en (d) tijd, (a) en (b) hangen direct samen met het doel en eventueel het analyseniveau van het onderzoek. Zulk een populatie wordt gewoonlijk aangeduid met de wat ongelukkige benaming „*theoretische populatie*”. Een voorbeeld van zulk een definitie is: (a) alle mannen, (b) in de agrarische sector werkzaam, (c) in Nederland, (d) in 1969.

Onder het administratief *bestand* zullen we verstaan de registratieplaats der populatie-elementen, bijv. een ledenlijst of een bevolkingsregister. De steekproef wordt uit het bestand getrokken en het bestand is de *statistische populatie*. Een *steekproef* is een deelverzameling uit de verzameling elementen welke in de populatie wordt gedefinieerd; het is de steekproef die de facto wordt onderzocht, en van waaruit op grond van de resultaten van het onderzoek conclusies getrokken gaan worden over

\*) Dit artikel berust op een deel van mijn skriptie „Bias tengevolge van nonresponse”, Sociologisch Instituut Nijmegen, 1968, gesuperviseerd door O. J. A. Janssen en door hem en A. Albinski van kritisch commentaar voorzien. Hen en H. W. van den Ende en P. G. Swanborn die het manuscript kritisch lazen, wil ik gaarne mijn dank betuigen.

de populatie. *Steekproefeenheden* zijn die elementen van de populatie welke in de steekproef zijn opgenomen.

Men generaliseert naar de statistische populatie. Voor de gedragswetenschappelijke theorie is het echter meer interessant te weten dat de resultaten gelden voor de theoretische populatie. Voor generalisatie naar de theoretische populatie zal men dan ook als noodzakelijke voorwaarde een *congruentie* tussen theoretische en statistische populatie moeten stellen. Bij eventuele incongruentie zal men dan genoegen moeten nemen met generalisatie naar een opnieuw gedefinieerde of *operationele populatie*. Deze herdefiniëring is een versmalling van een of meer van de constituerende delen van de theoretische populatie in de richting van het bestand. Een voorbeeld van een (extreem versmalde) operationele populatie is (a) alle mannen  $\leq 21$  jaar, (b) werkzaam op landbouwbedrijven  $< 20$  ha, (c) in de N.O. polder, (d) per 31 december 1968. Wat overigens voldoende of onvoldoende congruentie is, is een criterium dat de onderzoeker voor elk onderzoek opnieuw moet vaststellen.

Een steekproef is *aselect* getrokken als de trekking op basis van toeval geschiedt en de procedure zodanig is dat de populatie-elementen een bekende niet-nul kans hebben om in de steekproef te worden opgenomen. Hier zal voortaan worden uitgegaan van de basisprocedure, de enkelvoudig aselechte steekproef, waarbij de procedure zodanig is dat de populatie-elementen een gelijke niet-nul kans hebben voor opname. Aselectiviteit is een tweede noodzakelijke voorwaarde voor generalisatie; op basis van aselechte trekking kan men steekproevenverdeling berekenen die essentieel is voor het opstellen van een generalisatieuitspraak.

Te onderzoeken karakteristieken van de populatie-elementen  $i$  worden via de meetoperaties van de steekproef omgevormd tot variabelen  $x_i, y_i, z_i \dots$  etc. Een steekproefwaarde  $t$  gebaseerd op een of meer variabelen noemen we *variate*; het is een schatting op basis van de  $n$  steekproefeenheden. We spreken van uni-, bi- en multivariate meting naargelang de kenmerkende numerieke expressie van de schatting gebaseerd is op een, twee of meer variabelen.

De populatiewaarde of ware waarde  $T$  is (analoog) de numerieke expressie gebaseerd op een of meer karakteristieken van de  $N$  populatie-elementen. Veelal wordt die geschat via inferentie uit de berekende variate.

Onder een *steekproefdesign* zullen we verstaan een steekproeftype van een bepaalde grootte, een bepaalde trekkingsprocedure en een bepaalde schattingsprocedure. De gemiddelde waarde van de theoretische verdeling van een oneindig aantal steekproeven met eenzelfde design (de steekproevenverdeling) is de verwachte waarde  $E(t)$  van de variate. Een design is gebiast als  $E(t) \neq T$ . De *bias*  $B$  is dan  $T - E(t)$ . Op basis van een ongebiast veronderstelde steekproef bestaan er een aantal procedures om de verwachte waarde van een variate te schatten en aldus te concluderen tot  $T$ . In een aantal gevallen wordt van de standaardfout van het design, de wortel uit de variantie van de verwachte waarde van variate, gebruik gemaakt; ook deze wordt meestal geschat. Een *betrouwbaar* design is een design met een kleine standaardfout.

Bias kan o.m. ontstaan door systematische fouten als artefact van de trekkingsprocedure, door nonresponse of als artefact van de dataverzamelings- en verwerkingsmethoden. De totaal fout van het design is de wortel uit de som van de variantie en de

gekwadrateerde systematische afwijkingen tengevolge van de verschillende biasbronnen. Een *geldig* design is een design met een kleine totaalfout<sup>1)</sup>. Geldigheid is een derde noodzakelijke voorwaarde voor generalisatie. Vanuit een ongeldig design generaliseert men niet met referentie aan  $T = E(t)$  maar aan  $= B + E(t)$ , met een gewoonlijk onbekende grootte B.

Op basis van een steekproef welke aan de drie bovengenoemde voorwaarden voldoet doet men uitspraken over de achterliggende populatie. Daarbij komt soms de term „representativiteit”<sup>2)</sup>, het afgespiegeld zijn van de populatie in de steekproef, ter sprake. Stelt men prijs op gebruikmaking van de term representativiteit dan moet men zich wel realiseren dat het een relatieve betekenis heeft. Een steekproef is representatief ten opzichte van iets, censusdata, samenhangen in de populatie e.d. Bij surveys is het de gewoonte een rapport vergezeld te doen gaan van een steekproefverslag, waarin de mate van congruentie, aselectiviteit en geldigheid aan de orde wordt gesteld. Gewoonlijk toetst men dan tegen (soms wel, soms nauwelijks) relevante censusdata als leeftijd, geslacht en religie; het lijkt me echter een voorschands minstens onvoorzichtige veronderstelling te menen dat met dit soort toetsing volstaan kan worden om plausibel te maken, laat staan te bewijzen dat een steekproef aan de eisen van congruentie, aselectiviteit en geldigheid voldoet, of dat de steekproef daarmee bijv. representatief is t.o.v. associaties en correlaties.

#### *Het biaseffect op uitspraken over de populatie*

Cochran<sup>3)</sup> beschrijft hoe bij univariate meting een opgetreden bias generalisatieuitspraken invalideert. De mate van invalidering hangt af van de ratio  $B/\sigma$  waarbij  $\sigma$  de standaardfout voor de verdeling is. Om zich voor te stellen hoe bias uitspraken over samenhangen kan vervormen kan men zien wat er gebeurt met de  $\chi^2$  in een test op onafhankelijkheid van twee variabelen. Hier wordt uitgegaan van een eenvoudige  $2 \times 2$  tabel. We zullen aannemen dat in één der cellen, zeg a, bias optreedt;  $\chi^2$  verwijst naar een berekening zonder,  $\chi_B^2$  naar een berekening met verwerking van het biaseffect. In onderstaande tabel verwijzen a, b, c, d, en n naar de absolute frekwenties.

De berekening wordt tweemaal uitgevoerd; beide malen wordt een gelijke n en gelijke marginales aangenomen.

	a	b	a + b
	c	d	c + d
	a + c	b + d	n

<sup>1)</sup> De hier besproken begrippen refereren vooral aan L. Kish, *Survey Sampling*, New York, 1965; W. G. Cochran, *Sampling Techniques*, New York, 1953; en G. U. Yule and M. G. Kendall, *An introduction to the theory of statistics*, London, 1965.

<sup>2)</sup> L. Kish, *op. cit.*, p. 26, stelt voor de term „representativiteit” als onduidelijk en overbodig te schrappen. Daar zijn goede argumenten voor maar het begrip is nogal ingeburgerd.

<sup>3)</sup> W. G. Cochran, *op. cit.*, p. 9.

Men kan nu een steekproevenverdeling voor de variatie  $a$  berekenen, een betrouwbaarheidsinterval voor generalisatie vaststellen<sup>4)</sup> en vervolgens de mate van vervorming bij optreden bias bezien. Dit komt echter in feite neer op reductie tot een soortgelijk probleem als bij univariate meting. Het heeft hier meer zin,  $\chi^2$  en  $\chi^2_B$  te vergelijken.

Voor  $2 \times 2$  tabellen gebruikt men de formule

$$(1) \chi^2 = \frac{n[|(ad - bc)| - \frac{1}{2}n]^2}{(a + b)(c + d)(a + c)(b + d)}$$

Bij een optredende bias  $B$  in cel  $a$  waardoor de „ware frekwentie”  $(a - B)$  wordt, is

$$(2) \chi^2_B = \frac{n[|(ad - bc) - nB| - \frac{1}{2}n]^2}{(a + b)(c + d)(a + c)(b + d)}$$

zodat, onder voorwaarde dat  $ad > bc$  dan en slechts dan  $\chi^2 \begin{matrix} < \\ = \\ < \end{matrix} \chi^2_B$

$$\text{als } B(B - 1) \left[ B - \frac{2(ad - bc)}{n} \right] \left[ B - \frac{2(ad - bc)}{n} - 1 \right] \begin{matrix} < \\ = \\ > \end{matrix} 0$$

We zien een analoog effect: invalidering van de uitspraak over samenhang. Een dergelijk effect blijft ook mogelijk wanneer men geen garanties heeft voor biasvrije multivariate meting, dit zowel voor associaties, schaal- en indexscores.

### Nonsresponse

We beperken ons tot het survey. Van *nonresponse* (uitval) bij een survey is sprake als het gaat over eenheden die wel in de steekproef werden getrokken maar van welke geen of onvolledige informatie werd verkregen. Men is er in geslaagd, daar een ruim scala van subtypen in aan te brengen<sup>6)</sup>. Het handigst lijkt me echter, gezien redenen voor en effect van nonresponse<sup>7)</sup>, een globale indeling te maken naar typen als „weigeraars” en „onbereikbaren”. Item-nonresponse, geen of onvolledig antwoord op een enkele variabele, kan dan afhankelijk van reden en effect als gewone nonresponse worden beschouwd. Overigens kunnen door middel van regressievergelijkingen, inlassen „op gezicht” of via controle- of filtervragen scores op een niet beantwoord item worden geschat of geforceerd.

De reden dat nonresponse als biasbron niet buiten beschouwing mag blijven is het feit dat een bepaalde groep eenheden systematisch uit de steekproef afwezig kan blijken hetgeen niet alleen op uni- maar ook op bi- en multivariate meting effect kan hebben. Dit is een probleem in de sfeer van het „correlated bias”-probleem<sup>8)</sup> voor

4) A. E. Maxwell, *Analysing qualitative data*, London, 1961, pp. 32-35.

5) zie ook K. Assakul and C. H. Proctor, Testing independence in two-way contingency tables with data subject to misclassification, *Psychometrika*, 32, 1967, pp. 67-76.

6) zo brengt M. B. Parten, *Surveys, polls and samples*, New York, 1950, het tot maar liefst 19 deeltypen.

7) C. S. Mayer and R. W. Pratt, A note on nonresponse in a mail survey, *Public Opinion Quarterly*, 30, 1966, pp. 637-646, konstateerden b.v. biaseffecten voor onbereikbaren en weigeraars naar tegengestelde richtingen.

8) een probleem, door H. C. Selvin, A critique of tests of significance in survey research, *American Sociological Review*, 22, 1957, pp. 519-527, aan de orde gesteld. Voor een bespreking van de daarop volgende discussie zie J. Galtung, *Theory and methods of social research*, Oslo, 1967, II, 5.

het survey in het algemeen: de onmogelijkheid van controle op relevante factoren die in interactie effect kunnen hebben in niet vooraf te voorspellen richting. Ook over eventuele bias tengevolge van nonresponse zal onder blijken dat evenmin over een te voorspellen richting van het effect kan worden gesproken.

Hoe weinig eenheden uit een steekproef verwijderd behoeven te zijn om effect op samenhangen te hebben laat een verkiezingsonderzoek in Nijmegen zien waar de relatie ontzuiling en stemgedrag werd getest met een aantal variabelen<sup>9)</sup>. Voor drie variabelen, leeftijd, opleiding en vervreemding werd significante interactie geconstateerd. Het betrof daar 519 respondenten die hun stem op een der vijf partijen: K.V.P., P.v.d.A., V.V.D., B.P. en D.66 hadden uitgebracht. Na toevoeging van 50 respondenten die op geen of een andere partij hadden gestemd verdween de interactie in twee gevallen, terwijl voor de derde variabele, leeftijd, een ander patroon ontstond. Dit is wellicht tevens een illustratie van het gevolg dat verwijdering van categorieën als „weet niet” of „anderszins” kan hebben.

In de onderzoeksliteratuur wordt soms de mening verdedigd dat men in een aantal gevallen de richting van de nonresponse kan voorspellen<sup>10)</sup>. Daarbij heeft men dan meestal karakteristieken als leeftijd, geslacht e.d. op het oog, hetgeen in het merendeel der gevallen nogal irrelevante informatie zou opleveren. De nonresponse hangt echter van een veelvoud van waarschijnlijk weer onderling correlerende factoren af, zodat voorspelbaarheid van richting minder voor de hand lijkt te liggen. Blijkens onderzoeksverslagen van een aantal auteurs<sup>11)</sup> wordt de proportie *onbereikbaren* o.m. beïnvloed door: de soort populatie (bijv. ongehuwden, bepaalde beroepsgroepen) eigenschappen van de individuele respondent (bijv. hobby's buitenshuis), het tijdstip van het onderzoek (bijv. vakanties, weekends, werktijden), de gebruikte waarnemingsmethode (schriftelijk, telefonisch of mondeling), en de beschikbare hoeveelheid tijd, staf en geld van de onderzoeksinstelling.

Voor *weigeraars* worden o.a. factoren genoemd als: survey-moeheid van respondenten (waarvan in een herinterview met Nijmeegse weigeraars weinig bleek), de soort populatie (bijv. leiders, devianten), de sociale positie van de respondenten (bijv. de marginaliteit van groepsleden), attitudes van de respondent op politiek en maatschappelijk terrein, de hoeveelheid tijd die de respondent ter beschikking staat, de motivatie van de respondent, eventuele vergoeding, de door de respondent begrepen relevantie van het onderzoek, wantrouwen en angst, de door de respondent gepercipieerde integriteit van de onderzoeker, de gebruikte waarnemingsmethode (in een aantal gevallen schijnt een combinatie van mondeling interview en schriftelijke enquête effectief ter voorkoming van nonresponse), de anonimiteit van de situatie, de aanwezigheid van anderen, de relatie van de respondent met de opdrachtgever, het prestige van de opdrachtgever en de onderzoeksinstelling, het aantal en de hardnekkigheid van de benaderingen (follow-ups bij postenquêtes schijnen zeer effectief), de status en de vriendelijkheid van de enquêteur, het onderwerp van de studie

<sup>9)</sup> zie A. L. MacDonald en D. Kruyt, Ontzuiling en stemgedrag bij katholieken, *Sociologische Gids*, 15, 1968, pp. 98-105.

<sup>10)</sup> zie b.v. F. F. Stephan and P. J. McCarthy, *Sampling Opinions*, London, 1958, p. 241.

<sup>11)</sup> de verschillende onderzoeksartikelen zijn te veel om hier te vermelden; geïnteresseerden wenden zich tot de schrijver.

(ook voor item-nonresponse; o.m. leveren moeilijkheden op: inkomen en geldzaken, politieke attitudes, de opleidingsgeschiedenis, politiek, sexueel en deviant gedrag, huwelijks- en geloofszaken, lichaamshygiëne en in het algemeen taboe-beladen onderwerpen), de lengte van de vragenlijst, de vermoedelijke duur van het interview en vooral de introductie en de introductiebrief. Het onderzoek naar redenen voor non-response neemt soms wel ridicule trekjes aan: zo zijn er experimentjes verricht naar de invloed van de kleur en het aantal postzegels op de enveloppe, en de kleur en de kwaliteit van het papier van de introductiebrief.

De voorspelbaarheid van de uitvalsrichting schijnt dus minder steekhoudend dan soms wordt aangenomen. Scott<sup>12)</sup>, die een uitgebreide secundaire analyse op post-enquêtedata heeft verricht, heeft dan ook weinig bemerkt van een voorspelbaarheid van uitvalsrichting. Zijn resultaten zijn onder in tabelvorm samengevat.

aantal studies	getoetst op karakteristiek	aantal dat nonresponse wel of niet kan localiseren
12	opleiding	9 lage opleiding valt uit 3 geen verschil in opleiding
8	geslacht	2 verschil in geslacht 6 geen verschil in geslacht
6	leeftijd/ geslacht	2 oude vrouwen vallen uit 1 jonge mannen vallen uit 3 geen verschil in leeftijd
3	kindertal	3 geen verschil in kindertal
4	status	2 lage status valt uit 1 hoge status valt uit 1 geen verschil in status
3	beroep	3 geen verschil in beroep
13	urbanisatie	7 suburbs valt uit 2 platteland valt uit 2 stad valt uit 1 kleine stad valt uit 1 geen verschil in urbanisatie
9	interesse in onderwerp	8 ongeïnteresseerd valt uit 1 geen verschil in interesse

#### *Schatting van nonresponse-effecten*

Bovenstaande opsomming van factoren die nonresponse beïnvloeden moge dienen ter overweging van hen die een onderzoek opzetten en de nonresponse zoveel mogelijk willen beperken. We zullen ons nu bezighouden met het effect van eventueel opgetreden nonresponse.

<sup>12)</sup> C. Scott, Research on mail survey, *Journal of the Royal Statistical Society*, series A, 1961, pp. 143-195.

Zij  $pr_1$  de proportie response (van een steekproeftotaal  $n$ ),  $pr_2 = 1 - pr_1 =$  de proportie nonresponse,  $\bar{x}_1 =$  het gemiddelde van een variabele dat op de response is gebaseerd, en  $\bar{x}_2 =$  het gemiddelde dat op de nonresponse is gebaseerd. De relatieve bias die ontstaat door substitutie van  $\bar{x}_1$  voor  $\bar{x}$  (het gemiddelde, gebaseerd op  $n$ ) is<sup>13)</sup>

$$(3) \quad B(\bar{x}_1) = pr_2 \cdot \frac{\bar{x}_1 - \bar{x}_2}{\bar{x}}$$

Dit soort bias kan aanzienlijke effecten hebben; nemen we als voorbeeld de (overigens vooral theoretisch relevante) bepaling van de steekproefgrootte voor generalisatie van een proportie van één gedichotomiseerde variabele. Men benadert  $n$  met<sup>14)</sup>

$$(4) \quad n = \frac{t^2 \alpha p q}{Tf^2}$$

waarbij  $Tf =$  de gewenste maximale totaalfout,  $t\alpha =$  de bij de grens behorende waarde, en  $p$  en  $q$  de proporties zijn van de variabele. Voor  $Tf = .05$ ,  $p = q = .5$  (waarbij  $p \cdot q$  voor de bepaling van  $n$  zo ongunstig mogelijk wordt), en  $\alpha = .05$  (zodat  $t\alpha = 1.96$ ) is de steekproefgrootte  $n \geq 384$  (afgerond).

Houdt men rekening met de bias tengevolge van nonresponse<sup>14)</sup> dan is

$$(5) \quad n = \frac{t^2 \alpha}{4 Tf (Tf - pr_2) \cdot pr_1} - 1$$

(waarbij dus is aangenomen dat  $p = q$ ).

Om een voorbeeld van het nonresponse-effect te geven: Bij  $\alpha = .05$  en  $Tf = .5$  zou  $n \geq 4$  moeten zijn volgens (4). Bij een  $pr_2 = .47$  moet bij gebruikmaking van (5) voor dezelfde  $\alpha$  en  $Tf$  de steekproefgrootte  $n \geq 127$  zijn<sup>15)</sup>.

Een aantal auteurs heeft zich beziggehouden met procedures om het effect van eventueel opgetreden nonresponse te schatten bij generalisatie. De eerste twee procedures zijn vooral toepasbaar op onbereikbaren, de andere ook op weigeraars.

1. *De weegfactormethode van Politz en Simmons.* Deze procedure heeft de intentie, tijd- en geldbesparend te zijn door het aantal benaderingen per respondent te verminderen<sup>16)</sup> en respondenten gedifferentieerd te wegen. Men gaat van de veronderstelling uit dat geïnterviewden die normaal weinig thuis zijn model kunnen staan voor onbereikbaren. De voorgestelde procedure komt op het volgende neer: Men benadert alle respondenten één maal en vraagt ze hoeveel maal ze uit  $j$  gelijke perioden thuis zijn. Is het antwoord  $i$  ( $i \leq j$ ) dan wordt de respondent met de factor

<sup>13)</sup> L. Kish, *op. cit.*, p. 585.

<sup>14)</sup> W. G. Cochran, *op. cit.*, p. 54 en 297.

<sup>15)</sup> D. Seigle, Some aids in the handling of missing data, *Social Science Information*, 6, 1967, pp. 132-150 geeft tabellen voor de bias die men kan meegeneneraliseren bij verschillende proporties nonresponse; ook deze zijn gebaseerd op univariate generalisatie.

<sup>16)</sup> overigens wijzen R. Goodman and C. F. Cannell, *Sampling errors and components of interview costs in relation to sample design*, z.p., z.j., pp. 7-8, er op dat de kumulatieve kosten per extra benadering relatief kleiner worden; Z. W. Birnbaum and M. G. Sirken, Bias due to non-availability in sampling surveys, *Journal of the American Statistical Association*, 45, 1950, pp. 98-110, hebben er een kostenfunctie voor opgesteld.

$j/i$  gewogen. Politz en Simmons stellen voor,  $j$  op 6 te stellen<sup>17)</sup>. Simmons die zich later gewaar is geworden dat op deze wijze de groep die nooit thuis is geheel buiten beschouwing blijft en dat een weegfactor 6 de resultaten een onbetrouwbare exactheid geeft, heeft in een later artikel voorgesteld, een combinatie van meer benaderingen en de weegfactormethode te gebruiken<sup>18)</sup>.

Tegen de weegfactormethode kan men als bezwaar aanvoeren dat de opgelegde assumpties (de  $j$  perioden worden aselekt gekozen en de respondenten worden op een aselekt tijdstip benaderd) te stringent zijn. Met name de tweede assumptie lijkt weinig realistisch.

2. *De vervangingsmethode van Kish*<sup>19)</sup>. Deze procedure heeft de intentie het aantal benaderingen, nodig om het aantal onbereikbaren en wellicht ook weigeraars te reduceren, zo efficiënt mogelijk te doen verlopen, maar vereist meerdere surveys achtereen om uiteindelijk het nonresponse-effect te schatten. Men doet het volgende: De nonresponse-adressen uit survey 1 worden gevoegd bij de steekproef-adressen voor survey 2 en gaan dienen ter vervanging van de aldaar optredende nonresponse. Als de response in survey 1 en 2 resp.  $k_1$  en  $k_2$  maal wordt benaderd, wordt de nonresponse van survey 1,  $m$  maal benaderd, zodat  $k_1 < m \leq (k_1 + k_2)$ . De aldus (door het groter aantal benaderingen) geïnterviewde nonresponse krijgt de vragenlijst voor survey 1 en survey 2 voorgelegd. In principe is dit proces iteratief: Zo kan de nonresponse van survey 1 en/of survey 2 weer worden benaderd in survey 3 enz.

De bezwaren tegen deze procedure zijn niet zozeer de onhoudbaarheid der assumpties als wel de beperkte toepassingsmogelijkheden; het is een procedure die pas op de lange duur resultaten oplevert en vereist eenzelfde wijze van steekproeftrekken voor surveys die dezelfde populatie moeten betreffen. Bovendien worden er nogal strenge eisen gesteld aan de consistentie van gedrag en attitudes van de populatie en/of een niet al te grote tijdsinterval tussen de surveys. Het zullen daarom waarschijnlijk slechts speciale bureau's zijn die er gebruik van kunnen maken; wellicht zou het een (gedeeltelijke) oplossing kunnen zijn voor de te verwachten uitval bij longitudinaal onderzoek.

3. *De maximum-bias-methode van Stephan en McCarthy*<sup>20)</sup>. Deze procedure bepaalt de maximale bias die kan ontstaan tengevolge van nonresponse en maakt gebruik van de Tf voor de opstelling van een betrouwbaarheidsinterval voor generalisatieuitspraken. Zij  $n_1$  het aantal response en  $n_2 = n - n_1$  het aantal nonresponse. Men dichotomiseert een variabele waarover men een generalisatieuitspraak wil doen in de percentages of proporties  $p$  en  $q = 1 - p$ . Heeft men de beschikking over  $n_1$  dan kan men slechts beschikken over de percentages of proporties voor  $n_1$ , zijnde  $p_1$  en  $q_1$ . Men berekent nu de waarde van  $p$  die zou ontstaan als in het ongunstigste

17) A. Politz and W. Simmons, An attempt to get the "not at homes" into a sample without callbacks, *Journal of the American Statistical Association*, 44, 1949, pp. 9-16, Further theoretical considerations regarding the plan for eliminating callbacks, idem, pp. 17-31.

18) W. R. Simmons, A plan to account for "not at homes" by combining weighting and callbacks, *Journal of Marketing*, 19, 1954, pp. 42-54.

19) L. Kish, *Op. cit.*, pp. 560-562.

20) F. F. Stephan and P. J. McCarthy, *op. cit.*, pp. 252-255.



geval de gehele  $n_2$  de ene klasse waarvan  $p$  de proportie is zou kiezen. Voor  $p$  zijn er dan twee mogelijkheden,  $p'$  en  $p''$ :

$$(6) \quad p' = \frac{n_1 \cdot p_1 + n_2 \cdot 0}{n} \quad \text{en} \quad p'' = \frac{n_1 \cdot p_1 + n_2 \cdot 1}{n}$$

De maximale bias  $B_{\max.}$  die dan zou ontstaan is de grootste der verschillen ( $p' - p$ ) en ( $p'' - p$ ). Men generaliseert met  $\pm t\alpha Tf$ , waarbij

$$(7) \quad Tf = \left[ \frac{P_1 \cdot Q_1}{n_1} + B_{\max.}^2 \right]^{1/2}$$

Het is een procedure die vooral van belang schijnt voor onderzoek dat in frekwentie-verdelingen is geïnteresseerd, polls en marktonderzoek e.d. Dit soort onderzoek maakt echter weer meestal gebruik van niet-a-selecte steekproeven.

4. *De dubbelsteekproefmethode van Hansen en Hurwitz*<sup>21)</sup>. Deze procedure beoogt het nonresponse-effect te evalueren maar kan eveneens worden gebruikt voor andere biasbronnen zoals enquêteursbias. Het effect van de relatieve bias (uit (3)) kan men compenseren door een gedeelte der nonresponse te herinterviewen en een weegfactor  $w$  te introduceren waarbij  $w = n_2/n'_2$ , en  $n'_2$  = het aantal herinterviewden via een steekproef uit het aantal nonresponse  $n_2$ .

Voor populatiegrootte  $N$  wordt de schatter voor  $x_{\text{pop.}}$  dan

$$(8) \quad x_{\text{st.}} = \frac{N}{n} (x_1 + wx'_2),$$

waarbij  $x_1 = n_1 \bar{x}_1$  en  $x'_2 = n'_2 \bar{x}'_2$ . Het accent refereert aan de herinterviewden en de subscripts  $\text{pop.}$  en  $\text{st.}$  resp. aan populatie en steekproef. Uit (8) valt af te leiden voor  $\bar{x}_{\text{pop.}}$  de schatter

$$(9) \quad \bar{x}_{\text{st.}} = \frac{1}{n} (n_1 \bar{x}_1 + n_n \bar{x}'_2),$$

en voor  $\text{pr.}_{\text{pop.}}$  ( $\text{pr.}$  staat voor proportie) de schatter

$$(10) \quad \text{pr.}_{\text{pop.}} = \frac{1}{n} (n_1 \text{pr.}_1 + n_2 \text{pr.}'_2)$$

Bij deze procedure moet echter gewezen worden op het gevaar van kunstmatige vergroting van de betrouwbaarheid tengevolge van de introductie van  $w$  en het feit dat bij een steekproef uit  $n_2$  wederom uitval kan optreden.

5. *De extrapolatiemethode van Reid en Ferber*<sup>22)</sup>. Reid constateerde een consistente afname van positieve antwoorden bij respondenten uit drie follow-up-waves. Ferber heeft daar aanleiding in gevonden er een min of meer gestandaardiseerde procedure uit de destileren voor post enquêtes. Hij gaat van de assumptie uit dat, hoe later en na hoeveel aanmaningen extra een respondent zijn antwoorden terugstuurt, des te meer hij begint te lijken op een echte uitvaller. In dat geval kan het de moeite lonen uit

<sup>21)</sup> M. H. Hansen and W. N. Hurwitz, The problem of non-response in sample surveys, *Journal of the American Statistical Association* 41, 1946, pp. 517-529; M. H. Hansen, W. N. Hurwitz and W. G. Madow, *Sampling Survey Methods and Theory*, New York, 1956. I, pp. 469-474, II, pp. 257-258 en 299-305; de variantieberekening komt hier niet ter sprake.

<sup>22)</sup> S. Reid, Respondents and non-respondents to mail questionnaires, *Educational Research Bulletin*, 21, 1942, pp. 87-96; R. F. Ferber, The problem of bias in mail returns: a solution, *Public Opinion Quarterly*, 12, 1948, pp. 669-676.

de waves een trend te extrapoleren. Lehman<sup>23)</sup> stelt voor dit doen met behulp van een grafische voorstelling met op de ene as het percentage antwoorden in een bepaalde categorie. Hij onderscheidt als mogelijkheden een lineaire of monotone toe- of afname, gelijkblijven of oscilleren, in welk geval geen conclusie valt te trekken. Een aantal auteurs<sup>24)</sup> bericht succesvol gebruik van deze procedure.

### *Uitbreiding*

Van bovenstaande procedures valt op dat ze alle bedoeld zijn voor univariate generalisatie. De mogelijkheid blijft daarnaast echter bestaan (cfr. (2)) dat bias ook inwerkt op associaties, indices, e.d. Wil men bijv. associaties generaliseren dan zal men bij nonresponse niet moeten controleren op biasvrijheid t.o.v. één variabele maar op biasvrijheid t.o.v. de associatie. Daartoe is een combinatie van bovenstaande procedures, waaraan ik bij gebrek aan beter zal refereren onder de benaming „gecombineerde weegmethode” wellicht handzaam. Het gebruik van de gecombineerde weegmethode kan vooral nuttig zijn bij grote  $n$  en/of relatief grote  $n_2$ .

Gaan we uit van het volgende: Er is een relatief grote uitval  $n_2$  opgetreden. Uit  $n_2$  wordt een steekproef  $n'_2$  getrokken op gelijke wijze als  $n$  uit  $N$ , die op gelijke wijze als bij  $n$  over soortgelijke enquêteurs met zo veel mogelijk gelijke instructies wordt verdeeld, dit om andersoortige bias dan die tengevolge van nonresponse te controleren. Op soortgelijke wijze kan men een steekproef  $n'_2$  bepalen ingeval men al tijdens de interviewperiode de respondenten niet meer dan een beperkt aantal malen wil benaderen (waardoor de kans op onbereikbaren groter wordt). Bij post enquêtes kan men besluiten, de na een sluitingsdatum overgebleven  $n_2$  in zijn geheel in een follow-up te betrekken.

Heeft men echter uit de weigeraars en/of onbereikbaren een steekproef getrokken dan bepaalt men een weegfactor  $w_2$  voor  $n'_2$ . Die voor  $n_1$  is 1. De procedure gaat van de veronderstelling uit dat een gewogen som  $n$  van response en moeilijker tot medewerking te bewegen respondenten (de teruggewogen  $n'_2$  of het restant, d.w.z.  $n'_2$  minus de in  $n'_2$  opgetreden nonresponse  $n_3$ ) een beter beeld geeft van de populatie dan de response alleen en dat, hoe moeilijker een respondent tot medewerking te bewegen is, des te meer hij een nonrespondent gelijkt. In verband met de vereiste consistentie van gedrag en attitudes in de bestudeerde populatie mag de tijdsinterval tussen interviewwave 1 en 2 niet te groot zijn.

In principe is deze procedure iteratief. Is  $n_3$  groot of relevant genoeg dan zou men voor een volgende interviewgolf weer een steekproef  $n'_3$  uit  $n_3$  kunnen trekken, die interviewen en terugwegen. Bij postenquêtes kan men  $n_3$  in een nieuwe follow-up-wave betrekken.

In een wat geformaliseerde gedaante:

$$(11) N = n_1 + w_2 [n'_2 + w_3 \cdot (n'_3 + w_4 \cdot (n'_4 + \dots))] \approx n$$

waarbij de weegfactor voor  $n'_i$  gelijk is aan  $w_i = n_i/n'_i$ . Men kan de richting van de bias t.a.v. b.v. een associatie, zeg een  $\gamma$  tussen variabele  $x$  en  $y$  wederom grafisch

<sup>23)</sup> E. C. Lehman, Test of significance and partial returns to mailed questionnaires, *Rural Sociology*, 28, 1963, pp. 284-289.

<sup>24)</sup> zie noot 11.

voorstellen op een assenkruis, waar op de ene as de 1e, 2e, ... i ... interviewwave en op de andere as de  $\Upsilon(x,y)$  wordt uitgezet, die voor  $n_1, n'_2, \dots, n'_i \dots$  telkens opnieuw wordt berekend. De associatie, berekend op basis van  $n$  wordt naar de populatie generaliseerd.

Zo men daar prijs op stelt valt de nonresponsebias op b.v. samenhangen, eventueel uitgesplitst naar type (weigeraars, onbereikbaren), te toetsen op significantie. Te denken valt aan toetsen, in Nederland geïntroduceerd door Stouthard<sup>25</sup>). In dit verband zij opgemerkt dat men hoopt, dat men de hypothese dat er geen effect bestaat van de 1e, 2e, ... i interviewfase op de samenhang in kwestie, niet behoeft te verwerpen.

---

<sup>25</sup>) Ph. C. Stouthard, Toetsing en schatting van interactie-effecten in contingentie-tabellen, *Sociale Wetenschappen*, 9, 1966, pp. 303-313.