A Structured Interview Guide for the Hamilton Depression Rating Scale

Janet B. W. Williams, DSW

• **The Hamilton Depression Rating Scale (HDRS) is the most widely used scale for patient selection and follow-up in research studies of treatments of depression. Despite extensive study of the reliability and validity of the total scale score, the psychometric characteristics of the individual items have not been well studied. In the only reliability study to report agreement on individual items using a test-retest interview method, most of the items had only fair or poor agreement. Because this is due in part to variability in the way the information is obtained to make the various rating distinctions, the Structured Interview Guide for the HDRS (SIGH-D) was developed to standardize the manner of administration of the scale. A test-retest reliability study conducted on a series of psychiatric inpatients demonstrated that the use of the SIGH-D results in a substantially improved level of agreement for most of the HDRS items.**

(*Arch Gen Psychiatry* 1988;45:742-747)

The Hamilton Depression Rating Scale (HDRS) was developed during the late 1950s as a standardized scale for the measurement of the severity of depressive symptoms.[1] The symptoms are defined by anchor-point descriptions that increase in intensity; clinicians are to consider both the intensity and frequency of a symptom when assigning it a rating value. The scale was initially designed to yield a total score based on 17 of its 21 items, although many investigators have used all 21 items.[2]

Since its initial publication, the HDRS has emerged as the most widely used scale for patient selection and follow-up in research studies of treatments for depression.[2,3] Undoubtedly, the success of this scale is due to its comprehensive coverage of depressive symptoms and related psychopathology, as well as its strong psychometric properties.[2,4] In numerous studies, the total HDRS score has proved reliable and to have a high degree of concurrent and differential validity.[3]

Accepted for publication Oct 16, 1987.

From the Department of Psychiatry, Columbia University College of Physicians and Surgeons, New York.

Reprint requests to Box 74, 722 W 168th St, New York, NY 10032 (Dr Williams).

Since the HDRS is commonly used to measure change over time, the individual items are often examined to study the differential effect of various treatments on specific symptoms or groups of symptoms of depression.[5-7] Therefore, reliability at the item level is important for research. Despite extensive study of the reliability and validity of the total HDRS score, however, the psychometric characteristics of the individual items have not been well studied.[4]

There are several studies reported in the literature in which the reliabilities of the individual items are examined. However, all but one of these studies report reliability data resulting from joint interviews, that is, interviews in which one clinician interviews the patient and makes ratings on the instrument, and another clinician, *observing the same interview*, also makes ratings. In some studies the live interview was observed[8]; in others, the reliability ratings were made from a videotape of the original interview.[4,9] However, because information variance is artificially eliminated with this joint observation procedure, it provides an inflated value for reliability if one is interested in generalizing to the real world in which different interviewers ask different questions to gather necessary information. Furthermore, with this joint reliability procedure, often the independence of the two clinicians' ratings is compromised when the rating decision of the interviewer inadvertently becomes known to the observing clinician because of its effect on the interviewer's inquiry.[10] For example, if in response to a question about suicidal ideation a patient describes thinking that sometimes he wishes he were dead, and the interviewer does not inquire further about any specific suicidal attempts, the observing clinician can assume that the interviewer will rate the severity of suicidal ideation as no more than mild.

Many researchers now regard the test-retest method as representing the "state of the art" of reliability assessment.[11] In this procedure, two clinicians independently perform and rate interviews of the same subject, as close together in time as possible. The advantage of this procedure is that it more closely approximates the reliability of judgments made in actual practice, in which independent assessments are the rule. As expected, reliability obtained by this method is generally lower than that obtained by joint assessment because of the increase in information

variance with the test-retest method.[10] Despite the fact that the test-retest method requires patients to undergo more interviews and may be more difficult to coordinate logistically, most researchers believe this expense is worth the increase in generalizability of the results.

The only published reliability study of the HDRS in which the test-retest method was used and item reliabilities were reported was conducted by Cicchetti and Prusoff.[11] In this study a series of patients with major depression were interviewed by two clinicians at each of two points in time: at randomization into a controlled trial of a tricyclic antidepressant and at the end of the clinical trial, in most cases 16 weeks after randomization. For most of the items, only fair or poor agreement was obtained, although agreement was significantly better at the end point than at randomization, probably because of greater variability in the extent of depressive symptoms present.

This general lack of item reliability of the HDRS may be due to one or both of two factors: varying interpretations of the meanings of the anchor-point descriptions and variability in the way in which the information is obtained to make the various rating distinctions. Many agree that certain items of the HDRS are problematic and should be revised to increase their usefulness.[4] Such a revision requires a major effort and is beyond the scope of this study. However, the current project grew out of an assumption that it should be possible to increase the reliability of the individual items by standardizing the way in which the rating information is obtained.

Since the availability of structured interview guides for rating various aspects of psychopathology, it has been amply demonstrated that the use of a structured interview guide generally increases the reliability of ratings.[12] The specified questions in such a guide ensure that raters obtain the same information from all patients, thus reducing the information variance. This study describes the result of an effort to improve the reliability of the HDRS at the item level by developing and testing a new instrument called the Structured Interview Guide for the HDRS (SIGH-D).

Test-Retest Item Reliabilities of Hamilton Depression Rating Scale*

Depression Scale Item	SIGH-D Study (n=23)	Cicchetti and Prusoff[11] At Randomization (n=86)	Cicchetti and Prusoff[11] At End (n=81)
Depressed mood	.80	.37	.72
Work and activities	.54	.33	.64
Genital sxs	.70	.39	.59
Somatic sxs GI	.59	.43	.51
Loss of weight	.58	.57	.06
Insomnia early	.80	.76	.57
Insomnia middle	.62	.57	.45
Insomnia late	.30	.42	.49
Somatic sxs general	.61	.30	.42
Feelings of guilt	.63	.18	.37
Suicide	.64	.59	.64
Anxiety psychic	.78	.19	.40
Anxiety somatic	.66	.34	.45
Hypochondriasis	.55	.29	−.04
Insight	.00	−.02	−.03
Psychomotor retardation	.32	.39	.26
Psychomotor agitation	.11	.20	.32
Diurnal variation	.52	.50	.59
Depersonalization and derealization	.70	.15	.24
Paranoid sxs	.74	.23	.32
Obsessional and compulsive sxs	.87	.47	.25
17 items	.81	...	...
21 items	.82	...	...
22 items†	...	.77	.89

*SIGH-D indicates Structured Interview Guide for the Hamilton Depression Rating Scale; sxs, symptoms; GI, gastrointestinal.

†Total score calculated on 22 items because diurnal variation and diurnal variation (severe) were considered separate items.

DEVELOPMENT OF THE INTERVIEW GUIDE

The effort to develop a structured interview guide for the HDRS began with observations of HDRS interviews given routinely by a number of experienced clinicians. Based on this and my own extensive experience with the scale, interview questions were drafted that were appropriate for gathering the information necessary to make the various item distinctions in a relatively standard way.

For the purpose of this project, the version of the HDRS that has come to be regarded as the more or less "standard" version was used.[13] Two minor changes were made: under somatic symptoms gastrointestinal, the anchor-point cues of "heavy feelings in abdomen" and "requests or requires laxatives or medication for bowels or medication for gastrointestinal symptoms" were eliminated since they were cumbersome and were almost never used, and a note to code "0" if the individual was not depressed was added to the "insight" and "diurnal variation" items for individuals who had recovered from their depression or who were in treatment for another mental disorder and were not depressed. Finally, the order of the items was changed to better conform to the order in which the information is obtained in most clinical interviews.

Once the initial interview guide was developed, it was pilot tested on a number of patients from convenience samples. These included both psychiatric patients and patients with Parkinson's disease and depression, and included, in some cases, repeat ratings over time. In addition, the interview guide was distributed to a number of researchers who use the HDRS as a routine instrument, urging them to try out the interview and asking for their comments. Finally, it was sent to Max Hamilton himself for his critical review. Revisions in the interview guide were subsequently made and it was again pilot tested on a small number of patients. The final instrument appears in the Figure.

THE SIGH-D

The interview guide is prefaced by an information page for raters instructing them to begin the query for each item with the first recommended SIGH-D question (appearing in bold for each item). Often this question will elicit enough information about the severity and frequency of a symptom for the clinician to rate the item with confidence. Follow-up questions are provided, however, for use when further exploration or additional clarification of symptoms is necessary. The questions provided in the interview guide should be asked until enough information has been obtained to rate the item. In some cases, raters may also have to add their own follow-up questions to obtain necessary information.

Whenever possible, each area of inquiry begins with an open-ended question to encourage patients to describe

This interview guide is based on the Hamilton Depression Rating Scale.[1]
Numbers in parentheses on far right are for computer data entry.

STRUCTURED INTERVIEW GUIDE FOR THE HAMILTON - DEPRESSION RATING SCALE (SIGH-D)

INTERVIEWER: The first question for each item should be asked exactly as written. Often this question will elicit enough information about the severity and frequency of a symptom for you to rate the item with confidence. Follow-up questions are provided, however, for use when further exploration or additional clarification of symptoms is necessary. The specified questions should be asked until you have enough information to rate the item confidently. In some cases, you may also have to add your own follow-up questions to obtain necessary information.

NOTES: Time period. Although the interview questions indicate that the ratings should be based on the patient's condition in the past week, some investigators using this instrument as a change measure may wish to base their ratings on the previous two to three days. If so, the questions may be preceded by "In the last couple of days..."

Loss of weight item. It is recommended that this item be rated positively whenever the patient has lost weight relative to their baseline weight (i.e., before their current episode of depression), provided that they have not begun to gain back lost weight. Once the patient has begun to gain weight, however, even if they are still below their baseline, they should no longer be rated positively on this item.

Referent of "usual" or "normal" condition. Several of the interview questions refer to the patient's usual or normal functioning. In some cases, such as when the patient has Dysthymia or Seasonal Affective Disorder, the referent should be to the last time they felt OK (i.e., not depressed or high) for at least a few weeks.

STRUCTURED INTERVIEW GUIDE FOR THE HAMILTON DEPRESSION RATING SCALE (SIGH-D)*

PT'S NAME: _____ PT'S ID: _ _ _ _ _ _ _ _ (1-7)

INTERVIEWER:_____ DATE: _ _ _ _ _ _ (8-13)

OVERVIEW: I'd like to ask you some questions about the past week. How have you been feeling since last (DAY OF WEEK)? IF OUTPATIENT: Have you been working? IF NOT: Why not?

What's your mood been like this past week?

Have you been feeling down or depressed?

Sad? Hopeless?

In the last week, how often have you felt (OWN EQUIVALENT)? Every day? All day?

Have you been crying at all?

DEPRESSED MOOD (sadness, hopeless, helpless, worthless):

0 - absent
1 - indicated only on questioning
2 - spontaneously reported verbally
3 - communicated non-verbally, i.e. facial expression, posture, voice, tendency to weep
4 - VIRTUALLY ONLY; this in spontaneous verbal and non-verbal communication (15)

IF SCORED 1-4 ABOVE, ASK: How long have you been feeling this way?

*Janet B.W. Williams, D.S.W., Biometrics Research Department, New York State Psychiatric Institute, 722 West 168th Street, New York, New York 10032

How have you been spending your time this past week (when not at work)?

Have you felt interested in doing (THOSE THINGS), or do you feel you have to push yourself to do them?

Have you stopped doing anything you used to do? IF YES: Why?

Is there anything you look forward to?

(AT FOLLOW-UP: Has your interest been back to normal?)

WORK AND ACTIVITIES:

0 - no difficulty
1 - thoughts and feelings of incapacity, fatigue or weakness related to activities, work or hobbies
2 - loss of interest in activity, hobbies or work - by direct report of the patient or indirect in listlessness, indecision and vacillation (feels he has to push self to work or activites)
3 - decrease in actual time spent in activities or decrease in productivity. In hosp, pt. spends less than 3 hrs/day in activities (hospital job or hobbies) exclusive of ward chores
4 - stopped working bec. of present illness. In hospital, no activities except ward chores, or fails to perform ward chores unassisted (16)

How has your interest in sex been this week? (I'm not asking about performance, but about your interest in sex - how much you think about it.)

Has there been any change in your interest in sex (from when you were not depressed)?

Is it something you've thought much about? IF NO: Is that unusual for you?

GENITAL SYMPTOMS (such as loss of libido, menstrual disturbances):

0 - absent
1 - mild
2 - severe (17)

How has your appetite been this past week? (What about compared to your usual appetite?)

Have you had to force yourself to eat?
Have other people had to urge you to eat?

SOMATIC SYMPTOMS GASTROINTESTINAL:

0 - none
1 - loss of appetite but eating without encouragement
2 - difficulty eating without urging (18)

Have you lost any weight since this (DEPRESSION) began? IF YES: How much?

IF NOT SURE: Do you think your clothes are any looser on you?

AT FOLLOW-UP: Have you gained any of the weight back?

LOSS OF WEIGHT (Rate either A or B):

A. When rating by history:
0 - no weight loss
1 - probable weight loss associated with present illness
2 - definite (according to patient) weight loss
3 - not assessed (19)

B. On weekly ratings by ward staff, when actual weight changes are measured:
0 - less than 1 lb. loss in week
1 - more than 1 lb. loss in week
2 - more than 2 lb. loss in week
3 - not assessed (20)

How have you been sleeping over the last week?

Have you had any trouble falling asleep at the beginning of the night? (Right after you go to bed, how long has it been taking you to fall asleep?)

How many nights this week have you had trouble falling asleep?

INSOMNIA EARLY:

0 - no difficulty falling asleep
1 - complains of occasional difficulty falling asleep - i.e., more than 1/2 hour
2 - complains of nightly difficulty falling asleep (21)

During the past week, have you been waking up in the middle of the night? IF YES: Do you get out of bed? What do you do? (Only go to the bathroom?)

When you get back in bed, are you able to fall right back asleep?

Have you felt your sleeping has been restless or disturbed some nights?

INSOMNIA MIDDLE:

0 - no difficulty
1 - complains of being restless and disturbed during the night
2 - waking during the night - any getting out of bed (except to void) (22)

What time have you been waking up in the morning for the last time, this past week?

IF EARLY: Is that with an alarm clock, or do you just wake up yourself? What time do you usually wake up (that is, before you got depressed)?

INSOMNIA LATE:

0 - no difficulty
1 - waking in early hours of morning but goes back to sleep
2 - unable to fall asleep again if gets out of bed (23)

How has your energy been this past week?

Have you been tired all the time?

This week, have you had any backaches, headaches, or muscle aches?

This week, have you felt any heaviness in your limbs, back or head?

SOMATIC SYMPTOMS GENERAL:

0 - none
1 - heaviness in limbs, back or head. Backaches, headache, muscle aches. Loss of energy and fatiguability.
2 - any clear-cut symptom (24)

Have you been especially critical of yourself this past week, feeling you've done things wrong, or let others down? IF YES: What have your thoughts been?

Have you been feeling guilty about anything that you've done or not done?

Have you thought that you've brought (THIS DEPRESSION) on yourself in some way?

Do you feel you're being punished by being sick?

FEELINGS OF GUILT:

0 - absent
1 - self-reproach, feels he has let people down
2 - ideas of guilt or rumination over past errors or sinful deeds
3 - present illness is a punishment. Delusions of guilt
4 - hears accusatory or denunciatory voices and/or experiences threatening visual hallucinations (25)

This past week, have you had any thoughts that life is not worth living, or that you'd be better off dead? What about having thoughts of hurting or even killing yourself?

IF YES: What have you thought about? Have you actually done anything to hurt yourself?

SUICIDE:

0 - absent
1 - feels life is not worth living
2 - wishes he were dead or any thoughts of possible death to self
3 - suicidal ideas or gesture
4 - attempts at suicide (26)

Have you been feeling especially tense or irritable this past week?

Have you been worrying a lot about little unimportant things, things you wouldn't ordinarily worry about? IF YES: Like what, for example?

ANXIETY PSYCHIC:

0 - no difficulty
1 - subjective tension and irritability
2 - worrying about minor matters
3 - apprehensive attitude apparent in face or speech
4 - fears expressed without questioning (27)

RATING BASED ON OBSERVATION DURING INTERVIEW

AGITATION:

0 - none
1 - fidgetiness
2 - playing with hands, hair, etc.
3 - moving about, can't sit still
4 - hand-wringing, nail biting, hair-pulling, biting of lips (32)

TOTAL 17-ITEM HAMILTON DEPRESSION SCORE: ___ ___ (33-34)

In this past week, have you had any of these physical symptoms? READ LIST, PAUSING AFTER EACH SX FOR REPLY.

How much have these things been bothering you this past week? (How bad have they gotten? How much of the time, or how often, have you had them?)

NOTE: DON'T RATE IF CLEARLY DUE TO MEDICATION (E.G., DRY MOUTH AND IMIPRAMINE)

ANXIETY SOMATIC (physiologic concomitants of anxiety, such as
GI - dry mouth, gas, indigestion, diarrhea, cramps, belching
C-V - heart palpitations, headaches
Resp - hyperventilating, sighing
Having to urinate frequently
Sweating):

0 - absent
1 - mild
2 - moderate
3 - severe
4 - incapacitating (28)

In the last week, how much have your thoughts been focused on your physical health or how your body is working (compared to your normal thinking)?

Do you complain much about how you feel physically?

Have you found yourself asking for help with things you could really do yourself? IF YES: Like what, for example? How often has that happened?

HYPOCHONDRIASIS:

0 - not present
1 - self-absorption (bodily)
2 - preoccupation with health
3 - frequent complaints, requests for help, etc.
4 - hypochondriacal delusions (29)

RATING BASED ON OBSERVATION

INSIGHT:

0 - acknowledges being depressed and ill OR not currently depressed
1 - acknowledges illness but attributes cause to bad food, climate, overwork, virus, need for rest, etc.
2 - denies being ill at all (30)

RATING BASED ON OBSERVATION DURING INTERVIEW

RETARDATION (slowness of thought and speech; impaired ability to concentrate; decreased motor activity):

0 - normal speech and thought
1 - slight retardation at interview
2 - obvious retardation at interview
3 - interview difficult
4 - complete stupor (31)

This past week have you been feeling better or worse at any particular time of day - morning or evening?

IF VARIATION: How much worse do you feel in the (MORNING OR EVENING)?

IF UNSURE: A little bit worse or a lot worse?

DIURNAL VARIATION:

A. Note whether symptoms are worse in morning or evening. If NO diurnal variation, mark none:
0 - no variation OR not currently depressed
1 - worse in A.M.
2 - worse in P.M. (35)

B. When present, mark the severity of the variation:
0 - none
1 - mild
2 - severe (36)

In the past week, have you ever suddenly had the feeling that everything is unreal, or you're in a dream, or cut off from other people in some strange way? Any spacey feelings? IF YES: How bad has that been? How often this week has that happened?

DEPERSONALIZATION AND DEREALIZATION (such as feelings of unreality and nihilistic ideas):

0 - absent
1 - mild
2 - moderate
3 - severe
4 - incapacitating (37)

This past week, have you felt that anyone was trying to give you a hard time or hurt you?

IF NO: What about talking about you behind your back?

IF YES: Tell me about that.

PARANOID SYMPTOMS:

0 - none
1 - suspicious
2 - ideas of reference
3 - delusions of reference and persecution (38)

In the past week, have there been things you've had to do over and over again, like checking the locks on the doors several times? IF YES: Can you give me an example?

Have you had any thoughts that don't make any sense to you, but that keep running over and over in your mind? IF YES: Can you give me an example?

OBSESSIONAL AND COMPULSIVE SYMPTOMS:

0 - absent
1 - mild
2 - severe (39)

TOTAL 21-ITEM HAMILTON DEPRESSSION SCORE: ___ ___ (40-41)

60 (79-80)

their experience in their own words. In this spirit, the interview begins with "I'd like to ask you some questions about the past week. How have you been feeling since last (DAY OF WEEK)?" The "depressed mood" item, then, begins with "What's your mood been like this past week?" and the insomnia items start with "How have you been sleeping over the last week?"

The interview guide was developed and tested for all 21 items of Hamilton's original scale, although Hamilton recommended that only the first 17 be used in calculating the total score.[1]

METHOD

To assess the effect of the use of the interview guide on the reliability of the individual items, a test-retest reliability study was conducted. Twenty-three patients (eight men, 15 women; mean age, 40 years) were selected from the inpatient services of the New York State Psychiatric Institute, New York. Patients were selected catch-as-catch-can on the basis of the availability of raters and patients. Since several of the services at the psychiatric institute specialize in specific diagnostic areas, many patients were included in the study whose primary areas of psychopathology were eating disorders and personality disorders. Although not its originally intended use,[1] the use of the HDRS in this study with patients whose primary complaint is not depression conforms to current use of the scale as a screening measure for severity of depression in the presence of other mental disorders.[14,15] The basic diagnostic data for this sample of patients are as follows:

DSM-III Diagnoses	No.
Alcohol hallucinosis and dependence	1
Schizophrenia	3
Schizoaffective	1
Paranoid disorder	1
Bipolar disorder, depressed	2
Major depression	7
Atypical depression	1
Bulimia	6
No axis I disorder (borderline personality disorder)	1

Nine raters participated in this study: four psychiatrists, three senior psychiatric social workers (I participated in nine of the paired interviews), and two psychologists, both students in doctorate programs. All raters received 1½ hours of didactic training that included practice rating and discussion of a demonstration videotaped interview, plus the supervised administration of the interview to at least one patient.

All test and retest interviews were conducted within four days of each other by clinicians blind to the complementary interview. Each set of interviews was followed within three days of the final interview by a discussion between the two raters to determine sources of disagreement. Of course, no individual ratings were changed on the basis of such discussion, even if it became clear that one rater had made a clerical error.

RESULTS

The average SIGH-D scores given by the "test" interviewers were 13.5 for the 17-item version and 16.5 for the 21-item scale; the corresponding retest scores were 12.5 and 15.1.

The item reliabilities obtained in this study are presented in the Table. Also presented for comparison are the reliabilities obtained in the Cicchetti and Prusoff study described above.[11] As can be seen, nearly all of the SIGH-D item reliabilities are higher than those obtained in the Cicchetti and Prusoff study, in which an interview guide was not used. It is most appropriate to compare the SIGH-D results on inpatients with those that Cicchetti and Prusoff obtained at randomization into their drug trial, since that group would be more acutely ill than at the end of the trial. This comparison reveals that all but three (late insomnia and psychomotor retardation and agitation) of the 21 SIGH-D items show better agreement. Compared with the Cicchetti and Prusoff results obtained at the end of the treatment period, only four (work and activities, late insomnia, psychomotor agitation, and diurnal variation) of the 21 SIGH-D items had a lower degree of reliability.

Of the 21 SIGH-D items tested, 12 showed good reliability ($R = .6$ or above). Of the remaining items, only two (work and activities and hypochondriasis) had adequate variance in this sample to determine reliability. This lack of variance is undoubtedly due to the limitation of the subjects in this study to psychiatric inpatients. The total scores of both the 17-item and 21-item versions of the SIGH-D show excellent reliability, comparable with that found by Cicchetti and Prusoff.

The HDRS scale alone requires at least a half hour to administer.[16] Raters in this study were asked to note the amount of time each SIGH-D interview took. The average amount of time was 28 minutes, indicating that the use of the SIGH-D does not increase the amount of time necessary to administer the scale over routine use.

COMMENT

This study demonstrates that the use of a structured interview guide for the HDRS results in generally increased reliability at the item level. This is similar to what was demonstrated by Endicott et al[8] in a comparison of agreement obtained by joint interviews on items from the Schedule for Affective Disorders and Schizophrenia–Change Version that were similar to the HDRS items and agreement on items from the actual HDRS scale. That study also demonstrated better agreement using the structured interview guide, although it was not developed specifically for rating the HDRS items.

All but two of the raters in the SIGH-D study had not had any experience with the HDRS prior to this study. The increase in item reliability with the SIGH-D is all the more impressive given the minimal training the interviewers received, the fact that they were of disparate backgrounds, and that for the most part they had not previously worked together. This suggests that the SIGH-D may be an efficient way to incorporate new raters quickly into a research program that uses the HDRS, without sacrificing reliability. This study also demonstrates that the use of this interview guide does not increase the amount of time necessary to administer the scale over routine clinical use.

Ideally, this study would have compared the test-retest reliability of the HDRS as usually administered (ie, without an interview guide) with the test-retest reliability of the SIGH-D on the same sample of subjects. However, such a study design would have involved administering the scale four times to each subject, a plan that is obviously not without its own logistic and scientific drawbacks. For this reason, the Cicchetti and Prusoff study was used as the comparison measure, with the recognition that since it involved a different sample of subjects, it is not the ideal control group.

Many critiques of the HDRS have cited difficulties with the items, ranging from lack of specificity of the item descriptions to poor discriminative validity of the individual items.[11,17] Unfortunately, although they were improved, in general the item reliabilities even using the SIGH-D were still not what one would hope. Only half of them were in the excellent to good range, with the rest ranging from fair to poor. This study was not designed to improve on the HDRS, but rather to improve the reliability of the original scale items. Future efforts must be devoted to improving the scale itself or developing a new one.

This study was supported in part by Biomedical Research Support grant 903-E759S, from the Research Foundation for Mental Hygiene Inc.

Miriam Gibbon, MSW, Mary Sano, PhD, Joan Carrillo, MS, Jo Ellen Loth, MSW, Robert Aranow, MD, Michael First, MD, Lawrence Jacobsberg, MD, and Ronald Winchell, MD, rated the patients. Raymond Henriques assisted with the research. Mark Davies was a statistical consultant. Robert Spitzer, MD, critically reviewed the data and manuscript.

References

1. Hamilton M: A Rating Scale for Depression. *J Neurol Neurosurg Psychiatry* 1960;23:56-62.
2. Hedlund JL, Vieweg BW: The Hamilton Rating Scale for Depression: A comprehensive review. *J Operational Psychiatry* 1979;10:149-165.
3. Carroll BJ, Fielding JM, Blashki TG: Depression rating scales: A critical review. *Arch Gen Psychiatry* 1973;28:361-366.
4. Rehm LP, O'Hara MW: Item characteristics of the Hamilton Rating Scale for Depression. *J Psychiatr Res* 1985;19:31-41.
5. Prusoff BA, Weissman M, Tanner J, Lieb J: Symptom reduction in depressed outpatients treated with amitriptyline or maprotiline: Repeated measurement analysis. *Compr Psychiatry* 1976;17:749-754.
6. Bech P, Allerup P, Reisby N, Gram LF: Assessment of symptoms change from improvement curves on the Hamilton depression scale in trials with antidepressants. *Psychopharmacology* 1984;84:276-281.
7. Zimmerman M, Coryell W, Pfohl B, Stangl D: Validity of the Hamilton endogenous subscale: An independent replication. *Psychiatry Res* 1986;18:209-215.
8. Endicott J, Cohen J, Nee J, Fleiss J, Sarantakos S: Hamilton Depression Rating Scale: Extracted from regular and change versions of the Schedule for Affective Disorders and Schizophrenia. *Arch Gen Psychiatry* 1981;38:98-103.
9. Ziegler VE, Meyer DA, Rosen SH, Biggs JT: Reliability of video taped Hamilton ratings. *Biol Psychiatry* 1978;13:119-122.
10. Spitzer RL, Williams JBW: Classification in psychiatry, in Kaplan H, Sadock B (eds): *Comprehensive Textbook of Psychiatry*, ed 4. Baltimore, Williams & Wilkins, 1985, pp 591-613.
11. Cicchetti DV, Prusoff BA: Reliability of depression and associated clinical symptoms. *Arch Gen Psychiatry* 1983;40:987-990.
12. Endicott J, Spitzer RL: A diagnostic interview: The Schedule for Affective Disorders and Schizophrenia. *Arch Gen Psychiatry* 1978;35:837-844.
13. Guy W (ed): *ECDEU Assessment Manual for Psychopharmacology*, publication No. ADM 76-336. Rockville, Md, US Dept of Health, Education, and Welfare, 1976.
14. Halmi KA, Eckert E, LaDu TJ, Cohen J: Anorexia nervosa: Treatment efficacy of cyproheptadine and amitriptyline. *Arch Gen Psychiatry* 1986;43:177-181.
15. Langer G, Koinig G, Hatzinger R, Schonbeck G, Resch F, Aschauer H, Keshavan MS, Sieghart W: Response of thyrotropin to thyrotropin-releasing hormone as predictor of treatment outcome: Prediction of recovery and relapse in treatment with antidepressants and neuroleptics. *Arch Gen Psychiatry* 1986;43:861-868.
16. Hamilton M: Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 1967;6:278-296.
17. Miller IW, Bishop S, Norman WH, Maddever H: The Modified Hamilton Rating Scale for Depression: Reliability and validity. *Psychiatry Res* 1985;14:131-142.