Assessing full remission

Roger McIntyre, MD; Sidney Kennedy, MD; R. Michael Bagby, PhD; David Bakish, MD

McIntyre, Kennedy, Bagby — Mood and Anxiety Disorders Program, Department of Psychiatry, University of Toronto, Toronto, Ont.; Bakish — Department of Psychiatry, University of Ottawa, Ottawa, Ont.

The 17-item Hamilton Rating Scale for Depression (HAM-D_{17}) has been used for 4 decades as the "gold standard" instrument to assess the severity of depression and response to therapy in clinical research. The clinical utility of the HAM-D_{17} is hampered, in part, by the length of time required to administer the interview and by concern about a lack of inter-rater reliability. Several groups have developed shorter versions of the HAM-D_{17} for use in clinical practice. However, despite extensive research highlighting the importance of achieving full remission in minimizing the risk of relapse and recurrence, these shortened questionnaires have not been validated for the task of distinguishing between remission and response. A shortened form of the HAM-D_{17} with cut-off scores for full remission would offer a useful tool that physicians could readily employ in clinical practice. On the basis of the responses of a sample of 292 patients with major depression who received standard clinical treatment at a tertiary university affiliated hospital (Depression Clinic, Centre for Addiction and Mental Health, Toronto, Ont.) we derived a shortened versionof the HAM-D. Seven items with the greatest frequency of occurrence and sensitivity to change with treatment were identified and designated as the Toronto HAM-D_7. A score of 3 or less on the Toronto HAM-D_7 was found to correlate with the 17-item HAM-D definition of full remission (i.e., score of 7 or less).

L'échelle de dépression de Hamilton (HAM-D_{17}) à 17 éléments sert depuis quatre décennies, en recherche clinique, comme «étalon-or» afin d'évaluer la gravité de la dépression et la réponse au traitement. L'utilité clinique de l'échelle HAM-D_{17} est en partie entravée par le temps nécessaire pour réaliser l'entrevue et par les préoccupations que soulève le manque de fiabilité entre les évaluateurs. Plusieurs groupes ont mis au point des versions abrégées de l'échelle HAM-D_{17} pour la pratique clinique. En dépit de recherches poussées qui mettent en évidence l'importance de réaliser une rémission complète pour réduire au minimum le risque de rechute et de récidive, ces questionnaires abrégés n'ont pas été validés pour la tâche qui consiste à distinguer la rémission de la réponse. Une version abrégée de l'évaluation HAM-D_{17} et des résultats limites dans le cas de la rémission complète constitueraient un outil utile que les médecins pourraient facilement employer en pratique clinique. En nous fondant sur les réponses d'un échantillon de 292 patients

Correspondence to: Dr. Roger McIntyre, Mood and Anxiety Disorders, Centre for Addiction and Mental Health, 250 College St., Toronto ON M5T 1R8; fax 416 979-6864; roger_mcintyre@camh.net

Information about the HAM-D_7 and its development can be obtained from Dr. R. Michael Bagby, Acting Director, Clinical Research Department, Centre for Addiction and Mental Health, 250 College St., Toronto ON M5T 1R8; fax 416 979-6821; Michael_Bagby@camh.net

Medical subject headings: antidepressive agents; behavioral symptoms; depressive disorder; drug therapy; psychiatric status rating scales; recurrence; remission induction; treatment outcome.

J Psychiatry Neurosci 2002;27(4):235-9.

Submitted Jun. 11, 2001
Revised May 16, 2002
Accepted May 27, 2002

aux prises avec une dépression majeure qui ont reçu le traitement clinique normalisé à un hôpital de soins tertiaires affilié à une université (Clinique de la dépression, Centre de toxicomanie et de santé mentale, Toronto (Ontario)), nous avons dérivé une version abrégée de l'évaluation HAM-D. Nous avons défini les sept éléments les plus fréquents et les plus sensibles aux changements produits par le traitement et nous avons donné à notre sous-échelle l'appellation «HAM-D_7 de Toronto». On a constaté qu'un résultat de 3 ou moins sur l'échelle HAM-D_7 de Toronto correspondait à une rémission complète selon la définition HAM-D à 17 éléments (c.-à-d. un résultat de 7 ou moins).

Introduction

The lifetime prevalence of major depressive disorder (MDD) in industrialized countries is between 5% and 25%.[1] A debilitating and life-threatening illness, MDD is responsible for reduced productivity and social functioning and a suicide rate of up to 15%.[2] Despite the availability of a variety of antidepressant medications and established psychotherapies, the long-term outcome of depression remains rather disappointing.

The goal of antidepressant treatment is sustained and full remission of depressive symptoms to prevent relapse and recurrence, with a return to previous levels of occupational and social functioning. Failure to achieve full remission is associated with an increased risk of relapse and recurrence, higher rates of chronicity, readmission to hospital, with high service utilization and a reduced quality of life. Therefore, distinguishing response (i.e., symptomatic improvement with residual or subsyndromal depressive symptoms) from remission (i.e., virtually full symptom elimination) has important clinical significance and requires the systematic monitoring of the presence and severity of depressive symptoms.

The Hamilton Depression Rating Scale (HAM-D) was originally published in 1960.[3,4] Although widely used by psychiatric researchers, especially in clinical trials, this and other clinician rating scales are not widely used in clinical practice. The time required to administer the questionnaire is thought to be one deterrent to its use. To obtain more clinically useful measures of depression severity and response to treatment, several groups have developed brief versions of the HAM-D.[5–7] However, despite extensive research highlighting the importance of achieving full remission, it is not known whether these shortened questionnaires provide a means of distinguishing between remission and response. A brief HAM-D with cut-off scores for full remission would be a useful tool that physicians could readily employ in clinical practice.

We derived a shortened version of the HAM-D with a cut-off score for remission on the basis of responses of a sample of patients who met *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (DSM-IV),[1] criteria for major depression and were receiving standard clinical treatment at a tertiary university affiliated hospital (Depression Clinic, Centre for Addiction and Mental Health [CAMH], Toronto, Ont.). We then compared the predictive validity of this and other abbreviated versions of the HAM-D.

Methods

All subjects were outpatients with unipolar nonpsychotic depression who were treated at the Depression Clinic at the CAMH and consented to be part of a clinical database. Criteria for entry into the database were: (a) a diagnosis of nonpsychotic major depressive disorder according to the DSM-IV, (b) a HAM-D – 17 item (HAM-D_{17}) total score of 16 or greater, (c) no concurrent active medical illness and (d) absence of antidepressant medication for a minimum of 2 weeks before treatment initiation.

The treatment protocol for the clinical database requires patients to be treated and followed for at least 14 weeks but not more than 26 weeks.

HAM-D_{17} evaluations were available for baseline and endpoint. The items on the HAM-D_{17} that were most strongly associated with change in clinical status were used to develop a briefer scale. In addition, the score associated with full remission was determined. These data were then compared with other previously derived short forms of the HAM-D_{17} including the Bech Melancholia Scale (which uses items 1, 2, 7, 8, 10 and 13),[5] the Gibbons Global Depression Severity Scale (items 1, 2, 3, 7, 9, 10, 11 and 14)[6] and the Maier and Phillip Severity Subscale (items 1, 2, 7, 8, 9 and 10).[7]

Results

Ratings were obtained from a sample of 292 (107 men,

185 women) patients with MDD. Of these, 200 (79 men, 121 women) were also rated at the end of the treatment. The average time from treatment initiation to protocol termination for those who completed the study was 20.0 (standard deviation 5.0) weeks.

Table 1 outlines the frequency of occurrence at baseline and the magnitude of change at the end of treatment for each of the 17 HAM-D items for the patients in the database. Depressed mood, guilt, suicide, insomnia (middle), difficulty with work and interests, psychic and somatic anxiety, as well as general somatic symptoms were reported by more than 70% of patients. Loss of insight and weight change were infrequent. With the exception of item 5 (middle insomnia), these items were also those that were most sensitive to change with treatment, exhibiting change scores (calculated as effect sizes [Cohen's *d*]) between 0.83 and 1.84. Insomnia was relatively less sensitive to change with treatment and was therefore not included in the final Toronto HAM-D_7.

The items that were reported most frequently and were the most sensitive to change (i.e., 1, 2, 3, 7, 10, 11 and 13) were included in the Toronto HAM-D_7 (Table 2). These items overlap considerably with those included in previous unidimensional subscales, with depressed mood, work and interests, guilt and psychic anxiety being included in all subscales. The items in the HAM-D short forms were tested for reliability and internal consistency and found to be comparable across the various shortened versions and the full HAM-D_{17}.

Frank and colleagues[8] defined full remission of depression as an HAM-D_{17} of 7 or less. The cut-off scores that would define a full remission comparable to that determined by the HAM-D_{17} are presented in Table 3. All scales demonstrated high rates of sensitivity and specificity. The positive predictive power was over 90%, and the negative predictive power over 80% in all cases.

Table 1: Frequency of symptoms at baseline and change scores at treatment termination for items on the HAM-D_{17}

Item no.	HAM-D_{17} symptom	Sex of patient, % of patients reporting symptom			Change score,* Cohen's *d*
		Male, *n* = 107	Female, *n* = 185	Total, *n* = 292	
1	Depressed mood	98.1	97.3	97.6	1.81
2	Guilt	86.0	84.9	85.3	0.86
3	Suicide	77.6	71.8	73.3	0.88
4	Insomnia (early)	60.7	62.2	61.6	0.37
5	Insomnia (middle)	69.2	72.4	71.2	0.59
6	Insomnia (delayed)	53.3	55.1	54.5	0.29
7	Work and interests	99.1	97.8	99.7	1.84
8	Retardation	57.0	45.4	49.7	0.24
9	Agitation	53.3	54.1	53.8	0.58
10	Psychic anxiety	58.8	89.7	89.4	0.83
11	Somatic anxiety	86.0	92.4	90.1	1.03
12	Gastrointestinal	44.9	47.0	46.2	0.31
13	General somatic	98.1	94.1	95.5	0.88
14	Loss of libido	64.5	68.6	67.1	0.29
15	Hypochondriasis	49.5	60.5	56.5	0.66
16	Loss of weight	24.3	27.6	26.4	0.24
17	Loss of insight	9.3	9.2	9.2	0.04

Note: Occurrence of symptom was defined as at least "1" or more on any given item.
*Change scores, presented as Cohen's *d* effect sizes, are based on the sample of 200 patients who received HAM-D ratings at both baseline and the end of treatment.

Discussion

The 17-item HAM-D measures a set of symptoms with face validity in major depression, including anxiety, sleep problems, impact on work and activities and hypochondriasis. Although the clinician-rated HAM-D_{17} and the longer 21-, 24- and 29-item versions have wide acceptance in research settings for measuring efficacy outcomes, the tool has been criticized for its inadequate reliability, lack of internal and external validity and overemphasis on somatic complaints.[5,9] Other observer tools, such as the 10-item Montgomery–Asberg Depression Rating Scale (MADRS), are also available and may offer improved validity.[10] However, none of

Table 2: Items included in shortened forms of the HAM-D_{17}

Item no.	HAM-D_{17} symptom	Subscales			
		Toronto	Bech[5]	Gibbons[6]	Maier[7]
1	Depressed mood	X	X	X	X
2	Guilt	X	X	X	X
3	Suicide	X		X	
4	Insomnia (early)				
5	Insomnia (middle)				
6	Insomnia (delayed)				
7	Work and interests	X	X	X	X
8	Retardation		X		X
9	Agitation			X	X
10	Psychic anxiety	X	X	X	X
11	Somatic anxiety	X		X	
12	Gastrointestinal				
13	General somatic	X	X		
14	Loss of libido			X	
15	Hypochondriasis				
16	Loss of weight				
17	Loss of insight				
Total no. of items		7	6	8	6

these rating instruments are popular in the clinical setting. This is primarily because of the length of time required to administer the interview, the lack of training for clinicians and the uncertain value of a given severity score and change across time for different populations.

The briefer unidimensional versions of the HAM-D_{17}, which assess "core depressive symptoms" commonly reported in clinical practice (e.g., the Bech Melancholia Scale, Maier and Phillip Severity Subscale and the Gibbons Global Depression Severity Scale)[5–7] share considerable symptom overlap in that they all include items 1, 2, 7 and 10. The items in the Toronto HAM-D_7, selected on the basis of their frequency of occurrence at baseline and their sensitivity to change with treatment, also included items 1, 2, 7 and 10.

These brief scales have been shown to correlate with the HAM-D_{17} assessment of both severity of symptoms and sensitivity to change over time. A study of 164 depressed outpatients with and without atypical features demonstrated that the Bech HAM-D_6 was as sensitive to symptom changes as the 17-, 21- and 24-item versions of the scale.[11] Furthermore, the different versions of the HAM-D were strongly correlated with each other at baseline and endpoint in both depression subtypes. It was concluded that the 6-item version of the HAM-D allowed the assessment of severity of depression with comparable sensitivity to the standard and more elaborate versions of the same scale. Hooper and Bakish[12] compared the sensitivity of the HAM-D_6 with the HAM-D_{17} and the MADRS in a retrospective analysis of 4 clinical trials (3 double-blinded, 1 open study) comprising 143 outpatients receiving treatment for major depressive disorder, with or without melancholia and/or dysthymic disorder. The briefer version strongly correlated with the longer version at baseline and termination. The HAM-D_6, HAM-D_{17}, and MADRS demonstrated equal sensitivity to change over the course of treatment, both in the full sample and in the dysthymic and melancholic subgroups. The ability of the shorter version to show comparable results supports the assertion that the HAM-D_6 measures "core" features of depression.

Table 3: Cut-off scores on subscales of the HAM-D_{17} (comparable to the HAM-D_{17} cut-off of 7) for predicting full remission of depressive symptoms

HAM-D subscale	Cut-off score	Sensitivity	Specificity	Positive predictive power	Negative predictive power
Toronto	3.04	0.95	0.84	0.94	0.86
Bech[5]	2.74	0.92	0.92	0.97	0.81
Gibbons[6]	3.27	0.96	0.89	0.96	0.89
Maier[7]	2.34	0.94	0.81	0.93	0.83

Faries et al[13] conducted 2 meta-analyses (n = 2899) to compare the sensitivity of the multidimensional HAM-D_{17} with the unidimensional briefer scales (Bech,[5] Maier[7] and Gibbons[6]) for detecting treatment differences. In both meta-analyses, the unidimensional core subscales outperformed the HAM-D_{17} at detecting treatment differences. With the improved responsiveness and increased effect size, studies based on these subscales would require one-third fewer subjects to detect drug treatment differences. The HAM-D_6 appears to be as (or more) sensitive to change during treatment as the HAM-D_{17} and the MADRS.

One potential limitation of the shorter form is that, statistically, the presence of fewer items typically results in lower reliability. However, our data indicate that the shorter forms have comparable reliability estimates to the HAM-D_{17}. In addition, all of these shortened versions have been extracted from the same parent HAM-D_{17}. Development of the original scale was guided by clinical experience and logic rather than by empirical testing and re-evaluation.[6] It is confounded by extraneous items that do not reflect severity of depression; it is vulnerable to the influence of antidepressant side effects, and the clinical value of the total score is not clear.[6,12] Moreover, the HAM-D_7 was not validated in patients with known concurrent medical disorders. It is well established that many people with depression in primary care settings present with multiple medical conditions and somatic complaints. The HAM-D_7 includes 2 items that assess somatic symptoms (somatic anxiety, energy). It behooves the clinician to ascertain if somatic symptoms are part of a confluence of depressive symptoms or due to a general medical condition; this scale does not replace everyday clinical decision making.

The question is, does a shortened version of a flawed scale have clinical utility? A prospectively designed study to investigate factors that are indicative of the severity of depression and are sensitive to change with antidepressant therapy would be ideal. A prospective study to validate the Toronto HAM-D_7 in general practice is planned.

The clinical utility of the shorter version is increased by the determination that a score of approximately 3 or less is comparable to a HAM-D_{17} score of less than 8,

which is considered a full remission. A cut-off score for "response" was not derived, because it is not considered an acceptable endpoint in clinical practice. A caution is that the cut-off scores derived in this study were based on discriminant function analysis, which employs an algorithm that maximizes a balance between sensitivity (in this instance the presence of remission) and specificity (the absence of remission). Different cut-off scores might be applied if the clinician is more concerned about misidentifying a patient who is not in remission as being in remission (undertreating) at the expense of misidentifying a patient who is in remission as not (overtreating).

Another caution is that the items that compose the HAM-D_7 were derived from a single sample and, therefore, need to be replicated in other samples before widespread use, especially in instances where important clinical decisions are to be made. Similarly, the cut-score proposed to detect full remission was derived using discriminant function analysis (DFA) in this sample only. As DFA procedures capitalize on "chance" effects, the cut-score derived in this sample must be replicated before widespread use in either clinical or research settings. Pending replication and cross-validation of these items and the cut-score for determining full remission, the use of the HAM-D_7 may have a role in clinical practice and antidepressant trials.

Competing interests: Dr. McIntyre has received research support from Janssen-Ortho, Eli Lilly, GlaxoSmithKline, the Centre for Addiction and Mental Health Foundation and Wyeth-Ayerst Canada; is on the speakers' bureaus of GlaxoSmithKline, Lundbeck, Wyeth-Ayerst Canada, Organon, Janssen-Ortho, Eli Lilly, Pfizer, Astra-Zeneca Canada and Boehringer Ingelheim; and is a consultant for Bristol Myers Squibb, GlaxoSmithKline, Janssen-Ortho, Astra-Zeneca Canada and Wyeth-Ayerst Canada. Dr. Kennedy has received research support from Pfizer, Astra-Zeneca, Organon and Boehringer Ingelheim; is on the speakers' bureaus of Lundbeck, Organon, Wyeth-Ayerst and GlaxoSmithKline; and serves on advisory boards for Pfizer, the Lundbeck Foundation, Eli Lilly, GlaxoSmithKline and Servier. Dr. Bagby received an honorarium to develop the 7-item Hamilton Rating Scale for Depression and derive cut-off scores and travel assistance to attend a conference where the data were discussed from CMED. Dr. Bagby also received research support and financial assistance for his work on the conceptual and statistical procedures used to develop the HAM-D_7 from Eli Lilly Canada. Dr. Bakish has received research support from Merck, Pharmacia, Astra-Zeneca, Pfizer, Wyeth-Ayerst and Boehringer Ingelheim and travel assistance and speaker's fees from Wyeth-Ayerst.

References

1. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 4th ed. Washington: American Psychiatric Association; 1994. p. 339-45.
2. Akiska HS. Mood disorders: introduction and overview. In: Kaplan HI, Saddock BJ, editors. *Comprehensive textbook of psychiatry*. 6th ed. Baltimore: Williams and Wilkins; 1995. p. 1067-78.
3. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;23:56-62.
4. Hamilton M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 1967;6:276-96.
5. Bech P, Gram LF, Dein E, Jacobsen O, Vitger J, Bolwig TG. Quantitative rating of depressive states. *Acta Psychiatr Scand* 1975;51:161-70.
6. Gibbons RD, Clark DC, Kupfer DJ. Exactly what does the Hamilton Depression Rating Scale measure? *J Psychiatr Res* 1993;27:259-73.
7. Maier W, Phillip M. Improving the assessment of severity of depressive states: a reduction of the Hamilton Depressive Scale. *Pharmacopsychiatry* 1985;18:114-5.
8. Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, Lavori PW, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: Remission, recover, relapse and recurrence. *Arch Gen Psychiatry* 1991;48:851-5.
9. Linden M, Borchelt M, Barnow S, Gesielmann B. The impact of somatic morbidity on the Hamilton Depression Rating Scale in the very old. *Acta Psychiatr Scand* 1995;92:150-4.
10. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;134:382-9.
11. O'Sullivan RL, Fava M, Agustin C, Baer L, Rosenbaum JF. Sensitivity of the six-item Hamilton Depression Rating Scale. *Acta Psychiatr Scand* 1997;95:379-84.
12. Hooper CL, Bakish D. An examination of the sensitivity of the six-item Hamilton Rating Scale for Depression in a sample of patients suffering from major depressive disorder. *J Psychiatry Neurosci* 2000;25:178-84.
13. Faries D, Herrera J, Rayamajhi J, DeBrota D, Demitrack M, Potter WZ. The responsiveness of the Hamilton Depression Rating Scale. *J Psychiatric Res* 2000;34:3-10.