Rating Scales for Depression in the Elderly: External and Internal Validity

Alex Kørner, M.D., Ph.D.; Lise Lauritzen, M.D., Ph.D.; Kirsten Abelskov, M.D.; Nils C. Gulmann, M.D.; Anne-Marie Brodersen, M.D.; Torben Wedervang-Jensen, M.D.; and Karen Marie Kjeldgaard, M.D.

Objective: The aim of the study was to assess the external and internal validity of the 6- and 17-item versions of the Hamilton Rating Scale for Depression (HAM-D_6 and HAM-D_{17}), the Bech-Rafaelsen Melancholia Scale, the 15- and 30-item versions of the Geriatric Depression Scale, and the Cornell Scale for Depression in Dementia in a population of depressed demented and nondemented Danish elderly.

Method: Two clinicians performed independent, blinded assessments of the study population, which was drawn from psychogeriatric outpatient clinics, and a control group of elderly subjects. Concurrent and convergent validity were assessed using correlation coefficient analyses, and to evaluate the internal validity, item response analysis using the Mokken coefficient and Rasch analysis was performed. A coefficient of homogeneity of 0.40 or higher indicated scalability. Data collection took place between October 2001 and April 2004.

Results: 145 subjects were included; 102 were female (mean age = 78.6 ± 6.8 years), and 43 were male (mean age = 72.4 ± 5.6 years). In the study group (N = 109), 73 subjects had depression only, and 36 had both depression and dementia; in the control group (N = 36), 11 subjects had dementia. The item-response analysis made a clear distinction between the scales. The HAM-D_6 was the only scale that fulfilled the criterion of total scalability in both the cognitively intact and the impaired populations. In terms of standardization according to the Clinical Global Impressions-Severity of Illness scale (CGI-S), the HAM-D_6 had the most convincing external validity overall. In terms of general correlation to the CGI-S, only small differences were shown between the scales.

Conclusion: The HAM-D_6 should be separately considered even when longer HAM-D versions are used for the measurement of depression in elderly persons.

(J Clin Psychiatry 2007;68:384–389)

Received Jan. 24, 2006; accepted July 7, 2006. From the Psychogeriatric Department, Frederiksborg General Hospital, Hillerød (Drs. Kørner, Lauritzen, and Kjeldgaard); the Psychogeriatric Department D, Psychiatric Hospital Risskov, Århus (Drs. Abelskov and Gulmann); and the Psychiatric Center Gentofte, Copenhagen University Hospital, Gentofte (Drs. Brodersen and Wedervang-Jensen), Denmark.

The study was supported by an unrestricted grant from Lundbeck Pharma A/S.

The authors express their gratitude to Professor Per Bech for critical advice and suggestions in preparing the manuscript.

The authors report no additional financial or other relationship relevant to the subject of this article.

Corresponding author and reprints: Alex Kørner, M.D., Psychogeriatric Department, Psychiatric Hospital, Dyrehavevej 48, DK-3400 Hillerød, Denmark (e-mail: alko@fa.dk).

In randomized clinical trials with antidepressants in depressed elderly patients, the Hamilton Rating Scale for Depression (HAM-D)[1] is the most frequently used outcome scale for demonstrating clinical effect.[2] It is customary to use the change in the total score of the first 17 items (HAM-D_{17}) to measure antidepressive efficacy of the treatment under examination. A reduction in the baseline score of 50% or more at endpoint, e.g., after 6 to 8 weeks of therapy, is often used as the definition of response to therapy.

However, the use of the total score of the HAM-D_{17} as a sufficient statistic in trials of antidepressants has recently been criticized by Bagby et al.[3] Their systematic review of all psychometric studies with the HAM-D_{17} showed that the scale is multidimensional, implying that the various symptom or cluster profiles should be shown rather than the global total score. Bagby et al.,[3] however, identified the 6-item HAM-D subscale (HAM-D_6), which includes the core items of depressive states, as a unidimensional scale, implying that its total score is a sufficient statistic.

The HAM-D_6 was originally developed from the HAM-D_{17}, using a global assessment of depressive states performed by experienced psychiatrists as an index of validity.[4] Using the item response theory models, it was then shown[5] that the HAM-D_6, but not the HAM-D_{17}, was a unidimensional scale with no bias from external issues such as age and gender. However, the age of subjects in these studies was below 65 years, as it was in most of the studies examined by Bagby et al.[3]

In the present study, we investigated various aspects of the validity of depression rating scales including the HAM-D in a group of subjects 65 years or older that included subjects with mild-to-moderate dementia.

METHOD

Data collection took place between October 2001 and April 2004. Seven experienced psychogeriatricians from 3 psychogeriatric centers participated. Prior to inclusion and during active enrollment in the study, co-rating sessions took place to establish the interrater coefficients and to ensure that the clinicians upheld the proper procedure throughout the study.

Two clinicians examined each patient or control subject independently. One clinician performed the diagnostic evaluation of both depression and dementia according to the tenth revision of the *International Classification of Diseases* (ICD-10)[6] and the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV).[7] In this article, we report DSM-IV diagnoses. The diagnosing clinician also performed a global judgment of severity of depression (Clinical Global Impressions-Severity of Illness scale [CGI-S]).[8] Blinded to this assessment, the other clinician interviewed the patient or control subject according to the psychometric instruments under investigation. Both evaluations took place on the same day as consecutive interviews or at the same time of the day on consecutive days.

The interviews were performed using the guidelines originally published by Hamilton.[1] Thus, the interviews were done in an unstructured way that allowed the interviewers to be flexible in not asking the same questions several times. The order of administration was not randomized. As advised by Hamilton,[1] we also used information from a knowledgeable caregiver.

Study Population

All subjects were 65 years or older. The patients were recruited from 3 psychogeriatric clinics. Controls were recruited from among relatives of patients in the participating clinics and among members of the local organizations for the elderly.

Patients with other major psychiatric illnesses and patients suffering from aphasic disorder were excluded. However, patients with dementia were eligible and no predefined cutoff on the Mini-Mental State Examination (MMSE)[9] was applied in order to approximate the study population as much as possible to a standard clinical psychogeriatric population. Thus, patients were included or excluded based on the clinician's judgment of the possible participant's ability to meaningfully perform the required tests and take part in the interview. In this respect, inclusion in the study was not based on degree of dementia after a randomization factor. Finally, the patients had to have a caregiver willing to cooperate. Informed consent had to be given to participate in the study.

Rating Scales

Hamilton Rating Scale for Depression. The HAM-D_{17} was used in the 17-item format accepted by Hamilton,[10] but with the original instructions for the 21-item version.[1] Thus, relevant information from caregivers was used for the scoring of the individual items. In the analysis, both the HAM-D_{17} and the HAM-D_6, which includes the core items of depression, were considered. The 6 core items are depressed mood, guilt feeling, work and interests, psychomotor retardation, psychic anxiety, and general somatic symptoms. The HAM-D_{17} has a theoretical score range of 0 to 52, and the HAM-D_6 has a theoretical score range of 0 to 22.

Bech-Rafaelsen Melancholia Scale. The Bech-Rafaelsen Melancholia Scale (MES)[2] was developed in a similar way to the HAM-D_6,[5] by subdividing retardation into verbal, motor, emotional, and cognitive retardation. The theoretical score range of the 11 MES items is 0 to 44.

Geriatric Depression Scale. We administered the original 30-item Geriatric Depression Scale (GDS)[11] in a Danish translation[12] as a self-rating scale. Responses were coded as 1 = yes, symptom present or 0 = no, symptom not present. Items are summed, and higher scores indicate a greater number of depressive symptoms. Originally intended for screening of depression in a population of nondemented elderly, it has satisfactory validity in that group. Shorter versions of the GDS have been developed; the most commonly used is the 15-item version (GDS_{15}),[13] but even shorter versions exist.[14] However, when the GDS_{15} is applied to a group of cognitively impaired elderly, its validity has been shown to decrease.[15] The GDS does not contain items concerning somatic complaints. We computed GDS_{15} scores based on the GDS_{30} scores.

Cornell Scale for Depression in Dementia. The Cornell Scale for Depression in Dementia (CSDD)[16] consists of 19 items each rated as 0 = absent, 1 = mild or intermittent, or 2 = severe. The scores of the individual items are summed, and a score of 8 or more suggests significant depressive symptoms. We recently published the sensitivity and specificity in this population of elderly with and without dementia and found the best cutoff score to be 6.[17] The CSDD is specifically designed to screen for depression in a population of demented elderly. It is an interviewer-administered scale relying on information from caregivers and from the clinician's observations during the interview. It comprises items concerning physical well-being, sleep, appetite, and other vegetative symptoms. Some studies have shown it to be valid for screening of depression in a nondemented population as well.[18]

Clinical Global Impressions scale. The Clinical Global Impressions scale[8] consists of 3 different observer-rated scales, i.e., the CGI-S, measuring severity of illness;

the CGI-C, measuring therapeutic response (change); and an efficacy index. In this study, we used the CGI-S for depression with collateral information from caregivers. Depression severity was assessed using 7 ratings: 1 = no depression, 2 = doubtful depression, 3 = mild depression, 4 = moderate depression, 5 = marked depression, 6 = severe depression, and 7 = extremely severe depression. With the CGI-S,[8] a depressed patient is judged in the context of the rater's experience or acquaintance with elderly depressed patients, i.e., against a clinical background, which may include demented as well as nondemented subjects. In this study, the CGI-S was used by experienced psychogeriatric clinicians. The CGI-S was applied as the external index of severity.

Mini-Mental State Examination. The MMSE[9] measures the cognitive impairment of the subjects. The MMSE consists of 11 items with a total score of 30. Lower score indicates poorer cognitive performance. The MMSE is used in a Danish validated version.[19]

We collected basic demographic data. In this study, we report the subjects' gender and age.

Statistics

Interrater reliability. The interrater reliabilities were analyzed using the intraclass coefficients (ICCs) according to Bartko and Carpenter[20] and Shrout and Fleiss.[21] Correlations between 0.80 and 1.00 are considered perfect interrater correlations, correlations between 0.60 and 0.79 are considered almost perfect, and correlations between 0.40 and 0.59 are considered adequate, while coefficients below 0.40 are considered inadequate.

Psychometric analysis of internal validity. To identify the unidimensionality of the various depression scales, item response theory models were used.[22]

Item response theory models have been used to evaluate the unidimensionality of the various depression rating scales as this type of analysis takes into account bias in the individual scale items as well as bias in the selection of subjects.[23] Both the nonparametric Mokken analysis[24] and the 1-parameter Rasch analysis[25] were used. The Mokken analysis is based on a Loevinger coefficient of homogeneity, which assesses the degree of rank-ordering of the items in terms of transmitted information about the dimension of symptom severity of depression.[26] The level of acceptance is a Loevinger coefficient of homogeneity of 0.40 or higher, implying that the total score is a sufficient statistic of symptom severity of depression. The 1-parameter Rasch analysis was carried out to evaluate bias in the selection of subject, using male versus female gender as the external criterion. The Rasch analysis was performed using the method described by Allerup.[27] In this analysis, the rejection level of homogeneity or unidimensionality was $p < .01$.

External validity. The CGI-S[8] was used as the index of clinical (external) validity. The different symptom depression scales were then standardized for each of the categories of the CGI-S, i.e., no depression, doubtful depression, mild depression, moderate depression, marked depression, and severe depression.

Ethics. The study was registered and approved by the local Scientific Ethics Committee under the registration number 2001-2-05. Each participating patient and control subject as well as the caregivers received written and verbal information prior to the study and gave written consent. When appropriate, a legal guardian gave written consent on behalf of patients or control subjects.

RESULTS

One hundred forty-five subjects were included; 102 were female (mean ± SD age = 78.6 ± 6.8 years), and 43 were male (mean age = 72.4 ± 5.6 years). One hundred nine (75%) of these were in the group under examination (79 women, mean age = 79.1 ± 6.4 years; 30 men, mean age = 72.2 ± 5.2 years), and 36 (25%) were in the control group (23 women, 76.1 ± 9.0 years; 13 men, 74.1 ± 6.1 years). Seventy-three subjects (67%) in the group under examination were depressed only, while 36 (33%) suffered from both depression and dementia. Of the 36 control subjects, 11 (31%) were demented.

Of the 109 subjects in the group under examination, 95 (87%) had major depressive disorder (36 severe depression, 37 moderate depression, and 22 mild depression), 3 (3%) had dysthymia, and 11 (10%) had depression not otherwise specified. Thirty-six (33%) were diagnosed with Alzheimer's disease, 10 (9%) had vascular dementia, 6 (6%) had mixed Alzheimer's disease and vascular dementia, and 11 (10%) had other dementia disorders. The mean MMSE score of the demented subjects in the control group was 18.9 ± 5.7, and the mean MMSE score of the demented subjects in the group under examination was 18.6 ± 6.7. The subjects in the group under examination without dementia had a mean MMSE score of 27.3 ± 2.1, while the subjects in the control group without dementia or depression had a mean score of 28.3 ± 2.1. The mean $HAM\text{-}D_{17}$ score of the subjects in the group under examination with depression only was 17.8 ± 6.0, and with comorbid dementia 16.4 ± 5.8. The mean $HAM\text{-}D_{17}$ scores of the control subjects with and without dementia were 3.4 ± 2.0 and 2.0 ± 1.9, respectively.

Interrater Reliability

In total, 15 subjects were interviewed and videotaped, and these interviews were assessed in co-rating sessions both prior to and during the study to evaluate interrater reliability of the clinician-based symptom rating scales and the CGI-S. All results in terms of ICCs were statistically significant: $HAM\text{-}D_{17}$: 0.90, $HAM\text{-}D_{6}$: 0.81, MES: 0.88, CSDD: 0.84, and CGI: 0.88. In a subgroup of patients (N = 9), the MMSE was analyzed. The ICC was 0.98.

Table 1. Mokken Analysis of the Depression Symptom Scales in Demented and Nondemented Subjects[a]

Group	HAM-D_{17}	HAM-D_6	MES	GDS_{30}	GDS_{15}	CSDD
All subjects (N = 145)	0.33	0.49	0.42	0.40	0.46	0.27
Nondemented subjects (N = 98)	0.37	0.54	0.49	0.45	0.55	0.29
Demented subjects (N = 47)	0.28	0.43	0.34	0.31	0.29	0.27

[a]A coefficient of homogeneity of 0.40 or higher signifies unidimensionality.
Abbreviations: CSDD = Cornell Scale for Depression in Dementia, GDS = Geriatric Depression Scale, HAM-D = Hamilton Rating Scale for Depression, MES = Bech-Rafaelsen Melancholia Scale.

Table 2. Standardization of the Depression Symptom Scales According to the CGI-S

	CGI-S Depression Category				
Scale	1 or 2 (no or doubtful)	3 (mild)	4 (moderate)	5 (marked)	6 (severe)
HAM-D_6					
All subjects (N = 145)	3.2	8.5	10.1	11.8	13.5
Nondemented (N = 98)	2.9	8.7	10.4	12.3	13.8
Demented (N = 47)	3.8	8.3	9.7	11.0	NA[a]
HAM-D_{17}					
All subjects (N = 145)	4.7	15.0	18.2	22.2	26.3
Nondemented (N = 98)	4.6	15.4	18.8	23.5	25.8
Demented (N = 47)	5.1	14.1	17.1	19.8	NA[a]
MES					
All subjects (N = 145)	4.5	13.8	17.0	20.3	24.3
Nondemented (N = 98)	3.5	13.5	16.6	20.5	24.0
Demented (N = 47)	6.9	14.1	17.6	19.8	NA[a]
CSDD					
All subjects (N = 145)	3.5	11.1	14.3	17.9	22.5
Nondemented (N = 98)	3.1	11.2	14.5	18.8	20.8
Demented (N = 47)	4.5	10.8	13.9	16.2	NA[a]
GDS_{15}					
All subjects (N = 145)	2.4	7.0	9.3	10.6	11.0
Nondemented (N = 98)	2.2	7.6	10.0	11.0	12.0
Demented (N = 47)	3.1	5.9	8.1	9.8	NA[a]
GDS_{30}					
All subjects (N = 145)	5.5	14.7	19.6	21.5	22.8
Nondemented (N = 98)	5.4	15.5	21.1	22.9	24.6
Demented (N = 47)	5.7	13.0	17.1	19.0	NA[a]

[a]Not applicable; 1 observation only.
Abbreviations: CGI-S = Clinical Global Impressions-Severity of Illness scale, CSDD = Cornell Scale for Depression in Dementia, GDS = Geriatric Depression Scale, HAM-D = Hamilton Rating Scale for Depression, MES = Bech-Rafaelsen Melancholia Scale.

Internal Validity

Table 1 shows the coefficient of homogeneity as obtained for each depression scale based on all subjects (N = 145) and on the nondemented (N = 98) and demented (N = 47) subjects separately. Only the HAM-D_6 obtained a coefficient of homogeneity above 0.40 for all 3 groups of patients (see Table 1). However, the MES and to a lesser extent the GDS_{30} had acceptable coefficients in all 3 groups of patients.

External Validity

Table 2 shows the standardization of the depression rating scales using the CGI-S categories 1 and 2 (no or doubtful depression), 3 (mild), 4 (moderate), 5 (marked), and 6 (severe) as an index of validity.

Of all 145 subjects included, according to CGI-S definitions, 35 had mild depression, 38 had moderate depression, 17 had marked depression, and 6 had severe depression. Of the 98 nondemented subjects, 23 had mild depression, 24 had moderate depression, 11 had marked depression, and 5 had severe depression. Finally, of the 47 demented subjects, 12 had mild depression, 14 had moderate depression, 6 had marked depression, and 1 had severe depression.

As only 1 demented subject was found to be in the category "severe depression," this category was not analyzed further.

Generally, all of the scales differentiated between the categories no/doubtful depression and mild depression and had an incremental rise in relation to depression severity. The group of demented subjects scored lower than the nondemented subjects on all scales with the exception of the MES. For all categories except marked depression, demented subjects scored higher on the MES than non-

Table 3. Intercorrelation Between CGI-S and the Depression Symptom Scales by Spearman Coefficients

Group	HAM-D_{17}	HAM-D_6	MES	CSDD	GDS_{30}	GDS_{15}
All subjects (N = 145)	0.85	0.80	0.84	0.82	0.79	0.75
Nondemented (N = 98)	0.87	0.83	0.87	0.86	0.84	0.80
Demented (N = 47)	0.78	0.71	0.76	0.72	0.69	0.64

Abbreviations: CGI-S = Clinical Global Impressions-Severity of Illness scale, CSDD = Cornell Scale for Depression in Dementia, GDS = Geriatric Depression Scale, HAM-D = Hamilton Rating Scale for Depression, MES = Bech-Rafaelsen Melancholia Scale.

demented subjects; however, the difference between the groups was relatively small.

When the HAM-D_6 is compared with the HAM-D_{17} and the MES, the mean scores for moderately depressed patients should be 9 or more, 18 or more, and 15 or more, respectively.[28] The mean scores for marked-severe depression on the 3 scales should be 12 or more for the HAM-D_6 and 25 or more for both the HAM-D_{17} and the MES. These mean scores were obtained on the HAM-D_6 for all subject groups, while the demented subjects did not obtain a sufficient mean score on the HAM-D_{17}. On the MES, the mean scores were within the expected range except for the category of severely depressed, in which the mean scores tended to be a bit lower. We are unaware of any report of a similar validated categorization of the CSDD and the GDS. However, the CSDD may, based on its many coinciding items with the HAM-D, be theoretically categorized as follows: moderate depression = mean score of 15 or more and marked-severe depression = mean score of 20 or more. A score of 20 or more on the GDS_{30} is reported to indicate severe depression.[11] Based on this, mean CSDD scores were too low in both the moderate and the marked-severe groups, while the GDS_{30} had sufficient mean scores.

Table 3 shows the intercorrelation between the CGI-S and the symptom depression scales. All scales had lower correlation with the CGI-S in the demented group, but only the GDS_{30} and GDS_{15} obtained Spearman coefficients below 0.70, i.e., 0.69 and 0.64, respectively.

DISCUSSION

The internal validity in terms of the coefficient of homogeneity showed that both the HAM-D_{17} and CSDD, and to a lesser degree the GDS, were multidimensional, which indicates that the total scores on these scales are not sufficient statistics. Our results confirmed that the HAM-D_6 as well as the MES are unidimensional scales[22] in elderly depressed subjects as well as in other populations. In a recent study, Bech et al.[23] found that the HAM-D_6 was more sensitive than the HAM-D_{17} in demonstrating a dose-response relationship of citalopram or escitalopram. As stated earlier, this finding is in accordance with the findings of Bagby et al.[3] in their review of the HAM-D.

The assessment of external validity of these commonly used symptom scales in elderly depressed patients is based on a global rating scale that must be performed by experienced psychogeriatricians because the scoring of depression severity relies on knowledge by acquaintance, i.e., clinical experience, as discussed by Borsboom.[29] It is essential to refer to the clinical validity of a rating scale in the validation procedure. We were able to demonstrate a high interobserver reliability of the CGI-S,[8] which proves that the standardization of the various symptom scales for depression is meaningful. Moreover, we were also able to demonstrate an acceptable interobserver reliability of the depression symptom scales and of the MMSE, which was used to give a quantitative measure of the cognitive impairment of the subjects.

The HAM-D_6 also had the overall most convincing external validity in terms of standardization according to the CGI-S. It had the least variation between the demented and nondemented subjects, and it differentiated the subjects by severity (Table 2), as defined by Bech.[30] The same was not the case for the HAM-D_{17}, the MES, the GDS, or the CSDD. Among these, however, the MES seems to perform better, as it sufficiently differentiated between the groups as defined by Bech,[30] although the scale showed a tendency toward higher scores in the demented group, which can certainly be attributed to the item concerning intellectual retardation, an item on which demented subjects will invariably have a high score. In terms of general correlation to an external global score (Table 3), all scales had values above 0.80 in the nondemented group, while the values in the demented group were between 0.70 and 0.80 in the clinician-rated scales and slightly below 0.70 in the 2 versions of the GDS. However, these differences are in themselves not substantial enough to warrant discarding any of the scales.

The investigators of this study have their daily clinical practice in 3 independent wards with different approaches to the use of rating scales in general and different preferences in relation to depression rating scales. In our opinion, no bias exists in terms of expecting better performances in particular scales. In addition, it should be pointed out that the item response theory model we

used in our statistical analysis is a test for item response bias.

Depressive disorders in a population of elderly depressed persons with and without dementia are heterogenous in origin and presentation, but it remains important to develop tools for assessing the severity of these disorders accurately. O'Sullivan et al.[31] have demonstrated that the HAM-D_6 was able to cover the core items of depression across different subtypes of depression when compared with the HAM-D_{17} and the enlarged HAM-D_{28}. The population in this study was also heterogeneous in terms of both depressive subtypes and cognitive impairment; therefore, it seems justified to state that the HAM-D_6 has a higher content validity than the other variants of HAM-D in relation to depression in the elderly.

This study has evaluated some of the more well-known depression rating scales and found it reasonable to conclude that among the evaluated scales the HAM-D_6 is psychometrically the most relevant observer-rated scale to use in daily clinical practice and possibly also in clinical studies. It has the further advantage of comprising only 6 items. However, as we have not administered it in isolation, we shall limit our conclusion to emphasizing the importance of calculating the HAM-D_6 score and not relying solely on the HAM-D_{17} score.

Drug names: citalopram (Celexa and others), escitalopram (Lexapro and others).

REFERENCES

1. Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 1967;6:278–296
2. Bech P, Tanghøj P, Andersen HF, et al. Citalopram dose-response revisited using an alternative psychometric approach to evaluate clinical effects of four fixed citalopram doses compared to placebo in patients with major depression. Psychopharmacology (Berl) 2002; 163:20–25
3. Bagby MR, Ryder AG, Schuller DR, et al. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? Am J Psychiatry 2004;161:2163–2177
4. Bech P, Gram L, Dein E, et al. Quantitative rating of depressive states. Acta Psychiatr Scand 1975;51:161–170
5. Bech P, Allerup P, Gram LF, et al. The Hamilton Depression Scale: evaluation of objectivity using logistic models. Acta Psychiatr Scand 1981;63:290–299
6. International Classification of Diseases and Related Health Problems, 10th Revision. Geneva, Switzerland: WHO; 1992
7. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Washington, DC: American Psychiatric Association; 1994
8. Guy W. Clinical Global Impressions (CGI). In: ECDEU Assessment Manual for Psychopharmacology. US Dept Health, Education, and Welfare publication (ADM) 76-338. Rockville, Md: National Institute of Mental Health; 1976:218–222
9. Folstein MF, Folstein SE, McHugh PR. "Mini-Mental State": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189–198
10. Bech P, Kastrup M, Rafaelsen OJ. Mini-compendium of rating scales. Acta Psychiatr Scand 1986;73(suppl 326):7–37
11. Yesavage JA, Brink TL, Rose TL, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1982–1983;17:37–49
12. Black C, Auerbach H, Mogensen F. Danish translation of Geriatric Depression Scale. Clin Gerontol 1995;16:60–61
13. Sheik JI, Yesavage JA. Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. Clin Gerontol 1986;37: 819–820
14. Shah A, Herbert R, Lewis S, et al. Screening for depression among acutely ill geriatric inpatients with a short Geriatric Depression Scale. Age Ageing 1997;26:217–221
15. Burke WJ, Houston MJ, Boust SJ, et al. Use of the Geriatric Depression Scale in dementia of the Alzheimer type. J Am Geriatr Soc 1989;37: 856–860
16. Alexopoulos GS, Abrams RC, Young RC, et al. Cornell Scale for Depression in Dementia. Biol Psychiatry 1988;23:271–284
17. Kørner A, Lauritzen L, Abelskov K, et al. The Geriatric Depression Scale and the Cornell Scale for Depression in Dementia: a validity study. Nord J Psychiatry 2006;60:360–364
18. Alexopoulos GS, Abrams RC, Young RC, et al. Use of the Cornell Scale in nondemented patients. J Am Ger Soc 1988;36:230–236
19. Lauritzen L, Kørner A, Bech P. A comparison between psychogeriatric rating scales: inter-rater reliability and validity. In: Bergener M, Finkel SI, eds. Treating Alzheimer's Disease and Other Dementias. New York, NY: Springer Publishing Corporation Inc; 1995:269–276
20. Bartko JJ, Carpenter WT. On the methods and theory of reliability. J Nerv Ment Dis 1976;163:307–317
21. Shrout PE, Fleiss JL. Intra-class correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–428
22. Licht RW, Quitzau S, Allerup P, et al. Validation of the Bech-Rafaelsen Melancholia Scale and the Hamilton Depression Scale in patients with major depression: is the total score a valid measure of illness severity? Acta Psychiatr Scand 2005;111:144–149
23. Bech P, Tanghøj P, Cialdella P, et al. Escitalopram dose-response revisited: an alternative psychometric approach to evaluate clinical effects of escitalopram compared to citalopram and placebo in patients with major depression. Int J Neuropsychopharmacol 2004;7:283–290
24. Sijtsma K, Molenaar W. Introduction to Nonparametric Item Response Theory. London, UK: Sage Publications; 2002
25. Bond TG, Fox CH. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. London, UK: Lawrence Erlbaum; 2001
26. Allerup P. Statistical Analysis of Rating Scales. Copenhagen, Denmark: Danish Institute of Educational Research; 1986
27. Allerup P. Rasch measurement. In: Husen T, Postlethwaite N, eds. International Encyclopedia of Education, vol 8. London, UK: Pergamon; 1994:4902–4913
28. Bech P. Meta-analysis of placebo-controlled trials with mirtazapine using the core items of the Hamilton Depression Scale as evidence of a pure antidepressive effect in the short-term treatment of major depression. Int J Neuropsychopharmacol 2001;4:337–345
29. Borsboom D. Measuring the Mind: Conceptual Issues in Contemporary Psychometrics. Cambridge, Mass: Cambridge University Press; 2005
30. Bech P. Rating Scales for Psychopathology, Health Status, and Quality of Life: A Compendium on Documentation in Accordance With the DSM-III-R and the WHO Systems. Berlin, Germany: Springer; 1993
31. O'Sullivan RL, Fava M, Agustin C, et al. Sensitivity of the six-item Hamilton Depression Rating Scale. Acta Psychiatr Scand 1997;95: 379–384