

Het Follow-up Project Psychotherapeutische Gemeenschappen (2): het non-respons probleem

door R. F. Koning en G. W. Tremonti

Inleiding

De dataverzameling in het kader van het follow-up-project psychotherapeutische gemeenschappen geschiedde met behulp van een questionnaire: de Nieuwkoopse Vragenlijst (NVL). Deze lijst werd op vijf momenten aan de cliënten voorgelegd: bij opname, bij ontslagen 1, 3 en 5 jaar na ontslag (zie voor uitvoerige informatie over het project, de onderzoeksopzet en de gebruikte vragenlijst: Wagenborg c.s., 1984). De NVL werd bij de drie laatste meetmomenten per post aan de ex-clieënten toegezonden. Als voorbereiding hierop werden de cliënten ten tijde van hun ontslag geïnformeerd dat zij de NVL zouden ontvangen. De toegezonden NVL werd vergezeld van een kort schrijven, waarin de relevantie van het onderzoek werd gememoreerd. Een gefrankeerde antwoordenveloppe werd bijgevoegd. Indien binnen twee weken geen respons was ontvangen, kreeg de ex-client de lijst opnieuw aangeboden, nu begeleid door een dringend verzoek om alsnog te reageren. Bleef ook op dit verzoek een reactie uit, dan werd verdere toezending gestaakt.

Het zal duidelijk zijn dat niet alle verzonden lijsten geretourneerd werden. De onderzoekers zagen zich hierdoor geconfronteerd met het zogenaamde attritieprobleem. De attritie – dat is het inkrimpen van het respondentenbestand – neemt bovendien toe naarmate de periode die verstrijkt na beëindiging van de behandeling, langer wordt. De responspercentages die in dit project werden geboekt tonen duidelijk deze trend (zie figuur 1).

Door de longitudinale onderzoeksopzet, waarbij gebruik gemaakt werd van herhaalde metingen, werd het attritieprobleem extra klem-

Schrijvers zijn beiden als klinisch psycholoog verbonden aan het Psychiatrisch Centrum St. Bavo te Noordwijkerhout.

Deze publikatie vindt plaats onder auspiciën van de Stichting Wetenschappelijk Onderzoek Psychotherapeutische Gemeenschappen (SWOPG).

mend. Bij het onderzoek werd gestreefd naar een reeks van vijf metingen per proefpersoon. Een niet-coöperatieve cliënt moest dus meermalen als non-respondent worden geboekt: te weten voor het meetmoment van weigering en voor de complete metingenreeks.

Het geleidelijk inkrimpen van een respondentenbestand brengt een interpretatieprobleem met zich mee. Er dient nagegaan te worden in hoeverre de mening van de respondenten representatief is voor het gehele onderzoekscohort.

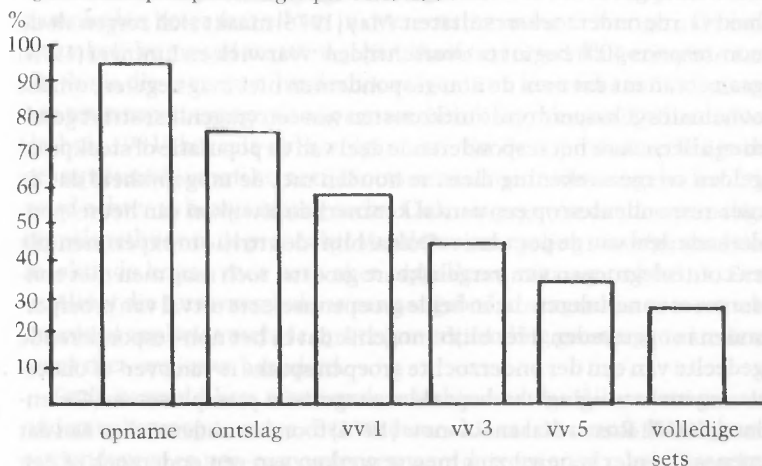
In dit artikel worden achtereenvolgens besproken: literatuurgegevens omtrent aard en omvang van het attritieprobleem; suggesties voor de aanpak ervan, de beschrijving van een aantal onderzoeken naar de effecten van attritie, en de invloed ervan op de resultaten van ons follow-up-project.

Probleemstelling

Non-respons—In diverse onderzoeksverslagen worden responspercentages verstrekt, die onderling sterk verschillen. Sommige auteurs geven aan dat 35% vaak niet wordt gehaald (Larsen c.s., 1979). Anderen daarentegen maken melding van 100% respons bij een follow-up-termijn van diverse jaren (Munby en Johnston, 1980). Deze verschillen zijn gedeeltelijk terug te voeren op de wijze van dataverzameling, maar voor een belangrijk deel ook op verschillen in de attitude van de onderzoekers, vooral wat betreft de inspanning die zij zich willen getroosten om de data te verkrijgen en hun ethische overwegingen.

Illustratief voor dit verschil zijn de opvattingen van Dillman (Nederhof, 1981) en Liberman (1978). Eerstgenoemde bepleit een zogenaamde Total Design Method, wat onder andere herhaaldelijk aanbieden van vragenlijsten alsmede het up to date houden van het adressenbestand

Figuur 1: Responspercentages per 01.09.1984



inhoudt. Deze methode berust onder meer op de opvatting dat non-respons voor een aanzienlijk deel veroorzaakt wordt door toevallige omstandigheden en zelden voortvloeit uit een principiële afwijzing. Past men deze methode toe, dan mag men hoge responspercentages verwachten: 65 à 90% blijkt ook bij delicate onderwerpen haalbaar. Daarentegen spreekt Liberman over 'detective-like behavior', dat vereist zou zijn om meer ex-cliënten te bereiken. Hij wijst er tegelijkertijd op dat men ter wille van de bescherming van de persoonlijke levenssfeer zeer terughoudend moet zijn in het uitvoeren van dergelijk spoorwerk. In het onderhavige project is gekozen voor een wijze van dataverzameling, waarbij een zekere terughoudendheid in acht is genomen.

De onderzochte cliënten werden in acht verschillende therapeutische instituten behandeld. Ondanks het feit dat van meet af aan duidelijk werd gemaakt hoe belangrijk het is dat op uniforme wijze zou worden gewerkt, ontwikkelden zich in de loop der jaren verschillen in de wijze waarop de ex-cliënten benaderd werden. Deze verschillen betroffen de gevolgde administratieve procedures; het opsporen van verloren gegane adressen; de tijdsinvestering van degene die de lijsten verzamelde en de wisselingen in de groep van dataverzamelaars. Ondanks het feit dat de wijze van dataverzamelen voor een belangrijk deel vastgelegd was in nauwkeurige instructies, bleven er kennelijk genoeg interpretatieproblemen en organisatievraagstukken over waardoor een uniforme handelwijze niet tot stand kon komen.

Een en ander heeft tot gevolg gehad dat de responspercentages, die door de diverse klinieken op de verschillende meetmomenten werden geboekt, een forse spreiding te zien geven. Zij variëren van 90 tot 45 procent bij de 1 jaar follow-up meting en van 80 tot 30 procent bij de 3 jaar follow-up-meting.

Generaliseerbaarheidsprobleem – Een te omvangrijke attritie mag worden opgevat als een ernstige bedreiging van de generaliseerbaarheid van de onderzoeksresultaten. May (1976) maakt zich zorgen als de non-respons 40% begint te overschrijden. Warwick en Lininger (1975) gaan ervan uit dat men de non-respondenten niet mag negeren, omdat conclusies gebaseerd op de uitkomsten van een vragenlijst strikt genomen alleen voor het responderende deel van de populatie of steekproef gelden en men rekening dient te houden met de mogelijkheid dat de niet-respondenten op een aantal kenmerken afwijken van het responderende deel van de populatie. Ook al blijft de attritie in experimentele en controlegroepen van vergelijkbare grootte, toch mag men niet zonder meer concluderen dat in beide groepen aselechte uitval van proefpersonen is opgetreden. Het blijft mogelijk dat in het non-responderende gedeelte van een der onderzochte groepen sprake is van over- of ondervertegenwoordiging van bepaalde categorieën proefpersonen (Greenland, 1977). Rosenthal en Rosnow (1975) toonden onder andere aan dat mensen eerder geneigd zijn mee te werken aan een onderzoek of een

enquête naarmate het onderwerp interessanter is of naarmate zij verwachten gunstiger te worden geëvalueerd door de onderzoeker. Nederhof (1981) is van mening dat met name het laatstgenoemde verschil een ernstige bedreiging vormt van evaluatief onderzoek. Baseert men zich hierbij uitsluitend op de respons van coöperatieve ex-cliënten, dan loopt men het risico dat de onderzoeksresultaten in gunstige zin geflatteerd worden, aangezien zich onder non-respondenten ex-cliënten kunnen bevinden die minder gunstige resultaten boekten (Kish en Hermann, 1971).

Een tweetal onderzoeken bij ex-cliënten uit psychotherapeutische gemeenschappen, die in het begin van de jaren zeventig waren opgenomen, leverden resultaten op die inderdaad nopen tot voorzichtigheid.

Zo meldt Bolten (1976) dat er een voor een clinicus coherent en herkenbaar beeld ontstaat van de groep niet-respondenten: 'het zijn de jongere, vaak psychopathiforme cliënten die drop-outs zijn qua behandeling'. Van Kalsbeek c.s. (1978) rapporteren: 'de respondenten zijn aan te duiden als de personen die het meest baat hebben gehad bij hun verblijf in de therapeutische gemeenschap – afgemeten aan de afname van klachten – en die ook tevredener zijn over hun verblijf dan de niet-respondenten. Anders gezegd: 'de tevreden cliënten schreven ons, de minder tevreden cliënten zwegen'.

Aanpak van het attritie- en generaliseerbaarheidsprobleem – De geschetste problemen zijn niet onoverkomelijk. Lage responspercentages bijvoorbeeld kunnen door een gedegen methode van dataverzameling worden vermeden. Zoals eerder gemeld varieerde de gedegenheid echter per kliniek. Tijdens de looptijd van het follow-up-project werd bovendien duidelijk dat de pogingen om de responspercentages die bij de drie vervolgmetingen werden behaald te verbeteren door middel van een halfjaarlijkse bijeenkomst van de dataverzamelaars, onvoldoende resultaat hadden. Die bijeenkomsten hadden tot doel om hun werkzaamheden beter met elkaar in overeenstemming te brengen. De oorspronkelijke procedure werd gehandhaafd en er werd afgezien van een methode die men met 'hinderlijk volgen' zou kunnen aanduiden, om het responspercentage alsnog aanmerkelijk te verhogen. Dillman (Nederhof, 1981) beveelt zulk volgen aan. Bij deze procedure krijgt de geënquêteerde in vrij korte tijd tot viermaal toe de lijst aangeboden. Dit werd echter te bezwaarlijk geacht. Ook het gegeven dat de eenmalige rappèlmethode in de procedure van dataverzameling was afgesproken, maakte de keuze voor het vaker rappelleren niet aantrekkelijk. Een ex-cliënt die tot tweemaal toe een aangeboden vragenlijst niet retourneerde of expliciet meldde op verdere toezending geen prijs te stellen, werd niet opnieuw benaderd.

Zoals vermeld kon niet worden verhinderd dat de attritie van de onderzoeksgroep toenam. Een onderzoek naar de verschillen tussen respondenten en non-respondenten werd daarom noodzakelijk geacht.

Om het ontstane generaliseerbaarheidsprobleem te ondervangen staat de onderzoekers een aantal methoden ter beschikking.

a. Bij het vaststellen van de onderzoeksresultaten wordt expliciet rekening gehouden met de groep niet-respondenten. Zo wordt bijvoorbeeld het gemiddelde van de groep respondenten gecorrigeerd door rekening te houden met fictieve scores van niet-respondenten (Moors en Muylwijk, 1975).

b. Indien de dataverzameling voorziet in meervoudig rappelleren, kan men de verschillen tussen gemaande respectievelijk niet gemaande proefpersonen benutten om een indicatie te verkrijgen over de verschillen tussen responderende en niet-responderende proefpersonen (Moors en Muylwijk, 1975).

c. Tenslotte kan men werken volgens een methode die bestaat uit het verzamelen van zoveel mogelijk andersoortige gegevens over de non-respondenten om deze vervolgens te vergelijken met die van de respondenten.

Methode c werd in eerste instantie toegepast op de patiëntengroep uit het vooronderzoek (Van Kalsbeek c.s., 1978), beschreven in de paragraaf over de resultaten. Recent werd deze methode herhaald op een deel van de cliëntengroep uit het hoofdonderzoek, aangevuld met methode b (Koning en Tremonti, 1984), beschreven in de paragraaf over de non-respons.

Non-respons bij het vooronderzoek

Opzet – In het kader van het vooronderzoek naar de psychometrische kwaliteiten van de Nieuwkoopse Vragenlijst, werden 835 vervolgljsten verzonden. Van de aangeschreven cliënten zond slechts 60% een bruikbare lijst retour. Van zowel de groep respondenten als de groep niet-respondenten waren gegevens uit de periode van hun opname beschikbaar, zodat het mogelijk was de verschillen tussen beide groepen te bestuderen. Er kon dus onderzocht worden of de groep niet-respondenten al tijdens het verblijf in een psychotherapeutische gemeenschap verschilden van de groep respondenten.

Bij het onderzoek werd gebruikt gemaakt van gegevens die op drie tijdstippen werden verzameld:

1. opnamegegevens (administratieve en testgegevens);
2. verblijfsgegevens (opnameduur, benutten van therapieaanbod);
3. ontslaggegevens (stafbeoordelingen, testgegevens).

Resultaten – Onder verwijzing naar het onderzoeksverslag (Van Kalsbeek c.s., 1978) wordt hier volstaan met een korte samenvatting van de resultaten. Bij de vergelijking van de respondenten met de niet-respondenten wat betreft hun psychische conditie bij *opname*, bleken zich geen verschillen voor te doen. Op andere aspecten verschilden de groepen echter wel. De mensen van wie later niets werd vernomen

waren jonger, minder intelligent, gingen meer af op hun eigen mening, waren minder gewetensvol en hadden meer weerstand tegen psychologische vragenlijsten dan de respondenten.

Voor zover daar gegevens over verzameld waren bleken de respondenten actiever gebruik te hebben gemaakt van het therapieaanbod gedurende hun *verblijf*. Bovendien bleken zij ook meer groepsgeoriënteerd te zijn. In enkele kliniepopulaties bleken zich onder de niet-respondenten vooral kortverblijvers te bevinden. Respondenten en niet-respondenten bleken ook van elkaar te onderscheiden te zijn indien werd gekeken naar hun conditie bij *ontslag* uit de klinieken. De respondenten konden worden aangeduid als personen die het meeste baat hadden bij hun verblijf in de psychotherapeutische gemeenschap – afgemeten aan de afname van klachten – en die ook meer tevreden waren over hun verblijf dan de niet-respondenten. Ook verschillen in de beoordeling door stafleden wezen in die richting.

Voor de volledigheid moet worden opgemerkt dat deze conclusies niet steeds gebaseerd zijn op de resultaten van dezelfde kliniepopulaties, zodat niet zonder meer mag worden aangenomen dat zij gegeneraliseerd kunnen worden voor de totale onderzoeksgroep. De verschillen in de conditie bij ontslag werden uitsluitend gevonden bij de klinieken van de Viersprong.

Non-respons bij het hoofdonderzoek

Opzet – De zojuist beschreven resultaten van het vooronderzoek, alsmede de bij het project geboekte attritie, noopten tot een herhaling van de onderzoeksprocedure bij de populatie van het hoofdonderzoek ($N=1536$). Daarbij is als volgt te werk gegaan. Er is allereerst nagegaan of de niet-respondenten ($N=35$) van een *tweetal* aan het project deelnemende klinieken verschilden van de patiënten die hun lijst 1 jaar na hun ontslag wel retourneerden ($N=143$), waarbij ook deze keer onderzocht werd of het mogelijk is om achteraf verschillen te ontdekken tussen respondenten en niet-respondenten ten tijde van hun verblijf in de therapeutische gemeenschap. Daartoe werd gebruik gemaakt van de volgende gegevens:

1. opnamegegevens (administratieve gegevens als sekse en leeftijd, en testdata als MMPI-scores en NVL-gegevens),
2. verblijfsgegevens (opnameduur),
3. ontslaggegevens (beoordelingen door therapeuten en NVL-gegevens).

Ditmaal werd ook gebruik gemaakt van:

4. follow-up-gegevens (frequentie van heropname, administratieve gegevens, ontleend aan de Patiëntenregistratie Intramurale Geestelijke Gezondheidszorg van de Hoofinspectie van de Geestelijke Volksgezondheid).

Bovendien werd nagegaan of er verschillen waren tussen cliënten die

direct hun vervolglst retourneerden en cliënten die daartoe eerst gemaand moesten worden.

Er werd besloten onderzoek uit te voeren bij de cliënten van een tweetal klinieken. De keuze viel op deze klinieken vanwege hun niet al te grote cliëntenpopulaties en de beschikbaarheid van veel psychologische testdata. Deze klinieken vertegenwoordigden elk een bepaald type therapeutische gemeenschap, de ene is een neurosekliek (Spiegelberg) en de andere is een kliniek voor zwakgeïntegreerde en psychotische cliënten (Jozef). De vraag doet zich voor of deze op praktische gronden gemaakte keuze het trekken van generaliserende conclusies niet bemoeilijkt. Het gaat namelijk om twee klinieken die zich qua responspercentages in positieve zin onderscheiden van de andere klinieken.

Een beschouwing van de administratieve procedures van alle acht deelnemende klinieken leidde tot de stelling dat de gevonden verschillen in responspercentages voor een belangrijk deel zijn te herleiden tot verschillen in wijze van dataverzameling. Uitgaande van deze stelling lijkt het gerechtvaardigd om het onderzoek te verrichten bij een tweetal klinieken dat de dataverzameling gedurende de gehele looptijd volgens de oorspronkelijke richtlijnen uitvoerde. Men vergroot hierdoor de kans dat de groep niet-respondenten voornamelijk uit 'echte' weigeraars (Nederhof, 1981) zal hebben bestaan.

Een andere bron van informatie over de generaliseerbaarheid van de resultaten zijn de verschillen tussen respondenten en non-respondenten van het *totale* onderzoekscohort. Er werd, weliswaar minder gedetailleerd, nagegaan of er verschillen aantoonbaar waren tussen cliënten die één of drie jaar na hun ontslag hun lijst wel of niet retourneerden. Daartoe werden uitsluitend NVL-gegevens gebruikt.

De nulhypothese die in beide onderzoekszetters getoetst werd, luidde: er is geen verschil tussen respondenten en non-respondenten. Het aanvaarden van deze hypothese zou sterk pleiten voor de generaliseerbaarheid van de resultaten van het follow-up-project.

Omdat het in dit geval gaat om een fundamentele uitspraak, werd een voorzichtige opstelling gekozen en werd het significantie-niveau op .10 gesteld. Dat betekent dat de kans op het constateren van verschillen als ze er in feite niet zijn (fout van de tweede soort) groter wordt. Er werd met andere woorden de voorkeur aan gegeven de kans dat men ten onrechte besluit dat er geen verschillen zijn, te verkleinen, zodat de onderzoekers met betrekking tot de generaliseerbaarheid van de resultaten liever ten onrechte twifelen dan dat zij er ten onrechte positieve uitspraken over doen. Het eerste werd minder bezwaarlijk geacht, omdat het tot de mogelijkheden behoort om in het geval van significante verschillen het eindoordeel over de generaliseerbaarheid op te schorten en nader onderzoek te verrichten.

Respondenten versus niet-respondenten van twee klinieken – Het

onderzoek werd verricht op de gegevens van 178 cliënten. Van de totale groep retournerden 35 cliënten hun vragenlijsten niet (19%). Voor de goede orde werd onderzocht bij wie van deze 35 sprake was van uitval door adreswijziging en door mortaliteit. Daarbij bleek onder andere dat de groep expliciete weigeraars – dat is de groep die schriftelijk of telefonisch te kennen had gegeven niet aan het onderzoek te willen meewerken – niet te verwaarlozen is; bij één kliniek betrof het bijvoorbeeld 6 van de 27 niet-respondenten. Er is geen verschil tussen respondenten en niet-respondenten wat betreft leeftijd en sekse.

De bij opname afgenomen MMPI (boekjesvorm, B. Nuttin en Beuten, 1969) laat op de meeste klinische schalen en aanvullende schalen geen verschil zien tussen hen die later wel en niet zouden responderen. Voor zover er wel verschillen aanwijsbaar zijn, wijzen zij op een relatief sterke geneigdheid tot het geven van deviante antwoorden, benadrukken van problemen en non-conformisme bij de latere niet-respondenten. Tevens is er sprake van een verhoogd wantrouwen. Deze feiten liggen in het verlengde van de bevindingen uit het vooronderzoek ($F:t=-1.78$, $p<.10$, $K:t=1.96$, $p<.10$, $Pd:t=-1.86$, $p<.10$, $Pa:t=1.69$, $p<.10$). De geringe verschillen wijzen op een verschil tussen respondenten en niet-respondenten, dat – gelet op het feit dat het een opname-meting betreft – niet gerelateerd kan worden aan mogelijke onvrede met de genoten behandeling. Een nadere inspectie van de gegevens van de afzonderlijke klinieken leverde verschillen op die bij beide klinieken wijzen op het meer benadrukken van klachten en symptomen bij hen die later niet-respondent blijken te worden ($F:t=-1.99$, $p<.10$, $HS:t=-1.74$, $p<.10$, $K:t=2.17$, $p<.10$).

De schaalscores op de Nieuwkoopse Vragenlijst ondersteunen het door bovengenoemde verschillen ontstane vermoeden dat niet-respondenten bij opname een problematischer beeld van zichzelf presenteren dan respondenten. Bij opname zijn verschillen tussen respondenten en niet-respondenten aantoonbaar, gelet op:

- welbevinden (lager voor de latere niet-respondenten); $t=1.95$, $p<.10$.
- moeilijkheden ondervonden in relaties met familieleden (meer); $t=-1.97$, $p<.10$.
- moeilijkheden ondervonden op de verschillende leefgebieden (meer); $t=-2.24$, $p<.05$. (Zie voor een verklaring van de verschillende variabelen Wagenborg c.s., 1984.)

Wat betreft de *opnameduur* vallen geen verschillen te constateren tussen degenen die later wel of niet hun lijst retournerden.

Bij *ontslag* zijn er nauwelijks verschillen tussen latere respondenten en niet-respondenten aanwijsbaar. Allereerst werd gekeken naar het oordeel van stafleden over de betrokken patiënten. Het betrof retrospectieve stafoordelen over onder meer de mate waarin met de betrokken cliënt productief gewerkt werd. Het oordeel van stafleden over het gedrag van de groep respondenten en de groep niet-respondenten tij-

dens de behandeling bleek niet te verschillen. De NVL-gegevens laten bij ontslag op schaalniveau geen verschillen zien. Op de vraag of de behandeling aansloot bij de problemen waarvoor men kwam, antwoorden de latere niet-respondenten kritischer ($t=1.86$, $p<.10$). Deze bevinding is consistent met de bevindingen uit het vooronderzoek, die erop wijzen dat 'tevreden klanten schrijven'. In tegenstelling tot hetgeen bij opname gold, kon geconstateerd worden dat de verschillen bij ontslag gering zijn.

Met betrekking tot de *follow-up*-gegevens is allereerst onderzocht of er verschillen zijn tussen de groep respondenten en de groep niet-respondenten wat betreft het aantal heropnames in instellingen voor de geestelijke gezondheidszorg. De gegevens over heropnamefrequentie zijn ontleend aan de Patiënten Registratie Intramurale Geestelijke Gezondheidszorg (PIGG). De hierboven beschreven data werden verzameld tijdens de behandelfase. De PIGG-gegevens daarentegen zijn afkomstig uit een periode die grotendeels parallel loopt aan de periode waarin de vervolgljsten werden verzameld. Er is nagegaan of de cliënten na ontslag uit een van de klinieken nog één of meer keren werden heropgenomen. Het blijkt dat de groep niet-respondenten op dit punt niet afwijkt van de groep respondenten. Dat geldt voor de totale onderzoeksgroep, maar ook na opdeling per therapeutische gemeenschap. Tevens werd gekeken of er in de maatschappelijke omstandigheden van de ex-clieënten zoals geregistreerd op het moment van heropname, verschillen waren. Zo werd nagegaan of er verschillen waren met betrekking tot burgerlijke staat, opleidingsniveau, bron van inkomsten, woonomstandigheden, diagnose en juridische status. Ook de lengte van de eerste heropname werd bij de berekeningen betrokken. De frequentieverdeling van de verschillende variabelen over de beide groepen liet in de meeste gevallen geen statistische bewerking toe, zodat volstaan moest worden met een globale impressie. In het algemeen kan gesteld worden dat er voor de genoemde variabelen geen verschil bestaat tussen de groep respondenten en de groep niet-respondenten. Alleen wat betreft het opleidingsniveau zijn enige verschillen te constateren. De categorie basisonderwijs is in de niet-respondentengroep licht oververtegenwoordigd.

Gerappelleerden versus niet-gerappelleerden van twee klinieken – In het bovenstaande werd verslag gedaan van het zoeken naar verschillen tussen respondenten en non-respondenten.

Binnen de groep respondenten is echter ook een onderscheid aan te brengen tussen degenen die in eerste instantie reageerden en degenen die na het ontvangen van een rappel hun lijst inzonden. Van deze twee groepen werden op de door hen ingezonden vervolgljsten schaal- en itemscores berekend.

Van de 143 respondenten reageerden er 96 op het eerste verzoek en 47 (33%) na een tweede oproep. Het bleek dat de ex-clieënten van de

Spiegelberg meer gemaand moesten worden dan hun collegae van Jozef ($X^2=8.21$, $df=1$, $p<.01$). Op schaalniveau manifesteren zich tussen beide groepen geen verschillen. Naast de vergelijkingen op schaalniveau vonden ook vergelijkingen op itemniveau plaats. Daartoe werden items uitgezocht die binnen de context van een non-responsonderzoek relevant zouden kunnen zijn. Het betreft onder andere 27 items die informeren naar de behoefte aan (geprolongeerde) behandeling; die inventariseren welk effect de onderzochten aan de behandeling toeschrijven en welke consequenties die behandeling voor hun maatschappelijk functioneren had. Slechts twee items vertonen verschillen. Het blijkt dat de gerappelleerde cliënten vaker behoefte uitspreken aan voortgezette behandeling (voornamelijk in de vorm van gesprekken) en dat zij frequenter aangeven dat hun omgeving hen discrimineert als gevolg van hun opname ($t=-2.08$, $p<.05$; $t=-1.91$, $p<.10$). Analooq aan de opzet van het respondenten/non-respondentenonderzoek is onderzocht in hoeverre er sprake was van specifieke kliniekverschillen bij de vergelijking van gerappelleerden en niet-gerappelleerden op de bovenbeschreven variabelen. Daarbij bleek dat de gerappelleerden van de 'psychose'-kliniek zich profileren als personen die kennelijk minder baat bij de behandeling hadden, maar dat zulks voor de gerappelleerden van de 'neurose'-kliniek zeker niet geldt.

Respondenten versus non-respondenten van het gehele cohort – Tot slot is onderzocht in hoeverre de respondenten op de een en drie jaar follow-up-meting van het gehele cohort verschillen van de non-respondenten wat betreft de NVL-gegevens die bij opname, ontslag en follow-up werden verzameld. Dit laatste is mogelijk omdat een non-respondent na drie jaar bij de eerste follow-up-meting nog tot de respondenten kan behoren.

Allereerst werd nagegaan hoe de non-respondenten zich per kliniek verdelen over de categorieën: aperte weigering, overleden, onbestelbaar retour, onbewerkbaar retour en reden niet te achterhalen. Het bleek dat bij ongeveer 70% van de gehele groep non-respondenten niet achterhaald kon worden wat de reden van het non-responderen was. Rond 8% weigerde expliciet deel te nemen, een kleine 15% der lijsten werd als onbestelbaar retour gezonden. De bekende mortaliteit bedroeg 7%. Globaal genomen leverde de dataverzameling per afzonderlijke kliniek vergelijkbare percentages op. Er is echter ook een klein aantal kliniekverschillen in weigerings- en mortaliteitspercentages aanwijsbaar. Het aantal onbestelbare lijsten steeg bij de dataverzameling na 3 jaar. Deze stijging manifesteerde zich het sterkst bij de jeugdklinieken.

Er kon voorts worden vastgesteld dat bij cliënten uit de neuroseklinieken bij de eerste vervolgmeting meer respons kon worden genoteerd dan bij cliënten uit de andere klinieken ($X^2=4.49$, $df=1$, $p<.05$) en dat het verschil in de responspercentages bij de tweede vervolgmeting niet

meer aantoonbaar was. Tevens bleken cliënten die drie maanden of korter in de diverse klinieken verbleven (drop-outs) zowel op de eerste vervolgmeting als de tweede meting minder te responderen dan cliënten die een langere periode in klinieken vertoefden (vervolg na 1 jaar: $X^2=7.76$, $df=1$, $p<.011$; vervolg na 3 jaar: $X^2=4.76$, $df=1$, $p<.05$).

Vervolgens werd onderzocht of er in het in een eerder stadium verzameld NVL-materiaal verschillen aantoonbaar waren tussen respondenten en non-respondenten. Voor de totale groep werd voor de non-respondenten van de eerste vervolgmeting bij opname geen verschil gevonden met de respondenten. Uit de verblijfsgegevens bleek dat de non-respondenten van de eerste vervolgmeting een kortere verblijfsduur hadden ($t=2.88$, $p<.01$). De non-respondenten van de tweede vervolgmeting bleken bij opname jonger te zijn ($t=2.60$, $p<.01$), boekten eveneens een korte verblijfsduur ($t=3.42$, $p<.01$), maar verschilden evenals de non-respondenten van de eerste vervolgmeting niet wat betreft hun scores op de schalen welbevinden en nuchterheid (zowel bij opname als bij ontslag). Op kliniekniveau waren op de schaal welbevinden wel verschillen aantoonbaar.

Non-respondenten van de eerste vervolgmeting van Jozef scoorden bij opname lager dan respondenten ($t=2.01$, $p<.05$) en die van Rijnland hoger ($t=-2.23$, $p<.05$). Amstelland non-respondenten scoorden bovendien hoger op de schaal nuchterheid ($t=-2.05$, $p<.05$). Dergelijke verschillen per kliniek deden zich ook voor bij de non-respondenten van de tweede vervolgmeting. Bij opname scoorden Spiegelbergers hoger op de schaal nuchterheid ($t=-2.37$, $p<.01$). Onder de respondenten van Zuideroord bevonden zich meer vrouwen ($X^2=5.69$, $df=1$, $p<.05$). Op het eerste vervolgmeetmoment waren er voor de tweede vervolgmeetmoment-respondenten en non-respondenten geen verschillen aantoonbaar op de schalen welbevinden en nuchterheid.

Kanttekeningen bij de significante verschillen – In de verschillende deelonderzoeken werd steeds de hypothese getoetst dat er geen verschil bestaat tussen respondenten en non-respondenten. Van een grote hoeveelheid variabelen werd onderzocht in hoeverre zij discrimineren tussen de genoemde groepen. Bij een dergelijke procedure moet, louter op grond van toeval, verwacht worden dat er op z'n minst een aantal variabelen significante verschillen oplevert. Van de meer dan 70 op de ongedeelde onderzoeksgroepen van de twee klinieken en het hoofdcohort uitgevoerde toetsen vielen er 12 significant uit. Van den Ende en Verhoef (1973) merken op dat wanneer men, zoals in dit onderzoek, een netwerk van samenhangende hypothesen toetst, men pas uitspraken kan doen als een bepaald aantal toetsen het significantieniveau te boven gaat. Bij deze werkwijze loopt men echter het risico dat inhoudelijk interessante verschillen van afzonderlijke variabelen als toevalsfluctuaties moeten worden beschouwd. Grove en Andreasen (1982) suggereren een andere oplossing voor het probleem van de

simultane toetsing van meerdere hypothesen. Zij stellen voor de individuele significantieniveaus scherper te stellen op geleide van het aantal variabelen of hypothesen. Het voordeel is dat specifieke variabelen in de analyse blijven. Het nadeel is echter dat de verschillen tussen de groepen heel groot moeten zijn om de aangescherpte significantiegrenzen te overschrijden. Omdat het in dit onderzoek gaat om het opsporen respectievelijk uitsluiten van verschillen tussen respondenten en non-respondenten op een veelheid van variabelen en omdat het niet op voorhand bekend is op welke variabelen deze verschillen zich zullen voordoen, is gekozen voor de methode zoals beschreven door Grove en Andreasen. Wanneer deze correctieregel (de Bonferroni inequality) toegepast wordt op de verschillen zoals die gevonden werden, dan vindt de volgende herwaardering plaats. Bij het onderzoek op de twee geselecteerde kliniepopulaties werd alpha per groep variabelen op .10 gesteld. Voor de 24 variabelen 'conditie bij opname' leverde dat een individuele alpha op van .004. Voor de groep van 25 variabelen 'conditie bij ontslag' was dat eveneens .004. En voor de 'conditie bij follow-up' werd alleen de heropnamefrequentie getoetst en bleef alpha derhalve .10. Geen enkele variabele of cluster van variabelen leverde na deze hertoetsing verschillen op. Voor de hoofdonderzoeksgroep werd alpha per cluster variabelen (3) eveneens op .10 gesteld, hetgeen individuele alpha's brengt op .03. Uit de groep opnamevariabelen van de (non-)respondenten op de eerste vervolgmeteren na hertoetsing een verschil in 'welbevinden' (Rijnland) en in 'nuchterheid' (Amsteland). Voor de ontslagvariabelen resteert een verschil in verblijfsduur (Rijnland). Voor de (non-)respondenten van de tweede vervolgmeteren liggen de verschillen weer anders. Er is bij opname een verschil in nuchterheid (Spiegelberg), leeftijd (totale groep) en sekse (Zuideroord). Bij ontslag is er verschil in verblijfsduur (Viersprong).

Discussie

Op grond van gegevens zoals die in het voorgaande gerapporteerd werden, mag gesteld worden dat er betrekkelijk weinig verschillen bestaan tussen zowel de respondenten en non-respondenten van de kliniepopulaties die gedetailleerd onderzocht zijn, als de respondenten en non-respondenten van het hele onderzoekscohort. Voor belangrijke variabelen zoals heropnamefrequentie en conditie aan het eind van de behandeling zijn geen verschillen te constateren.

Toch is het beeld niet helemaal rimpelloos. Bij de opnamemeteren zijn bij sommige PTG-en enige verschillen aantoonbaar in de psychische conditie. Deze bevinding komt slechts gedeeltelijk overeen met de resultaten uit het vooronderzoek waarbij geen verschil gevonden werd in psychische conditie maar wel in leeftijd, intelligentie en attitude ten opzichte van de vragenlijsten. In het vooronderzoek werden bovendien verschillen aangetoond in de conditie bij ontslag. Een ver-

klaring van deze tegenstrijdige resultaten kan voor een deel gevonden worden in het feit dat de conclusies betrekking hebben op steeds andere onderzoekssubgroepen. Zo bleek bij nadere inspectie van de vooronderzoeksgegevens dat de conclusies omtrent de conditie aan het eind van de behandeling hoofdzakelijk gebaseerd zijn op gegevens van slechts een tweetal der deelnemende klinieken, die overigens wel een groot percentage vormen van het onderzoekscohort (Viersprong Jeugd- en Viersprong Volwassenenkliniek). Dit gegeven relativiseert de slotconclusie van het vooronderzoek dat er bij ontslag grote verschillen zouden bestaan tussen respondenten en non-respondenten. Bovendien laten de non-respondenten van het hoofdonderzoek op de twee vervolgmetingen – die afkomstig zijn uit alle aan het project deelnemende klinieken – bij het ontslagmoment geen verschil zien met de respondenten. Ook uit de resultaten van de gedetailleerde analyses op de twee geselecteerde klinieken blijken geen verschillen voor wat betreft de psychische conditie aan het eind van de behandeling.

De verschillen tussen respondenten en non-respondenten doen zich vooral voor bij de variabelen leeftijd en lengte van het verblijf in de PTG. Soms doen zich ook verschillen voor in psychische conditie bij opname, maar deze verschillen zijn niet consistent. Het is niet zo verwonderlijk dat de gevonden verschillen zo inconsistent zijn. Het indelingscriterium 'non-respondent' leidt in de verschillende onderzoeken tot groepen van wisselende omvang en samenstelling. De verschillen die in meer onderzoeken terugkomen (zoals verblijfsduur), zijn kennelijk minder gevoelig voor de specifieke samenstelling van de groepen en hebben mogelijk een meer algemene betekenis. De overige verschillen doen vermoeden dat sprake is van weinig kenmerkende verschillen tussen respondenten en non-respondenten. Desalniettemin roepen zij de vraag op in welke mate zij van invloed zijn op de generaliseerbaarheid van de resultaten. Voor een deel zijn de gevonden verschillen verklaarbaar. Dat non-respondenten jonger zijn heeft mogelijk te maken met de grotere sociale mobiliteit die kenmerkend is voor deze leeftijdsfase. Dat zou betekenen dat de non-respondenten geen weigeraars zijn, maar slecht traceerbare respondenten. De verschillen in conditie bij opname en de verschillen in verblijfsduur zijn minder makkelijk te verklaren. Zij zijn weliswaar inconsistent en specifiek voor klinieken, maar omdat gekozen is voor het benadrukken van verschillen mogen zij niet zonder meer buiten beschouwing gelaten worden. Het verschil in lengte van verblijf werd wel meermalen gevonden, zij het niet voor alle klinieken. Deze bevindingen vormen een belemmering voor de generaliseerbaarheid op kliniekniveau. Het gegeven dat zich onder de non-respondenten relatief meer drop-outs bevinden, rechtvaardigt deze voorzichtige opstelling.

In het kader van het onderzoek naar de lange-termijneffecten van klinische psychotherapie is het gegeven dat er op het moment van de follow-up-meting geen verschillen aangetoond konden worden in her-

opnamefrequentie en het gegeven dat eveneens geen verschillen aantoonbaar waren in de psychische conditie op het moment van ontslag, van doorslaggevend belang. De slotconclusie van dit non-responsonderzoek moet luiden dat wanneer men afziet van verschillen die specifiek zijn voor klinieken, de bevindingen van de respondentengroep gegeneraliseerd mogen worden naar de totale onderzoeksgroep. Daarbij dient men te bedenken dat algemene uitspraken niet zonder meer geldig zijn voor alle therapeutische gemeenschappen. Voor twee van de acht aan het project deelnemende klinieken is aannemelijk gemaakt dat er geen verschillen zijn tussen respondenten en non-respondenten. Volledigheidshalve zou daarom een onderzoek bij andere kliniepopulaties op zijn plaats zijn. Vooralsnog mag ervan worden uitgegaan dat de representativiteit van de onderzoeksbevindingen niet al te zeer lijkt te worden bedreigd door het attritie-fenomeen.

Literatuur

- Bolten, M. P. (1975), *Meten is weten*. Werkgroep Follow-up Project PTG, rapport nr. 1, 1976.
- Dillman, D. A., Mail and telephone surveys. The total design method. New York 1978. In: A. I. Nederhof, *Beter Onderzoek*. Dienst sociaal-wetenschappelijk onderzoek, Rijksuniversiteit Leiden, Den Haag 1981.
- Ende, H. van den, M. Verhoef (1973), *Inductieve statistiek voor gedragswetenschappen*. Agon Elsevier, Amsterdam.
- Greenland, S. (1977): Response and follow-up bias in cohort studies. *American Journal of Epidemiology*, 6, no. 3, p. 184-187.
- Grove, W. M., N. C. Andreasen (1982), Simultaneous Tests of Many Hypotheses in Exploratory Research. *Journal of Nervous and Mental Diseases*, 170, 1, p. 3-8.
- Kalsbeek, A. van, G. Lantink en J. E. A. Wagenborg (1978), *Een vergelijking van respondenten met niet-respondenten bij het follow-up project psychotherapeutische gemeenschappen*, rapport 7.
- Kish, G. B., H. T. Hermann, The fort meade alcoholism treatment program: a follow-up study. *Quart. Journal stud. Alcohol*, 32, p. 628-635.
- Koning, R. F., G. W. Tremonti (1984): *Het non-responsprobleem*, Werkgroep 'Follow-up project PTG', rapport 18.
- Koster, A. M., J. E. A. Wagenborg, A. J. Hesselink, *Gegevens van ex-patiënten 3 jaar na hun ontslag uit de PTG*. Werkgroep 'Follow-up project PTG', rapport 20 (in voorbereiding).
- Larsen, D. J., C. C. Attkisson, T. D. Hargreaves (1979): Nguyeh, Assessment of client-patient satisfaction: development of a general scale. *Eval. and Prog. Planning*, 2, p. 197-207.
- Liberman, B. L. (1978): The maintenance and persistance of Change-Long-term follow-up Investigations of Psychotherapy. In: J. D. Frank (red.), *Effective ingredients of succesful psychotherapy*, New York, Bruner/Mazel.
- May, R. A., A. Hussain Tuema, W. J. Dixon (1976): Schizophrenia - A follow-up study of results of treatment I design and other problems. *Archives Gen. Psychiatry*, 33, p. 474-478.

- Moors, I. J. A., J. Muylwijk (1975), *Steekproeven, een inleiding tot de praktijk*. Agon Elsevier.
- Moser, C. A., G. Kalton (1971), *Survey methods in social investigation*. Heinemann Educational Books, LTD.
- Munby, M., D. W. Johnston (1980), Agoraphobia: the long term follow-up of behavioural treatment. *British Journal Psychiatry*, 137, p. 418-427.
- Nederhof, A. J. (1981), *Beter onderzoek*, Dienst Sociaal Wetenschappelijk Onderzoek, RU Leiden, Den Haag.
- Nuttin, J., L. B. Beuten (1969), *Handleiding bij de persoonlijkheidsinventaris MMPI*, Leuven Universitaire Uitgaven, Leuven.
- Rosenthal, R., R. L. Rosnow (1975), *The Volunteer Subject*, New York.
- Wagenborg, J. E. A., M. A. Koster (1984), *Gegevens van ex-patiënten 1 jaar na hun ontslag uit de PTG*. Werkgroep Follow-up project PTG', rapport 19.
- Wagenborg, J. E. A., G. W. Tremonti, A. J. Hesselink & R. F. Koning (1984), Het Follow-up Project Psychotherapeutische Gemeenschappen, 1: Vraagstelling, opzet en eerste resultaten, *Nederlands Tijdschrift voor Psychiatrie*, 26-6, p. 377-401.
- Warwick, D. P., C. A. Lininger (1975), *The sample survey: theory and practice*. McGraw-Hill.