

Focus Betrouwbaarheid van een semi-gestructureerd interview voor de Hamilton-depressieschaal

door R.W. Kupka, F. de Jonghe, M. Koeter en H.D.B. Vermeulen

Gepubliceerd in 1996, no. 10

Samenvatting

Wij onderzochten de tussenbeoordelaarsbetrouwbaarheid van een Nederlandstalig semi-gestructureerd interview voor de Hamilton-depressieschaal. Er werd gebruik gemaakt van de interviewer-observatormethode en de resultaten werden vergeleken met een recent gepubliceerd onderzoek. Uit de bevindingen blijkt dat het gebruik van een semi-gestructureerd interview het betrouwbaar toepassen van de Hamilton-schaal, ook door beoordelaars zonder klinisch-psychiatrische ervaring, mogelijk maakt.

Inleiding

De depressieschaal van Hamilton (1960, 1967) is de meest gebruikte observatieschaal voor het vaststellen van de ernst van een depressie. Max Hamilton schrijft in 1967 dat de eerste versie van zijn depressieschaal 'must be regarded as little more than a preliminary report, intended to arouse interest and encourage the use and development of such a scale'. Dat hij dit doel ruimschoots bereikt heeft, blijkt uit het grote aantal versies dat inmiddels van deze meetschaal bestaat, zowel in het Engels als in andere talen.

De Hamilton-schaal is door meerdere auteurs gemodificeerd en/of voorzien van toelichtingen, zoals Guy (1976), Bech e.a. (1986), Rehm en O'Hara (1985) en Williams (1988). Een inventarisatie in ons land (Zitman e.a. 1989) leverde zes Engelse en zeven Nederlandse versies op, waarvan een deel nooit is gepubliceerd.

Dat verschillende versies van een meetschaal uiteenlopende meetresultaten opleveren, ligt voor de hand. Zo vonden Hooijer e.a. (1991) bij het gebruik van de in het Nederlands vertaalde versie van Bech e.a. (1989) significant lagere somscores dan bij het gebruik van de vertaling van Dijkstra (1974). Verder hing de betrouwbaarheid ervan onder meer samen met de klinische ervaring en/of training van de beoordelaar: onervaren beoordelaars komen tot lagere somscores. Ook indien meerdere beoordelaars dezelfde meetschaal gebruiken, kunnen er verschillen optreden. Wanneer een bepaalde patiënt meerdere malen met eenzelfde meetinstrument onderzocht wordt, zijn er vijf mogelijke bronnen van variantie in de resultaten (Van den Brink 1989): subjectvariantie, situatievariantie, informatievariantie, observatievariantie en criteriumvariantie. Het structureren van een meetinstrument leidt tot vermindering van de informatie- en observatievariantie, standaardisering ervan reduceert de criteriumvariantie.

Williams (1988) beschrijft een gestructureerd interview voor de 21-item Hamilton-schaal. Zij vergeleek de test-hertestbetrouwbaarheid van dit interview op item-niveau met de test-hertestbetrouwbaarheid zoals die werd vastgesteld door Cicchetti en Prusoff (1983). De betrouwbaarheid was bij 18 van de 21 items groter bij het gestructureerde interview. Voor het Nederlandse taalgebied is door De Jonghe (1994) een semi-gestructureerd interview opgesteld voor de 17-item Hamilton Depression Rating Scale (HDRS), gebaseerd op de versie uit 1967 en de toelichtingen die daarbij werden gegeven. In het hier beschreven onderzoek wordt de

tussenbeoordelaars-betrouwbaarheid van dit semi-gestructureerde interview beschreven.

Wij onderzochten hoe de betrouwbaarheid van het semi-gestructureerde interview zich verhoudt tot die van een niet-gestructureerd afgenomen Hamilton-schaal, en of de betrouwbaarheid samenhangt met de ernst van de depressie. Ook onderzochten wij of getrainde beoordelaars zonder klinisch-psychiatrische ervaring met behulp van dit instrument even betrouwbaar kunnen beoordelen als onderzoekers met klinische ervaring.

Methode

Design en steekproef - In de Leidraad voor het scoren van de Hamilton Depression Rating Scale (De Jonghe 1994) worden de zeventien items zo nauwkeurig mogelijk omschreven, evenals de ankerpunten en de beslisregels die leiden tot het toekennen van een score. De tijdsperiode waarop de vraag betrekking heeft, betreft de zeven dagen voorafgaande aan het interview, inclusief de dag van het interview zelf. Tevens worden voor elk item openingsvragen geformuleerd en suggesties gedaan voor verdere vragen. De volgorde waarin de items aan de orde komen, ligt vast en wijkt af van de nummering van de items, waardoor het gesprek met de minst belastende onderwerpen begint. Waar nodig geeft de leidraad aanvullende instructies over het item en de wijze van scoren, om een zo groot mogelijke betrouwbaarheid te bereiken. Behalve een gedetailleerde omschrijving, staan de criteria op het scoringsformulier nogmaals samengevat weergegeven.

In analogie met de studie van Williams (1988), wilden wij onze uitkomsten vergelijken met de resultaten van anderen. Wij namen daarvoor het onderzoek van Hoencamp e.a. (1992), die gebruik maakten van de in het Nederlands vertaalde versie van Bech e.a. (D'Haenen en Verhoeven 1989). Evenals deze onderzoekers kozen wij derhalve voor een interviewer-observator-design. Drie psychiaters (FdJ, HV, RK), met ruime klinische ervaring en getraind in de semi-gestructureerde afname van de Hamilton-depressieschaal, interviewden respectievelijk acht, acht en zes patiënten. Deze interviews werden op videoband vastgelegd en beoordeeld door de beide andere psychiaters en door vier klinisch psychologen met relatief weinig klinische ervaring, maar eveneens getraind in het afnemen van het interview. Alle 22 patiënten voldeden aan de criteria voor een DSM-III-R-diagnose depressie in engere zin, vastgesteld op basis van psychiatrische anamnese en onderzoek. Om een ruime variatie in ernst van de depressie te verkrijgen zijn zowel patiënten van een polikliniek (8), van een open opnameafdeling (8) als van een gesloten opnameafdeling (6) in het onderzoek betrokken. Patiënten werden geïnterviewd in volgorde van aanmelding respectievelijk opname.

Statistische bewerking - Als maat voor de overeenstemming wordt kappa' gebruikt. In deze maat wordt de geobserveerde proportie-overeenstemming (Po) gecorrigeerd 'voor toevalsovereenstemming (Pe):

$$\text{kappa} = (\text{Po} - \text{Pe}) / (1 - \text{Pe}).$$

Kappa kan worden geïnterpreteerd als dat deel van de maximaal mogelijke overeenstemming (na correctie voor toeval) dat werkelijk is bereikt. Bij een kappa van .60 kan van voldoende en bij een kappa van .80 van een goede interbeoordelaarsbetrouwbaarheid worden gesproken. Bij de berekening van de kappa maakt de omvang van de discrepantie tussen de scores van de standaard en de beoordelaars niets uit. Dit maakt de kappa niet geschikt als overeenstemmingsmaat bij een ordinale variabele met een relatief grote range, zoals de totaalscore op de HDRS. Een 'gewogen' versie van de kappa, waarbij grote discrepanties zwaarder tellen dan kleine discrepanties, ligt hier meer voor de hand. De kappa is berekend middels het computerprogramma AGREE. Voor de items wordt de ongewogen kappa gepresenteerd, voor de totaalscore ook de gewogen kappa. In dit onderzoek is steeds (arbitrair) de score van de auteur van het semi-gestructureerde interview (F. de Jonghe) als standaard genomen, waartegen de beoordelingen van de overige psychiaters en psychologen werden afgezet.

Resultaten

De ernst van de depressie varieerde bij de onderzochte personen van matig tot zeer ernstig. De gemiddelde somscore was 26.6, met een spreiding van .15 - .40. De resultaten van de kappawaarden ten opzichte van de standaard (gemiddelden en spreiding) staan vermeld in tabel 1, waarin tevens de vergelijking met de resultaten van Hoencamp e.a. (1992) is opgenomen.

Tabel 1:

Gemiddelde kappawaarden en spreiding, vermeld voor elk HDRS-item en voor de somscore; de bovenste items hebben het grootste verschil met het vergelijkingsonderzoek

Op itemniveau is er in het algemeen een hoge overeenstemming tussen de beoordelaars. De kappa's van alle beoordelaars ten opzichte van de standaard liggen allemaal boven de .60. De gemiddelde kappa ligt voor elk item boven de .70. De items 8 (geremdheid), 9 (agitatie), 14 (seksualiteit) en 17 (inzicht) leveren de grootste verschillen op. De kappawaarden liggen bij elf van de zeventien items hoger dan in het vergelijkingsonderzoek. In dit verschil komt mogelijk het effect van een grotere mate van structurering van het interview en standaardisering van de scoring tot uiting. De kappa is bij drie items gelijk en bij drie items lager (tabel).

Op het niveau van de somscore is de overeenstemming voldoende (gemiddelde ongewogen kappa .63 met een spreiding van .39 - .74), maar beduidend lager dan op itemniveau. Dit is vergelijkbaar met die van het vergelijkingsonderzoek (kappa .65, spreiding .36 - .83). Overigens is een lineair gewogen kappa een meer realistische maat voor overeenstemming over de somscore; hierbij wegen grote verschillen zwaarder dan kleine verschillen. De gemiddelde lineair gewogen kappa met betrekking tot de somscore bedraagt .92 met een spreiding van .85 - .93.

De door ons onderzochte populatie scoorde gemiddeld hoger op de HDRS dan de door Hoencamp e.a. onderzochte groep (resp. 26.6 en 21.95). Het is voorstelbaar dat meer uitgesproken psychopathologie leidt tot eenduidiger beoordeling. Wij onderzochten of er een correlatie bestond tussen de mate van overeenstemming over de somscore en de ernst van de depressie (HDRS-somscore). Deze correlatie was laag: gemiddeld .22 (spreiding .09 - .34). Dit betekent dat de betrouwbaarheid van het interview niet toe- of afneemt naarmate de onderzochte patient ernstiger depressief is.

Ten slotte bleek dat getrainde beoordelaars zonder klinische ervaring zeker zo betrouwbaar scoren als eveneens getrainde psychiaters. Zowel op item- als op somscoreniveau waren hun kappa's vaak gelijk en soms zelfs hoger.

Bespreking

Indien men niet alleen de globale verandering van een depressie wil vaststellen door middel van de HDRS-somscore, maar daarnaast inzicht wil krijgen in de relatieve veranderingen in het symptoomprofiel, is betrouwbaarheid op itemniveau noodzakelijk. Zowel uit het onderzoek van Williams (1988) als uit ons onderzoek blijkt dat een semi-gestructureerd interview deze betrouwbaarheid doet toenemen in vergelijking tot een vrij interview. Wij vonden bij alle items een hoge tussenbeoordelaarsbetrouwbaarheid. De overeenstemming is het grootst bij items waarbij de verbale uitingen van de patiënt bepalend zijn voor de score. Bij typische observatie-items (8, remming, en 9, agitatie) vonden wij een lagere overeenstemming. Dit verschijnsel is ook door andere onderzoekers vastgesteld (Slooff e.a. 1983). Een mogelijke verklaring voor de relatief lage overeenstemming bij de items 14 (seksualiteit) en 17 (inzicht) is van geheel andere aard. Het betreft hier de enige items waarbij niet de aanwezigheid maar de afwezigheid van een verschijnsel beoordeeld

wordt. In de nabespreking bleek dat er meerdere malen foutief was gescoord: 'afwezigheid van' scoren als 0 in plaats van 2. Bij training dient op de mogelijkheid van deze fout gewezen te worden.

Conclusie

Het toepassen van de Hamilton-depressieschaal in de vorm van het semi-gestructureerde interview, waarbij itemdefinities, openingsvragen, ankerpunten en de wijze van beoordelen expliciet beschreven worden, vergroot de betrouwbaarheid op itemniveau, onafhankelijk van de ernst van de depressie. Bovendien blijkt dat ook getrainde onderzoekers zonder klinische ervaring het interview betrouwbaar kunnen toepassen, waardoor de bruikbaarheid van de schaal bij wetenschappelijk onderzoek vergroot wordt.

Literatuur

Bech, P., M. Kastrup en O. Rafaelsen (1986), Mini-compendium of rating scales for states of anxiety, depression, mania, schizophrenia with corresponding DSM-III syndromes. *Acta Psychiatrica Scandinavica*

, 73 (supplement 326).

Bech, P., M. Kastrup en O. Rafaelsen (1989), *Mini-compendium van beoordelingsschalen in de psychiatrie*. In het Nederlands vertaald door H. D'Haenen en W. Verhoeven. VUB-press.

Brink, W. van den (1989), *Meting van DSM-III persoonlijkheidspathologie. Betrouwbaarheid en validiteit van de SIDP-P en as II van de DSM-III*. Proefschrift Rijksuniversiteit Groningen.

Cichetti, D.V., en B. Prusoff (1983), Reliability of depression and associated clinical symptoms. *Archives of General Psychiatry*, 40, 987-990.

Dijkstra, P. (1974), Classificatie en statistiek. In: *Stemming en ontstemming*. De Erven Bohn, Amsterdam.

Guy, W. (1976), *ECDEU Assessment Manual for Psychopharmacology*. National Institute of Mental Health, Rockville.

Hamilton, M. (1960), A rating scale for depression. *Journal of Neurology, Neurosurgery, Psychiatry*, 23, 56-62.

Hamilton, M. (1967), Development of a rating scale for primary depressive illness. *British Journal of Social Clinical Psychology*, 6, 278-296.

Hoencamp, E., I. Betten en P. Haffmans (1992), De tussenbeoordelaarsbetrouwbaarheid van de Hamilton Rating Scale for Depression. *Acta Neuropsychiatrica*, 4, 86-89.

Hooijer, C., F. Zitman, E. Griez e.a. (1991), The Hamilton Depression Rating Scale (HDRS): changes in scores as a function of training and version used. *Journal of Affective Disorders*, 22, 21-29.

Jonghe, F. de (1994), *Leidraad voor het scoren van de Hamilton Depression Rating Scale*. Benecke, Amsterdam.

Rehm, L.P., en W. O'Hara (1985), Item characteristics of the Hamilton Rating Scale for Depression. *Journal of Psychiatric Research*, 19, 31-41.

Slooff, C.J., W. Mulder-Hajonides van de Meulen en R. van den Hoofdakker (1985), De Nederlandse vertaling van de Present State Examination negende editie I: betrouwbaarheidsaspecten. *Tijdschrift voor Psychiatrie*, 25, 151-163.

Williams, J.B.W. (1988), A structured interview guide for the Hamilton Depression Rating Scale. *Archives of General Psychiatry*, 45, 742-747.

Zitman, F.G., E. Griez en C. Hooijer (1989), Standaardisering depressievragenlijsten. *Tijdschrift voor Psychiatrie*, 31, 114-123.

Summary: Reliability of a semi-structured interview for the Hamilton Depression Rating Scale

We investigated the interrater reliability of a Dutch semi-structured interview for the Hamilton Depression Rating Scale. An interviewer-observer design was used, and the results were compared with a recent publication. The results suggest that the use of a semi-structured interview facilitates the reliable application of the Hamilton scale, especially by raters without clinical experience.

De auteurs zijn respectievelijk psychiater (ten tijde van dit onderzoek werkzaam bij het Academisch Medisch Centrum te Amsterdam), hoogleraar psychiatrie (Universiteit van Amsterdam en Psychiatrisch Ziekenhuis Amsterdam Centrum/Oud-West/Noord), psycholoog en psychiater (Academisch Medisch Centrum). Correspondentie-adres: R.W. Kupka, H.C. Rümke Groep, Willem Arntsz Huis, Postbus 61, 3500 AB Utrecht.

Het artikel werd voor publicatie geaccepteerd op 20-6-1996.