

HAMILTON DEPRESSION RATING SCALE (HDRS OF HAMD)

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

Meetinstrument	Hamilton Depression Rating Scale
Afkorting	HDRS of HAMD
Auteur	Hamilton (1967)
Thema	Depressie
Doel	De ernst van depressie en de veranderingen in depressie beoordelen
Populatie	Patiënten met een diagnose van depressie
Gebruikers	Een getraind clinicus
Aantal items	17
Aanwezigheid van de patiënt vereist	Ja
Vindplaats van het meetinstrument	Hamilton, M. (1967). Development of a rating scale for primary depressive illness. <i>British Journal of Social and Clinical Psychology</i> , 6, 278-296. Engelstalige versie: http://www.strokecenter.org/trials/scales/hamilton.html Franstalige versie: http://www.paris-nord-sftg.com/outils.hamilton.0312.php3 http://www.echelles-psychiatrie.com/echelle-hdrs.php of

DOEL

Het doel van de Hamilton Depression Rating Scale is de evaluatie van de ernst van depressie, nadat deze gediagnosticeerd werd.

DOELGROEP

De doelgroep van de HDRS bestaat uit patiënten waarbij de diagnose van depressie gesteld werd.

BESCHRIJVING

De Hamilton Depression Rating Scale is een meetinstrument dat de nadruk legt op de lichamelijke en gedragsymptomen. Ze bestaat uit 17 items die de symptomen opgetreden in de voorbije week in rekening brengen. Bepaalde items worden gedefinieerd als een reeks waarbij de intensiteit verhoogt

en andere items worden gedefinieerd als een aantal omstandigheden met gelijke waarde. De 17 items zijn depressieve gemoedstoestand, schuldgevoel, zelfdoding, slapeloosheid bij het inslapen, slapeloosheid tijdens de nacht, slapeloosheid 's morgens, werk en activiteiten, vertraging (vertraging van denken en spreken, daling van het concentratievermogen, daling van de beweegkracht), onrust, psychische angst, lichamelijke angst, gastro-intestinale symptomen, hypochondrie, gewichtsverlies en aandacht. De Hamilton Depression Rating Scale bevat naast de 17 items nog vier bijkomende variabelen: variatie tijdens de dag, vervreemding, paranoïde symptomen en obsessionele symptomen. Deze vier items worden niet in de meetschaal opgenomen aangezien de variatie per dag geen depressie of ernst van depressie meet maar wel de soort depressie aangeeft. De andere drie items zijn zo uitzonderlijk dat het niet nuttig is om ze op te nemen. Om deze schaal te gebruiken is training van de interviewers nodig.

De evaluatie van de ernst van de depressie door de Hamilton Depression Rating Scale gebeurt via een semi-gestructureerd interview. De HDRS maakt gebruik van Lickertschalen van drie of vijf punten. De schaal met drie punten wordt enkel gebruikt als de kwantificering van de variabelen moeilijk of onmogelijk is, wat het geval is voor acht items. Er wordt geen onderscheid gemaakt tussen de intensiteit en de frequentie van de symptomen, het is de beoordelaar die belang hecht aan elk element. Voor de Lickerschalen van 3 en 5 punten, staat score 0 voor niet significante of afwezige symptomen. Voor de schaal van 5 punten staat score 1 voor « licht of triviaal », scores 2 en 3 staan voor « matig » en score 4 betekent « ernstig ». Bij de schaal met 3 punten staat 1 voor « licht of twijfelachtig » en score 2 betekent « duidelijk aanwezig ». De optelling van de scores geeft de totaalscore. Voor item 16, « gewichtsverlies », kiest de beoordelaar de quotatie waar hij rekening mee wil houden: volgens wat de patiënt zegt of beoordeeld volgens gewicht. Meestal wordt het verslag van de patiënt gevolgd, maar soms is het belangrijk om volgens het gewicht te beoordelen. De scores gaan van 0 (minimum) tot 52 (maximum). Een score tussen 0 en 7 betekent de afwezigheid van depressie, een score tussen 8 en 17 staat voor een lichte depressie, een score tussen 18 en 25 wijst op een matige depressie en een score tussen 26 en 52 duidt een ernstige depressie aan. De auteur (Max Hamilton, 1960) raadt aan om de evaluatie uit te voeren met twee onafhankelijke interviewers die dezelfde vragenlijst afnemen. De score van de patiënt wordt verkregen door het gemiddelde te nemen van de scores van de twee interviewers.

Er zijn verschillende problemen opgedoken in verband met de specifieke symptomen. Bij bepaalde items is het moeilijk om een onderscheid te maken tussen de verschillende beschrijvingen en beoordelingen (depressieve gemoedstoestand, zelfdoding, werk en verlies van interesse, vertraging, onrust, gastro-intestinale symptomen, algemene lichamelijke symptomen, hypochondrie, aandacht, gewichtsverlies). Bijvoorbeeld, voor het item « zelfmoord » behaalt een zelfmoordpoging een score

4, maar alse en dergelijke poging plots plaatsvindt, in een omgeving waar dit niet verwacht werd, krijgt dit item score 3. Het is niet evident om het onderscheid te maken tussen een poging die bedoeld is om te doden en een poging die de nood om hulp uitdrukt. De beoordelaar zal de situatie moeten inschatten.

BETROUWBAARHEID

De equivalentie (*equivalence*) wordt uitgedrukt door waarden hoger dan 0.80, ongeacht de studie. Hamilton (1960) ontwikkelt waarden variërend in functie van het aantal patiënten: een stijging van het aantal patiënten zorgt voor een stijging van de interbeoordelaarsbetrouwbaarheid (*interrater reliability*). Bagby *et al.* (2004) berichten over waarden voor de interrater betrouwbaarheid die variëren tussen 0.82 en 0.98. De interrater betrouwbaarheid behaalt een score van 0.92 wanneer de schaal gebruikt wordt via een gestructureerd interview. Bowling (1997) wijst op een score van 0.96 en bevestigt dat verschillende studies deze score onderschrijven. Hij vermeldt scores tussen 0.84 en 0.98, de resultaten van onderzoek van Hamilton (1976), Knesevitch *et al.* (1977) et Rehm (1981). McDowell (2006) rapporteert dat Montgomery en Asberg (1979) een correlatie van 0.89 gevonden hebben. Hedlung en Vieweg (1979) vinden bij een bepaalde studie correlatiewaarden lager dan 0.84, hoewel de meeste correlatiecoëfficiënten hoger zijn dan 0.88 en de hoogste waarden 0.96 en 0.98 omvatten. Morriss *et al.* (2004) vinden op hun beurt een correlatie van 0.95. Deze waarden liggen allen een stuk hoger dan 0.80. Ze drukken dus een goede betrouwbaarheid tussen verschillende beoordelaars uit.

De interne consistentie (*internal consistency*) bereikt waarden tussen 0.46 en 0.97 (Bagby *et al.*, 2004) waarbij er tien onderzoeken een interne consistentie van meer dan 0.70 rapporteren. McDowell (2006) brengt verslag uit van de resultaten van verschillende onderzoeken (Bech *et al.*, 1990; Rehm *et al.*, 1985; Hedlung *et al.*, 1979; Diefenbach *et al.*, 2001, Caroll *et al.*, 1981) waarbij de waarden variëren tussen 0.48 en 0.95. Morriss *et al.* (2004) op hun beurt berichten over waarden rond 0.60 of hoger, met uitzondering van de variabelen « hypochondrie » en aandacht ». De items met een waarde hoger dan 0.60 zijn goed, maar de items die geen 0.60 halen, hebben een zwakke interne consistentie. De algemene interne consistentie wordt bijgevolg verzwakt.

Bagby *et al.* (2004) rapporteren een stabiliteit (*stability*) die varieert tussen 0.81 en 0.98 voor de totale schaal. De items behalen een score tussen 0.00 en 0.85. De items verhogen niettemin gemiddeld de stabiliteit tot 0.54. Bowling (1997) vindt in zijn onderzoek een test-retest correlatie van 0.65 voor de totaalscore en van 0.04 tot 0.77 voor de afzonderlijke items. McDowell (2006) vindt op zijn beurt een stabiliteitswaarde van 0.72 voor de totale schaal. De waarden voor de volledige schaal

zijn zeer goed, maar wanneer de analyse per variabele gebeurt, zijn sommige waarden zeer zwak. De stabiliteit is goed op voorwaarde dat alle variabelen samen in rekening gebracht worden.

VALIDITEIT

Bagby *et al.* (2004) bekijken verschillende studies waarin de constructvaliditeit (*construct validity*) van de Hamilton Depression Rating Scale geanalyseerd wordt. Deze is meestal matig tot goed, hoewel negen op 71 studies een correlatie hebben die lager is dan 0.50 en terwijl twee studies een inverse correlatie vaststellen (-0.86, -0.47, - 0.65).

De concurrente validiteit (*criterion-related validity*) blijkt goed te zijn. De Hamilton Depression Rating Scale werd met een groot aantal instrumenten vergeleken. De correlatie met de Beck Depression Inventory bedraagt 0.70 (Hamilton, 1976). Schwab *et al.* (1967) stelden een correlatie vast tussen de Hamilton Depression Rating Scale en de Beck Depression Inventory met een coëfficiënt van 0.75. Tussen de Montgomery-Asberg Depression Rating Scale en de Hamilton Depression Rating Scale werd een concurrente validiteit vastgesteld met correlatiecoëfficiënt $r=0.71$ (Maier *et al.*, 1988). Voor andere schalen bedroeg de correlatie met de HDRS tussen 0.63 en 0.90. Minder goede waarden worden gevonden voor de correlatie tussen de Hamilton Depression Rating Scale en de Zung's Self-rating Depression Scale: van 0.22 tot 0.95 (Hedlung *et al.*, 1979); dit geldt ook voor de correlatie met de Minnesota Multiphasic Personality Inventory depression scale: van 0.27 tot 0.34 (Hedlung *et al.*, 1979).

De sensitiviteit (*sensitivity*) bedraagt 0.76 volgens Bagby *et al.* (2004) en 0.88 volgens Bowling (1997). De specificiteit (*specificity*) bedraagt 0.91. De positieve predictieve waarde (*predictive positive value*) bedraagt 0.77 (Bagby *et al.*, 2004) en 0.80 (Bowling, 1997). De negatieve predictieve waarde (*negative predictive value*) bedraagt 0.92. Al deze waarden wijzen op een zeer goede validiteit.

GEBRUIKSVRIENDELIJKHEID

Het duurt 20 tot 30 minuten om het meetinstrument af te nemen.

VARIANTEN

De Hamilton Depression Rating Scale heeft een heel aantal varianten. De varianten bestaan uit uitbreiding van de schaal, vermindering van het aantal items, aanpassing in een schaal voor zelfafname, geïnformatiseerde versies en/of toevoeging van een handleiding voor een gestructureerd interview. Deze varianten zijn HDRS 29, HDRS 8, HDRS 6, HDRS 24, HDRS 7, HDRS 31 en de Hamilton Depression Inventory (HDI), Interactive Voice Response version (IVR), Structured Interview Guide for the Hamilton Depression Scale (SIGH-D), Seasonal Affective Disorder version

(SIGH-SAD), de Structured Interview Version (HDS-SIV) en GRID-Hamilton Depression Rating Scale (GRID-HAMD). Al deze varianten bevatten een groot deel van de items uit de oorspronkelijke HDRS met 17 items. Naargelang de versie werden er elementen uit de originele Hamilton Depression Rating Scale bijgevoegd of weggelaten.

OPMERKINGEN

De Hamilton Depression Rating Scale is een meetinstrument dat ontworpen werd om de ernst van een depressie te meten, maar de schaal is niet geschikt om de veranderingen ten gevolge van therapie-effecten op te volgen.

Een bepaald onderzoek (Zimmerman *et al.*, 2005) belicht de verschillende voordelen van de HDRS, zoals de continuïteit in de evaluatie, wat toelaat om vergelijkingen te maken, en de capaciteit om een onderscheid te maken tussen placebomedicatie in klinische proeven en om veranderingen vast te stellen. Desalniettemin menen verschillende onderzoekers dat een revisie en een standaardisatie van de schaal opportuun is. Een studie van Bagby *et al.* (2004) besluit met het voorstel om de Hamilton Depression Rating Scale te vervangen door de Inventory of Depressive Symptomatology of de Montgomery-Asberg Depression Rating Scale, aangezien deze schalen gebaseerd zijn op de hedendaagse kennis over depressie en rekening houden met vooruitgang in de psychometrie en statistiek. Bagby *et al.* (2004) stellen dat de Hamilton Depression Rating Scale tekortschiet op het conceptueel niveau en op het niveau van de psychometrie in de psychiatrische context.

Als de schaal gebruikt wordt in de eerstelijns hulp, wordt een betere interrater betrouwbaarheid vastgesteld dan wanneer de HDRS vergezeld wordt van een gestructureerd interview.

RÉFÉRENTIES

Hamilton, M. (1960). A rating scale for depression. *J. Neurol. Neurosurg. Psychiat.*, 23, 56-62.

Bagby, R.M., Ryder, A.G., Schuller, D.R., & Marshall, M.B. (2004). The Hamilton Depression Rating Scale: Has the Gold Standard Become a Lead Weight? *AM J Psychiatry*, 161, 2163-2177.

Bouvard, M., & Cottraux, J. (2005). Protocoles et échelles d'évaluation en psychiatrie et en psychologie. *Collection Pratiques en psychothérapie, Edition Masson, 4^e édition*, 171-173, 192-195.

Bowling, A. (1997). A review of quality of life measurement scales. *Measuring Health, Open University Press, second edition*, 75-76.

McDowell, I. (2006). A Guide to Rating Scales and Questionnaires. *Measuring Health, Oxford University Press, third edition, 369-378.*

Morriss, R., Leese, M., Chatwin, J., Baldwin, D., & the TREAD Study Group (2004). Inter-rater reliability of the Hamilton Depression Rating Scale as a diagnostic and outcome measure of depression in primary care. *Journal of Affective Disorders, 111, 204-213.*

Steer, R.A., Beck, A.T., Riskind, J.H., & Brown, G. (1987). Relationships Between the Beck Depression Inventory and the Hamilton Psychiatric Rating Scale for Depression in Depressed Outpatients. *Journal of Psychopathology and Behavioral Assessment, 9(3), 327-339.*

VINDPLAATS VAN HET MEETINSTRUMENT

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

Engelstalige versie: <http://www.strokecenter.org/trials/scales/hamilton.html>

Franstalige versie: <http://www.paris-nord-sftg.com/outils.hamilton.0312.php3> of
<http://www.echelles-psychiatrie.com/echelle-hdrs.php>

Hamilton Depression Rating Scale (HDRS of HAMD)

Hamilton (1960)

Author (year)	Setting	Sample (n)	Design	Reliability	Validity
1. Hamilton, M. (1960)			Prospective study	E	
2. Bagby, R.M., Ryder, A.G., Schuller, D.R., & Marshall, M.B. (2004)			Systematic review	IC E S	CtV CsV Sen NPV Sp PPV
3. Bowling, A. (1997)			Book – description of the scale and of the studies over the scale	E IC S	CrV Sen PPV
4. McDowell, I. (2006)			Book – description of the scale and of the studies over the scale	IC E S	CrV
5. Morriss, R., Leese, M., Chatwin, J., Baldwin, D., & the TREAD Study Group (2004)	Primary care setting	n=42 patients		E IC	

Betrouwbaarheid/ fiabilité: Stability (S), Internal Consistency (IC), Equivalence (E)

Validiteit/ validité: Face Validity (FV), Content Validity (CtV), Criterion Validity (CrV), Construct Validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR), Odds Ratio (OR), Area Under the Curve (AUC)

Results reliability	Results validity	Commentary
<p>1. E interrater reliability For two raters, the correlation between summed scores for the first 10 patients was 0.84. Adding successively 10 patients at the time, the correlation changed to 0.84, 0.88, 0.89, 0.89, 0.90, 0.90. (last N=70)</p>		
<p>2. IC Estimates ranged from 0.46 to 0.97, and 10 studies reported estimates ≥ 0.70. E interrater reliability Pearson's r ranged from 0.82 to 0.98 and the intraclass r ranged from 0.46 to 0.99. E interrater reliability Using a structured interview version of the Hamilton depression scale, they found an overall intraclass coefficient of 0.92. S test-retest Retest reliability for the Hamilton depression scale ranged from 0.81 to 0.98. Retest reliability at the item level ranged from 0.00 to 0.85. With the Structured Interview Guide for the Hamilton Depression Rating Scale, this increased the mean retest reliability across individual items to 0.54.</p>	<p>CtV Because of its wide use and long clinical tradition, the Hamilton depression scale seems to both define as well as measure depression. CsV Studies reporting estimates of convergent validity of the 17-item Hamilton Depression Rating Scale, compared with other depression measures. HDRS – depression measures < 0.50: 9 studies HDRS – depression measures between 0.50 and 0.70: 12 studies HDRS – depression measures > 0.70: 14 studies HDRS – Global Assessment Scale: - 0.86 and – 0.47 HDRS – Visual Analogue Scale: - 0.65 Sen Mean: 0.76 with a range from 0.45 to 0.88 Sp Mean: 0.91 with a range from 0.75 to 1.00 PPV Mean: 0.77 with a range from 0.37 to 1.00 NPV Mean: 0.92 with a range from 0.86 to 0.99</p>	

Results reliability	Results validity	Commentary
<p>3. E interrater reliability Inter-rater reliability with two psychiatrists rating 20 subjects was good (Pearson's $r=0.96$), a finding confirmed by other studies (Korner et al., 1990)</p> <p>E interrater reliability Its inter-rater reliability is reported to be good: correlations are high ranging from 0.84 to 0.98 (Hamilton, 1976; Knesevitch et al., 1977; Rehm, 1981)</p> <p>IC The α correlations for internal consistency were 0.82 to 0.83</p> <p>S test-retest The test-retest correlations were high at 0.65 for the total score, although the item correlations were variable at 0.04 to 0.77 (15-day retest)</p>	<p>CrV The scale is reported to have high concurrent validity with good agreement with other scales, particularly the Beck, with correlations reported of over 0.70 (Hamilton, 1976). Schwab et al. (1967) compared the Hamilton with the Beck scale on 153 medical in-patients. The correlation between these two scales was 0.75.</p> <p>Sen - PPV Hamer et al. (1991) reported that a threshold score of height gave a sensitivity of 88 per cent and a positive predictive value of 80 per cent in comparison with diagnoses made with the DSM-III.</p>	

Results reliability	Results validity	Commentary
<p>4. IC A internal consistency was reported from a World Health Organization study in five countries and showed figures of $\alpha=0.48$ at baseline, rising to 0.85 after 11 days of treatment (Bech et al., 1990). Other estimates include 0.76 (Rehm et al., 1985), 0.83, 0.94 and 0.95 (Hedlung et al., 1979), and 0.84 (Diefenbach et al., 2001). Carroll et al. (1981) obtained a median item-total correlation of 0.54 (range 0.19 to 0.78).</p> <p>E interrater reliability Hamilton originally reported a correlation of 0.90 between pairs of ratings for 70 patients (Hamilton, 1960). Montgomery and Asberg (1979) found an inter-rater correlation of 0.89 before treatment. Hedlung and Vieweg (1979) reviewed ten reliability studies: in only one small study did inter-rater correlation fall below 0.84; most coefficients fell above 0.88. Even higher values include 0.96 and 0.98 in two samples of long-term care patients (Foster, J.R. et al., 1988). Intraclass coefficients include 0.70 to 0.72 (Maier, W. et al., 1988), 0.83 (Korner, A. et al., 1990), 0.85 (Winokur, A. et al., 1982) and 0.92 (Endicott, J. et al., 1981).</p> <p>S test-retest Intraclass rest-retest reliability at 3 weeks was 0.72 for the 17-item version, and 0.69 for an abbreviated five-item version (Maier, W., in: Bech, P. et al., 1990)</p>	<p>CrV The Hamilton scale has frequently been tested against clinical ratings of severity. The total score of the 17-item version was significantly related to globally assessed severity of depression on the Raskin three-item rating scale for depression ($r=0.65$); correlations for the Montgomery-Asberg Depression Rating Scale (0.71) and BRMS (0.70) were slightly higher (Maier, W. et al., 1988). Other figures include a correlation of 0.81 between the HDRS and the Raskin scale, and -0.86 between the HRDS and overall severity rated by the Global Assessment Scale (GAS) (Rehm, L.P. et al., 1985). Hedlung and Vieweg's review cited correlations of 0.84, 0.89 and 0.90 with clinical ratings of severity; they also listed a lower correlation of 0.67 for a study that included depressed and nondepressed patients (Hedlung, J.L., Vieweg, B.W., 1979). Other low correlations with a clinical severity rating include 0.68 (Maier, W., in: Bech, P. et al., 1990). In nine comparable studies, correlations between the HRDS and Zung's Self-rating Depression Scale ranged from 0.22 to 0.95 (Hedlung, J.L. et al., 1979). Correlations with a range of other depression scales covered in Hedlung and Vieweg's review ranged from 0.63 to 0.87; correlations with the Minnesota Multiphasic Personality Inventory depression scale, however, were lower, at 0.27 and 0.34 (Hedlung, J.L. et al., 1979).</p> <p>CrV Change scores on the HRDS were compared with a clinician's classification of 35 patients under treatment into improved or not improved; the point biserial correlation was 0.59, comparing unfavourably with a figure of 0.70 for the MADRS (Montgomery, S.A. et al., 1979).</p>	

Results reliability	Results validity	Commentary
<p>5. E interrater reliability Intraclass correlation for the total scores in session 1: 0.968, in session 2: 0.937, in session 3: 0.896, in session 4: 0.938. Intraclass correlation for overall: 0.947</p> <p>IC Weighted kappa coefficients for individual items were close to 0.6 or above for all items except hypochondriasis and insight.</p>		

Betrouwbaarheid/ fiabilité: Stability (S), Internal Consistency (IC), Equivalence (E)

Validiteit/ validité: Face Validity (FV), Content Validity (CtV), Criterion Validity (CrV), Construct Validity (CsV)

Sensitivity (Sen), Specificity (Sp), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Receiver Operating Curve (ROC), Likelihood Ratio (LR),

Odds Ratio (OR), Area Under the Curve (AUC)

Gelieve bij gebruik van dit rapport als volgt te refereren :

Bulteel L., Gobert M., Piron C., de Bonhome, A., De Waegeneer, E., Vanderwee K., Verhaeghe S., Van Hecke, A., Defloor T. (2010) Actualiseren van de bestaande BeST-databank & aanvullen van de bestaande BeST-databank met nieuwe schalen. Brussel: Federale Overheidsdienst Volkgezondheid van de voedselketen en leefmilieu.

Comment citer ce rapport ?

Bulteel L., Gobert M., Piron C., de Bonhome, A., De Waegeneer, E., Vanderwee K., Verhaeghe S., Van Hecke, A., Defloor T. (2010) Actualisation de la base de données BeST & ajout de nouvelles échelles dans la base de données BeST. Bruxelles: Service Public Fédéral Santé Publique, Sécurité de la Chaîne alimentaire et Environnement.