BJPsych The British Journal of Psychiatry (2008) 192, 52–58. doi: 10.1192/bjp.bp.106.032532

Development and reliability of a structured interview guide for the Montgomery–Åsberg Depression Rating Scale (SIGMA)

Janet B. W. Williams and Kenneth A. Kobak

Background
The Montgomery–Åsberg Depression Rating Scale (MADRS) is often used in clinical trials to select patients and to assess treatment efficacy. The scale was originally published without suggested questions for clinicians to use in gathering the information necessary to rate the items. Structured and semi-structured interview guides have been found to improve reliability with other scales.

Aims
To describe the development and test–retest reliability of a structured interview guide for the MADRS (SIGMA).

Method
A total of 162 test–retest interviews were conducted by 81 rater pairs. Each patient was interviewed twice, once by each rater conducting an independent interview.

Results
The intraclass correlation for total score between raters using the SIGMA was r=0.93, P<0.0001. All ten items had good to excellent interrater reliability.

Conclusions
Use of the SIGMA can result in high reliability of MADRS scores in evaluating patients with depression.

Declaration of interest
None. Funding detailed in Acknowledgements.

Observer-rated depression rating scales are used in clinical trials of antidepressants to select patients for study, and to assess the efficacy of the treatment being tested. The importance of the quality of ratings in clinical trials has recently been emphasised.[1] Accumulating evidence suggests that the quality of ratings can make the difference between a failed trial and one in which drug separates from placebo.[2] Therefore, any method that improves the quality of clinical trial ratings may improve our ability to conduct successful antidepressant trials. We describe the development of a structured interview guide for the Montgomery–Åsberg Depression Rating Scale (MADRS) and its test–retest reliability, assessed in a sample of 51 persons with varying levels of depressive symptoms.

Montgomery–Åsberg Depression Rating Scale

The MADRS was developed in the late 1970s from items that were found in several studies to be sensitive to change with antidepressant treatment.[3] Since its publication the scale has become increasingly popular worldwide. Dissatisfaction with the leading alternative, the Hamilton Rating Scale for Depression (HRSD), has further contributed to the popularity of the MADRS.[4]

The importance of reliability of assessments in a clinical trial cannot be overestimated. Without good interrater agreement the chances of detecting a difference in effect between drug and placebo are significantly reduced. Muller & Szegedi demonstrated that as the reliability of a rating scale decreases from 0.8 to 0.5, the power of the test to detect a significant difference between drug and placebo drops from 71% to 51%, increasing the risk of type II error.[5] In general, total scale score reliability of the most commonly used depression rating scales such as the MADRS and HRSD is high, with or without the use of a structured interview guide.[3,6] However, as compounds have become targeted to specific symptoms and clinical trials have revealed specific drugs' effects on clusters of symptoms,[7,8] it has become more important to be able to depend on the reliable measurement of an individual symptom or a subgroup of symptoms. Self-report versions of clinician-administered scales have been developed[9,10] that show high degrees of correlation with the clinician versions; however, there are limited empirical data on their signal detection relative to the clinician in placebo-controlled trials.

There is minimal information available concerning the interrater reliability of the MADRS. The original article[4] reported excellent agreement between rater pairs, but only as 'conjoint' reliability, and in only 11 patients. Maier *et al* reported total score intraclass coefficients (ICCs) of 0.73, 0.66 and 0.82 in three separate samples, using joint interviews in the first sample, and independent interviews in the second and third samples (which were actually the same patients pre- and post-treatment).[11] Unfortunately item reliability was not provided, although the authors did report that three of the MADRS items (inner tension, lassitude and suicidal thoughts) had ICCs lower than 0.60 in all three samples. Davidson *et al* tested the reliability of the MADRS in 44 people receiving in-patient treatment for depression, using an experienced research nurse and a psychiatrist as joint interviewers.[12] Overall agreement was 'acceptable' and ranged from 'fair' to 'good' on individual items. More recently, a Japanese version was developed and tested in Japan in joint interviews on a sample of seven patients with DSM–IV major depressive disorder.[13] Individual-item ICCs were in the 'very good' to 'excellent' range; however, the weakness of the testing method (small sample size and repeated assessment of the same patients by the same three raters in joint interviews) compromised the significance of the results. Therefore, there is reason to believe that the interrater reliability of the MADRS in a typical research study could be improved.

The MADRS was originally published without suggested questions for clinicians to use in gathering the information necessary to rate the ten items. However, several studies have found that using a structured or semi-structured interview guide improves

reliability on similar rating scales.[14,15] Moberg *et al* compared independent interviews using the standard unstructured HRSD with the Structured Interview Guide for the HRSD (SIGH–D) and concluded that the SIGH–D 'produced uniformly higher item- and summary scale reliabilities than the unstructured HDRS'.[16] Further, in one placebo-controlled antidepressant trial, raters who more closely adhered to a semi-structured interview guide were found to have better signal detection than raters who did not.[2] Such an interview guide provides some assurance that raters across clinical trial sites administer the scale in approximately the same way. Structured interview guides also facilitate training in the use of a scale by providing new raters with explicit instructions and specific interview questions that have been derived from expert interviews. Structured interview guides have become fairly standard for diagnostic interviews,[16] as well as for many rating scales, including the Hamilton scales for depression (SIGH–D)[14] and anxiety (SIGH–A).[17,18] In general, they are designed to approximate an expert administration of the scale.

Development of SIGMA probes and conventions

A semi-structured interview guide similar to the SIGH–D was originally developed by J.B.W.W. for the MADRS in 1988 and has undergone several revisions since then, based on user experience and feedback from raters. More recently, K.A.K. joined as co-author in a major overhaul of the interview guide. The Structured Interview Guide for the MADRS (SIGMA) provides structured probes to ensure standardisation of administration and comprehensiveness of coverage of the ten items of the scale. The SIGMA questions were developed to obtain the information needed to assess each of the items' anchor points (see Appendix). Each item begins with questions in bold type that should be asked exactly as written. Often these questions will elicit enough information about the severity and frequency of a symptom for the item to be rated 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 the rater has enough information to rate the item confidently. Raters are also encouraged to add their own probes as necessary to obtain enough information to rate each item accurately.

In the SIGMA the original MADRS appears on the right-hand side of the page and the structured interview guide questions appear on the left. The interview guide begins with the 'overview', which is a brief explanation of the time period to be covered, and initial questions to allow some rapport to develop and to give the interviewer some sense of the context of the interviewee's current situation. The interview guide then follows, with questions for each of the ten MADRS items.

In the SIGMA the only change that was made to the original MADRS was to reverse the order of administration of the first two items (apparent sadness and reported sadness). There was consensus from users that asking about reported sadness first made for a more logical flow to the interview. Direct probes were added to the apparent sadness item to supplement the rater's observation (e.g. 'In the past week, do you think you have looked sad or depressed to other people?) The rationale for these additional probes was that without the aid of an informant who has seen the patient over the past week it is difficult to rate the persistence and depth of this item based on observation during the interview alone. This technique has been used successfully in self-report and telephone-administered versions of the MADRS,[10,19] as well as in computerised and paper-and-pencil self-report versions of the HRSD[20,21] and the Hamilton Rating Scale for Anxiety.[22] Raters are instructed to consider both sources of information (direct observation and self-report) in rating this item.

In the interview guide there is an emphasis on open-ended questions, to encourage respondents to describe their experience in their own words, and to avoid raters' 'putting words in the person's mouth'. Thus, for example, instead of asking the person at the beginning of the interview, 'Have you been feeling sad or unhappy?', the enquiry begins, 'How have you been feeling since last [day of week]?' Likewise, instead of asking whether the person has had trouble sleeping in the past week, the sleep item begins, 'How has your sleeping been in the past week?' Some items are assessed more directly, to improve the efficiency of the interview. For many responses the person is asked to provide examples; for instance, if there is a positive response to the question, 'Have you had trouble concentrating or collecting your thoughts in the past week?' the interviewer is instructed to ask, 'Can you give me some examples?' Once the person has described the symptom in his or her own words, the interviewer can then decide whether concentration difficulty is truly present, which would be rated in this item.

Once the revised SIGMA was completely drafted, revisions were made based on feedback from a number of users in the field, and the instrument was finalised. This report describes a formal assessment of the test–retest reliability of the 2006 version of the SIGMA, which is presented in full in the Appendix.

Method

To test the interrater reliability of the SIGMA, 162 test–retest interviews (81 dyads) were conducted. Within each dyad each patient was interviewed twice, once by each rater. Interviews were conducted independently from each other, with raters masked to the results of the other rater's interview. Conducting independent (*v.* joint) evaluations of the same patient mitigates the artificial inflation of reliability coefficients that occurs when one rater interviews the patient and the second rater simply observes the first rater's interview. Conducting independent evaluations is thus a more rigorous test of interrater reliability, and addresses the question of whether you would achieve the same result if the person were interviewed by a different rater using the same instrument. Interviews were conducted on the same day, in order to control for changes in participants'conditions due to time. A brief (5–15 min) distracting task was given between interviews to minimise memory effects.

There is growing interest in the use of remote technologies for delivering assessments in clinical trials.[23] Therefore, of the 81 pairs of interviews, 30 pairs were done using two face-to-face interviews, 30 pairs were done using one face-to-face and one videoconference interview, and 21 pairs were done using one face-to-face and one telephone interview. To control for the confounding influence of participant differences on mode of administration, the same 30 people were used in the 'face-to-face *v.* face-to-face' and the 'video *v.* face-to-face' cohorts. To minimise memory effects these cohorts were interviewed on different days (1–3 days apart) and no rater ever saw the same patient twice.

Fifty-one participants (14 men and 37 women; mean age 43 years, s.d.=12.35, range 20–72) with a mood disorder diagnosed according to DSM–IV criteria were included.[24] The diagnoses were major depression, n=27; major depression in partial remission, n=15; minor depression, n=2; dysthymia, n=1; bipolar disorder (depressed), n=2; and depression not otherwise specified, n=4. Diagnoses were determined using a modified version of the mood module of the Mini International Neuropsychiatric Interview (MINI)[25] and the overview section of the Structured Clinical Interview for DSM–IV (SCID).[16] Since previous versions

https://doi.org/10.1192/bjp.bp.106.032532 Published online by Cambridge University Press

of the SIGMA are already widely used in clinical trials, a range of mood disorders was included in order to evaluate the reliability of the SIGMA across a wide range of symptom severity, including patients in partial recovery. The sample was 82% White, 10% African American, 2% Hispanic and 6% 'other'. About half (47%) had a college degree. Participants were recruited from the Madison, Wisconsin area in response to advertisements in a local newspaper looking for people who were currently or had recently experienced symptoms of depression. Respondents were screened over the telephone by a research assistant for gross exclusions (current or lifetime schizophrenia, current psychosis or acute suicidal ideation) and were then scheduled for further follow-up screening with a clinician. All participants signed informed consent statements approved by the Allendale institutional review board, and were paid US $50.

The rater cohort consisted of two male and four female interviewers. Five had doctoral degrees (four in psychology and one in social work), and one a master's degree in counselling psychology. Prior to the study, raters underwent reliability training on the scale, consisting of a didactic review of the scale and scoring conventions, followed by at least three practice interviews observed by a trainer (two group and one individual). Raters also observed each other's training sessions in order to enhance learning. Raters had a range of prior experience with the MADRS, ranging from extensive (J.B.W.W., K.A.K.) to minimal. All raters' interviewing skills were evaluated using the Raters Applied Performance Scale (RAPS)[26] and all raters scored at least 'good' on all dimensions before the study began. Raters were paired using numerous permutations of dyads, to maximise generalisability. Order was counterbalanced, so that raters conducted an equal number of first and second interviews.

Results

The sample represented a wide range of depression severity, with scores on the MADRS ranging from 0 to 38. Overall, the level of depression in the sample was moderately severely (mean MADRS score 20.5, s.d.=10.35). Half the sample had a MADRS score over 22.

The intraclass correlation for total score between raters conducting MADRS interviews using the SIGMA was r=0.93 (P<0.0001, 95% CI 0.89–0 .95). In addition, there was no significant difference between the mean MADRS scores obtained by the first interviewer and the second interviewer: 20.49 (s.d.=10.5) *v.* 20.65 (s.d.=10.6); mean difference 0.16 points ($t_{(80)}$=0.36, P=0.72). Internal consistency reliability (coefficient alpha) was also examined. Similar levels of internal consistency were found for interviews administered by the first interviewer (r=0.90) and those done by the second interviewer (r=0.90), z=–0.19, P=0.85; the 95% confidence interval for the difference ($d_{(r)}$=0.0057) was –0.50 to 0.12. An examination of the interrater reliability (ICC) at the item level is presented in Table 1. As Blacker & Endicott have indicated, 'it is sometimes said that an [ICC] above 0.8 can be considered excellent, 0.7–0.8 good, 0.5–0.7 fair, and less than 0.5 poor' (page 9).[27] All of the ten MADRS items using the SIGMA had good to excellent interrater reliability, with more than half of them in the excellent range, as was the total score.

The correlation (ICC) between the SIGMAs administered by videoconference and those administered face-to-face was r=0.95 (P<0.0001, 95% CI 0.89–0.97). The correlation between SIGMAs administered by telephone and those given face-to-face was r=0.90 (P<0.0001, 95% CI 0.78–0.96) and that between the 30 pairs of face-to-face interviews was r=0.93 (P<0.0001, 95% CI 0.87–0.97). There was no statistically significant difference among these correlations: z=0.45, P=0.66, 95% CI for the difference ($d_{(r)}$=0.0141) −0.09 to 0.98 for two face-to-face *v.* video and face-to-face; z=0.07, P=0.95, 95% CI for the difference ($d_{(r)}$=0.0048) –0.59 to 0.729 for two face-to-face *v.* telephone and face-to-face.

We also took this opportunity to examine the diagnostic accuracy of the MADRS in differentiating major depression from the other diagnoses. Using a cut-off score of 18 on the MADRS, the sensitivity of the scale for the diagnosis of major depression was 87%, its specificity was 61%, the positive predictive value was 74% and the negative predictive value 79%.

Interview length

The mean length of interviews conducted with the SIGMA was 25.8 min (s.d.=10.04, range 5–56). The mean interview length for the interview conducted second was 2.4 min shorter than the mean length of the interview conducted first (26.9 min *v.* 24.5 min; $t_{(78)}$=2.38, P=0.020).

Discussion

In our development and testing of a structured interview guide for the MADRS, the most stringent test of interrater reliability was

Table 1 Intraclass correlations between raters on individual items of the Montgomery–Åsberg Depression Rating Scale, by mode of administration

	Intraclass correlation[a]			
	All modes of administration[b] (n=81)	Telephone *v.* face-to-face (n=21)	Video *v.* face-to-face (n=30)	Face-to-face *v.* face-to-face (n=30)
Apparent sadness	0.82	0.80	0.80	0.84
Reported sadness	0.93	0.84	0.84	0.86
Inner tension	0.75	0.80	0.72	0.75
Reduced sleep	0.86	0.77	0.86	0.94
Reduced appetite	0.85	0.83	0.86	0.87
Concentration difficulties	0.85	0.68	0.85	0.95
Lassitude	0.74	0.67	0.72	0.82
Inability to feel	0.79	0.86	0.79	0.76
Pessimistic thoughts	0.71	0.61	0.69	0.78
Suicidal thoughts	0.89	0.92	0.86	0.92
Total	0.93	0.90	0.95	0.93

a. All correlations significant at P<0.001.
b. Independent interviews, six raters.

https://doi.org/10.1192/bjp.bp.106.032532 Published online by Cambridge University Press

used: independent test–retest interviews. This method best approximates the interrater agreement that would be achieved in many clinical trials in which patients are assessed by a different rater at each visit. A strength of our study is that six different raters participated. This suggests that the positive results were not due to a single pair of raters who work closely together; rather, these results are likely to be generalisable to a pool of similarly trained raters. Although the training was fairly rigorous, it is replicable. All raters in this study were experienced mental health clinicians, although they had varying degrees of prior experience with the MADRS scale. The level of agreement that can be achieved with raters who have less clinical experience is unknown, although it is likely that the structured interview guide facilitates the reliable use of this scale by less-experienced raters because it standardises the questions used to elicit the information necessary to rate the items.

To our knowledge, this is the first assessment of the reliability of the MADRS in which all test–retest interviews were independent, and in which agreement at the item level was reported. In this study agreement on the total MADRS score was in the excellent range, and the reliability of all ten of the items was good to excellent. Our results also support the equivalence of remote administration of the MADRS using the SIGMA, by both telephone and videoconference, to face-to-face interviews. This finding has important implications for the way in which clinical trial assessments are conducted: remote assessments can offer more efficient and flexible administration paradigms than face-to-face assessments.

This study has demonstrated that with the use of the SIGMA, a group of interchangeable raters can achieve high reliability of MADRS total and item scores in a range of patients with depression. The extent to which this improvement in interrater agreement translates into improved signal detection in trials using the MADRS remains to be demonstrated.

Janet B. W. Williams, DSW, Biometrics Research Unit, Columbia University, New York, New York; **Kenneth A. Kobak**, PhD, MedAvante, Inc., Madison, Wisconsin, USA

Correspondence: Janet B. W. Williams, MedAvante, Inc., 100 American Metro Blvd., Suite 106, Hamilton, NJ 08619, USA. Email: jwilliams@medavante.net

First received 18 October 2006; final revision 28 February 2007; accepted 13 March 2007

Acknowledgements

The authors would like to acknowledge the contributions of Drs Elizabeth Jeglic, Ian Sharp and Karen Rybechenko, and Donna Salvucci, MEd, for conducting the clinical interviews in the study and for feedback on the items; Drs Alan Feiger and Norm Rosenthal for suggestions on the probes; Dr Kia Crittenden and Kristy Harris, MA, for patient screening; and Don Borque for assistance with study coordination. Support for this study was provided by MedAvante, Inc.

Appendix

Structured Interview Guide for the Montgomery–Åsberg Depression Rating Scale (SIGMA)

Interviewer

The questions in bold type for each item should be asked exactly as written. Often these questions 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. Note that questions in parentheses are optional, for instance if information is unknown.

Notes

Time period. The ratings should be based on the patient's condition in the past week.

Change from baseline. In general, a symptom is rated as present only when it reflects a change from before the depression began. The interviewer should try to identify a 2-month period of non-depressed functioning and use this as a reference point. In some cases, such as when the patient has dysthymia, the referent should be to the last time the person felt all right (i.e. not depressed or high) for at least a few weeks.

This interview guide is based on the Montgomery–Åsberg Depression Rating Scale (MADRS) (Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979; **134**: 382–9). The scale itself has been retained in its original form, except for reversing the order of the first two items. This guide adds interview questions to aid in the assessment and application of the MADRS. Previous versions of this guide appeared in 1988, 1992, 1996 and 2005.

https://doi.org/10.1192/bjp.bp.106.032532 Published online by Cambridge University Press

Structured Interview Guide for the Montgomery–Åsberg Depression Rating Scale (SIGMA)

PT'S INITIALS: _____ PT'S ID: _____ INTERVIEWER: _____ TIME BEGAN SIGMA: _____ DATE: _____

OVERVIEW: I'd like to ask you some questions about the past week. How have you been feeling since last (DAY OF WEEK)? IF OUT-PATIENT: Have you been working? (What kind of work do you do?) IF NOT: Why not?

In the past week, have you been feeling sad or unhappy?
(Depressed at all?) IF YES: Can you describe what this has been like for you?
(IF UNKNOWN: How bad has that been?)
IF DEPRESSED: Does the feeling lift at all if something good happens
How much does your mood lift? Does the feeling ever go away completely?
(What things have made you feel better?)
How often did you feel (depressed/OWN EQUIVALENT) this past week?
(IF UNKNOWN: How many days this week did you feel that way? How much of each day?)
In the past week, how have you been feeling about the future? (Have you been discouraged or pessimistic?) What have your thoughts been? How (discouraged or pessimistic) have you been? How often have you felt that way? Do you think things will ever get better for you?
IF ACKNOWLEDGES DEPRESSED MOOD, TO GET CONTEXT ASK:
How long have you been feeling this way?

1. REPORTED SADNESS. Representing reports of depressed mood, regardless of whether it is reflected in appearance or not. Includes low spirits, despondency or the feeling of being beyond help and without hope. Rate according to intensity, duration and the extent to which the mood is reported to be influenced by events.

0 Occasional sadness in keeping with the circumstances
1
2 Sad or low but brightens up without difficulty
3
4 Pervasive feelings of sadness or gloominess. The mood is still influenced by external circumstances
5
6 Continuous or unvarying sadness, misery or despondency

RATING BASED ON OBSERVATION DURING INTERVIEW AND THE FOLLOWING QUESTIONS.
In the past week, do you think you have looked sad or depressed to other people? Did anyone say you looked sad or down?
How about when you've looked in the mirror? Did you look gloomy or depressed?
IF YES: How sad or depressed do you think you have looked?
How much of the time over the past week do you think you have looked depressed or down?
IF APPEARANCE WAS DEPRESSED IN PAST WEEK: Have you been able to laugh or smile at all during the past week? IF YES: How hard has it been for you to laugh or smile, even if you weren't feeling happy inside?

2. APPARENT SADNESS. Representing despondency, gloom and despair. (More than just ordinary transient low spirits) reflected in speech, facial expressions and posture. Rate by depth and inability to brighten up.

0 No sadness
1
2 Looks dispirited but does brighten up without difficulty
3
4 Appears sad and unhappy most of the time
5
6 Looks miserable all the time. Extremely despondent

Have you felt tense or edgy in the past week? Have you felt anxious or nervous?
IF YES: Can you describe what that has been like for you? How bad has it been? (Have you felt panicky?)
What about feeling fearful that something bad is about to happen?
How hard has it been to control these feelings? (What has it taken to help you feel calmer? Has anything worked to calm you down?)
How much of the time have you felt this way over the past week?

3. INNER TENSION. Representing feelings of ill-defined discomfort, edginess, inner turmoil, mental tension mounting to either panic, dread or anguish. Rate according to intensity, frequency, duration and the extent of reassurance called for.

0 Placid. Only fleeting inner tension
1
2 Occasional feelings of edginess and ill-defined discomfort
3
4 Continuous feelings of inner tension or intermittent panic which the patient can master with some difficulty
5
6 Unrelenting dread or anguish. Overwhelming panic

How has your sleeping been in the past week? (How many hours have you been sleeping, compared with usual?)

Have you had trouble falling asleep? (How long has it been taking you to fall asleep this past week?)

Have you been able to stay asleep through the night? (Have you been waking up at all in the middle of the night? How long does it take you to go back to sleep?)

Has your sleeping been restless or disturbed?

4. REDUCED SLEEP. Representing the experience of reduced duration or depth of sleep compared to the subject's own normal pattern when well.

0 Sleeps as usual
1
2 Slight difficulty dropping off to sleep or slightly reduced, light, or fitful sleep
3
4 Sleep reduced or broken by at least 2 hours
5
6 Less than 2 or 3 hours sleep

How has your appetite been this past week?
(What about compared with your usual appetite?)
Have you been less interested in food? (How much less?)
Does food taste as good as usual? IF LESS: How much less?
Have you had to force yourself to eat?
Have other people had to urge you to eat?

5. REDUCED APPETITE. Representing the feeling of a loss of appetite compared with when well. Rate by loss of desire for food or the need to force oneself to eat.

0 Normal or increased appetite
1
2 Slightly reduced appetite
3
4 No appetite. Food is tasteless
5
6 Needs persuasion to eat at all

https://doi.org/10.1192/bjp.bp.106.032532 Published online by Cambridge University Press

Appendix (continued)

Have you had trouble concentrating or collecting your thoughts in the past week? (How about at home or at work?) IF YES: Can you give me some examples? (Have you been able to concentrate on reading a newspaper or magazine? Do you need to read things over and over again?)

How often has that happened in the past week? Has this caused any problems for you? IF YES: Can you give me some examples?

Has your trouble concentrating been so bad at any time in the past week that it has been difficult to follow a conversation? (IF YES: How bad has that been? How often has that happened this past week?)

NOTE: ALSO CONSIDER BEHAVIOUR DURING INTERVIEW.

6. CONCENTRATION DIFFICULTIES. Representing difficulties in collecting one's thoughts mounting to incapacitating lack of concentration. Rate according to intensity, frequency, and degree of incapacity produced.

0 No difficulties in concentration
1
2 Occasional difficulties in collecting one's thoughts
3
4 Difficulties in concentrating and sustaining thought which reduces ability to read or hold a conversation
5
6 Unable to read or converse without great difficulty

Have you had any trouble getting started at things in the past week? IF YES: What things?

Have you had to push yourself to do things? IF YES: What things? How hard have you had to push yourself? Are you OK once you get started or is it still more of an effort to get something done? What about getting started at simple routine everyday things (like getting dressed)?

Have you done everyday things more slowly than usual? (Have you been sluggish?) IF YES: Like what, for example? How bad has that been?

7. LASSITUDE. Representing a difficulty getting started, or slowness initiating and performing everyday activities.

0 Hardly any difficulty in getting started. No sluggishness
1
2 Difficulties in starting activities
3
4 Difficulties in simple routine activities, which are carried out with effort
5
6 Complete lassitude. Unable to do anything without help

Have you been less interested in things around you, or in activities you used to enjoy? IF YES: What things? How bad has that been? How much less interested in (those things) are you now compared with before?

Have you been less able to enjoy the things you usually enjoy? Has there been any change in your ability to feel emotions? (Do you feel things less intensely than you used to, things like anger, grief, pleasure?) IF YES: Can you tell me more about that? (IF UNKNOWN: Are you able to feel any emotions at all?)

How do you feel towards your family and friends? Is that different from usual? IF REDUCED: Do you feel less than you used to towards them?

8. INABILITY TO FEEL. Representing the subjective experience of reduced interest in the surroundings, or activities that normally give pleasure. The ability to react with adequate emotion to circumstances or people is reduced.

0 Normal interest in the surroundings and in other people
1
2 Reduced ability to enjoy usual interests
3
4 Loss of interest in the surroundings. Loss of feelings for friends and acquaintances
5
6 The experience of being emotionally paralysed, inability to feel anger, grief or pleasure, and a complete or even painful failure to feel for close relatives and friends

Have you been putting yourself down, or feeling that you're a failure in some way, over the past week? (Have you been blaming yourself for things that you've done, or not done?) IF YES: What have your thoughts been? How often have you felt that way?

Have you been feeling guilty about anything in the past week? What about feeling as if you have done something bad or sinful? IF YES: What have your thoughts been? How often have you felt that way?

ALSO CONSIDER RESPONSES TO QUESTIONS ABOUT PESSIMISM FROM ITEM 1.

9. PESSIMISTIC THOUGHTS. Representing thoughts of guilt, inferiority, self-reproach, sinfulness, remorse and ruin.

0 No pessimistic thoughts
1
2 Fluctuating ideas of failure, self-reproach or self-depreciation
3
4 Persistent self-accusations, or definite but still rational ideas of guilt or sin. Increasingly pessimistic about the future
5
6 Delusions of ruin, remorse, or unredeemable sin. Self-accusations which are absurd and unshakeable

This past week, have you felt like life isn't worth living? IF YES: Tell me about that. IF NO: What about feeling as if you're tired of living?

This week, have you thought that you would be better off dead? IF YES: Tell me about that.

Have you had thoughts of hurting or even killing yourself this past week? IF YES: What have you thought about? How often have you had these thoughts? How long have they lasted? Have you actually made plans? IF YES: What are these plans? Have you made any preparations to carry out these plans? (Have you told anyone about it?)

10. SUICIDAL THOUGHTS. Representing the feeling that life is not worth living, that a natural death would be welcome, suicidal thoughts, and preparation for suicide. Suicidal attempts should not in themselves influence this rating.

0 Enjoys life or takes it as it comes
1
2 Weary of life. Only fleeting suicidal thoughts
3
4 Probably better off dead. Suicidal thoughts are common, and suicide is considered as a possible solution, but without specific plans or intention
5
6 Explicit plans for suicide when there is an opportunity. Active preparations for suicide

TOTAL MADRS SCALE SCORE: _____

https://doi.org/10.1192/bjp.bp.106.032532 Published online by Cambridge University Press

References

1 Kobak KA, Engelhardt N, Williams JB, Lipsitz JD. Rater training in multicenter clinical trials: issues and recommendations. *J Clin Psychopharmacol* 2004; **24:** 113–17.

2 Kobak KA, Feiger AD, Lipsitz JD. Interview quality and signal detection in clinical trials. *Am J Psychiatry* 2005; **162**: 628.

3 Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979; **134**: 382–9.

4 Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am J Psychiatry* 2004; **161**: 2163–77.

5 Muller MJ, Szegedi A. Effects of interrater reliability of psychopathologic assessment on power and sample size calculations in clinical trials. *J Clin Psychopharmacol* 2002; **22**: 318–25.

6 Hedlund JL, Vieweg BW. The Hamilton Rating Scale for Depression: a comprehensive review. *J Oper Psychiatry* 1979; **10**: 149–61.

7 Gibbons RD, Clark DC, Kupfer DJ. Exactly what does the Hamilton Depression Rating Scale measure? *J Psychiatr Res* 1993; **27**: 259–73.

8 Faries D, Herrera J, Rayamajhi J, DeBrota D, Demitrack M, Potter WZ. The responsiveness of the Hamilton Depression Rating Scale. *J Psychiatr Res* 2000; **34**: 3–10.

9 Kobak KA, Greist JH, Jefferson JW, Mundt JC, Katzelnick DJ. Computerized assessment of depression and anxiety over the telephone using interactive voice response. *MD Comput* 1999; **16**: 64–8.

10 Mundt JC, Katzelnick DJ, Kennedy SH, Eisfeld BS, Bouffard BB, Greist JH. Validation of an IVRS version of the MADRS. *J Psychiatr Res* 2006; **40**: 243–6.

11 Maier W, Philipp M, Heuser I, Schlegel S, Buller R, Wetzel H. Improving depression severity assessment: I. reliability, internal validity and sensitivity to change of three observer depression scales. *J Psychiatr Res* 1988; **22**: 3–12.

12 Davidson J, Turnbull CD, Strickland R, Miller R, Graves K. The Montgomery–Åsberg Depression Scale: reliability and validity. *Acta Psychiatr Scand* 1986; **73**: 544–8.

13 Takahashi N, Tomita K, Higuchi T, Inada T. The inter-rater reliability of the Japanese version of the Montgomery–Asberg depression rating scale (MADRS) using a structured interview guide for MADRS (SIGMA). *Hum Psychopharmacol* 2004; **19**: 187–92.

14 Williams JBW. A structured interview guide for the Hamilton Depression Rating Scale. *Arch Gen Psychiatry* 1988; **45**: 742–7.

15 Moberg PJ, Lazarus LW, Mesholam RI, Bilker W, Chuy IL, Neyman I, Markvart V. Comparison of the standard and structured interview guide for the Hamilton Depression Rating Scale in depressed geriatric inpatients. *Am J Geriatr Psychiatry* 2001; **9**: 35–40.

16 Spitzer RL, Williams JBW, Gibbon M, First MB. The Structured Clinical Interview for DSM–III–R (SCID). II. Multisite test–retest reliability. *Arch Gen Psychiatry* 1992; **49**: 630–6.

17 Shear MK, Vander Bilt J, Rucci P, Endicott J, Lydiard B, Otto MW, Pollack MH, Chandler L, Williams J, Ali A, Frank DM. Reliability and validity of a structured interview guide for the Hamilton Anxiety Rating Scale (SIGH–A). *Depress Anxiety* 2001; **13**: 166–78.

18 Williams JBW. *Structured Interview Guide for the Hamilton Anxiety Scale (SIGH–A).* New York State Psychiatric Institute, 1996.

19 Hermens ML, Ader HJ, van Hout HP, Terluin B, van Dyck R, de Haan M. Administering the MADRS by telephone or face-to-face: a validity study. *Ann Gen Psychiatry* 2006; **5**: 3.

20 Kobak KA, Reynolds WM, Rosenfeld R, Greist JH. Development and validation of a computer-administered version of the Hamilton Depression Rating Scale. *Psychol Assess* 1990; **2**: 56–63.

21 Reynolds WM, Kobak KA. Reliability and validity of the Hamilton Depression Inventory: a paper-and-pencil version of the Hamilton Depression Rating Scale clinical interview. *Psychol Assess* 1995; **7**: 472–83.

22 Kobak KA, Reynolds WM, Greist JH. Development and validation of a computer-administered version of the Hamilton Anxiety Scale. *Psychol Assess* 1993; **5**: 487–92.

23 Kobak KA, Kane JM, Thase ME, Nierenberg AA. Why do clinical trials fail? The problem of measurement error in clinical trials: time to test new paradigms? *J Clin Psychopharmacol* 2007; **27**: 1–5.

24 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* (4th edn) (DSM–IV). APA, 1994.

25 Sheehan DV, Lecrubier Y, Sheehan K, Amorim P, Janavs J, Weiller E, Baker R, Dunbar G. The mini international neuropsychiatric interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM–IV and ICD–10. *J Clin Psychiatry* 1998; **59**: 22–3.

26 Lipsitz J, Kobak K, Feiger A, Sikich D, Moroz G, Engelhard A. The Rater Applied Performance Scale: development and reliability. *Psychiatry Res* 2004; **127**: 147–55.

27 Blacker D, Endicott J. Psychometric properties: concepts of reliability and validity. In *Handbook of Psychiatric Measures* (ed. AJ Rush): 7–14. American Psychiatric Association, 2000.

100 words

Insight

Anthony S. David

Insight was once described as ‘academically nourishing but clinically sterile’. Yet few worthwhile discussions in clinical psychiatry omit consideration of insight. It is worthwhile because, in people with psychosis, it predicts clinical and functional outcome, coercion and capacity, mood and cognition. Some dismiss it as mere agreeing with the doctor; more ‘us and them’. Poet Robbie Burns (1759–1796) asks God to give us the gift ‘to see oursels as others see us’, ending denial of our imperfections and unwillingness to turn our gaze upon them. Thinking about insight demands we view ourselves and our flawed humanity critically. We are all a bit ‘us’ and ‘them’.

doi: 10.1192/bjp.192.1.58

https://doi.org/10.1192/bjp.bp.106.032532 Published online by Cambridge University Press