

De waarde van de GHQ als toestandsmaat

door M.W.J. Koeter, J. Ormel, W. van den Brink,
W. Dijkstra, J. Schoenmacker en J. Staal

Samenvatting

In dit artikel staan de volgende twee vragen centraal: (1) Kan de General Health Questionnaire (GHQ) als maat voor de ernst van psychopathologie worden gebruikt, of anders gezegd, volgt ze de veranderingen in de ernst van het toestandsbeeld voldoende nauwkeurig, (2) leidt de door Goodchild en Duncan-Jones voorgestelde alternatieve scoringswijze van de GHQ tot betere resultaten. In een longitudinaal design werden bij 175 nieuwe psychiatrische polikliniekpatiënten driemaal de GHQ-28 en de Present State Examination (PSE) afgenomen. De GHQ-28 is een selfreport instrument voor de meting van psychopathologie, de PSE is een gestructureerd psychiatrisch interview waarmee het toestandsbeeld nauwkeurig wordt vastgelegd. De eerste meting vond plaats kort na de intake, de beide andere respectievelijk vier en tien maanden later. Op basis van de verkregen gegevens kunnen de volgende conclusies worden getrokken. De GHQ-28 blijkt, mits gecorrigeerd voor het hertesteffect, een goede toestandsmaat, die veranderingen in de ernst van het toestandsbeeld (vastgesteld aan de hand van de PSE-Index of Definition) goed weergeeft. Maar de door Goodchild en Duncan-Jones (1985) voorgestelde alternatieve scoringswijze van de GHQ leidt niet tot betere resultaten.

Inleiding

De General Health Questionnaire (GHQ, Goldberg 1972) is een oorspronkelijk uit 60 items bestaande selfreport vragenlijst voor de detectie van niet psychotische psychiatrische aandoeningen in de bevolking en patiënten in de eerstelijns gezondheidszorg. De hoogte van de GHQ-score geeft de waarschijnlijkheid aan dat er bij betrokkene sprake is van geestelijke gezondheidsproblemen. Kenmerkend voor de GHQ is het individugebonden referentiekader van de items. In de formulering en antwoordcategorieën wordt onderscheid gemaakt tussen chronische, al lang bestaande klachten die door betrokkene min of meer als egosyntoon worden ervaren en het optreden van nieuwe symptomen, dan wel de toename in intensiteit van bestaande symptomen. Een score wordt alleen als pathologisch beschouwd als er sprake is van een afwijking van of discontinuïteit in het door de respondent

als 'normaal' ervaren functioneren. Dit in tegenstelling tot andere vragenlijsten die de geestelijke gezondheidstoestand proberen te meten, (vaak 'state' vragenlijsten genoemd), waar de individuele norm geen rol speelt en de aan- of afwezigheid van een symptoom ongeacht de chroniciteit wordt gescoord, zoals de SCL-90 (Derogatis, Lipman, Covis 1973). Uit de oorspronkelijke 60 itemversie zijn een aantal kortere versies, de GHQ-30, GHQ-20, GHQ-12, en een geschaalde versie, de GHQ-28 afgeleid.

De validiteit van de GHQ als case finding en screeningsinstrument is onderzocht aan de hand van de overeenstemming tussen de indeling in gevallen en niet-gevallen volgens de GHQ en een op een gevalideerd psychiatrisch interview gebaseerd klinisch oordeel. Meestal werd hiervoor de Present State Examination (PSE, Wing, Cooper en Sartorius 1974) of het Clinical Interview Schedule (CIS, Goldberg e.a. 1970) gebruikt. In het algemeen waren de sensitiviteit en specificiteit bevredigend, beide rond de .80 (Banks 1983; Bridges en Goldberg 1986; Finlay-Jones en Murphy 1979; Goldberg 1972; Goldberg e.a. 1976; Goldberg en Hillier 1979; Goodchild en Duncan-Jones 1985; Rabin en Brooks 1981; Tarnapolsky e.a. 1979). Alleen Benjamin e.a. (1982) rapporteerden een lage sensitiviteit (.52). Het misclassificatiepercentage varieerde van 8% bij een groep multiple sclerose-patiënten (Rabin en Brooks 1981) tot 26% bij twee groepen huisartspatiënten (Finlay-Jones & Murphy 1979). Daarnaast werd door verschillende onderzoekers een substantiële synchronische correlatie gevonden tussen de GHQ-totaalscore en de PSE- en CIS-scores. De synchronische correlaties van de GHQ-totaalscore met de PSE-totaalscore of PSE-ID varieerden van 0.58 (Goodchild & Duncan-Jones 1985) tot 0.83 (Rabin & Brooks 1981) en met de CIS van 0.61 (Benjamin e.a. 1982) tot 0.86 (Goldberg e.a. 1976).

Twee punten met betrekking tot de GHQ vragen nog om een nader onderzoek.

a. *Gevoeligheid van de GHQ voor veranderingen in het toestandsbeeld.* Naast case-finding en screeningsinstrument wordt de GHQ ook gebruikt als maat voor de ernst van het toestandsbeeld (Henderson 1981; Henderson, Byrne en Duncan-Jones 1981). Rabins en Brooks (1981) concluderen dat de GHQ een goede ernstmaat is, terwijl Goldberg & Hillier (1979) suggereren dat verandering in de GHQ-score als maat voor de effectiviteit van behandelingen kan worden gebruikt. De substantiële synchronische correlaties en de hoge sensitiviteit en specificiteit lijken het gebruik als ernstmaat te ondersteunen. Toch is hier enige voorzichtigheid op zijn plaats en nader onderzoek geboden. Een hoge synchronische correlatie is wel een noodzakelijke, maar nog geenszins voldoende voorwaarde voor de validering van de GHQ als maat voor de ernst van het toestandsbeeld. Een goede ernstmaat moet ook veranderingen in de ernst van het toestandsbeeld betrouwbaar registreren. Een trait-neuroticisme schaal bij voorbeeld, blijkt in de

praktijk ook redelijk goed als screeningsinstrument voor affectieve stoornissen te functioneren (Hodiamont 1986). Per definitie zijn trait-neuroticisme schalen echter niet in staat om veranderingen in de ernst van het toestandsbeeld betrouwbaar te registreren. Eerder werden door ons (Koeter e.a. 1986) bij voorbeeld correlaties van rond de .20 gevonden tussen veranderingen in de scores op de Inadequatie-, Verongelijkheid-, en Zelfwaarderingsschaal van de Nederlandse Persoonlijkheds Vragenlijst (Luteijn 1974) en veranderingen in de ernst van het toestandsbeeld. De validiteit van de GHQ als ernstmaat voor het toestandsbeeld hangt met andere woorden (mede) af van het antwoord op de vraag in hoeverre de GHQ in staat is veranderingen in de ernst van het toestandsbeeld te weerspiegelen. Voor zover ons bekend werd dit nog niet eerder in een longitudinaal onderzoek, met een gestandaardiseerd klinisch psychiatrisch interview als criterium, nagegaan.

b. *De gevoeligheid van de GHQ voor chronische stoornissen en de effectiviteit van een alternatieve scoringsmethode.* Vanwege het individugebonden referentiekader van de items kan de GHQ relatief ongevoelig zijn voor chronische psychopathologie (Benjamin e.a. 1982; Goldberg 1976). Dit kan onder meer de correlatie van de GHQ-score met andere maten voor de psychopathologie (PSE-ID etc.) drukken. Recentelijk is door Goodchild & Duncan-Jones (1985) een alternatieve scoringswijze voor de GHQ voorgesteld in een poging om de gesignaleerde ongevoeligheid voor chronische problematiek op te heffen.

Dit artikel probeert op basis van een longitudinaal onderzoek bij 175 nieuwe psychiatrische polikliniekpatiënten antwoord te geven op de volgende vragen: (a) Hoe gevoelig is de GHQ voor veranderingen in het toestandsbeeld die zijn vastgesteld met behulp van een gevalideerd psychiatrisch interview, de PSE. (b) In hoeverre is de alternatieve scoringswijze op dit punt superieur ten opzichte van de oorspronkelijke. Het artikel beperkt zich tot de GHQ-28.

Steekproef

Het onderzoek werd uitgevoerd bij een groep psychiatrische polikliniekpatiënten die in leeftijd varieerden van 16 tot 60 jaar. Patiënten met een psychose, verslavingsproblemen, zwakbegaafdheid en/of een organisch psychosyndroom werden uitgesloten omdat deze stoornissen een adequate invulling van de GHQ kunnen belemmeren, of omdat de GHQ niet geacht wordt ze op te sporen. Van de 460 patiënten die gedurende een aanengesloten periode van 6 maanden door de psychiatrische poliklinieken van het Academisch Ziekenhuis en de twee algemene ziekenhuizen van de stad Groningen via de normale intake-procedure (geen spoedopnames en keuringen) in behandeling werden genomen, moesten er 121 worden uitgesloten op grond van hun pathologie en werden 46 anderen om organisatorische redenen niet benaderd. De resterende 293 patiënten werd gevraagd aan het onderzoek

deel te nemen. Tweeënzeventig van hen weigerden (24%), 23 vielen om diverse praktische redenen af (zouden gaan verhuizen, te lange reistijd etc.). Uiteindelijk zeiden 198 patiënten hun medewerking aan het onderzoek toe. Tussen de verschillende metingen vielen 23 patiënten uit (12%). De analyses zijn gebaseerd op de resterende 175 patiënten. Deze groep bestaat voor 58% uit vrouwen en de gemiddelde leeftijd is 32 jaar (standaarddeviatie 10).

Methode en gebruikte instrumenten

Er werd gebruik gemaakt van een longitudinale onderzoeksofzet met drie meetmomenten. De eerste meting vond binnen een week na de intake plaats, de tweede en derde meting respectievelijk vier en tien maanden later. Bij de 41 patiënten van de algemene ziekenhuizen werd alleen de eerste en derde meting verricht. Op ieder meetmoment werd op de polikliniek een semi-gestructureerd psychiatrisch interview afgenomen. Enkele dagen voor dit interview werd de patiënt de GHQ-28 toegestuurd, met het verzoek deze in te vullen en mee te nemen naar de polikliniek.

De GHQ-28

De GHQ-28 is een door patiënt zelf in te vullen schriftelijke vragenlijst. De vragen hebben betrekking op het optreden van ongewone en onaangename psychische verschijnselen en op het onvermogen om normaal te blijven functioneren, verschijnselen die men doorgaans aantreft bij mensen met psychiatrische aandoeningen. Bij elk item dient de respondent zijn huidige toestand (in de afgelopen 4 weken) te vergelijken met de voor hem gewone toestand en vervolgens een van de vier volgende antwoordcategorieën te kiezen: geen, niet meer dan gewoonlijk, meer dan gewoonlijk, veel meer dan gewoonlijk. De GHQ-28 is door Goldberg (1979) ontwikkeld uit de originele 60 item-versie en heeft vier schalen a) somatische symptomen, b) angst en slapeloosheid, c) sociaal dysfunctioneren en d) ernstige depressie. De GHQ-28 is speciaal bedoeld om gebruikt te worden in onderzoek waarin men meer informatie over het toestandsbeeld nodig heeft dan alleen de algemene ernstscore. Als drempelwaarde voor een psychiatrische casus wordt de totaalscore '5' genomen.

De GHQ-28 werd door ons op twee manieren gescoord. Allereerst op de oorspronkelijk door Goldberg (1978) voorgestelde wijze, waarbij de antwoordcategorieën 'geen' of 'minder dan anders' en 'niet meer dan anders' als normaal worden beschouwd en de antwoordcategorieën 'wat meer dan gewoonlijk' en 'veel meer dan gewoonlijk' als pathologisch. Daarnaast zijn de items ook op de door Goodchild en Duncan-Jones (1985) voorgestelde alternatieve manier gescoord. Hiertoe werd eerst door drie onafhankelijke beoordelaars vastgesteld voor welke

items de antwoordcategorie 'niet meer dan anders' kon duiden op de aanwezigheid van chronische pathologie. Dit bleek voor 20 items het geval, verder aangeduid als negatieve items. Bij deze 20 items werd derhalve de antwoordcategorie 'niet meer dan anders' ook als pathologisch gescoord. Over de alternatieve versie van de GHQ is nog weinig bekend, als drempelwaarde nemen Goodchild en Duncan-Jones (1985) de score '13'. In dit onderzoek wordt hierbij aangesloten. Samengevat:

- originele scoringswijze (Goldberg): 0 0 1 1 voor alle 28 items. Drempelwaarde '5';
- alternatieve scoringswijze (Goodchild en Duncan-Jones): 0 1 1 1 voor de 20 negatieve items en 0 0 1 1 voor de overige acht items. Drempelwaarde '13'. Deze alternatieve versie is verder aangegeven als de CGHQ.

De Present State Examination

De aanwezigheid, aard en ernst van de psychiatrische problematiek werd bij alle patiënten vastgesteld met behulp van de Present State Examination (Wing, Cooper en Sartorius 1974) en het hiermee verbonden CATEGO-computerprogramma. De PSE is een semi-gestructureerd psychiatrisch interview waarin systematisch naar de aanwezigheid in de voorafgaande vier weken van 140 psychiatrische symptomen wordt gevraagd. Voor de ernst van de aanwezige pathologie werd gebruik gemaakt van de PSE-totaalscore en de PSE Index of Definition (ID). De PSE-totaalscore is een ongewone somscore die een indruk geeft van het totaal aantal psychiatrische symptomen. De ID is een soort gewogen somscore die de waarschijnlijkheid aangeeft dat er sprake is van een specifieke psychiatrische diagnose. De hoogte van de ID kan variëren van 1 tot en met 8, waarbij de scores de volgende betekenis hebben: ID = 1 geen psychiatrische symptomen aanwezig; ID = 2 tot en met 4, wel psychiatrische symptomatologie, maar een specifieke psychiatrische diagnose is onwaarschijnlijk; ID = 5, de drempelwaarde, en ID = 6 tot en met 8 veel (specifieke) psychiatrische symptomen aanwezig, een specifieke psychiatrische diagnose is waarschijnlijk tot zeker. Gezien de wijze waarop de PSE-totaalscore en de ID tot stand komen, hun correlatie (.80), en de aard van de onderzoekspopulatie (geen psychosen), komen beide als ernstmaat in aanmerking. De PSE werd afgenomen door vier ervaren klinisch psychiaters en een klinisch psycholoog die allen een grondige PSE-training hadden genoten.

Op basis van het beloop van hun klachten tussen twee metingen werden de patiënten in vier beloopsgroepen onderverdeeld.

- a. de volledig herstelden: die respondenten van wie de PSE-ID tussen $t(x)$ en $t(x+1)$ is gedaald en op $t(x+1)$ de waarde 1 of 2 heeft;
- b. de verbeterden maar (nog) niet geheel herstelden: de respondenten die hoewel hun PSE-ID tussen $t(x)$ en $t(x+1)$ is gedaald op $t(x+1)$ nog een PSE-ID groter of gelijk aan 3 hebben;

- c. de onveranderden: die respondenten voor wie geldt dat hun PSE-ID op $t(x)$ precies gelijk is aan die op $t(x+1)$;
- d. de verslechterden: die respondenten bij wie de PSE-ID tussen $t(x)$ en $t(x+1)$ is gestegen.

Deze indeling werd driemaal gemaakt. Op basis van het beloop tussen t_1 en t_2 , tussen t_2 en t_3 en tussen t_1 en t_3 . Tabel 1 geeft voor ieder meetinterval de verdeling van de patiënten over de vier beloopsgroepen aan.

Resultaten

Relatie tussen de GHQ-totaalscore en veranderingen in de ernst van het toestandsbeeld – Over het algemeen is er tussen de verschillende metingen sprake van een statistisch significante daling in de gemiddelde GHQ-totaalscore (tabel 1). Deze daling doet zich met uitzondering van de volgende groepen in alle beloopsgroepen en over elk meetinterval voor. De uitzonderingen zijn: de groep verslechterden tussen t_2 en t_3 waar sprake is van een statistisch significante stijging en de groepen verslechterd tussen t_1 en t_3 en onveranderd tussen t_2 en t_3 waar geen statistisch significante veranderingen konden worden vastgesteld.

Twee punten springen in het oog: (1) de groep geheel herstelden blijkt gemiddeld op $t(x+1)$ ook in GHQ-termen onder het casusniveau te komen. De gemiddelde totaalscore ligt ruim onder de drempelwaarde van vijf; (2) behalve bij het interval t_2 - t_3 daalt de GHQ-score ook in de groep onveranderden. Omdat dit verschijnsel zich alleen voordoet bij de vergelijking van de eerste met de andere twee metingen en niet tussen de tweede en de derde meting vermoeden wij dat het zogenaamde hertesteffect een rol heeft gespeeld. Dit is het tot op heden onvoldoende verklaarde fenomeen van verandering in gemiddelde scores bij groepen waarvan men mag verwachten dat hun ware positie op het gemeenten construct niet is veranderd, het lijkt zich vooral voor te doen bij vergelijking met de eerste afname en bij negatief getinte vragen (Henderson e.a. 1981; Windle 1954). Ook Goodchild en Duncan-Jones (1985) rapporteren substantiële hertesteffecten, maar voor de GHQ-30.

Onder de assumptie dat dit effect gemiddeld binnen alle groepen gelijk is en de rangorde niet beïnvloedt, kan het op eenvoudige wijze van de overige effecten worden gescheiden, door van de geobserveerde verklaring in de groepen waarbij veranderingen in de ernst van het toestandsbeeld optreden, de geobserveerde verandering in de groep qua toestandsbeeld gelijkgebleven patiënten (= hertesteffect) af te trekken. Deze correctie werd voor alle drie de intervallen uitgevoerd.

Tabel 2 geeft de veranderingen in gemiddelde GHQ-totaalscores weer nadat het hertesteffect is gecorrigeerd. Nu blijkt de richting van de GHQ-totaalscoreverandering duidelijker te corresponderen met

Tabel 1: De gemiddelde GHQ-scores per beloopsgroep en per interval

Beloopsgroep	Meting				
	t1		t2		
	gemiddelde	st. dev.	gemiddelde	st. dev.	
Hersteld (n = 31)	14.85	8.36	1.72	2.53	***
Verbeterd nog niet hersteld (n = 44)	17.51	6.79	7.15	6.52	***
Onveranderd (n = 39)	15.68	8.38	9.62	7.04	***
Verslechterd (n = 20)	15.64	7.09	13.87	7.49	ns

Beloopsgroep	Meting				
	t2		t3		
	gemiddelde	st. dev.	gemiddelde	st. dev.	
Hersteld (n = 26)	8.48	7.18	2.96	4.14	***
Verbeterd nog niet hersteld (n = 33)	10.21	8.15	7.21	7.22	***
Onveranderd (n = 41)	6.41	6.90	6.48	8.29	ns
Verslechterd (n = 32)	6.15	6.04	9.38	8.48	**

Beloopsgroep	Meting				
	t1		t3		
	gemiddelde	st. dev.	gemiddelde	st. dev.	
Hersteld (n = 56)	15.13	7.92	2.61	3.83	***
Verbeterd nog niet hersteld (n = 62)	16.80	7.09	6.88	7.15	***
Onveranderd (n = 36)	14.02	8.71	10.32	8.24	***
Verslechterd (n = 19)	13.72	8.39	13.57	9.38	ns

* $p \leq .10$ ** $p \leq .05$ *** $p \leq .01$

Tabel 2: Veranderingen in de gemiddelde GHQ-score per beloopsgroep en per interval gecorrigeerd voor hertesteffect

Tijdsinterval	Beloopsgroep		
	Hersteld	Verbeterd nog niet hersteld	Verslechterd
t1-t2	-7.13 ***	-4.37 ***	+4.23 ***
t2-t3	-5.60 ***	-3.07 ***	+3.20 **
t1-t3	-8.83 ***	-6.22 ***	+3.55 ***

*** $p \leq .01$ ** $p \leq .05$

Tabel 3: De gemiddelde GHQ-scores per beloopsgroep en per interval

Beloopsgroep	Meting				
	t1		t2		
	gemiddelde	st. dev.	gemiddelde	st. dev.	
Hersteld (n = 31)	20.07	5.68	7.60	5.39	***
Verbeterd nog niet hersteld (n = 44)	21.26	4.66	15.79	5.77	***
Onveranderd (n = 39)	20.02	6.09	16.23	6.57	***
Verslechterd (n = 20)	19.55	5.68	19.52	5.80	ns

Beloopsgroep	Meting				
	t2		t3		
	gemiddelde	st. dev.	gemiddelde	st. dev.	
Hersteld (n = 26)	15.16	6.04	8.24	6.12	***
Verbeterd nog niet hersteld (n = 33)	17.81	5.99	14.09	6.56	***
Onveranderd (n = 41)	13.61	7.71	11.79	8.56	**
Verslechterd (n = 32)	12.47	7.26	15.91	8.36	***

Beloopsgroep	Meting				
	t1		t3		
	gemiddelde	st. dev.	gemiddelde	st. dev.	
Hersteld (n = 56)	19.94	5.65	7.70	5.95	***
Verbeterd nog niet hersteld (n = 62)	21.30	4.29	14.80	6.46	***
Onveranderd (n = 36)	19.59	5.64	16.06	7.62	***
Verslechterd (n = 19)	18.86	5.97	19.45	7.53	ns

* $p \leq .10$ ** $p \leq .05$ *** $p \leq .01$

Tabel 4: Veranderingen in de gemiddelde GHQ-score per beloopsgroep en per interval gecorrigeerd voor hertesteffect

Tijdsinterval	Beloopsgroep		
	Hersteld	Verbeterd nog niet hersteld	Verslechterd
t1-t2	-8.68 ***	-1.77 **	+3.76 ***
t2-t3	-5.09 ***	-1.90 **	+5.27 ***
t1-t3	-8.72 ***	-2.98 ***	+4.12 ***

*** $p \leq .01$ ** $p \leq .05$

Tabel 5: Spearman rangcorrelaties tussen de verandering in gemiddelde (C)GHQ-score en verandering in de PSE-ID en PSE-totaalscore

		V-ID	V-TPSE
T1/T2	V-GHQ	.45	.54
	V-CGHQ	.53	.53
T1/T3	V-GHQ	.54	.60
	V-CGHQ	.55	.52
T2/T3	V-GHQ	.45	.45
	V-CGHQ	.52	.47

V-ID = verschil tussen de PSE-ID-score op $t(x)$ en $t(x+1)$

V-TPSE = verschil tussen de PSE-totaalscore op $t(x)$ en $t(x+1)$

V-GHQ = verschil in GHQ-totaalscore tussen $t(x)$ en $t(x+1)$

V-CGHQ = verschil tussen CGHQ-totaalscore tussen $t(x)$ en $t(x+1)$

die van de PSE-ID. Over ieder tijdsinterval vertonen de herstelden en verbeterden een significante daling terwijl de groep verslechterden een significante stijging laten zien.

In tabel 3 zijn veranderingen in CGHQ-scores gepresenteerd en in tabel 4 de voor het hertesteffect gecorrigeerde veranderingen. Vergelijking van de voor hertest gecorrigeerde gemiddelde GHQ- en CGHQ-scores (tabel 2 en 4) laat een belangrijk verschil zien. Als in het meetinterval de pathologie verdwijnt (de groep herstelden) is er tussen de GHQ en CGHQ geen verschil in de grootte van de verandering. Indien er aan het eind van het meetinterval nog steeds sprake is van pathologie (de groep verbeterd maar nog niet hersteld), vertoont de CGHQ relatief een minder grote daling.

De relatie tussen veranderingen in de ernst van het toestandsbeeld en veranderingen in de (C)GHQ-totaalscore kan ook worden vastgesteld aan de hand van de correlaties tussen beide. Tabel 5 geeft van zowel de GHQ- als de CGHQ-totaalscore de correlatie met de verandering in de PSE-ID- respectievelijk PSE-totaalscore. Deze correlaties liggen alle rond de .50. In verband met de onbetrouwbaarheid van verschilsscores vormen ze waarschijnlijk een onderschatting van de ware correlaties (Fergusson 1971).

Samenvatting en discussie

Op basis van de gepresenteerde gegevens kan ten aanzien van de GHQ-28 het volgende worden geconcludeerd (1) De GHQ-28 lijkt, mits gecorrigeerd voor het hertesteffect, een goede toestandsmaat die veranderingen in de ernst van het toestandsbeeld goed weergeeft. (2) De alternatieve scoringswijze (CGHQ) leidt ten aanzien van de onderzoeksvraag niet tot superieure resultaten.

Bij gebruik van de PSE-ID- en PSE-totaalscore als extern criterium blijkt de GHQ-28 veranderingen in de ernst van psychiatrische stoor-

nissen goed te volgen. Nadat de patiënten op basis van de verandering in hun PSE-ID tussen $t(x)$ en $t(x+1)$ in vier beloopsgroepen waren verdeeld, bleek na een correctie voor het hertesteffect, dat de groepen herstelde en verbeterde patiënten een statistisch significante daling en de groep verslechterde patiënten een statistisch significante stijging in gemiddelde GHQ-28-totaalscore vertoonden. De correlatie tussen de verandering in de PSE-ID- en PSE-totaalscore enerzijds, en veranderingen in de GHQ-totaalscore anderzijds, liggen rond de .50, wat gezien de onbetrouwbaarheid van verschillscores alleszins acceptabel is. Bovenstaande resultaten vormen naast de reeds in ander onderzoek gerapporteerde synchronische correlaties tussen GHQ- en PSE-scores een extra validering van de GHQ als toestandsmaat. De hoogte van de GHQ-score geeft kennelijk niet alleen antwoord op de vraag hoe waarschijnlijk het is dat iemand al dan niet een casus is, maar is ook een goede indicator van de ernst van de stoornis.

De door ons toegepaste correctie voor het hertesteffect is gebaseerd op twee assumpties waarbij op deze plaats enige kanttekeningen geplaatst dienen te worden. De eerste aanname was dat binnen de groep patiënten die tussen $t(x)$ en $t(x+1)$ qua ernst van hun toestandsbeeld onveranderd zijn, de daling in gemiddelde GHQ-score uitsluitend door het hertesteffect wordt veroorzaakt. Als kritiek kan men aanvoeren dat de geconstateerde daling mede veroorzaakt kan zijn door de insensitiviteit van de GHQ voor chronische klachten. Respondenten wordt gevraagd hun huidige toestand te vergelijken met hun 'usual self'. Bij een aantal patiënten kunnen de chronische symptomen na verloop van tijd deel gaan uitmaken van dit 'usual self', zodat symptomen die op $t(x)$ als afwijkend van hun 'usual self' worden ervaren (en dus gescoord) dat op $t(x+1)$ niet worden (en dus niet meer worden gescoord). Dit verschijnsel zou met name in de groep onveranderden spelen. Om de volgende redenen lijkt dit proces hier geen rol van betekenis te spelen:

(1) Het lijkt onwaarschijnlijk dat het proces waarin klachten deel gaan uitmaken van het 'usual self' en als egosyntoon worden ervaren zich in de relatief korte periode tussen de metingen (resp. 4 en 10 maanden) heeft kunnen voltrekken.

(2) Als het wel een rol zou spelen, hadden wij in de groep van veranderden ook tussen de tweede en derde meting een daling van de GHQ-score moeten vinden. Deze daling heeft zich echter niet voorgedaan, wat weer correspondeert met aanwijzingen dat het hertesteffect vooral optreedt tussen de eerste en daaropvolgende metingen (Henderson e.a. 1981).

(3) Bovendien zou in het geval dat het proces een rol speelt, een verschuiving hebben moeten optreden van (veel)-meer-dan-gewoonlijk antwoorden naar niet-meer-dan-anders antwoorden. Een verschuiving die tevens sterker diende te worden naarmate de lengte van het meetinterval toeneemt. Deze verschuiving kan worden vastgesteld

aan de hand van het verschil tussen de verandering in de GHQ- en de verandering in de CGHQ-scores. Tussen t1 en t2 was deze verschuiving er inderdaad, maar niet tussen t2 en t3.

(4) In een ander onderzoek (Koeter e.a. 1987), bleek dat de CGHQ-totaalscore niet beïnvloed wordt door de duur van de klachten. De daling van de CGHQ-totaalscore tussen t1 en t2 in de groep onveranderden, kan dus moeilijk worden toegeschreven aan de eventuele insensitiviteit van de GHQ voor chronische klachten. Na correctie voor het hertesteffect laten de GHQ en CGHQ ten aanzien van veranderingen in de ernst van het toestandsbeeld bovendien een vergelijkbaar beeld zien. Een tweede aanname was dat het hertesteffect in alle vier de beïnvloeds-groepen even sterk is. Hier doet zich het volgende probleem voor. Als verklaring voor het hertesteffect wordt wel aangevoerd dat respondenten geneigd zijn bij een vervolgmeting sociaal wenselijk te antwoorden (Henderson e.a. 1981). De mate waarin deze tendens zich kan manifesteren is afhankelijk van de mate waarin men klachten heeft op t(x+1). Voor de groep herstelden, die is samengesteld uit patiënten die op t(x+1) volgens een extern criterium nauwelijks tot geen symptomen meer hebben, is het bijna onmogelijk om op de GHQ sociaal wenselijk te scoren. Dit zou betekenen dat de hertestcorrectie het best alleen kan worden uitgevoerd op groepen waarbij ook op t(x+1) nog sprake is van pathologie. Het niet toepassen van een hertestcorrectie bij de groep herstelden, zou echter geen invloed hebben op de gepresenteerde conclusies van dit onderzoek. Al deze overwegingen leiden ertoe dat wij menen dat onze schatting van het hertesteffect redelijk betrouwbaar is.

Ten slotte willen we nog ingaan op de vraag naar de generaliseerbaarheid van de resultaten. Deze vraag valt uiteen in twee deelvragen. De eerste vraag betreft de generaliseerbaarheid naar andere versies van de GHQ. Hoewel er een groot aantal GHQ-versies in omloop zijn (GHQ-12, GHQ-20, GHQ-28, GHQ-30, GHQ-60), is de itemoverlap tussen de verschillende versies aanzienlijk. De GHQ-12 en GHQ-20 bestaan volledig uit GHQ-28-items, terwijl de GHQ-30 vijftien items van de GHQ-28 bevat. Dit maakt ons inziens generalisatie naar deze GHQ-versies mogelijk. Voor de GHQ-60 ligt het anders. Hoewel wij geen redenen zien waarom deze oorspronkelijke versie zich afwijkend zou gedragen van de verkorte versies, is voor een definitief antwoord verder onderzoek nodig. De tweede generalisatievraag betreft de generalisatie naar andere populaties dan nieuwe psychiatrische polikliniekpatiënten. Er is gekozen voor een psychiatrische polikliniekpopulatie omdat voor de beantwoording van de onderzoeksvraag substantiële verschuivingen in de ernst van het toestandsbeeld noodzakelijk waren. Ons inziens is er echter geen reden om aan te nemen dat de GHQ veranderingen in de ernst van het toestandsbeeld bij huisartspatiënten of in de bevolking minder goed zou volgen.

De GHQ-28 lijkt al met al een goede toestandsmaat op te leveren die

veranderingen in de ernst van het toestandsbeeld goed volgt. De door Goodchild en Duncan-Jones (1986) voorgestelde alternatieve scoringswijze levert geen duidelijke winst op.

Literatuur

- Banks, M. (1983), Validation of the General Health Questionnaire in a young community sample. *Psychological Medicine*, 13, 349-353.
- Benjamin, S., P. Decalmer en D. Haran (1982), Community screening for mental illness: a validity study of the General Health Questionnaire. *British Journal of Psychiatry*, 140, 174-180.
- Bridges, K.W., en D.P. Goldberg (1986), The validation of the GHQ-28 and the use of MMSE in neurological in-patients. *British Journal of Psychiatry*, 148, 548-553.
- Derogatis, L.R., R.S. Lipman en L. Covi (1973), SCL-90 an outpatient psychiatric ratingscale. *Preliminary Report Psychopharmacology bulletin* 9, 13-27.
- Fergusson, G.A. (1971), *Statistical analysis in psychology and education*. Londen, McGraw-Hill.
- Finlay-Jones, R.A., en E. Murphy (1979), Severity of psychiatric disorder and the 30-item GHQ. *British Journal of Psychiatry*, 134, 609-616.
- Goldberg, D.P., B. Cooper, M.R. Eastwood, H. Kedward en M. Sheperd (1970), A standardised psychiatric interview suitable for use in community surveys. *British Journal of Preventive and Social Medicine*, 24, 18-23.
- Goldberg, D. (1972), *The detection of general illness by questionnaire*. Maudsley monograph no 21. O.U.P.
- Goldberg, D.P., K. Rickels, R.Q. Downing en P. Hesbacher (1976), A comparison of two psychiatric screening tests. *British Journal of Psychiatry*, 129, 61-67.
- Goldberg, D. (1978), *Manual of the General Health Questionnaire*. NFER publishing company.
- Goldberg, D.P., en V.F. Hillier (1979), A scaled version of the GHQ. *Psychological Medicine* 9, 139-145.
- Goodchild, M.E., en P. Duncan-Jones (1985), Chronicity and the GHQ. *British Journal of Psychiatry*, 146, 55-61.
- Henderson, S.H., D.G. Byrne en P. Duncan-Jones (1981), *Neurosis and the social environment*. Academic Press Australia.
- Henderson, S.H. (1981), Social relationships, adversity and neurosis: an analysis of prospective observations. *British Journal of Psychiatry*, 138, 391-398.
- Hodiamont, P.P.G. (1986), *Het zoeken van zieke zielen*. Hoofdstuk 8, dissertatie, Katholieke Universiteit Nijmegen.
- Koeter, M.W.J., J. Ormel, W. van de Brink, en W. Dijkstra (1986), Toestandsbeeld en persoonlijkheidsmeting: de stabiliteit van de Nederlandse Persoonlijkheids vragenlijst in een psychiatrische polikliniek populatie. Aanboden aan het *Nederlands Tijdschrift voor de Psychologie*.
- Koeter, M.W.J., W. van de Brink en J. Ormel (1987), De sensitiviteit van de GHQ voor chronische klachten. Aanboden aan het *Tijdschrift voor Psychiatrie*.
- Luteijn, F. (1974), *De constructie van een persoonlijkheidsvragenlijst (NPV)*. Lisse, Swets en Zeitlinger.

- Rabin, P.V., en B.R. Brooks (1981), Emotional disturbances in multiple sclerosis patients. Validity of the GHQ. *Psychological Medicine*, 425-427.
- Tarnapolsky, A., D.J. Hand, E.K. McLean, H. Roberts en R.D. Wiggins (1979), Validity and the uses of a screening questionnaire (GHQ) in the community. *British Journal of Psychiatry*, 134, 508-515.
- Windle, C. (1954), Test-retest effect on personality questionnaires. *Educational and Psychological measurement* 14, 617-633.
- Wing, J.K., J.E. Cooper en N. Sartorius (1974), *Measurement and classification of psychiatric symptoms: an instruction manual for the PSE and Catego program*. Londen, Cambridge University Press.

Schrijvers zijn respectievelijk: sociaal-wetenschappelijk onderzoeker, vakgroep Psychiatrie RUG; socioloog, vakgroepen Psychiatrie en Medische Sociologie RUG; psychiatrisch epidemioloog, vakgroep Psychiatrie RUG; klinisch psycholoog, vakgroep Psychiatrie RUG; psychiater, Diaconessenhuis Groningen en psychiater, Rooms Katholiek Ziekenhuis Groningen. Correspondentie-adres: M.W.J. Koeter, vakgroep Psychiatrie Rijksuniversiteit Groningen.

De gegevens voor dit artikel zijn verkregen in een door de Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek onder nummer 135703 gesubsidieerd onderzoek naar de validiteit van de GHQ.