

Een model voor het bepalen van psychiatrische praevalentie: de relatie GHQ-PSE

door P. P. G. Hodiament en S. H. J. Veling

Inleiding

De Wereldgezondheidsorganisatie stelt de psychiatrische epidemiologie primair ten doel: 'to assess the prevalence of different types of mental-ill health in a population, as a basis for the prevention, treatment and control of these diseases'. (Lin and Standley, 1972). Praevalentieonderzoek wordt verricht om een indruk te krijgen van het aantal mensen binnen een populatie, dat op een zeker moment lijdt aan een zekere mate van psychische gestoordheid. (Reid, 1960). Als men zich beperkt tot de studie van patiënten die onder behandeling zijn, wordt het beeld van de psychiatrische morbiditeit vertroebeld door factoren die te maken hebben met ziekte- en hulpzoekgedrag. Voor een zuiver beeld van de problematiek is men dus aangewezen op (vanwege het grote aantal niet-zieken) tijdrovend onderzoek van de onbehandelde populatie. Een dergelijke praevalentie-survey kan bijdragen tot een antwoord op diverse vragen, bijvoorbeeld naar de last van langdurige stoornissen voor een gemeenschap, of naar het mogelijke effect van therapeutische maatregelen op het chronisch worden van psychiatrische aandoeningen. Bovendien vormen gelokaliseerde verschillen in morbiditeitsniveaus een belangrijke leidraad voor de planning van voorzieningen, zoals de formatie van veldteams in de ambulante geestelijke gezondheidszorg. In associatie met andere variabelen liggen praevalantiegegevens ten grondslag aan allerlei hypothesen over causale relaties, die, indien bevestigd, een aanknopingspunt bieden voor preventief en therapeutisch handelen. Vergelijking van de praevalentie der verschillende populaties is de eerste stap op weg naar

Schrijvers zijn respectievelijk (1) psychiater, werkzaam in het RIAGG en aan het Instituut voor Psychiatrie van de Katholieke Universiteit Nijmegen, tevens verbonden aan de afdeling sociale psychiatrie van de GG & GD Eindhoven en (2) statisticus, hoofd Technisch Wetenschappelijke Toepassing van het Computerinstituut Woudenstein van de Erasmus Universiteit Rotterdam.

het leggen van oorzakelijke verbanden. Vergelijkbaarheid veronderstelt echter een duidelijke definitie en identificatiemethode voor psychiatrische gevallen.

Tot voor kort waren deze condities eerder uitzondering dan regel. Gedurende het onderzoekstijdperk, waarin het concept van een geval bijvoorbeeld bepaald werd door het subjectieve oordeel van de respondent – the perceived morbidity – liep de gevonden praeventie uiteen van 1 tot 69%. (Dohrenwend, 1974). Toepassing van gestandaardiseerde interviews met welomschreven psychiatrische criteria had een drastische reductie van de variatie in praeventiecijfers tot gevolg. Daarmee was een belangrijke voorwaarde vervuld voor vergelijkingen in eerdergenoemde zin. (Goldberg, 1980). Uit het oogpunt van doeltreffendheid worden deze gestandaardiseerde meetinstrumenten tegenwoordig gebruikt in combinatie met screeningstechnieken.

Screening is de ingeburgerde term voor een zeefprocedure, bedoeld om waarschijnlijke ziektegevallen* te selecteren uit het totaal van de bevolking. Diagnostiek is in deze fase niet aan de orde. (Wilson & Jungner, 1968). De twee-fasen-strategie biedt de psychiatrische epidemiologie een kader voor screening en vervolgens bevestiging en beschrijving van haar casuïstiek. (Eastwood, 1971; Goldberg, 1974; Henderson, Duncan-Jones, Byrne, Scott and Adcock, 1979). In de eerste fase wordt (een steekproef uit) de bevolking met behulp van een vragenlijst gescreend om potentiële gevallen te identificeren, die in tweede instantie door een psychiater worden onderzocht ter bevestiging van hun 'geval-zijn' en voor beschrijving van hun problematiek. (Williams e.a., 1980). De hoofdrol die de psychiater speelt in het geheel, hangt samen met de overtuiging, dat zijn klinische oordeel vooralsnog de beste maatstaf is voor geestesziekte en als zodanig '... the ultimate criterium against which other means for identifying psychiatric disorder are validated'. (Blum, 1962). De betrouwbaarheid van het psychiatrische oordeel, die lang te wensen over heeft gelaten, is aanzienlijk verbeterd door standaardisatie van de interviews en opzet van gemeenschappelijke trainingen voor interviewers.

Dit artikel bechrijft onze ervaring met de combinatie van een vragenlijst, the General Health Questionnaire, en een gestandaardiseerd psychiatrisch interview, the Present State Examination, in een regionaal praeventieonderzoek. We behandelen de instrumenten afzonderlijk en noemen enige condities voor hun gezamenlijk gebruik in een bevolkingssurvey. De samenhang van GHQ en PSE werd bestudeerd aan de hand van een logistisch model, waarvan de toepassing het mogelijk maakt praeventiecijfers, en bovendien sensitiviteit en specificiteit van het screeningsinstrument, voor de betreffende populatie te bepalen op basis van de gevonden GHQ-scores alleen. Voor de gezond-

* De termen (ziekte)geval, casus en case, worden in dit artikel als synoniemen beschouwd en gebruikt.

heidsregio Nijmegen zullen we beide bewerkingen uitvoeren en de resultaten presenteren.

The *General Health Questionnaire* is een door de betrokkene zelf in te vullen vragenlijst, door Goldberg (1972) ontworpen voor de opsporing van psychiatrische aandoeningen bij respondenten in de eerste lijn. De vragen hebben betrekking op het optreden van ongewone en onaangename psychische verschijnselen en op onvermogen om normaal te (blijven) functioneren, kenmerken die men aantreft bij het merendeel van de psychiatrische syndromen. Door haar opzet mist de vragenlijst in principe defensieve en demente respondenten, karakterafwijkingen en chronische symptoomneurosen. Bepaalde psychosen, waarvoor de GHQ niet bedoeld is, worden wel geregistreerd, wanneer zij de eerder genoemde kenmerken van psychofysiologisch ongemak vertonen. Hoewel de psychiatrische problematiek, waarop de GHQ is gericht, algemeen gezien wordt als een kwantitatieve afwijking van de norm en de GHQ-score dus een plaats betekent op een as, die loopt van normaliteit tot geestesziekte, zijn er argumenten aan te voeren voor een dichotomie in dit continuüm tussen non-casus enerzijds en casus anderzijds. Artsen, epidemiologen en beleidsmakers gaat het immers om gevallen die ze respectievelijk willen behandelen, vergelijken en tellen. (Goldberg, 1978). In dit verband geldt de GHQ-score als een schatting van de waarschijnlijkheid dat iemand een psychiatrische casus is.

Als basis voor zijn vragenlijst zocht Goldberg 140 items op het gebied van depressie, angst, hypochondrie en waarneembaar (sociaal inadequate) gedrag, waarvan hij veronderstelde dat ze relevant waren voor het onderscheid tussen de groep psychiatrische patiënten en de 'normalen'. Bij elk item dient de respondent zijn huidige ('the past few weeks') situatie te vergelijken met de voor hem gewone toestand en vervolgens één van de antwoordcategorieën – minder dan anders, of helemaal niet; hetzelfde als anders; meer dan anders; veel meer dan anders – te onderschrijven. Het aantal malen, dat één van de laatste twee categorieën is aangekruist, vormt de totale score van de GHQ. De scoringsmethode en de formulering van items en antwoorden zijn zo gekozen, dat fouten als gevolg van de neiging om het einde of midden aan te kruisen en in te stemmen of te ontkennen, tot een minimum zijn gereduceerd. (Goldberg, 1972, 1978). Door selectie van de vragen die het best discrimineerden tussen niet-zieken, licht psychiatrisch gestoorde, en ernstig psychiatrisch gestoorde patiënten en die bovendien op een aantal klinisch relevant geachte factoren uit de principale componentenanalyse de hoogste lading vertoonden, ontstond de oorspronkelijke 60 item-versie van de GHQ.

Zowel in een split-half als in een test-hertest evaluatie bleken de betrouwbaarheidscoëfficiënten van deze versie + 0.90 of hoger. Zoals eerder gesteld, geldt het psychiatrische oordeel als het uiteindelijke criterium waaraan (concurrente) validiteit van vragenlijsten wordt

afgemeten. Het screenend vermogen van dergelijke lijsten hangt verder af van het percentage psychiatrische gevallen, dat door de test correct wordt geïdentificeerd (sensitiviteit) en van het percentage niet-gevallen, dat negatief op de test scoort (specificiteit).

Met de hulp van gestandaardiseerde psychiatrische interviews (het Clinical Interview Schedule en de Present State Examination) zijn de validiteit, de sensitiviteit en de specificiteit van zowel de 60 item-versie als van een meer handzame 30 item-versie, herhaaldelijk bepaald in een scala van behandelsituaties alsmede in veldonderzoek. (Goldberg & Blackwell, 1970; Tennant, 1977; Newson-Smith & Hirsch, 1978; Benjamin et al, 1982; Goldberg e.a., 1976; Tarnopolski, Hand, Mc Lean, Roberts and Wiggins, 1979). Daaruit blijkt dat de GHQ in het algemeen kan bogen op een redelijk screenend vermogen. Men realiseert zich echter, dat sensitiviteit en specificiteit geen absolute kenmerken zijn van een bepaalde test, doch grootheden die variëren met de situatie, waaraan het instrument is getoetst. Het is een redelijke veronderstelling, dat de mate van zorg die mensen ontvangen, positief gecorreleerd is met de nood waarin ze verkeren en dus met de gemiddelde GHQ-score. Zo zal de meerderheid van de respondenten in de kliniek hoog en in de open populatie laag scoren, terwijl patiënten uit een huisartsenpraktijk, wat betreft hun score-niveau, min of meer gelijkmatig verdeeld zijn. Voor elk praevaletieonderzoek zou de GHQ dus geijkt moeten worden aan een gestandaardiseerd psychiatrisch oordeel in een representatieve steekproef uit die betreffende bevolking. Dergelijke kostbare en tijdrovende pilot-studies zouden overbodig zijn als we de beschikking hadden over een model, waarmee sensitiviteit en specificiteit konden worden bepaald, gegeven de verdeling van de GHQ in die populatie.

In het kader van een Australische veldstudie over de relatie tussen neurotische problematiek en de sociale omgeving, zochten Duncan-Jones en Henderson (1978) naar een methode om de dichotoom uitgedrukte relatie tussen GHQ-score en PSE-resultaat te vervangen door een waarschijnlijkheidsrelatie. Ze stelden een logit regressie-vergelijking op, waaruit ze voor elke GHQ-score de kans konden berekenen dat de betreffende respondent een formeel geval was naar PSE-maatstaven. Deze formule bood veel meer mogelijkheden dan de drempelscore, waarboven een respondent beschouwd werd als een geval en waaronder als normaal. De vergelijking gold in principe echter alleen voor de onderzochte populatie in Canberra. Van een internationaal bruikbaar model in eerdergenoemde zin zou pas sprake zijn als de relatie tussen GHQ- en PSE-resultaat van een totaal verschillende populatie uitgedrukt zou kunnen worden in dezelfde formule.

De *Present State Examination* is een gestandaardiseerd psychiatrisch interview, door Wing e.a. (1967, 1974) ontwikkeld tot een betrouwbaar hulpmiddel bij de bepaling van symptomen op het gebied van de neurosen en de functionele psychosen. Als referentiekader

dienden met name de psychopathologieën van Jaspers en Schneider en het handboek van Mayer Gross, Slater en Roth (Wing, 1983). Ongeveer 20 jaar zijn verstreken sinds de eerste toepassing van het interview dat inmiddels in meer dan 40 talen ter beschikking staat. De PSE wordt (ook in Nederland) vooral gebruikt voor het onderzoek van klinische populaties (Pen e.a., 1977, Giel e.a., 1980, Verhey, 1981 e.a.). De geschiktheid van het instrument voor praevaletie-surveys in open populaties is door Wing (e.a., 1977) aannemelijk gemaakt, maar de feitelijke toepassing op dat gebied bleef tot enkele onderzoeken beperkt (Henderson e.a., 1981).

Het PSE-CATEGO-ID-systeem bestaat onder meer uit de volgende componenten:

- een lijst met 140 psychiatrische symptomen, die bestaat uit 100 vraagitems naar aan- of afwezigheid en ernst van neurotische en psychotische symptomen in de maand, voorafgaand aan het interview en 40 beoordelingsitems, met name op het gebied van gedrag, affect en spraak tijdens het interview. Het interviewschema is gebaseerd op de traditionele klinische vraagtechniek met de mogelijkheid flexibel te reageren op uitingen van de respondent.

- een verklarende woordenlijst, waarin de 140 items zoveel mogelijk zijn geoperationaliseerd.

- een lijst van syndromen die gebruikt kan worden zowel voor het achterhalen van vroegere aandoeningen als voor het samenvatten van symptoomprofielen.

- het Catego-computerprogramma dat een aantal regels bevat voor samenvatting van PSE-gegevens. Van de 140 symptomen worden 114 gegroepeerd tot 36 syndromen die enerzijds leiden tot vier zogenaamde PSE-subscores en een totale PSE-score, en anderzijds tot 12 klassen waaruit een tentatieve diagnose in ICD-termen kan worden afgeleid. De sub-scores zijn gebaseerd op de scores van vier categorieën syndromen, die bestaan uit wanen en hallucinaties, stoornissen op het gebied van gedrag en spraak, specifiek neurotische symptomen en aspecifiek neurotische symptomen. ICD-codes zijn met name van toepassing op klasse S (schizofrene psychose – ICD 295.3), klasse P (paranoïde psychose – ICD 297.9), klasse M (manische en gemengd affectieve psychosen – ICD 296.1, 3), klasse D (depressieve psychosen – ICD 296.2), klasse R (geremde depressies – ICD 296.2/300.4), klasse N (neurotische depressie – ICD 300.4), klasse A (angstneurosen – ICD 300.0,2) en klasse B (dwangneurose – ICD 300.3).

- de Index of Definition, een reglement, waarmee de PSE-data naar niveaus van diagnostiseerbaarheid kunnen worden onderverdeeld. Niveau 1 is gekenmerkt door de afwezigheid van PSE-symptomen (de totale score is 0). De totale score van niveaus 2 en 3 bestaat uit aspecifieke neurotische symptomen: niveau 2 heeft een totale score van 1 tot 4, niveau 3 van 5 tot 9. Niveau 4 kan tot stand komen zowel door een totale score aan aspecifieke neurotische symptomen van 10 of meer,

als door de aanwezigheid van één enkel sleutelsymptoom, zoals depressie, zonder de daarbij behorende symptomen, zoals schuldgevoelens, etcetera. Niveau 5, de drempelwaarde, is een minimumvereiste voor een diagnostische classificatie. Op dit niveau zijn sleutelsymptomen aanwezig in combinatie met elkaar of met andere belangrijke symptomen (bijv. een depressieve stemming, psychomotore traagheid en pathologische schuldgevoelens). Met de niveaus 6, 7 en 8 neemt de zekerheid toe, dat de aanwezige symptomen kunnen worden geclassificeerd in één van de conventionele categorieën van neurosen of functionele psychosen.

In een praeventieonderzoek biedt de PSE dus informatie over de respondenten op twee belangrijke gebieden: met behulp van de Index of Definition kan de symptomatologie van de populatie worden gegroepeerd in 8 niveaus met toenemende diagnostiseerbaarheid, casu quo ernst, waarna toepassing van Catego-computerprogramma op niveau 5 en hoger bovendien diagnostische classificaties oplevert. De kwaliteit van beide resultaten is uiteraard afhankelijk van de betrouwbaarheid waarmee die symptomen zijn vastgesteld. Wing e.a. (1977) hebben aangetoond dat niet-medische interviewers een even hoge interviewer-observator betrouwbaarheid kunnen bereiken als psychiaters, wanneer ze adequaat getraind zijn. Omdat het punt van de betrouwbaarheid des te sterker geldt voor een combinatie van meetinstrumenten, hebben wij speciale aandacht besteed aan selectie en opleiding van de interviewers.

Een andere conditie voor een gezamenlijke toepassing van GHQ en PSE betreft het interval tussen het afnemen van beide instrumenten, dat, zo stelde Duncan-Jones (1979), niet langer dan een week mocht zijn. In het Australische twee-fasen-onderzoek was namelijk ontdekt dat de correlatie tussen de GHQ-score en de totale PSE-score daalde van .8 tot beneden .5, wanneer deze periode werd overschreden. De oorzaak van deze dramatische daling zou liggen in de snel voorbijgaande aard van het soort problematiek dat door de GHQ wordt gemeten.

Materiaal en methoden

In dit onderzoek is gebruik gemaakt van een 30 item GHQ-versie*. De 30 items (Goldberg, 1972) uit de vertaling van de oorspronkelijke 60 item vragenlijst, werden door twee groepen afzonderlijk gecorrigeerd. De ene groep stelde zich ten doel zo dicht mogelijk bij de Engelse versie te blijven, terwijl de andere groep de strekking van de items benadrukte, met de in Nederland gangbare vragenlijsten als referentiekader. Uit beide vertalingen kwam een compromisversie tot stand, die weer werd

* "Copyright: 1978. General Practice Research Unit. International copyright in all countries. All rights reserved by the NFER-Nelson Publishing Company Ltd., Windsor, England".

omgezet in het Engels. Na vergelijking met de oorspronkelijke versie en het aanbrengen van enige wijzigingen, was de vragenlijst voor onderzoek gereed*.

De gerapporteerde data zijn verzameld in het kader van het *Regioproject Nijmegen*. Voor de achtergronden en doelstellingen van dit project zij verwezen naar Tax e.a. [1983]. We presenteren hier alleen de voor dit artikel relevante gegevens. In de Gezondheidsregio Nijmegen (31 gemeenten, 400.000 inwoners) werd een representatieve steekproef van 3245 volwassenen (18-64 jaar, respons 75%) uit de open populatie geïnterviewd met betrekking tot problemen van somatische, psychopathologische en psychosociale aard, daarmee samenhangend gedrag en een aantal andere variabelen. Uitgezonderd waren geïnstitutionaliseerde personen en anderstaligen, waarmee geen interview mogelijk bleek. Voor de steekproef hadden we de populatie verdeeld in twee strata: Nijmegen en de overige 30 gemeenten. Uit het tweede stratum werd a-select getrokken, een naar grootte proportionele steekproef van 10 gemeenten en vervolgens, eveneens a-select, een gelijk aantal inwoners uit de registers van de burgerlijke stand, behorend bij elk van die 10 gemeenten. Nijmegen leverde een aparte, grotere steekproef. De aantallen waren zo gekozen, dat elk lid van de populatie dezelfde kans had om in de steekproef te worden opgenomen. Achteraf zijn de respondenten met de populatie vergeleken naar leeftijd, burgerlijke staat en geslacht. Wat betreft de leeftijd vormen de steekproefelementen een getrouwe afspiegeling van de populatie. De vrouwen zijn iets oververtegenwoordigd (51,3% in plaats van 49,6%), terwijl de ongehuwden iets ondervertegenwoordigd zijn (28,1% in plaats van 21,4%). In het vervolg zal worden aangenomen dat we te maken hebben met een representatieve steekproef uit de populatie.

Het interview duurde gemiddeld 90 minuten. De eerste waarneming werd verzorgd door de NV v/h de Nederlandse Stichting voor Statistiek met behulp van ervaren en getrainde interviewers. Tijdens het interview kreeg de respondent de 30 item GHQ voorgelegd. De interviewer berekende na afloop de somscore en stelde ter plekke vast of de respondent voor een PSE-interview in aanmerking kwam.

De keuze verliep als volgt:

- een willekeurig achtste deel van de vragenlijsten was voorzien van een aanwijzing, dat de respondent een PSE-interview zou krijgen, ongeacht de score op de GHQ
- de overige respondenten werden geselecteerd wanneer hun GHQ-score 10 of hoger was.

De reden voor deze selectie lag enerzijds in ons streven naar een bepaalde hoeveelheid gevallen met wie een vervolgonderzoek zou

* Het vertalerscollectief bestond uit respectievelijk Hermsen (psychologe), Hodiamont (zenuwarts), Bosma (socioloog), Sijben (researchpsycholoog), Clasens (psychologisch medewerkster) en Eskes (arts).

plaatsvinden en anderzijds in de noodzaak om niet alleen respondenten met een hoge GHQ-score psychiatrisch te interviewen (zoals normaal in een zeefprocedure), maar ook respondenten met een lage score, wilden we tenminste in staat zijn de relatie GHQ-PSE te analyseren.

De GHQ-score 10 werd gekozen op basis van een pilotstudy onder de eerste 350 respondenten, die uitwees, dat een dergelijke drempel zou leiden tot het afnemen van de 500, budgettair toegestane, PSE interviews.

Bij de analyse is gebruik gemaakt van *alle* PSE-respondenten, ongeacht de wijze van selecteren.

Wanneer de respondent aan het criterium voldeed, vroeg de interviewer toestemming voor een vervolgesprek binnen een week. Vierenzeventig procent van de respondenten stemde daarmee in. Het eerste interview, dat ook de GHQ omvatte, werd afgenomen in de periode februari tot april 1983, terwijl het tweede, het PSE-interview, kort daarna volgde (bij 73% binnen 8 dagen, bij 93% binnen 15 dagen na het eerste interview). Uiteindelijk werden 3232 volledig ingevulde GHQ-vragenlijsten verkregen. Met 485 respondenten vond een PSE-interview plaats. Honderdenéén van deze 485 interviews resulteerden in een Index of Definition van 5 of hoger. Deze zullen in het vervolg worden aangeduid als de *cases*.

Voor de PSE-interviews werden tien interviewers geselecteerd (9 vrouwen en 1 man), die allen ervaring hadden met psychiatrische problematiek. Acht van hen waren afgestudeerde of bijna afgestudeerde psychologen, één was maatschappelijk werkster met Voortgezette Opleiding en één was semi-arts. Zij werden gedurende een week in het Psychomedisch Studiecentrum Vijverdal (Maastricht) getraind door dr. F. Verhey, die zijn kennis van de PSE-methodiek heeft opgedaan in 'The Maudsley Institute' in Londen, het ontstaanscentrum van het instrument. In de twee maanden tussen de training en het begin van het veldwerk deden de interviewers ervaring op met 10 live-interviews en 10 van een audio-band te scoren interviews bij respondenten uit de open populatie, een huisartsenpraktijk, een neurosenkliniek en twee psychiatrische ziekenhuizen. Daarover zal elders worden gerapporteerd (Duine e.a., 1983). Tijdens het veldwerk vonden regelmatig 'booster-sessies' plaats om zowel de onderlinge consensus als de overeenstemming met de 'officiële' PSE-lijn te handhaven.

De tabellen 1 en 2 geven in detail de voornaamste gegevens weer, die in de analyse worden gebruikt. Tabel 1 toont de GHQ-score voor de 3232 respondenten van de eerste meting.

Tabel 2 weerspiegelt de verdeling van de GHQ-score bij de respondenten die een PSE-interview hebben gehad (opgesplitst naar *cases* en *non-cases*).

Tabel 1: De verdeling van de GHQ-score in de steekproef (N = 3232)

GHQ-				GHQ-			
score	aantal	%	cum. %	score	aantal	%	cum. %
0	1105	34.2	34.2	15	18	0.6	97.0
1	577	17.9	52.0	16	14	0.4	97.4
2	361	11.2	63.2	17	14	0.4	97.9
3	269	8.3	71.5	18	17	0.5	98.4
4	184	5.7	77.2	19	8	0.2	98.6
5	130	4.0	81.3	20	7	0.2	98.9
6	112	3.5	84.7	21	10	0.3	99.2
7	86	2.7	87.4	22	8	0.2	99.4
8	73	2.3	89.6	23	2	0.1	99.5
9	58	1.8	91.4	24	6	0.2	99.7
10	46	1.4	92.9	25	4	0.1	99.8
11	32	1.0	93.8	26	2	0.1	99.8
12	30	0.9	94.8	27	2	0.1	99.9
13	26	0.8	95.6	28	3	0.1	100.0
14	28	0.9	96.4				

Tabel 2: Verdeling van de GHQ-score over cases en non-cases

GHQ-			
score	non-cases	cases	totaal
0	72	1	73
1	36	1	37
2	29	1	30
3	31	1	32
4	26	2	28
5	21	2	23
6	13	2	15
7	16	0	16
8	6	2	8
9	9	2	11
10	31	5	36
11	20	7	27
12	17	7	24
13	13	8	21
14	7	9	16
15	8	5	13
16	9	5	14
17	5	4	9
18	6	8	14
19	2	4	6
20	1	4	5
21	3	4	7
22	3	4	7
23	0	2	2
24	0	6	6
25	0	2	2
26	0	1	1
28	0	2	2
totaal	384	101	485

Analyse en resultaten

We zullen nu het verband tussen de GHQ-score en het PSE-case-zijn nader bestuderen en deze relatie verwerken in een schatting van de PSE-case praevaletie in de hele populatie. Formeel gaat het daarbij om een periodepraevaletie van één maand (Goldbergen Huxley, 1980), de referentietermijn die expliciet geldt voor de PSE en min of meer impliciet voor de GHQ. Gezien de betrekkelijke stabiliteit van de problematiek, gemeten door de PSE, lijkt het verschil tussen deze éénmaands periodepraevaletie en het begrip puntpraevaletie beperkt.

Nagegaan zal worden of de praevaletie samenhangt met leeftijd en geslacht. Daarna preciseren we onze resultaten door het geven van 95% betrouwbaarheidsgrenzen. Vervolgens bepalen we de sensitiviteit en de specificiteit van de GHQ, opgevat als screeningsinstrument in de open populatie voor het aanwijzen van PSE-cases. Aansluitend zullen we de geschiktheid van de GHQ voor twee verschillende doeleinden ter discussie stellen, namelijk als instrument om de praevaletie te bepalen en als screeningstest.

Een elegante manier om het verband tussen het al dan niet case zijn (een dichtome variabele) en de score op de GHQ (een kwantitatieve variabele) aan te geven, is het logistische model. Voor de achtergronden daarvan zij verwezen naar Kleinbaum e.a. (1982). In concreto verloopt de redenering aldus. Geef met p de kans aan, dat iemand met een score g op de GHQ een case is. Dan is $1-p$ de kans, dat een dergelijke respondent geen case is. Het logistische middel stelt nu dat

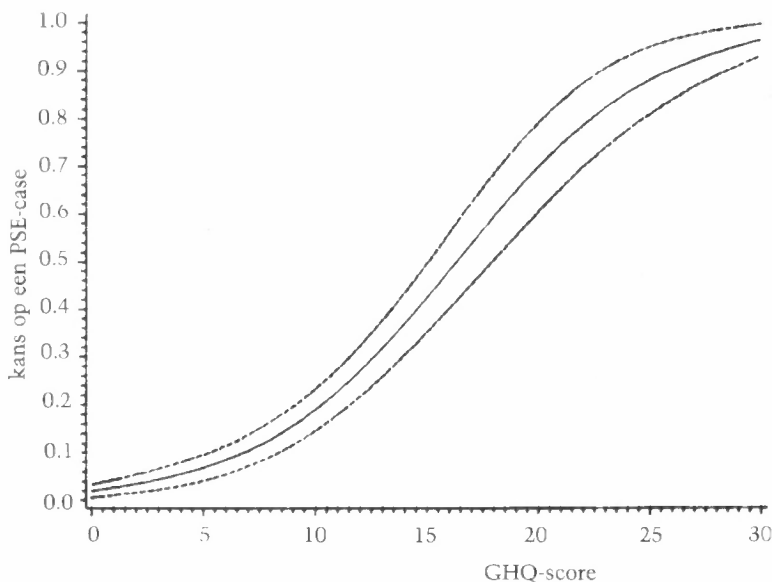
$$\log (p/1-p) = \alpha + \beta \times g$$

Inde formule zijn α en β te schatten parameters. Bij iedere waarde van g kan men zo $\log (p/1-p)$ bepalen en daarmee ook p . De grootheid $\log (p/1-p)$ duidt men ook wel aan met de notatie $\text{logit}(p)$. Toepassing van de gegevens uit tabel 2 levert als schattingen voor de parameters: $\alpha = -3.76$ en $\beta = 0.23$.

Een grafiek van p als functie van de GHQ-score g is gegeven in figuur 1. Met behulp van een zogenaamde goodness-of-fit analyse werd nagegaan of het model met de data overeenstemt. Het resultaat van deze analyse (waarvan de uitwerking buiten het kader van dit artikel valt) was bepaald positief.

Als volgende stap geldt de bepaling van de praevaletie van PSE-cases in de populatie van de gezondheidsregio Nijmegen. Gegeven zijn de GHQ-scores van 3232 respondenten die als een representatieve steekproef kunnen worden aangemerkt, en het logistische model (figuur 1), geschat op basis van zowel een GHQ-score als een PSE-ID-niveau (≥ 5 betekende een casus, < 5 een non-casus) bij 485 personen die echter geen representatieve steekproef vormen uit de hele populatie. We hebben al aangegeven dat de score op de GHQ een rol heeft gespeeld bij hun keuze. Maar ook bij de respondenten die ongeacht hun GHQ-score werden aangewezen, is selectie opgetreden, gezien het feit

Figuur 1: Het logistische model. Verband tussen GHQ-score en de kans op een PSE-case (met 95% betrouwbaarheidsgebied)



dat degenen met een hogere GHQ-score meer bereid waren tot een tweede (PSE-)interview dan de laag scorenden. Representativiteit van de groep PSE-responsdenten is echter niet noodzakelijk voor het bepalen van de (logistische) regressielijn. Deze ogenschijnlijk gewaagde bewering willen wij adstrueren met een voorbeeld uit de fysica. Iemand heeft 10.000 staven ijzer gelegeerd, elk met een lengte van 10 m, en hij wil, voor de planning van opslagruimte, weten hoeveel procent van die staven bij hun huidige temperatuur langer is dan 10.01 m (casus in onze terminologie).

Wanneer het meten van de lengte (naar analogie van de PSE) tijdrovend is, het meten van de temperatuur (naar analogie van de GHQ) daarentegen eenvoudig, zal de onderzoeker een klein aantal willekeurig gekozen staven van verschillende temperatuur (bijv. 5 van 0°C en 5 van 50°C) óók beoordelen op hun lengte (GHQ resp. PSE). Aan de hand van die gegevens bepaalt hij de lineaire uitzettingscoëfficiënt (d.w.z. de regressielijn). Vervolgens meet hij bij een representatieve steekproef van de staven de temperatuur (GHQ) waarna hij met behulp van de regressielijn het percentage te lange staven (cases) kan berekenen.

Als gevolg van het niet-lineaire verband tussen GHQ en PSE was in ons geval een veel groter aantal GHQ-PSE-combinaties nodig om een optimale curve te verkrijgen.

Dat de 10 staven, gebruikt voor de ijking, niet op lengte geselecteerd mogen worden (dus niet de voorkeur geven aan lang uitgevallen staven van 0°C of de cases onder de lage GHQ-scoorders) is de enige eis

waaraan de onderzoeker moet voldoen om een reële regressielijn te kunnen trekken. Om de praevaletie te berekenen, kunnen we dus volstaan met de aanname dat het verband tussen de GHQ en het al dan niet case-zijn, gevonden bij de 485 respondenten, ook geldt voor de open populatie. Bij eenzelfde GHQ-score veronderstellen we slechts dezelfde kans voor cases en non-cases dat iemand voor een psychiatrisch interview in aanmerking is gekomen. Met andere woorden, de cases en de non-cases onder de respondenten met een GHQ-score van bijvoorbeeld 12 hebben met dezelfde kans toestemming gegeven voor een PSE-interview. Deze veronderstelling is plausibel en onvermijdelijk in empirisch onderzoek. Op deze basis kennen we, per GHQ-score, de kans dat iemand een case is (figuur 1). In tabel 1 staat de verdeling van de GHQ in de hele steekproef van 3232. Combinatie van tabel 1 met figuur 1 levert zo een schatting voor de regionale praevaletie. De uitkomst van deze berekening is 7.3%, iets lager dan 9.0% en 10.6% (Henderson, 1981), gevonden in respectievelijk Canberra (Australië) en Camberwell (Engeland), praevaletiewaarden, die eveneens werden bepaald met behulp van de PSE, in een twee-fasen-onderzoek onder de open (d.w.z. niet-geïnstitutionaliseerde) populatie.

Als we aannemen dat de rest van het psychopathologische spectrum (organische aandoeningen, addictie, persoonlijkheidsstoornissen) ongeveer een derde tot de helft uitmaakt van de door de PSE gemeten beelden (Ormel en Giel, 1983), komt de totale psychiatrische praevaletie op $\pm 10\%$.

Om de invloed van de tijd tussen het bepalen van de GHQ-score en het houden van het PSE-interview na te gaan, zijn de correlaties berekend tussen de GHQ-score en de totaalscore van de PSE-symptomen. Deze correlatie was 0.68 voor respondenten met een tijdsinterval van maximaal 3 dagen en liep af tot 0.61 voor respondenten met een tijdsinterval van meer dan een week. Dit is geen dramatische daling te noemen. De problematiek is blijkbaar van minder voorbijgaande aard dan men op grond van eerder onderzoek zou verwachten.

Met het oog op de actieradius van het model willen we nagaan of de gevonden parameters onafhankelijk zijn van belangrijke achtergrondvariabelen, zoals leeftijd en geslacht. Herhalen we de analyse, met in het rechterlid van het model twee termen toegevoegd (een term voor het geslacht, met coëfficiënt γ en een term voor leeftijd met coëfficiënt δ), dan zien we geen betere overeenstemming van de data met het model: de parameters verschillen niet significant van 0 ($p > .10$ bij geslacht en $p > .60$ bij leeftijd). Dit heeft belangrijke consequenties. We kunnen het model, zoals gerepresenteerd in figuur 1, zonder meer toepassen op de GHQ-scores van alleen vrouwen, of alleen mannen. De praevaletie is dan respectievelijk 7.5% en 7.2%. Het verschil tussen beide percentages blijkt niet significant. Ook kunnen we leeftijdsafhankelijke praevaletiecijfers berekenen, door het model op de GHQ-scores van verschillende leeftijdscategorieën toe te passen.

De praevalentie neemt globaal toe met de leeftijd, tot ongeveer 57 jaar, en neemt daarna weer af. De analyse van praevalentieverschillen in diverse subpopulaties en van de kwalitatieve relatie tussen GHQ en PSE zal binnenkort worden gepubliceerd.

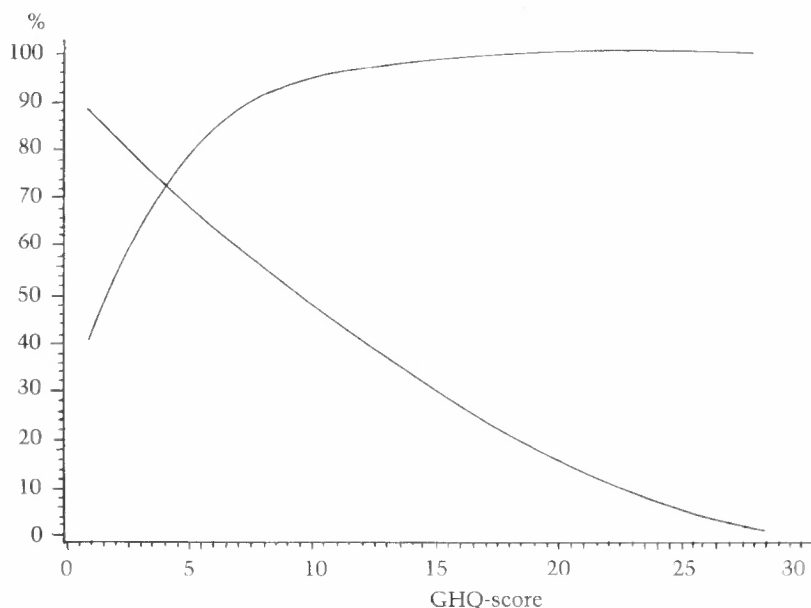
De nauwkeurigheid van onze resultaten valt af te lezen uit 95% betrouwbaarheidsgebieden.

Behalve de parameters α en β , kan men ook hun respectievelijke fout bepalen. Met behulp daarvan bestaat de mogelijkheid om, per GHQ-score, de kans dat iemand een case is, te voorzien van een 95% betrouwbaarheidsinterval. Wanneer we die in figuur 1 intekenen, en de 31 punten verbinden met een vloeiende lijn, krijgen we betrouwbaarheidsgebieden, zoals ze in de figuur zijn weergegeven. Ook voor de praevalentie kunnen we een benaderd 95% betrouwbaarheidsinterval (5.5%-9.2%) berekenen.

Gedetailleerde weergave van deze berekeningen valt buiten het kader van dit artikel.

Tenslotte gaan we na in hoeverre de GHQ opgevat kan worden als screeningstest in de open populatie. In figuur 2 zijn de sensitiviteit en de specificiteit van de GHQ weergegeven. Voor iedere drempelwaarde g van de GHQ is de sensitiviteit (het percentage van de totaal aantal cases dat positief op de test scoort) en de specificiteit (het percentage van het totaal aantal non-cases dat negatief op de test scoort) uit de grafiek af te lezen. Als drempel geldt die waarde waarop of waarboven

Figuur 2: Sensitiviteit en specificiteit voor verschillende drempelwaarden van de GHQ



de score op de test als positief wordt aangemerkt. Uit de grafiek blijkt een drempelwaarde van bijvoorbeeld 4 ongeveer overeen te komen met een sensitiviteit van 72% en een specificiteit van 75%. Aangezien (tabel 1) 28,5% van de populatie boven die drempelwaarde scoort, zal men bij 28,5% van de populatie een PSE moeten afnemen om 72% van alle in de populatie aanwezige cases te treffen.

Discussie

Voor een praevalentiebepaling van PSE-cases in de open populatie dient men te beschikken over de verdeling van de GHQ in die populatie en de parameters van het logistisch model. Het vaststellen van de GHQ-score blijkt een relatief eenvoudige zaak. De lijst is kort en zeer snel af te nemen. Het bepalen van de parameters van het model eist veel meer inspanning, omdat een groot aantal PSE-interviews moet worden afgenomen. Willen we nu de praevalentie in een geheel andere populatie bepalen, dan geldt als noodzakelijk en voldoende voorwaarde de gelijkheid van de onderlinge relatie van de GHQ en de PSE ten aanzien van beide populaties, zoals weerspiegeld in de parameters α en β . Het gaat er dus om welke waarden α en β aannemen in populaties, die verschillen van de onze. Zoals gezegd, zijn ons tot nu toe alleen uit het genoemde Australische onderzoek parameters bekend, waarvan de berekening met behulp van de 30-item GHQ en de PSE is uitgevoerd. (Henderson e.a., 1981).

De door hen gerapporteerde $\alpha = -3.63$ en $\beta = 0.23$, liggen buitengewoon dicht bij de door ons gevonden waarden ($\alpha = -3,76$ en $\beta = 0,23$). Ons onderzoek heeft aangetoond, dat het model evengoed opgaat voor mannen als voor vrouwen en bovendien voor verschillende leeftijdscategorieën. Met het model in de hand blijkt de GHQ een uiterst efficiënte vragenlijst om de praevalentie te bepalen. Case-identificatie-instrumenten zoals de PSE, zijn dan overbodig voor het bepalen van de praevalentie. We weten wel *hoeveel* cases er zijn, niet wie die cases zijn.

Als we cases in de populatie daarentegen willen opsporen, dus willen weten *wie* de cases zijn, zijn we genoodzaakt ook PSE-interviews te houden. Met behulp van de GHQ kan men het aantal af te nemen PSE's drastisch beperken. Het zal van de concrete toepassing afhangen of de kwaliteit van de GHQ als screeningstest voldoende is. Men houde wel rekening met het feit dat de sensitiviteit en de specificiteit van de GHQ als screeningstest (weergegeven in figuur 2) populatie-afhankelijk zijn. In een populatie met bijvoorbeeld veel hoogscorders op de GHQ, zal de bruikbaarheid van het instrument als screeningstest veel gunstiger uitvallen.

Conclusie

Het beschreven logistische model lijkt zowel praktisch als theoretisch een belangrijke aanwinst. De toepasbaarheid op (in geslacht, leeftijd en cultuur) zeer uiteenlopende populaties maakt het mogelijk met een minimum aan onderzoeksinvestering – het afnemen van alleen de GHQ – een maximum aan rendement te behalen in casu een psychiatrisch gevalideerd praevaletiecijfer.

Op theoretisch gebied ligt de waarde van het model vooral in de brug, die geslagen wordt tussen enerzijds psychische nood, voor zover de direct betrokken leek die ervaart en weergeeft in zijn antwoord op de vragenlijst, en anderzijds psychopathologie, voor zover die wordt gezien en vastgelegd door de indirect betrokken specialist (Hodiamont, 1983).

Literatuur

- Benjamin, S., P. Decalmer, D. Haran (1982). Community Screening for mental illness: a validity study of the General Health Questionnaire, *British Journal of Psychiatry*, 140, 174-180.
- Blum, R. H. (1962). Case-identification in psychiatric epidemiology: methods and problems, *Millbank Memorial Fund Quarterly*, 40, 253-288.
- Dohrenwend, B. en Dohrenwend, B. (1974). Social and cultural influences on psychopathology, *Annual Review of Psychology*, 25, 417-452.
- Dohrenwend, B. en Dohrenwend, B. (1982). Perspectives on the past and future of psychiatric epidemiology, *American Journal of Public Health*, 72, 11, 1271-1279.
- Duine, T. J., P. P. G. Hodiamont, S. H. J. Veling, L. H. Boerma (1983). Betrouwbaarheid van de PSE in diverse bevolkingscategorieën. In *voorbereiding*.
- Duncan-Jones, P., S. Henderson (1978). The use of a two-phase design in a prevalence survey, *Social Psychiatry*, 13, 231-237.
- Duncan-Jones, P. (1979). Validity and uses of the GHQ, *British Journal of Psychiatry*, 135, 382.
- Eastwood, M. R., (1971). Screening of psychiatric disorder, *Psychological Medicine*, 1, 197-208.
- Giel, R., H. C. Sauer, C. J. Slooff, D. Wiersma (1980). Over epidemiologie van functionele psychosen en invaliditeit, *Tijdschrift voor Psychiatrie*, 22, 710-722.
- Goldberg, D. P., B. Blackwell (1970). Psychiatric illness in general practice; a detailed study using new methods of case-identification, *British Medical Journal*, 2, 439-443.
- Goldberg, D. P., (1972). *The detection of psychiatric illness by questionnaire*. Oxford University Press, London.
- Goldberg, D. P., (1974). Screening for disease, *The Lancet*, 23 november, 1245-1247.
- Goldberg, D. P., K. Richels, R. Downing, P. Hesbacher (1976). A comparison of two psychiatric screening tests, *British Journal of Psychiatry*, 129, 61-67.
- Goldberg, D. P. (1978). *Manual of the General Health Questionnaire*. NFER Publishing Company, Windsor.

- Goldberg, D. P., P. Huxley (1980). *Mental illness in the Community*. Tavistock Publications, London.
- Henderson, S., P. Duncan-Jones, D. G. Byrne, R. Scott, S. Adcock (1979). A standardised study of prevalence, *Acta Psychiatrica Scandinavica*, 60, 355-374.
- Henderson, S., D. G. Byrne, P. Duncan-Jones (1981). *Neurosis and the social environment*. Academic Press, Australia.
- Hodiament, P. P. G. (1983). Het 'wat' in de psychiatrische epidemiologie, *Tijdschrift voor Psychiatrie*, 25, 6, 391-406.
- Kleinbaum, D. G., L. L. Kupper, H. Morgenstern (1982). *Epidemiologic Research. Principles and quantitative methods*. Lifetime Learning, London etc.
- Lin, T. Y., C. C. Standley (1962). The scope of epidemiology in psychiatry, *Public Health Papers* (16), WHO Genève.
- Newson-Smith, J. G. B., S. R. Hirsch (1979). Psychiatric symptoms in self-poisoning patients, *Psychological Medicine*, 9, 493-500.
- Ormel, J. en R. Giel (1983). Omvang, beloop en behandeling van psychische stoornissen in de praktijk van de huisarts, *Tijdschrift voor Psychiatrie*, 10, 688-710.
- Pen, A., W. A. Arrindell, W. Lawant, J. Pols (1977). Betrouwbaarheid van de Present State Examination, *Tijdschrift voor Psychiatrie*, 6, 375-388.
- Reid, D. D. (1960). Epidemiological methods in the study of mental disorders, *Public Health Papers* (2), WHO Genève.
- Tarnopolski, A., D. Hand, E. K. McLean, H. Roberts, R. Wiggins (1979). Validity and uses of a screening questionnaire (GHQ) in the community, *British Journal of Psychiatry*, 134, 508-515.
- Tax, B., C. König-Zahn, P. H. J. M. Heydendaal, L. H. Boerma, J. W. Furer, E. J. H. ter Heine, P. P. G. Hodiament, J. M. G. Persoon, S. H. J. Veling (1983). Regioproject Nijmegen. Een longitudinaal onderzoek in de bevolking naar ziekte- en hulpzoekgedrag bij somatische klachten, psychiatrische symptomen en psychosociale problemen. In *voorbereiding*.
- Tennant C. (1977). The General Health Questionnaire, a valid index of psychological impairment in Australian Populations, *The Medical Journal of Australia*, 2, 392-394.
- Verhey, F. H. M. (1981). *Langzame hersenpotentialen en gemeten gedrag*. Leiter Nypels BV, Maastricht.
- Williams, P., A. Tarnopolski, D. Hand (1980). Case-definition and case-identification in psychiatric epidemiology: review and assessment, *Psychological Medicine*, 10, 101-114.
- Wilson, J. M. G., G. Jungner (1968). Principles and practice of screening for disease. *Public Health papers*, WHO Genève.
- Wing, J. K., J. L. T. Birley, J. E. Cooper, P. Graham, A. D. Isaac (1967). Reliability of a procedure of measuring and classifying 'Present Psychiatric State', *British Journal of Psychiatry*, 113, 499-515.
- Wing, J. K., E. Cooper, N. Sartorius (1974). *Measurement and classification of psychiatric symptoms*. Cambridge University press, Cambridge.
- Wing, J. K., J. Nixon, S. A. Mann, J. P. Leff (1977). Reliability of the PSE (ninth edition) in a population survey, *Psychological Medicine*, 7, 505-516.
- Wing, J. K. (1983). Use and Misuse of the PSE, *British Journal of Psychiatry*, 143, 11-117.

De gerapporteerde gegevens werden verzameld in het kader van het Regioproject Nijmegen, waarvoor het Praeventiefonds een aanzienlijke subsidie verleende. Het artikel kwam tot stand dank zij de inspanning van velen, zoals de regioprojectgroep, de interviewers en de respondenten. De auteurs zijn met name H. Straatman (statisticus) en P. Vreuls (beiden verbonden aan het Instituut voor Sociale Geneeskunde van de Katholieke Universiteit Nijmegen) erkentelijk respectievelijk voor de bijdrage in de methodologische discussie en de nauwgezette uitvoering van het programmeerwerk.