Validation of the Aphasic Depression Rating Scale

Charles Benaim, MD, PhD; Bruno Cailly, MD; Dominic Perennou, MD, PhD; Jacques Pelissier, MD

Background and Purpose—The Aphasic Depression Rating Scale (ADRS) was developed to detect and measure depression in aphasic patients during the subacute stage of stroke.

Methods—Six experts selected an initial sampling of behavioral items from existing depression rating scales. Stroke patients (aphasic and nonaphasic) were assessed with these items by the rehabilitation staff, with the Hamilton Depression Rating Scale (HDRS) for nonaphasic patients only, by a psychiatrist, and by the rehabilitation staff with Visual Analog Scales (VAS). A second item selection was conducted after a regression algorithm was run including VAS as independent variables (criterion validity) and after their factorial structure was analyzed with a principal component analysis (factorial validity). The construct validity was evaluated with respect to the other depression assessments. A threshold for the diagnosis of depression was computed with respect to the psychiatrist’s diagnosis. Interrater and test-retest reliability were assessed in 2 additional groups of aphasic patients.

Results—Eighty patients participated in the study (59 aphasic). Fifteen behavioral items from existing depression rating scales were selected, and 9 were retained after the validation process. ADRS correlated highly with VAS and HDRS (r=0.60 to 0.78, P=10^-4 to 10^-6). With respect to the psychiatrist’s diagnosis, the sensitivity and specificity of ADRS were 0.83 and 0.71, respectively, when the threshold was set at 9/32. Its factorial structure was comparable to HDRS structure. Interrater and test-retest reliability were high (average κ coefficient of the 9 items=0.69).

Conclusions—ADRS is a valid, reliable, sensitive, and specific tool for the evaluation of depression in aphasic patients during the stroke subacute phase. (Stroke. 2004;35:000-000.)

Key Words: depression ■ aphasia ■ stroke assessment ■ reproducibility of results

Because most standardized assessments of mood disorders are questionnaires,¹,² aphasic patients are not able to complete them. The Visual Analog Mood Scales (VAMS) have been developed for use in patients with communication (or other cognitive) deficiencies.³ They consist of 8 individual vertical VAMS measuring the following mood states: sad, happy, tense, afraid, confused, tired, energetic, and angry. Despite the assertions of Price et al.⁴ that stroke patients are unable to successfully complete Visual Analog Scales (VAS), it is the only self-assessment tool available for aphasic patients. Another way to assess depression in noncommunicant patients is to ask proxy raters. Recently, a 10-item questionnaire based on observable behavior thought to be associated with depression has been validated: the Stroke Aphasic Depression Questionnaire (SADQ). It was initially designed for aphasic patients living at home, completed by a spouse or a caregiver, then validated for aphasic patients in the hospital, completed by the nurse.⁵ SADQ, however, suffers from 1 methodological problem. To establish its validity, it has been compared with standardized questionnaires in nonaphasic patients. It was assumed that if the items from SADQ are correlated with questionnaire measures of depression in nonaphasic patients, then this will also be true of aphasic patients. However, there are some important behavioral differences between these 2 groups of patients. For example, patients with right-hemisphere stroke often exhibit reduced facial expression of emotions and indifference to entourage or exhibit euphoria and hypomania, which may influence an observer’s interpretation of their internal mood state.⁶ The aim of the present study was to construct a new behavioral depression rating scale for aphasic stroke patients, (1) the validity of which is established in a population including patients with severe aphasia, (2) the assessment of which relies on interviews of trained medical or paramedical staff, and (3) which is validated in subacute poststroke patients hospitalized in a neurorehabilitation unit.

To develop the Aphasic Depression Rating Scale (ADRS), we performed 3 consecutive studies corresponding to the usual steps of clinical scale validation.⁷

Study 1: Item Preselection

The purpose of study 1 was to select a pool of items able to detect and quantify depression from observable behavior in aphasic patients (content validity).
Methods
Eighteen members of the neurorehabilitation team were interviewed concerning typical depressed behavior in aphasic patients (1 psychiatrist, 1 neurologist, 1 physiatrist, 1 psychologist, 3 speech therapists, 5 physiotherapists, 2 hand therapists, 2 nurses, 2 auxiliary nurses), and the most frequently reported behaviors were noted. Six experts (1 psychiatrist, 1 neurologist, 1 physiatrist, 1 psychologist, 2 speech therapists) analyzed 3 existing depression scales that contain items describing observable behavior: the Hamilton Depression Rating Scale (HDRS), 1 the Montgomery & Asberg Depression Rating Scale (MADRS), 2 and the Salpetriere Retardation Rating Scale (SRRS). 8 Only items (1) that could be completed without interviewing patients, (2) that described depression behavior reported by the team, and (3) that were selected by at least 4 experts were retained.

Results
Fifteen items were selected by experts. Minor modifications were made for adaptation to hemiplegic patients. For instance, “slowness and paucity of movements—trunk and limbs” was replaced by “slowness and paucity of movements—trunk and nonaffected limbs.”

Study 2: Final Item Selection
The aims of study 2 were (1) to develop the final version of ADRS by refining the initial sampling of items (criterion and factorial validity), (2) to measure the correlation between ADRS and other assessments of depression (construct validity), and (3) to estimate a threshold for the diagnosis of depression.

Methods
After having given informed consent in accordance with the guidelines of the local ethics committee, every stroke patient admitted to our neurorehabilitation unit for a given period of time was examined by the same psychiatrist, who graded the severity of depression symptoms from 0 (no symptom of depression) to 100 (extremely severe depression), which we have termed dep-psychiatrist. Patients were assessed during the weekly rehabilitation staff meeting. For each patient, the members of the rehabilitation team (1 physiatrist, 1 resident, 1 psychologist, 1 speech therapist, 1 physiotherapist, 1 hand therapist, 1 nurse, and 1 auxiliary nurse) who were best acquainted with him or her (1) completed the 15 items selected in study 1 and (2) assessed the severity of depression on a basic visual analog scale of 0 to 100 (1 for each member). An average score was then calculated for the team, termed dep-rehab. Members of the team were unaware of the psychiatrist’s rating.

Criterion Validity
In the process of validating the clinical scale, the correlation of each item with a gold standard, if any, is used to select the best items and thereby refine draft versions of the scale. 7 There is no gold standard for assessing depression in an aphasic patient. In our study the psychiatrist’s assessment was important because the psychiatrist is an expert in mood disorders; the rehabilitation team’s assessment was also important because members of the team are well acquainted with the patient. We then chose both dep-psychiatrist and dep-rehab as gold standards, and we used the McHenry variable selection algorithm, which allows specification of 2 dependent variables (instead of 1 in classical regression methods) and usually yields the same subset as the “all possible regressions” procedure. 6 The algorithm searches for the best 1-variable model and produces an index, the Wilks λ, which is the multivariate extension of the R² and behaves like 1−R². It then searches for the best 2-variable model and calculates the Wilks λ decrease. It subsequently searches for the best 3-variable model, and so on. If the decrease of Wilks λ is small when 1 variable is added to the n-variable model, then the n-variable model is selected as the final model.

Factorial Validity
A principal component analysis (PCA) was performed to analyze the structure of the 15 items selected in Study 1; subsequently, some items were eliminated to avoid redundancies. For each set of redundant items, we retained the item (1) that was selected in the criterion validity section, or (2) that correlated with the greatest number of PCA factors, or (3) whose completion by the staff was the easiest. From that point on, the questionnaire was called ADRS.

Construct Validity
The correlation between ADRS and other constructs was measured with the product-moment correlation coefficient. We considered 3 constructs: dep-psychiatrist, dep-rehab (for all patients), and HDRS (for patients able to answer all HDRS questions, ie, all but severely aphasic patients).

Estimating a Threshold for the Diagnosis of Depression
A threshold was established after comparison of ADRS scores with the diagnosis made by the psychiatrist (depression or no depression) in the patients.

Results
Of the 52 patients admitted to the unit during the study period, 2 refused to take part; no patient was excluded. Fifty patients participated in the study (20 women and 30 men; mean ±SD age was 60±13 years (range, 28 to 80 years); 35 were left-hemisphere stroke patients (LHS), and 15 were right-hemisphere stroke patients (RHS); 29 LHS were aphasic, and 6 RHS were not. Twenty-five patients could not complete nonbehavioral items of HDRS: 21 LHS had severe aphasia, and 3 LHS and 1 RHS had other cognitive deficiencies. Mean dep-psychiatrist was 29.6/100 (±17.4), mean dep-rehab was 37.1/100 (±17.7), and mean HDRS was 10.8/58 (±6.6). Depression was diagnosed in 29 of 50 patients (58%) by the psychiatrist and in 17 of 25 patients (68%) by the HDRS (score >7/52). The time since admission was 60±45 days (range, 4 to 174 days).

Criterion Validity
The Wilks λ value was 0.121 for the 7-item model and 0.116 for the 8-item model, which is a minor increase (0.005). We then selected the 7-item model: Apparent Sadness; Insomnia; Gastrointestinal; Somatic; Weight; and Anxiety–Psychological; Somatic Symptoms–Gastrointestinal; Mimic–Slowness of Facial Mobility; Loss of Weight; and Anxiety–Somatic.

Factorial Validity
Four axes included 71% of the information. The 4-dimensional structure can be described as follows. For factor 1 (F1), all factor loadings were >0.50, except for Insomnia–Early (0.37), Insomnia–Middle (0.39), and Agitation (0.32). For factor 2 (F2), factor loadings of Mimic–Slowness of Facial Mobility–Nonaffected Side and Slowness and Paucity of Movements–Trunk, Nonaffected Limbs were the highest for this factor (0.59, 0.58). The correlation between these items was very high (r=0.77, P<10−6). Factor loadings of Anxiety–Psychic and Anxiety–Somatic were also relatively high (0.49, 0.45, respectively) on F2. For factor 3 (F3), the 3 insomnia items had the highest factor loadings (0.59 to 0.68) and were closely correlated with one another (r=0.49 to 0.84, P<10−3 to P<10−6). For factor 4 (F4), factor loadings of Somatic Symptoms–General and Somatic Symptoms–Gastrointestinal were the highest (0.46, 0.40, respec-
tively) for this factor. The correlation between these items was very high ($r=0.63$, $P<10^{-6}$). Compared with the other items, Anxiety–Somatic also had a relatively high factor loading on F4 (0.34).

The other item redundancies were Agitation–Hypochondriasis ($r=0.53$, $P<10^{-6}$) and Fatigability–Lassitude ($r=0.74$, $P<10^{-6}$).

These results suggest that F1 is a general factor of depressive illness, measuring the severity of the symptoms, F3 is an insomnia factor, and F2 and F4 are both anxiety axes, with F2 being a factor of retardation and F4 representing somatic symptoms.

To reduce redundancies among items, 6 items were eliminated according to predefined rules (see Methods, Factorial Validity). The 9 remaining items made up the ADRS (see Appendix).

Because the 3 depression rating scales from which ADRS items are derived yield similar scores (HDRS: 0 to 52; MADRS: 0 to 60; SRRS: 0 to 60), each of the 9 items has approximately the same weight in the total score with respect to the 8 others as in its original scale. For example, Apparent Sadness is a 0 to 6 ordinal item of MADRS, and therefore its weight is 0.10 (6/60); Mimic–Slowness of Facial Mobility is a 0 to 4 ordinal item of SRRS, and therefore its weight is 0.07 (4/60); the weight ratio for these 2 items is 1.5 (6/4). The weight of an item reflects the importance it represents for the assessment of depression according to the authors. In ADRS, which is a scale from 0 to 32, the weights of all items are increased by 60/32 (MADRS and SRRS items) or 52/32 (HDRS items). This is due to the fewer number of items in ADRS, which only contains 9 behavioral items. The preceding weights then become 0.19 (6/32) and 0.13 (4/32), but the ratio between the 2 items, which is the most important, remains the same. As a result, the weights of HDRS items are slightly underestimated in ADRS because their weights are increased by 52/32, whereas items from MADRS and SRRS are increased by 60/32. Because the ratio 60/52 is close to 1 (1.15), we decided not to change HDRS item scales.

**Construct Validity of ADRS**

The correlations between ADRS and dep-psy, dep-rehab (50 patients), and HDRS (25 patients) were high ($r=0.60$, $P<10^{-6}$; $r=0.78$, $P<10^{-6}$; $r=0.77$, $P<10^{-6}$, respectively). Correlation coefficients in RHS only and in LHS only were also high ($r=0.58$, $P<0.03$; $r=0.70$, $P<10^{-2}$; $r=0.84$, $P<10^{-3}$ for RHS; and $r=0.60$, $P<10^{-3}$; $r=0.86$, $P<10^{-6}$; $r=0.64$, $P<0.04$ for LHS).

When only the 25 patients able to complete HDRS were considered, ADRS correlated better with dep-psy and dep-rehab ($r=0.59$, $P<10^{-6}$; $r=0.85$, $P<10^{-6}$, respectively) than did HDRS ($r=0.40$, $P<0.05$; $r=0.59$, $P<10^{-2}$, respectively).

**Estimating a Threshold for the Diagnosis of Depression**

Mean ADRS was 10.1/32 ($\pm 5.6$). Thirty patients (60%) had a score $\geq 9/32$. With this value considered as a threshold, the sensitivity of ADRS compared with the diagnosis made by the psychiatrist was 0.83, and the specificity was 0.71. The value 9/32 was preferred to 8/32, which provided a poor specificity (0.52), and to 10/32, which provided a poor sensitivity (0.72). The sensitivity and the specificity of HDRS were poor (0.59 and 0.63, respectively).

**Study 3: Reliability**

The aim of study 3 was to assess the consistency of ADRS in 2 different rehabilitation teams (intrarater reliability), and from 1 administration of ADRS to another (test-retest reliability).

**Methods**

The intrarater reliability was assessed by comparing the ADRS scores of 15 new aphasic stroke patients, completed by 2 different rehabilitation teams within 24 hours. The test-retest reliability was assessed by comparing the ADRS scores of 15 additional patients assessed at a 2-week interval by the same team.

For each item, agreement between the teams’ ratings was assessed with $\kappa$ coefficients. For ADRS global scores, the association between teams’ ratings was assessed with correlation coefficients. Because of the relatively small number of observations at this stage, we used Spearman rank correlation coefficients.

**Results**

The average $\kappa$ coefficient over the 9 items was 0.69 (range, 0.37 to 1) for the intrarater reliability and 0.58 (range, 0.33 to 1) for the test-retest reliability. Items the completion of which does not require communicating with the patient at all (either verbal or nonverbal communication) were the more reliable ($\kappa >0.80$): Loss of Weight, Apparent Sadness, and Insomnia–Middle. On the other hand, the item Hypochondriasis was the least reliable ($\kappa=0.35$; 0.37).

When the global scores of ADRS were calculated, intrarater and test-retest reliabilities were very high (both $r=0.89$, $P<10^{-6}$).

**Discussion**

ADRIS is, to our knowledge, the only depression rating scale validated for the evaluation by the rehabilitation team of depression in poststroke aphasic patients in the hospital. Its final version contains 9 items taken from HDRS, MADRS, and SRRS. It is different from VAMS, which were validated for the self-evaluation of mood disorders in patients with cognitive disorders (including aphasia), and from SADQ, which was validated for the evaluation of aphasic patients by nurses.

The preceding results suggest that ADRS is a valid and reliable instrument. It has shown good content, criterion, factorial, and construct validity; high test-retest and intrarater reliability; and high sensitivity and specificity. A score $\geq 9/32$ strongly suggests the presence of depression. The factorial structure of ADRS is similar to the structure of HDRS. In a group of 172 healthy subjects, Hamilton described 3 main factors. The first measured the severity of the depression symptoms, which corresponds to F1 in the present study. The second was a bipolar factor with the symptoms of anxiety (psychic and somatic) and agitation counterbalancing retardation, suicide, depression, and loss of insight, which corresponds to F2 and F4. The third factor contained many items, including insomnia items, which are supported by F3. This provides support for the factorial validity of ADRS.

Despite these encouraging results, some remaining issues require attention. First, even if ADRS has high construct
validity, the best way to measure internal mood state in patients is to ask them. VAMS have been validated for self-administration in stroke patients. In its validation study, however, patients had to be able to complete the Profile of Mood States, and those with severely impaired language comprehension were thus excluded. The authors of VAMS indicated that in their clinical experience VAMS can be completed by most patients, even those with very limited communication and cognitive abilities. They also wrote that VAMS cannot be used to diagnose mood disorders but may be beneficial as part of a more extensive clinical evaluation. ADRS can be completed for all patients and may be used to diagnose depression. The association of ADRS and VAMS could then be very useful for assessing mood disorders in aphasic patients.

A second issue relates to the possibility of using ADRS in nonaphasic stroke patients. Fifteen RHS and 10 LHS without severe aphasia participated in the study because of the need to include the HDRS to assess the construct validity of ADRS. ADRS may then theoretically be used in all stroke patients. We should, however, closely examine the validity in RHS patients in whom reduced facial emotion expression or euphoria may influence an observer’s interpretation of the patient’s internal mood state.6 In our study, the correlations between ADRS and dep-psy, dep-rehab, and HDRS were very high in RHS, providing support for high validity in those patients, but this must be confirmed in a more extensive RHS population.

The fact that ADRS correlated better with dep-psy and dep-rehab than did HDRS in communicant stroke patients suggests that it is a better tool for examining these patients. This is probably due to the fact that ADRS items have been selected to mathematically explain dep-psy and dep-rehab. Nevertheless, it is evident that among the HDRS items that are not included in ADRS, Loss of Libido and Work and Interests are not suitable for subacute stroke inpatients. On the other hand, the items Depression, Guilt, and Loss of Insight should be pertinent in communicant stroke patients. A comparison of classic depression rating scales and ADRS remains to be made in these patients. Another explanation for this better correlation could be that both assessments (dep-rehab and ADRS) are based on patient observation, while HDRS relies on interviews.

Another important issue concerns the interpretation of symptoms described in ADRS items. First, it is difficult for the clinician to determine whether somatic symptoms (including insomnia, loss of appetite, constipation, fatigability) are due to depression or to stroke. It is probable that in many cases both depression and somatic symptoms are direct consequences of the stroke. In this study we have only demonstrated that the presence of these symptoms is often associated with depression (whatever the reason) and that ADRS displays a good sensitivity and a good specificity, which is most important. Second, although items of ADRS are completed without interviewing patients, it is of course easier with patients who are able to describe their symptoms or express their feelings with words. Although in our study this phenomenon had no negative effect on ADRS properties (the construct validity was good in both RHS and LHS patients), it should be studied in an extensive group of profoundly dysphasic patients.

A final issue pertains to the psychometric qualities of ADRS. The sensitivity of ADRS to fine clinical changes, as detected by a psychiatrist or the rehabilitation staff, should be assessed in a new group of poststroke patients. In conclusion, this preliminary study suggests that ADRS is a valid, reliable, sensitive, and specific rating scale for the evaluation of depression in aphasic poststroke inpatients hospitalized in a neurorehabilitation unit. It can be used for the diagnosis and the follow-up of depression, but this should be confirmed by other studies. A few methodological issues remain to be addressed, particularly the sensitivity to change.

Appendix

ADRS Items

Item 1. Insomnia—Middle
0: No difficulty
1: Patient indicates being restless and disturbed during the night/observed sleep disturbance
2: Waking during the night; any getting out of bed (except to go to the toilet)

Item 2. Anxiety—Psychic
0: No difficulty
1: Some tension and irritability
2: Worrying about minor matters
3: Apprehensive attitude apparent in patient’s face or speech
4: Fears indicated (verbal or nonverbal expression) without questioning

Item 3. Anxiety—Somatic
Symptoms can be gastrointestinal (dry mouth, flatulence, indigestion, diarrhea, cramps, belching); cardiovascular (palpitations, headaches); respiratory (hyperventilation, sighing). Other symptoms include urinary frequency, sweating.
0: Absent
1: Mild
2: Moderate
3: Severe
4: Incapacitating

Item 4. Somatic Symptoms—Gastrointestinal
0: None
1: Loss of appetite but continues to eat; heavy feelings in abdomen
2: Difficulty eating (not due to arm paresis); requests or requires laxatives or medication for bowels or for gastrointestinal symptoms

Item 5. Hypochondriasis
0: Not present
1: Self-absorption (bodily)
2: Preoccupation with health
3: Frequent complaints, requests for help, etc.
4: Hypochondriacal delusions

Item 6. Loss of Weight
Ratings is by measurement and does not take into account loss of weight that occurred during the acute poststroke phase.
0: <0.5 kg weight loss per week
1: 0.5 kg to 1 kg weight loss per week
2: >1 kg weight loss per week

Item 7. Apparent Sadness
Item represents despondency, gloom, and despair (more than just ordinary transient low spirits) reflected in speech (not available in case of severe aphasia), facial expression, and posture. Rating is by depth and inability to brighten mood.
0: No sadness
1: between 0 and 2
2: Looks dispirited but brightens without difficulty
3: between 2 and 4
4: Appears sad and unhappy most of the time
5: between 4 and 6
6: Looks miserable all the time; extremely despondent

**Item 8. Mimic–Slowness of Facial Mobility (concerns only the nonaffected side)**

0: The head moves freely, resting flexibly on the body with the gaze either exploring the room or fixed on the examiner or on other objects of interest in an appropriate manner.
1: There may be some reduction of mobility, not easily confirmed.
2: Reduction of mobility is definite but mild; the gaze, while often fixed, is still capable of shifting; mimic, although monotonous, is still expressive.
3: Patient does not move the head or explore the room and usually stares at the floor, seldom looking at the examiner; patient is slow to smile; expression is unchanging.
4: Face is completely immobile and painfully inexpressive.

**Item 9. Fatigability (takes into account motor deficiency, if any)**

0: Fatigability is not indicated spontaneously (by verbal or nonverbal communication) or after direct questioning.
1: Fatigability is not indicated spontaneously, but evidence of it emerges in the course of the interview.
2: Patient is distressed by fatigability in his everyday life (eating, washing, dressing, climbing stairs, or any other physical activity the patient is usually able to do despite motor deficiency).
3: Fatigability is such that the patient must curb some activities.
4: Near-total reduction of activities owing to an overwhelming sense of fatigue.

**References**

Validation of the Aphasic Depression Rating Scale
Charles Benaim, Bruno Cailly, Dominic Perennou and Jacques Pelissier

Stroke. published online May 13, 2004;
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2004 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/early/2004/05/13/01.STR.0000130591.95710.20.citation

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org//subscriptions/