

➤ The Hamilton Depression Rating Scale (HDRS ou HAMD)

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

Instrument de mesure	Hamilton Depression Rating Scale
Abréviation	HDRS ou HAMD
Auteur	Hamilton M. (1960)
Thème	Dépression
Objectif	Evaluer la sévérité et les changements dans la dépression
Population	Patients avec un diagnostic de troubles affectifs de type dépressif
Utilisateurs	Clinicien entraîné
Nombre d'items	17
Participation du patient	Oui
Localisation de l'instrument de mesure	Echelle en anglais : Hamilton, M., (1967). Development of a rating scale for primary depressive illness. <i>Br J Soc Clin Psychol</i> , 6, 278-296. Echelle en anglais : http://www.strokecenter.org/trials/scales/hamilton.html Echelle en français : http://www.paris-nord-sftg.com/outils.hamilton.0312.php3 ou http://www.echelles-psychiatrie.com/echelle-hdrs.php

OBJECTIF

L'objectif de la Hamilton Depression Rating Scale est l'évaluation de la sévérité de la dépression, une fois que celle-ci a été diagnostiquée.

PUBLIC CIBLE

Le public visé par la HDRS est la population des patients pour lesquels un diagnostic de troubles affectifs de type dépressif a été posé.

DESCRIPTION

L'échelle Hamilton Depression Rating Scale est une échelle qui met l'accent sur les symptômes somatiques et comportementaux. Elle se compose de 17 items prenant en compte des symptômes survenus au cours de la semaine écoulée. Certains items sont définis comme des séries d'augmentation d'intensité, et d'autres sont définis par un nombre de conditions de valeur égale. Les 17 items sont l'humeur dépressive, le sentiment de culpabilité, le suicide, l'insomnie au début de la nuit, l'insomnie au milieu de la nuit, l'insomnie du matin, le travail et les activités, le retard (ralentissement de la pensée et du langage, baisse de la faculté de concentration, baisse de l'activité motrice), l'agitation, l'anxiété psychique, l'anxiété somatique, les symptômes gastro-intestinaux, les symptômes somatiques généraux, les symptômes génitaux, l'hypocondrie, la perte de poids et la perspicacité. La Hamilton Depression Rating Scale comprend, en plus des 17 items, quatre variables supplémentaires : variation diurne, déréalisation, symptômes paranoïaques et symptômes obsessionnels. Ces quatre items sont exclus de l'échelle car la variation diurne ne mesure pas la dépression ou sa sévérité mais indique la sorte de dépression. Tandis que les trois autres items sont tellement peu fréquents qu'il n'est pas utile de les inclure. Pour utiliser cette échelle, l'entraînement des interviewers est nécessaire.

L'évaluation de la sévérité de la dépression par la Hamilton Depression Rating Scale se fait via une interview semi-structurée. La HDRS se cote avec une échelle de Likert à 3 points ou à 5 points. L'échelle à 3 points n'est utilisée que lorsque la quantification des variables est difficile, voire impossible, ce qui est le cas pour 8 items. Aucune distinction n'est faite entre l'intensité et la fréquence des symptômes, c'est l'évaluateur qui donne de l'importance à chaque élément. Pour les échelles de Likert à 3 et 5 points, le score 0 signifie des symptômes non significatifs ou absents. Pour l'échelle à 5 points, un score de 1 signifie « léger ou trivial », les scores 2 et 3 signifient « moyen » et le score 4 signifie « sévère ». Pour l'échelle à 3 points, un score de 1 signifie « léger ou douteux » et un score de 2 signifie « clairement présent ». L'addition des scores donne le score total. Pour l'item 16, « perte de poids », l'examineur choisit la cotation dont il veut tenir compte : selon les dires du patient ou apprécié par pesée. En général, la cotation retenue est celle selon les dires du patient mais parfois, il est important d'évaluer par la pesée. Les scores varient de 0 (minimum) à 52 (maximum). Un score entre 0 et 7 signifie une absence de dépression, un score entre 8 et 17 représente une légère dépression, un score de 18 à 25 indique une dépression modérée, tandis qu'un score situé entre 26 et 52 dévoile une dépression sévère. L'auteur (Max Hamilton, 1960)

conseille de réaliser l'évaluation via deux examinateurs évaluant indépendamment lors de la même interview. Le score du patient est obtenu par la moyenne des scores des deux examinateurs.

Plusieurs problèmes ont émergé en lien avec les symptômes spécifiques. Certains items sont tellement proches en description et en jugement qu'il est difficile de faire des distinctions (humeur dépressive, suicide, travail et perte d'intérêt, retard, agitation, symptômes gastro-intestinaux, symptômes somatiques généraux, hypocondrie, perspicacité, perte de poids). Par exemple, pour le suicide, une tentative de suicide obtient un score de 4, mais une telle tentative peut parfois se produire soudainement dans un contexte de très faible tendance suicidaire. Dans ce cas, le suicide a un score de 3. Il n'est pas évident de distinguer une vraie tentative de suicide d'un appel à l'aide, l'examineur sera juge de la situation.

FIABILITÉ

L'équivalence (*equivalence*) s'exprime par des valeurs supérieures à 0.80, quelles que soient les études. Hamilton (1960) développe des valeurs variant en fonction du nombre de patients : l'augmentation du nombre de patients induit une augmentation de la valeur de la fidélité inter-juges (*interrater reliability*). Bagby *et al.* (2004) cite des valeurs pour la fidélité inter-juges variant entre 0.82 et 0.98. La fidélité inter-juges atteint un score de 0.92 lorsqu'une version de la Hamilton Depression Rating Scale est utilisée via une interview structurée. Bowling (1997) indique un score de 0.96 et affirme que plusieurs études soutiennent ce score. Il cite un score entre 0.84 et 0.98, résultat d'études de Hamilton (1976), Knesevitch *et al.* (1977) et Rehm (1981). McDowell (2006) rapporte que Montgomery et Asberg (1979) ont trouvé une corrélation de 0.89, tandis que Hedlung et Vieweg (1979) évoquent qu'une seule étude obtient un score de corrélation inférieur à 0.84 alors que la plupart des coefficients de corrélation sont supérieurs à 0.88 et que les valeurs les plus élevées incluent 0.96 et 0.98. Morriss *et al.* (2004), quant à eux, trouvent une corrélation de 0.95. Toutes ces valeurs sont très élevées puisqu'elles sont supérieures à 0.80. Elles expriment donc une bonne fiabilité entre les examinateurs.

La consistance interne (*internal consistency*) obtient des valeurs qui varient de 0.46 à 0.97 (Bagby *et al.*, 2004) dont dix études avec une consistance interne supérieure à 0.70. McDowell (2006) rapporte des valeurs de diverses études (Bech *et al.*, 1990 ; Rehm *et al.*, 1985 ; Hedlung *et al.*, 1979 ; Diefenbach *et al.*, 2001, Carroll *et al.*, 1981) qui varient aussi entre 0.48 et 0.95. Morriss *et al.* (2004), quant à eux, expriment des valeurs proches ou supérieures à 0.6, à l'exception des variables

« hypochondrie » et « perspicacité ». Les items supérieurs à 0.60 sont bons, mais les items avec un score inférieur à 0.60 expriment une faible consistance interne. La consistance interne globale est donc mitigée.

Bagby *et al.* (2004) rapportent une stabilité (*stability*) variant entre 0.81 et 0.98 pour la totalité de l'échelle. En ce qui concerne les items, ceux-ci obtiennent un score entre 0.00 et 0.85. La moyenne des items augmente néanmoins la stabilité jusqu'à 0.54. Bowling (1997), lui, rapporte une corrélation test-retest de 0.65 pour le score total et variant entre 0.04 et 0.77 pour les items. De son côté, McDowell (2006) exprime une stabilité de 0.72 pour l'échelle totale. Les valeurs du total de l'échelle sont très bonnes mais lorsqu'elles sont analysées variable par variable, certaines valeurs sont très faibles. Nous retiendrons que la stabilité est bonne lorsque toutes les variables sont prises en compte ensemble.

VALIDITÉ

Bagby *et al.* (2004) reprend diverses études dans lesquelles la validité du construit (*construct validity*) de la Hamilton Depression Rating Scale est analysée. Celle-ci est généralement moyenne à bonne, même si neuf études sur trente-sept ont une corrélation inférieure à 0.50 et que deux études expriment une corrélation inverse (-0.86, -0.47, - 0.65).

La validité concurrente (*criterion-related validity*) apparaît être bonne. La Hamilton Depression Rating Scale a été comparée à de nombreuses échelles. Avec le Beck Depression Inventory, la corrélation est équivalente à 0.70 (Hamilton, 1976). Pour Schwab *et al.* (1967), la corrélation entre la Hamilton Depression Rating Scale et le Beck Depression Inventory est de 0.75. Entre la Montgomery-Asberg Depression Rating Scale et la Hamilton Depression Rating Scale, la validité concurrente est évaluée à 0.71 (Maier *et al.*, 1988). Pour les autres échelles, la corrélation avec la HDRS varie entre 0.63 et 0.90. Des valeurs moins bonnes sont trouvées lors de la corrélation entre la Hamilton Depression Rating Scale et la Zung's Self-rating Depression Scale : 0.22 à 0.95 (Hedlung *et al.*, 1979) ; ainsi que lors de la corrélation avec la Minnesota Multiphasic Personality Inventory depression scale : 0.27 et 0.34 (Hedlung *et al.*, 1979).

La sensibilité (*sensitivity*) est exprimée par des valeurs de 0.76 (Bagby *et al.*, 2004) et de 0.88 (Bowling, 1997). La spécificité (*specificity*) vaut 0.91. La valeur prédictive positive (*predictive positive value*) obtient une cote de 0.77 (Bagby *et al.*, 2004) et de 0.80 (Bowling, 1997). La valeur prédictive négative (*negative predictive value*), elle, équivaut à 0.92. Toutes ces valeurs sont très élevées et expriment une très bonne validité.

CONVIVALITÉ

L'échelle se complète en une durée de 20 à 30 minutes.

VARIANTE

La Hamilton Depression Rating Scale a un nombre non négligeable de variantes. Ces variantes proviennent de l'élargissement de l'échelle, d'une réduction du nombre d'items de l'échelle, de transformations en échelle d'auto-évaluation, en version informatisée et/ou de l'ajout de guides d'interviews structurés. Via ces différents procédés, la Hamilton Depression Rating Scale (HDRS) s'est notamment transformée en HDRS 29, HDRS 8, HDRS 6, HDRS 24, HDRS 7, HDRS 31 ou encore 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), la Structured Interview Version (HDS-SIV) et GRID-Hamilton Depression Rating Scale (GRID-HAMD). Toutes ces variantes incluent une majorité d'items provenant de la HDRS originale à 17 items. Selon la version, des éléments de la Hamilton Depression Rating Scale originale ont été ajoutés ou supprimés.

REMARQUES

La Hamilton Depression Rating Scale est une échelle qui a été créée pour mesurer la sévérité de la dépression mais elle n'est pas adéquate pour le suivi des changements dus aux effets du traitement.

Une étude (Zimmerman *et al.*, 2005) a mis en lumière plusieurs avantages de la HDRS, tels que la continuité dans l'évaluation qui permet des comparaisons entre des décennies, et la capacité à discriminer des médicaments de placebos dans les essais cliniques et à mesurer le changement. Néanmoins, plusieurs pensent qu'une révision et une standardisation de l'échelle est opportune. Une étude de Bagby *et al.* (2004) se conclut par la proposition de remplacer la Hamilton Depression Rating Scale par la Inventory of Depressive Symptomatology ou la Montgomery-Asberg Depression Rating Scale. Bagby *et al.* (2004) proposent une des deux échelles comme remplaçante de la Hamilton car ces dernières sont basées sur la connaissance actuelle de la dépression et tiennent compte des progrès psychométriques et statistiques. Bagby *et al.* (2004) définissent la Hamilton

Depression Rating Scale comme imparfaite au niveau conceptuel et au niveau psychométrique dans le contexte psychiatrique.

Si l'échelle est utilisée dans les soins de santé primaires, l'évaluation par la HDRS produit une meilleure fidélité inter-juges lorsque la HDRS est accompagnée d'une interview standardisée.

RÉFÉRENCES

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.

LOCALISATION DE L'INSTRUMENT DE MESURE

Echelle en anglais : Hamilton, M., (1967). Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol*, 6, 278-296.

Echelle en anglais : <http://www.strokecenter.org/trials/scales/hamilton.html>

Echelle en français : <http://www.paris-nord-sftg.com/outils.hamilton.0312.php3> ou <http://www.echelles-psychiatrie.com/echelle-hdrs.php>

Hamilton Depression Rating Scale

Hamilton, M. (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 Sp CsV PPV Sen NPV
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</p> <p>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</p> <p>Estimates ranged from 0.46 to 0.97, and 10 studies reported estimates \geq 0.70.</p> <p>E interrater reliability</p> <p>Pearson's r ranged from 0.82 to 0.98 and the intraclass r ranged from 0.46 to 0.99.</p> <p>E interrater reliability</p> <p>Using a structured interview version of the Hamilton depression scale, they found an overall intraclass coefficient of 0.92.</p> <p>S test-retest</p> <p>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</p>	<p>CtV</p> <p>Because of its wide use and long clinical tradition, the Hamilton depression scale seems to both define as well as measure depression.</p> <p>CsV</p> <p>Studies reporting estimates of convergent validity of the 17-item Hamilton Depression Rating Scale, compared with other depression measures.</p> <p>HDRS – depression measures < 0.50 : 9 studies</p> <p>HDRS – depression measures between 0.50 and 0.70 : 12 studies</p> <p>HDRS – depression measures > 0.70 : 14 studies</p> <p>HDRS – Global Assessment Scale : - 0.86 and – 0.47</p> <p>HDRS – Visual Analogue Scale : - 0.65</p> <p>Sen</p>	

<p>Scale, this increased the mean retest reliability across individual items to 0.54.</p>	<p>Mean : 0.76 with a range from 0.45 to 0.88</p> <p>Sp</p> <p>Mean : 0.91 with a range from 0.75 to 1.00</p> <p>PPV</p> <p>Mean : 0.77 with a range from 0.37 to 1.00</p> <p>NPV</p> <p>Mean : 0.92 with a range from 0.86 to 0.99</p>	
<p>3. E interrater reliability</p> <p>Inter-rater reliability with two psychiatrists rating 20 subjects was good (Pearson's $r = 0.96$), a finding confirmed by other studies (Korner <i>et al.</i>, 1990)</p> <p>E interrater reliability</p> <p>Its inter-rater reliability is reported to be good: correlations are high ranging from 0.84 to 0.98 (Hamilton, 1976 ; Knesevitch <i>et al.</i>, 1977 ; Rehm, 1981)</p> <p>IC</p>	<p>CrV</p> <p>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 <i>et al.</i> (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</p> <p>Hamer <i>et al.</i> (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>	

<p>The alpha correlations for internal consistency were 0.82 to 0.83</p> <p>S test-retest</p> <p>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>4. IC</p> <p>Alpha 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 <i>et al.</i>, 1990). Other estimates include 0.76 (Rehm <i>et al.</i>, 1985), 0.83, 0.94 and 0.95 (Hedlung <i>et al.</i>, 1979), and 0.84 (Diefenbach <i>et al.</i>, 2001). Caroll <i>et al.</i> (1981) obtained a median item-total correlation of 0.54 (range 0.19 to 0.78).</p> <p>E interrater reliability</p> <p>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-</p>	<p>CrV</p> <p>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. <i>et al.</i>, 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. <i>et al.</i>, 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. <i>et al.</i>, 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. <i>et al.</i>, 1979).</p> <p>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</p>	

<p>term care patients (Foster, J.R. <i>et al.</i>, 1988).</p> <p>Intraclass coefficients include 0.70 to 0.72 (Maier, W. <i>et al.</i>, 1988), 0.83 (Korner, A. <i>et al.</i>, 1990), 0.85 (Winokur, A. <i>et al.</i>, 1982) and 0.92 (Endicott, J. <i>et al.</i>, 1981).</p> <p>S test-retest</p> <p>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. <i>et al.</i>, 1990)</p>	<p>Personality Inventory depression scale, however, were lower, at 0.27 and 0.34 (Hedlung, J.L. <i>et al.</i>, 1979).</p> <p>CrV</p> <p>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. <i>et al.</i>, 1979).</p>	
<p>5. E interrater reliability</p> <p>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.</p> <p>Intraclass correlation for overall : 0.947</p> <p>IC</p> <p>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)