

University of Groningen

Wat is ons eigen oordeel waard?

Koeter, Maarten Wilhelmus Johannes

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

1992

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Koeter, M. W. J. (1992). *Wat is ons eigen oordeel waard? de vragenlijstmethode in de psychiatrie*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

In dit proefschrift wordt de betekenis van de zelfbeoordelingsvragenlijst in het psychiatrisch (epidemiologisch) onderzoek besproken. Achtereenvolgens wordt ingegaan op de validiteit van de General Health Questionnaire (GHQ-28) en de Symptom Check List (SCL-90) als algemene ernstmaat, de validiteit van de angst- en depressieschalen van de GHQ en de SCL en de waarde van de Nederlandse Persoonlijke Vragenlijst (NPV) in een psychiatrische populatie met wisselende psychopathologie.

De kern van het proefschrift bestaat uit een 6-tal artikelen over deze onderwerpen die in afgelopen 5 jaren in verschillende nationale en internationale tijdschriften zijn gepubliceerd. Deze artikelen worden voorafgegaan door een inleidend hoofdstuk waarin wordt ingegaan op de relatie tussen interview- en vragenlijstmethode in de psychiatrie. Het proefschrift eindigt met een discussie en conclusies.

Centraal in **hoofdstuk 1** staat de vergelijking van de zelfbeoordelingsvragenlijst met het klinische interview.

Zowel vragenlijst als interview zijn gericht op de fenomenologische beschrijving van de psychopathologie van de patiënt. Het verschil tussen beide is dat vragenlijsten zich richten op klachten, dat wil zeggen afwijkingen van de persoonlijke norm, terwijl het psychiatrisch interview zich richt op symptomen, dat wil zeggen afwijkingen van de klinische norm.

Er wordt uitvoerig stilgestaan bij de voor- en nadelen van de interview- en de vragenlijstmethode. De balans tussen voor- en nadelen hangt af van het te meten fenomeen, de expertise, de financiële ruimte en de tijd die men ter beschikking heeft. Zelfbeoordelingsvragenlijsten zijn sterk gestructureerd en gestandaardiseerd. Ze hebben in het algemeen een hoge betrouwbaarheid. Hetzelfde geldt voor volledig gestructureerde en gestandaardiseerde psychiatrische interviews als de DIS en de CIDI. Het klinisch interview is daarentegen meestal weinig gestructureerd en gestandaardiseerd en heeft dienengevolge een lagere betrouwbaarheid. Semi-gestructureerde interviews als de PSE en de SCAN nemen een tussenpositie in.

Hierbij is sprake is van een redelijke structurering en een sterke standaardisering met als gevolg een goede betrouwbaarheid.

Bij gebrek aan een gouden standaard in de psychiatrie, kan men in het geval van discrepanties tussen zelfbeoordeling (vragenlijst) en expert-rating (interview), de vraag wat de meest valide representatie is van de 'werkelijke' score, formeel niet beantwoorden. Het gegeven dat interviewdata meer aansluiten bij de klinische praktijk en de assumptie dat de klinische relevantie van vragenlijstgegevens die geen relatie vertonen met bestaande klinische syndromen, niet direct duidelijk is, maakt dat in de praktijk interview-data als meer valide worden beschouwd dan zelfbeoordelings-data. Dit uit zich in het gegeven dat vragenlijsten meestal worden gevalideerd aan een psychiatrisch interview. Deze traditie wordt in dit proefschrift op een kritische wijze voortgezet.

Aan het einde van dit hoofdstuk wordt ingegaan op de veranderde rol van zelfbeoordelingsvragenlijsten in het psychiatrisch epidemiologisch onderzoek.

Het hoofdstuk wordt besloten met een beschrijving van het 'Poli-project'; een longitudinaal onderzoek met 3 meetmomenten onder psychiatrische polikliniekpatiënten die gedurende een aaneengesloten periode van 6 maanden door het Academisch Ziekenhuis Groningen en de twee algemene ziekenhuizen van de stad Groningen in behandeling zijn genomen. Op dit onderzoek zijn alle in het proefschrift opgenomen artikelen gebaseerd.

Hoofdstuk 2 bestaat uit twee artikelen en beschrijft de waarde van de GHQ-28 en de SCL-90 als maat voor de ernst van psychopathologie. Cross-sectioneel worden er substantiële correlaties gevonden tussen de totaalscore van beide vragenlijsten en de totaalscore van de Present State Examination (PSE). Binnen het domein van de niet-psychotische psychiatrische stoornissen kunnen zowel de GHQ- als de SCL-totaalscore dus als indicator voor de ernst van psychopathologie, zoals gemeten met een psychiatrisch interview, worden gebruikt.

Gebruik van vragenlijsten als indicator voor **verandering** in psychopathologie vereist dat er naast een cross-sectionele correlatie, ook longitudinaal sprake is van een relatie tussen de vragenlijst- en de interviewscore. De GHQ- en SCL-totaalsco-

res vertonen naast substantiële cross-sectionele correlaties ook substantiële longitudinale correlaties met de PSE-totaal score en de PSE-index of definition. Het gebruik van de GHQ en SCL als maat voor verandering in de psychopathologie wordt echter bemoeilijkt door de aanwezigheid van het zogenaamde 'hertest-effect'. Dit hertest-effect wordt gedefinieerd als de gemiddelde verandering in vragenlijstscores in de groep patiënten die in termen van het psychiatrisch interview niet veranderd is. Correctie voor het hertest-effect is noodzakelijk om verandering in de vragenlijstscores te kunnen interpreteren als verandering in psychopathologie. Zonder deze correctie is er namelijk binnen de op basis van het psychiatrisch interview gedefinieerde beloopsgroepen 'onveranderd' en 'verslechterd' nog sprake van een gemiddelde verbetering op de vragenlijst.

Hoofdstuk 3 gaat in op het effect van de specifieke antwoordcategorieën van de GHQ op de gevoeligheid voor de detectie van chronische klachten. In tegenstelling tot bijvoorbeeld de SCL-90 vraagt de GHQ niet naar de aan- of afwezigheid van klachten, maar naar afwijkingen van het voor de patiënt 'normale' functioneren. Bij patiënten met chronische klachten is het niet denkbeeldig dat er een verschuiving van de persoonlijke norm optreedt, waardoor chronische klachten na verloop van tijd niet meer als ego-dystoon worden ervaren en dientengevolge niet meer als afwijkend van de 'normale' situatie worden gescoord. Personen met chronische pathologie zouden zo een verhoogde kans hebben tot de groep 'fout negatieven' te behoren. Deze potentiële ongevoeligheid voor chronische klachten kan worden verminderd door bij alle 'negatieve' items (dit zijn items waarbij de antwoordcategorie 'niet meer dan anders' kan duiden op de aanwezigheid van chronische pathologie) de antwoordcategorie 'niet meer dan anders' ook als pathologisch te scoren. Deze alternatieve scoringswijze wordt aangeduid met de term Changed General Health Questionnaire (CGHQ).

In onze onderzoekspopulatie bleek de chroniciteit van klachten samen te hangen met de hoogte van de GHQ-totaalscore en de kans om met behulp van de GHQ als 'geval' geïdentificeerd te worden. Bij gelijke ernst volgens het psychiatrisch interview scoren patiënten met chronische klachten gemiddeld lager op de GHQ

dan patiënten met recente klachten. De hoogte van de CGHQ-totaalscore bleek niet samen te hangen met de chroniciteit van de klachten. Bovendien bleek de invloed van chroniciteit op de sensitiviteit van de CGHQ kleiner dan op die van de GHQ. Hierbij moet wel de kanttekening gemaakt worden dat de sensitiviteit van de GHQ ook in de groep chronici nog zeer hoog is. De beste resultaten als instrument voor de opsporing van psychische stoornissen worden overigens bereikt door een combinatie van de originele en de alternatieve scoringswijze.

Hoofdstuk 4 bestaat uit twee artikelen en gaat in op het gebruik van de GHQ en SCL voor het beschrijven van de (veranderingen in de) ernst van specifieke psychopathologie, in het bijzonder angst en depressie. In het eerste deel wordt de constructie van een prototypische angst- en een prototypische depressieschaal besproken. De schaalconstructie procedure was er op gericht, door beperking van de itemoverlap, minimaal correlerende schalen te verkrijgen. De correlatie tussen de uiteindelijke schalen bedroeg op de drie meetmomenten respectievelijk .00, .23 en .28. Beide schalen toonden daarnaast voldoende betrouwbaarheid en validiteit. Zo bleken ze sterk gecorreleerd met gangbare gevalideerde dimensionele ernstmaten voor angst en depressie (de Hamilton Anxiety Rating Scale en de Hamilton Rating Scale for Depression) en was er sprake van een goede divergerende en convergerende validiteit ten aanzien van de DSM-III en Bedford College diagnose depressie. Het belang van dergelijke schalen voor het psychiatrisch (epidemiologisch) onderzoek is driedig: (1) Valideringsonderzoek van bestaande angst en depressieschalen leidt zelden of nooit tot een goede divergerende validiteit. Voor een belangrijk deel is dit te wijten aan het feit dat de criterium-instrumenten (bijvoorbeeld de HARS en HDRS) niet erg specifiek zijn. Depressieschalen zijn vaak 'gecontamineerd' met angst-symptomen, terwijl angstschalen gecontamineerd zijn met depressie symptomen. De prototypische schalen kunnen gebruikt worden als relatief specifieke criterium instrumenten. (2) Ondanks een uitgebreide literatuur over het onderwerp blijft de specifieke relatie tussen angst en depressie een onopgelost probleem. Het feit dat de klinische beelden een sterke symptoomoverlap vertonen, maakt het moeilijk om met behulp van bestaande instrumenten een

onderscheid te maken tussen werkelijke comorbiditeit en overlap in de definities en operationalisaties van beide syndromen. De prototypische angst en depressieschalen bieden hier mogelijk uitkomst. (3) Deze schalen maken het mogelijk het longitudinale beloop van angst/depressie comorbiditeit te bestuderen en bieden daarbij de mogelijkheid differentiële effecten van pharmacotherapeutische en andere behandelingen op angst en depressie te onderzoeken.

In het tweede deel van dit hoofdstuk wordt ingegaan op de validiteit van de GHQ- en SCL-angst en depressieschalen. De GHQ- en SCL-depressieschalen bleken in categoriaal opzicht duidelijk gerelateerd aan het concept 'affectieve stoornis' volgens DSM-III. Ook op dimensioneel niveau was er sprake van een goede divergerende en convergerende validiteit ten opzichte van de prototypische angst- en depressieschaal. Voor de SCL-angst- en agorafobieschaal werd eveneens een substantiële relatie met de relevante DSM-III categorieën (angststoornissen) gevonden. Bovendien was er voor beide SCL-schalen sprake van een goede convergerende en divergerende validiteit ten opzichte van de prototypische angst- en depressieschaal. De resultaten voor de GHQ-angst/slapeloosheid schaal waren echter problematisch. Bij deze schaal was er geen sprake van een relatie met DSM-III angststoornissen en bestond er ook géén convergerende en divergerende validiteit ten opzichte van respectievelijk de prototypische angst- en depressieschaal.

Het valideringsonderzoek op subschaalniveau betrof echter alleen de cross-sectionele relatie tussen de subschalen en de criterium scores. De vraag of de GHQ- en SCL-schalen ook gebruikt kunnen worden om veranderingen in angst en depressie vast te stellen (de longitudinale samenhang tussen de subschaalscores en veranderingen in psychopathologie) moet nog worden onderzocht.

Hoofdstuk 5 gaat in op de validiteit van de persoonlijkheidsmeting met behulp van de Nederlandse Persoonlijkheds Vragenlijst (NPV) bij psychiatrische polikliniek patiënten. Hoewel persoonlijkheidsmeting vaak een standaard procedure is bij de intake, zijn er sterke aanwijzingen voor een contaminatie van ten minste een aantal persoonlijkheidsdimensies met het psychiatrisch toestandbeeld van de

tiënt. Over een periode van 4 maanden (te kort om werkelijke veranderingen in de persoonlijkheid te verwachten), vonden wij in een groep psychiatrische polikliniek patiënten substantiële veranderingen in de meeste subschaalscores van de NPV. Bij de oorzaken van deze veranderingen wordt een onderscheid gemaakt tussen het hertest-effect en de instabiliteit van persoonlijkheidsschaalscores als gevolg van contaminatie met het toestandsbeeld.

Bij de Inadequatie-, Verongelijkheid- en Zelfwaarderingsschaal was er sprake van een substantieel hertest-effect. Bij de Inadequatie-, Sociale Inadequatie-, Verongelijkheid- en Zelfwaarderingsschaal was er tevens sprake van contaminatie met het toestandsbeeld. Voor de klinische praktijk betekent dit dat persoonlijkheidsdiagnostiek bij patiënten met tijdelijke psychiatrische symptomatologie mogelijk tot vertekende resultaten leidt. Men kan beter wachten tot het toestandsbeeld verdwenen, of tot sub-klinisch niveau gedaald is. Daarnaast moet men er bij onderzoek naar persoonlijkheidsverandering rekening mee houden dat veranderingen in NPV-vragenlijstcores mede beïnvloed kunnen zijn door het hertest-effect en/of veranderingen in het psychiatrisch toestandsbeeld.

Hoofdstuk 6 bevat de discussie en conclusies. Centraal in dit hoofdstuk staan een uitdieping van het hertest-effect en een bespreking van de potentiële beperkingen van het onderzoek.

Het hoofdstuk begint met een theoretische raamwerk waarbinnen de relatie vragenlijst-interview zinvol kan worden bestudeerd. In dit model wordt aangenomen dat de score op de vragenlijst samenhangt met de interviewscore, maar ook afhankelijk is van overige factoren die onafhankelijk of in samenhang met de interviewscore invloed hebben op de beantwoording van de vragenlijst. Bij deze overige factoren wordt onderscheid gemaakt tussen systematische en random-factoren. Een belangrijke systematische component is het hertest-effect. Er wordt daarbij verondersteld dat dit hertest-effect in alle beloopsgroepen even groot is. Het wordt met andere woorden beschouwd als een constante systematische component. Bij deze benadering van het hertest-effect worden vier kanttekeningen geplaatst. (1) Het hertest effect is relatief gedefinieerd ten opzichte van de PSE. Aannemende dat

ook de PSE onderhevig is aan een hertest-effect betekent dit dat er in feite sprake is van een onderschatting van het werkelijke hertest-effect. (2) Het hertest effect is op groepsniveau gedefinieerd. In onze benadering kan dus niet op individueel niveau voor het hertest-effect worden gecorrigeerd. (3) Het hertest-effect lijkt setting gebonden. Dit hangt mogelijk samen met twee setting-afhankelijke verschillende verklaringen: sociale wenselijkheid en legitimering, die ieder afzonderlijk of in combinatie kunnen werken. (4) De assumptie dat het hertest-effect in alle beloopsgroepen even sterk is vraagt met name uit het oogpunt van de sociale wenselijkheidshypothese enige nuancering. Mogelijk zou een gedifferentieerde correctie voor hertest-effect de voorkeur hebben.

Bij de in dit proefschrift beschreven methoden voor correctie van het hertest-effect is het noodzakelijk dat er een schatting gemaakt kan worden van de verandering in gemiddelde schaalcores bij patiënten die qua psychopathologie onveranderd zijn. Een absolute correctie van het hertest-effect vraagt om een valide vaststelling van de werkelijke verandering in psychopathologie, bijvoorbeeld door een psychiatrisch interview. Als men evenwel het interview afneemt, is het niet langer relevant de vragenlijst af te nemen. Een relatieve correctie van het hertest-effect kan plaatsvinden met behulp van een controlegroep. Onder de assumptie dat de groepen vergelijkbaar zijn op alle overige factoren die van invloed zijn op de vragenlijstscore, vindt de correctie voor het hertest effect in de t-toets tussen experimentele en controlegroep impliciet plaats. Hoewel de methode met een controlegroep geen volledige oplossing biedt in de situatie waarin men een gedifferentieerde correctie nodig heeft, wordt beargumenteerd dat een ongedifferentieerde correctie in deze situatie leidt tot een conservatieve schatting van het werkelijke effect.

In het tweede deel van hoofdstuk 6 wordt ingegaan op een aantal potentiële beperkingen van het dissertatie onderzoek: (1) beperkingen van de statistische validiteit ten gevolge van het gebruik van verschillcores en het ontbreken van een gouden standaard, (2) beperkingen van de interne validiteit als gevolg van de gevolgde steekproefprocedure en (3) beperkingen van de externe validiteit als gevolg van de keuze voor psychiatrische polikliniekpatiënten. Beargumenteerd

wordt dat er geen redenen zijn om te twijfelen aan de interne, externe en statistische validiteit van het dissertatie onderzoek.

Op basis van het in dit proefschrift beschreven onderzoek kan geconcludeerd worden dat er sprake is van een substantiële samenhang tussen gegevens verkregen met behulp van de vragenlijstmethode en gegevens verkregen met behulp van de interviewmethode. Deze samenhang werd zowel cross-sectioneel als longitudinaal gevonden. Bij het gebruik van vragenlijstcores als maat voor verandering in psychopathologie, moet echter wel rekening gehouden worden met een relatief sterk hertest-effect. Negeren van het hertest-effect leidt tot een te optimistisch beeld van de veranderingen in de ernst van de psychopathologie.