Psychological Medicine, 1996, **26**, 477–486.

The Inventory of Depressive Symptomatology (IDS): psychometric properties

A. J. RUSH,[1] C. M. GULLION, M. R. BASCO, R. B. JARRETT AND M. H. TRIVEDI

From the Department of Psychiatry and Mental Health Clinical Research Center, University of Texas Southwestern Medical Center, Dallas, TX., USA

SYNOPSIS The psychometric properties of the 28- and 30-item versions of the Inventory of Depressive Symptomatology, Clinician-Rated (IDS-C) and Self-Report (IDS-SR) are reported in a total of 434 (28-item) and 337 (30-item) adult out-patients with current major depressive disorder and 118 adult euthymic subjects (15 remitted depressed and 103 normal controls). Cronbach's α ranged from 0·92 to 0·94 for the total sample and from 0·76 to 0·82 for those with current depression. Item total correlations, as well as several tests of concurrent and discriminant validity are reported. Factor analysis revealed three dimensions (cognitive/mood, anxiety/arousal and vegetative) for each scale. Analysis of sensitivity to change in symptom severity in an open-label trial of fluoxetine ($N = 58$) showed that the IDS-C and IDS-SR were highly related to the 17-item Hamilton Rating Scale for Depression. Given the more complete item coverage, satisfactory psychometric properties, and high correlations with the above standard ratings, the 30-item IDS-C and IDS-SR can be used to evaluate depressive symptom severity. The availability of similar item content for clinician-rated and self-reported versions allows more direct evaluations of these two perspectives.

INTRODUCTION

The 28-item Inventory of Depressive Symptomatology, Clinician-Rated (IDS-C) and Self-Report (IDS-SR) scales were reported to have reasonable psychometric properties (Rush *et al.* 1986). Since then, two additional items to cover atypical symptom features have been added. This report describes the psychometric properties of both the 28- and 30-item versions of the IDS-C and IDS-SR in a new patient sample.

The rationale for the development of the IDS-C rests on evaluation of the Hamilton Rating Scale for Depression (HRS-D) (Hamilton, 1960, 1967). The original 17-item HRS-D was developed on psychiatric in-patients and its symptom coverage is inadequate. For example, the 17-item scale does not cover all criterion symptoms for major depressive disorder (MDD) or for depressive subtypes (e.g. melancholic or atypical features). The content coverage has changed over the years, due in part, to the evolving diagnostic criterion symptoms for major depressive and other mood disorders (Prien *et al.* 1991; Grundy *et al.* 1994). Moreover, the content of selected items is a mixture of symptoms (e.g. 'Work and Interest' covers fatigue, energy, motivation, interest, and the capacity for pleasure, which overlaps with 'Somatic Energy' and 'Psychic Anxiety' also rates irritability). The expanded versions of the HRS-D have been criticized for including items infrequently encountered in depression (e.g. obsessions and compulsions, paranoia, depersonalization/derealization) (Nelson & Mazure, 1990). The psychometric properties of most of the expanded versions have not been established. However, the 17-item HRS-D has been subjected to many psychometric evaluations. A key criticism is that it has more than one factor (for a review, see Hedlund & Vieweg, 1979), which implies that the total score is a mixture of several factors. An extensive item evaluation (Rehm & O'Hara, 1985) revealed that 6 of the 17 items had such poor performance with out-patients that deleting them from the instrument would improve its overall psycho-

[1] Address for correspondence: Dr A. John Rush, Department of Psychiatry at St. Paul, POB 1, Suite 600, 5969 Harry Hines Boulevard, Dallas, Texas 75235-9101, USA.

metric performance. More recently, Gibbons *et al.* (1993) attempted to identify the content of the primary HRS-D factor and found that most of the items were discarded as not contributing to information about the severity of depression.

The rationale for developing the IDS-SR rests on the limitations of the available self-report measures. The Carroll Rating Scale (CRS) (Carroll *et al.* 1981) is a 52-item yes/no self-report form of the 17-item HRS-D. Therefore, content-related limitations of the HRS-D are likely to apply to the CRS. A commonly used self-report is the Beck Depression Inventory (BDI) (Beck *et al.* 1961, 1979). While the items were not chosen to reflect a particular theory of depression, 11 of 21 items (52%) measure the cognitive symptoms of depression (for a review, see Beck *et al.* 1988). Thus, it heavily weights cognitive, as compared to vegetative and other depressive symptoms. It does not assess some of the DSM-IV (APA, 1994) criterion symptoms (e.g. overeating, oversleeping, weight gain, psychomotor agitation, and early and middle insomnia) and does not include symptoms necessary for subtype evaluation (melancholic or atypical features).

The IDS-C and IDS-SR both cover the same symptoms with corresponding response scales so that a direct comparison between these two perspectives can be undertaken in the evaluation of depressive symptoms. Both the IDS-C and IDS-SR were developed to provide a more sensitive measure than the HRS-D for out-patients with MDD, since our clinical experience suggested that some significant depressions lead to only modest total scores on the 17-item HRS-D (e.g. 10 to 16). All criterion diagnostic symptoms for a major depressive episode are included. The wording of scale choices reflects the DSM focus on frequency of symptoms rather than the focus on intensity found in the HRS-D and BDI. Selected items were not included because of their infrequent presence in depressed patients or their relative non-specificity to depressive syndromes (e.g. loss of insight, depersonalization, obsessive–compulsive symptoms and paranoid ideation). All items are equally weighted using a 0- to 3-point scale. Finally, the 30-item versions include all melancholic and atypical criterion symptoms for DSM-IV, which provides for symptom measurement in various forms of depression.

This report provides an analysis of the psychometric properties of both the 28- and 30-item versions of the IDS-C and IDS-SR in a large, previously unreported out-patient sample. In order to assure that variation in MDD symptomatology is the primary factor characterizing our sample, all subjects had to be medication-free, and subjects with concurrent Axis I diagnoses were excluded.

METHOD

Subjects

Subjects were drawn from a large database maintained by the Mental Health Clinical Research Center (MHCRC), which includes patients who underwent extensive evaluation in the Mood Disorders Program at UT Southwestern Medical Center in Dallas, Texas (1984–1994). Subjects were recruited by word of mouth, self-referral, referral from practitioners, and advertisements for various psychotherapeutic and pharmaceutical trials, as well as laboratory-based investigations of mood disorders. Subjects who passed initial screening with no obvious general medical conditions and apparently significant levels of depressive symptoms, and who appeared to meet criteria for one or more ongoing protocols then participated in a structured clinical interview to establish diagnoses by DSM-III-R (APA, 1987) and Research Diagnostic Criteria (RDC) (Spitzer *et al.* 1978).

To be included in the present sample, each subject had to have a complete 28- or 30-item IDS-C and IDS-SR collected within 2 days of each other. The 17-item HRS-D had to be collected within ± 2 days of the IDS-C and the BDI had to be collected within ± 2 days of the IDS-SR. Depressed subjects included in these analyses met criteria for MDD (by DSM-III-R), were free of unstable general medical conditions and any general medical disorder thought to cause mood symptoms (e.g. untreated thyroid disorder) and had no other current Axis I diagnoses. Also excluded were subjects taking any psychotropic medication within 14 days prior to the full evaluation.

All normal controls were evaluated by a screening battery identical to that used for patients. To enter the normal control group,

subjects were required to have no lifetime personal history or history in first-degree relatives of any Axis I psychiatric disorder, no substance abuse or dependency diagnosis, no current general medical condition and had to be free of all medications for at least 2 weeks.

The remitted depressed group was composed of patients who presented for initial evaluation with a Structured Clinical Interview for DSM-III-R (SCID) (Spitzer *et al.* 1992) diagnosis of one or more prior episodes of major depression, but who were currently in the remitted or recovered state. All remitted subjects were medication-free for at least 2 weeks prior to evaluation with the IDS.

A subgroup of the depressed patients was used to evaluate the sensitivity of the IDS-C and IDS-SR to change during treatment. These subjects ($N = 58$) differ from the larger sample in that the age range was restricted to 18–50 years and a minimum of 16 on the 17-item HRS-D was required. Excluded were subjects with DSM-III-R diagnoses of concurrent organic mental disorders, substance abuse or dependence disorders within the last year, lifetime delusional or psychotic disorders (e.g. schizophrenia), antisocial personality disorder, or a history of three or more suicide attempts without concurrent MDD, melancholic type. Finally, all ratings had to be obtained within ± 3 days of the scheduled weekly visit.

Analysis procedures

The two forms of the IDS were developed to: (*1*) assess depressive symptom severity; (*2*) discriminate between depressed and euthymic states; and (*3*) provide sensitive measures of symptom change during treatment.

The primary use of these two instruments is to measure overall depressive symptom severity. The ideal instrument would be a unifactorial measure of this construct. This is evaluated by factor analysing each instrument and computing Cronbach's α (Cronbach, 1951) (a measure of the homogeneity of items) for both the whole instrument and for subsets of items defining any separate factor. It should also have satisfactory concurrent validity, which was evaluated by correlations with other commonly used measures of depressive symptom severity (i.e. the HRS-D and the BDI), and the depressive severity rating coded as the fifth digit on the DSM-III-R diagnostic code. Finally, there should be high interrater reliability, indicating that raters are measuring the same construct. This was assessed via intraclass correlations (Lahey *et al.* 1983) among trained clinical raters on a sample of independently rated videotaped interviews using the IDS-C and HRS-D.

The total score should distinguish between currently symptomatic patients and euthymic controls, which is tested by *t* test. Also, the total score should be independent of demographic features, except in so far as these appear to be general associations with syndrome features. These were evaluated by analysing the associations of total scores with age, education, gender and marital status.

Evaluation of depressive severity in treatment trials often requires deciding whether a patient is above, or below, a cut-point indicating, for instance, recovery, relapse, or recurrence (Frank *et al.* 1991). We determined the optimal threshold for distinguishing currently symptomatic from euthymic subjects using a receiver operating characteristic (ROC) analysis (Kraemer, 1988).

Finally, the sensitivity of each measure to change during treatment was evaluated in the fluoxetine-treated patient group. Sensitivity to change is more relevant in a measure of variable symptom state than is test–retest reliability. Data included between-visit differences on total score for each instrument and pair of visits for 10 weekly visits. The differences for a given pair of visits were correlated between the IDS-C and HRS-D over patients, and between the IDS-SR and HRS-D. Parallel analyses were carried out on the IDS-C and on the IDS-R. For comparison purposes, some of these analyses were also conducted on the HRS-D and BDI in the same sample of subjects.

RESULTS

Of 1306 subjects in the MHCRC database, 552 met inclusion criteria; 455 of these were excluded because of Axis I co-morbidity. All of the 552 subjects completed the first 28-items on the IDS-C and IDS-SR; of these, 456 also completed both items 29 and 30 of the 30-item IDS-C and IDS-SR. One hundred and three subjects qualified as normal controls and 15 qualified as euthymic, formerly depressed patients. All

Table 1. *Clinical and demographic features of the 28-item IDS patient sample*

Measure	Symptomatic MDD ($N = 434$) Measure or % (S.D.)	Remitted MDD ($N = 15$) Mean or % (S.D.)	Controls ($N = 103$) Mean of % (S.D.)
Age (years)	39·4 (10·8)	34·2 (11·6)	31·6 (9·8)
Mean IDS-C Score (28-item)	33·3 (7·3)	9·1 (6·3)	2·3 (2·4)
Mean IDS-SR Score (28-item)	36·4 (8·9)	16·1 (10·2)	4·4 (3·7)
Mean HRS-D Score	19·3 (4·4)	4·7 (3·0)	1·5 (1·6)
Mean BDI Score	25·6 (8·4) ($N = 421$)	7·9 (7·6)	0·9 (1·7)
Caucasian (%)	89·8 ($N = 432$)	100·0	83·5
Married or cohabitating (%)	48·5 ($N = 431$)	46·7	50·5
Female (%)	64·6 ($N = 432$)	73·3	50·5
Age at onset (years)	28·2 (12·1) ($N = 424$)	22·9 (6·6)	NA
Number of episodes	2·2 (1·4) ($N = 411$)*	3·0 (3·9) ($N = 14$)†	NA
Length of illness (years)	11·1 (10·4) ($N = 424$)	11·3 (9·4)	NA

* 18 reported too many to count; † 1 reported too many to count.
IDS-C = Inventory of Depressive Symptomatology-Clinician-rated; IDS-SR = Inventory of Depressive Symptomatology–Self-Report; HRS-D = Hamilton Rating Scale for Depression; BDI = Beck Depression Inventory; MDD = major depressive disorder; NA = not available.

controls and remitted subjects completed both versions of the IDS.

Sample description

Table 1 shows the clinical and demographic features of the sample. Of the currently symptomatic ($N = 434$), 23·1% met DSM-IV criteria for atypical symptom features, 34·6% met RDC for definite endogenous features (both diagnosed at the nadir of the episode), and 17·1% met DSM-III (APA, 1980) criteria for melancholia. Nearly two-fifths (39·6%) were in their first major depressive episode and 90·6% had primary major depression by RDC. Subjects ranged in age from 12 to 71; 90% fell in the age range of 22 to 57.

Internal consistency

Cronbach's α was used to estimate the internal consistency of the 28- and 30-item versions for both the IDS-C and IDS-SR forms, as well as for the BDI and 17-item HRS-D (see Table 2). Results for both the total sample ($N = 552$) and currently symptomatic patients ($N = 434$) are shown. The IDS-C, IDS-SR and BDI have approximately equivalent internal consistency when measured in all subjects, while α for the HRS-D was noticeably worse in every sample examined. The lower part of the table gives α for all four instruments in the two subgroups that completed the 30-item versions of the IDS.

Inter-rater reliability

The inter-rater reliability of the two clinician-rated measures was calculated using intraclass correlations with raters retained in the error term. This resulted in intraclass correlations of 0·92 for the HRS-D and 0·96 for the 30-item IDS-C.

Factor structure

Two principal components analyses with VARIMAX rotation, one each on the IDS-C and IDS-SR, were run in a subset of patients who completed the 30-item versions ($N = 353$).

Table 2. *Cronbach's α for each rating instrument and sample*

	IDS-C	IDS-SR	HRS-D	BDI
28-item version				
All subjects ($N = 552$)	0·92	0·93	0·88	0·93*
Symptomatic only ($N = 434$)	0·67	0·77	0·56	0·83†
30-item version				
All subjects ($N = 456$)	0·94	0·94	0·89	0·94‡
Symptomatic only ($N = 338$)	0·67	0·77	0·53	0·83§

* $N = 539$; † $N = 421$; ‡ $N = 447$; § $N = 329$.
IDS-C = Inventory of Depressive Symptomatology–Clinician-Rated; IDS-SR = Inventory of Depressive Symptomatology–Self-Report; HRS-D = Hamilton Rating Scale for Depression; BDI = Beck Depression Inventory.

Table 3. *Results of principal components analysis of the 30-item IDS-C in symptomatic and remitted major depressive disorder* (N = 353)*

	Factor 1	Factor 2	Factor 3
Interest in people/activities (19)	**77**	2	13
Pleasure/enjoyment (not sex) (21)	**75**	4	−6
Reactivity of mood (8)	**68**	0	−12
Feeling sad (5)	**65**	12	18
Energy/fatiguability (20)	**61**	13	19
Concentration/decision making (15)	**57**	12	21
Interest in sex (22)	**48**	−19	−7
Quality of mood (10)	**45**	−23	−24
Future pessimism (17)	**43**	28	18
Suicidal thoughts (18)	**42**	25	0
Psychomotor retardation (23)	**40**	10	12
Self-criticism and blame (16)	**36**	23	29
Interpersonal sensitivity (29)	**35**	16	29
Feeling irritable (6)	**35**	23	4
Sympathetic arousal (26)	12	**55**	12
Psychomotor agitation (24)	4	**47**	−12
Constipation/diarrhoea (28)	6	**47**	3
Panic/phobic symptoms (27)	1	**44**	26
Middle insomnia (2)	**34**	**44**	**−43**
Feeling anxious or tense (7)	19	**38**	1
Aches and pains (25)	6	**38**	29
Initial insomnia (1)	25	**37**	−29
Diurnal variation of mood (9)	−5	17	−2
Sleeping too much (4)	4	−21	**66**
Weight change (13, 14)	9	13	**46**
Leaden paralysis/physical energy (30)	26	9	**45**
Appetite disturbance (11, 12)	26	11	**42**
Early morning awakening (3)	21	29	**−41**
Cronbach's α for factor	0·80	0·54	0·33

* Printed values × 100 and rounded to the nearest integer. Item number appears in parentheses after each item label. Bold-face type indicates those items with higher factor loadings.

IDS-C = Inventory of Depressive Symptomatology – Clinician-Rated.

The item correlations in both instruments appear to be well accounted for by three factors (Tables 3 and 4). The first factor on both instruments appears to be a cognitive/mood factor. The second factor agrees less clearly between instruments, but in each, may be interpreted as an anxiety/arousal factor. The third factor is not identical in the two instruments and, although both appear to include sleep regulation, the IDS-SR also includes appetite regulation and leaden paralysis.

Concurrent validity

The total scores on the IDS-C, IDS-SR, HRS-D and BDI were correlated with clinical severity as gauged by the DSM-III-R fifth digit. The fifth digit takes into account both symptom severity and overall disability of the current episode, while the most recent 1 week is rated by all four rating scales. In the 448 subjects with current or remitted MDD and a fifth digit severity code there were 15 fully remitted, 13 partially remitted, 11 mildly ill, 303 moderately ill, 101 severely ill and five psychotic patients. In the subsample with 30-item IDS scores ($N = 352$) there were 224 moderately depressed, 85 severely depressed and four psychotic patients. Correlations with the fifth digit in the 352 subjects were 0·68 for the HRS-D, 0·48 for the BDI, 0·62 for the IDS-C (30-item), and 0·54 for the IDS-SR (30-item). In the 448 subjects with 28-item scores, the correlations were 0·57 for the IDS-C and 0·49 for the IDS-SR.

A second concurrent validation was provided by obtaining Pearson product moment correlations between the two IDS total scores and the HRS-D and BDI scores in the whole sample (Table 5). These two different perspectives (clinician ad self-report) shared slightly less variance than either single perspective measured by two different instruments. On the other hand, there was a strong enough relationship between each pair of rating scale totals to suggest concurrent validity. Since the IDS-C and IDS-SR

Table 4. *Results of principal components analysis of the 30-item IDS-SR in symptomatic and remitted major depressive disorder* (N = 353)*

	Factor 1	Factor 2	Factor 3
Interest in people/activities (19)	**70**	13	−3
Pleasure/enjoyment (not sex) (21)	**70**	9	11
Feeling sad (5)	**61**	29	18
Future pessimism (17)	**55**	17	20
Reactivity of mood (8)	**54**	2	16
Self-criticism and blame (16)	**52**	24	14
Appetite disturbance (11, 12)	**49**	14	−27
Energy/fatiguability (20)	**47**	**42**	−32
Suicidal thoughts (18)	**44**	2	32
Interest in sex (22)	**42**	1	0
Quality of mood (10)	**39**	5	9
Weight change (13, 14)	**36**	1	−14
Sympathetic arousal (26)	−4	**67**	11
Feeling anxious or tense (7)	6	**60**	19
Aches and pains (25)	1	**60**	−7
Panic/phobic symptoms (27)	2	**54**	20
Feeling irritable (6)	23	**53**	−6
Psychomotor retardation (23)	25	**51**	−5
Psychomotor agitation (24)	3	**50**	**36**
Leaden paralysis/physical energy (30)	25	**50**	−27
Concentration/decision making (15)	**38**	**44**	−16
Interpersonal sensitivity (29)	**37**	**43**	3
Constipation/diarrhoea (28)	14	**32**	1
Diurnal variation of mood (9)	−16	21	5
Early morning awakening (3)	16	8	**64**
Middle insomnia (2)	20	4	**59**
Initial insomnia (1)	14	24	**42**
Sleeping too much (4)	14	8	−47
Cronbach's α for factor	0·77	0·75	0·39

* Printed values × 100 and rounded to the nearest integer. Item number is in parentheses after each item label. Bold-face type indicates those items with higher factor loadings.

IDS-SR = Inventory of Depressive Symptomatology – Self-Report.

are based on identical item content, the self-report and clinician perspectives are somewhat different when item content is controlled. The IDS-SR total score tends to be 3–4 points higher than the IDS-C total in general.

Differentiation between symptomatic patients and euthymic subjects

Another test of the validity of the 28-item IDS-C and IDS-SR is the degree to which the total score discriminates between symptomatic depressed patients and euthymic subjects. Those symptomatic with major depressive disorder ($N = 434$) were compared to euthymic subjects ($N = 118$) utilizing the 28-item IDS-C and IDS-SR. Both scales significantly discriminated between the two groups ($P < 0{\cdot}0001$). Similarly for the 30-item version, a total of 338 symptomatic patients were compared to 118 euthymic subjects. Again, both scales significantly discriminated between the groups ($P < 0{\cdot}0001$).

Association with patient characteristics

No instrument showed a significant relationship with either age or marital status. Very small, although significant, negative correlations were found with education on the HRS-D (−0·22), BDI (−0·17) and 28-item IDS-C (−0·15). Men, compared with women, scored lower on both the BDI and 28-item IDS-SR. The 30-item IDS-C and IDS-SR ratings were unrelated to age, gender, marital status, or education.

Sensitivity to treatment effects

To determine whether the 30-item IDS-C and IDS-SR detect symptom change during acute phase treatment, we compared each with the most widely accepted measure, the HRS-D. Fig. 1 shows the weekly means for the IDS-C, HRS-D and IDS-SR over 10 weeks of acute treatment in the fluoxetine trial (BDI not available). As indicated at the bottom of the figure, the

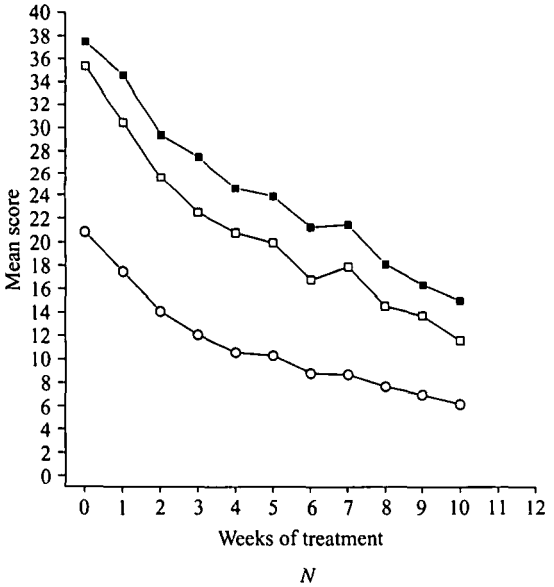
Table 5. *Correlations between rating scale totals in the total samples for the* 28*-item* (N = 552) *and* 30*-item* (N = 456) *versions of the IDS**

	28-item IDS-C	28-item IDS-SR	HRS-D	BDI
28-item IDS-C	1·0	—	—	—
28-item IDS-SR	0·88	1·0	—	—
HRS-D	0·94	0·85	1·0	—
BDI†	0·83	0·92	0·83	1·0
	30-item IDS-C	**30-item IDS-SR**	**HRS-D**	**BDI**
30-item IDS-C	1·0	—	—	—
30-item IDS-SR	0·91	1·0	—	—
HRS-D	0·95	0·88	1·0	—
BDI‡	0·86	0·93	0·85	1·0

* All correlations significant at $P < 0{\cdot}0001$.
† $N = 539$, ‡ $N = 447$.
IDS-C = Inventory of Depressive Symptomatology–Clinician-Rated; IDS-SR = Inventory of Depressive Symptomatology–Self-report; HRS-D = Hamilton Rating Scale for Depression; BDI = Beck Depression Inventory.

Table 6. *Association between diagnosis of current symptomatic major depressive disorder* (*D*+ *or D*−) *and each test's assignment of patients to symptomatic* (*T*+) versus *euthymic* (*T*−) *groups*

Test		Diagnosis D+	D−
30-item IDS-C			
Optimal cut ⩾ 13			
Specificity = 0·96	T+	337	5
Sensitivity = 1·0	T−	1	113
30-item IDS-SR			
Optimal cut ⩾ 18			
Specificity = 0·94	T+	337	7
Sensitivity = 1·0	T−	1	111
HRS-D			
Optimal cut ⩾ 8			
Specificity = 0·99	T+	335	1
Sensitivity = 0·99	T−	3	117
BDI ($N = 447$)			
Optimal cut ⩾ 6			
Specificity = 0·93	T+	329	8
Sensitivity = 1·0	T−	0	110



Week	0	1	2	3	4	5	6	7	8	9	10
N											
HRS-D	59	59	55	53	52	51	48	46	49	44	43
IDS-C	59	59	55	53	52	51	48	46	49	44	43
IDS-SR	58	58	55	53	52	51	48	45	49	45	42

FIG. 1. Symptom change over the course of acute phase treatment with fluoxetine. (○, HRS-D 17-item; □, IDS-C 30 item; ■, IDS-SR 30-item.)

sample size declines over weeks of treatment. The three curves are nearly parallel, differing primarily in level, which reflects the different total score ranges of these instruments.

A more stringent evaluation of similarity is whether change on the HRS-D and IDS-C or IDS-SR from one visit to the next is correlated over subjects. We estimated slope and intercept from the regressions of change in IDS-C or IDS-SR on change in HRS-D for each pair of weeks, and the correlation coefficient for the same data. The 95% confidence interval for both sets of slope estimates ranged over weeks from $\pm 0{\cdot}30$ to $\pm 0{\cdot}44$. The IDS-C and HRS-D were more strongly related than were the IDS-SR and the HRS-D. In general, the regression slopes were smaller for the IDS-SR than for the IDS-C. The IDS-C score dropped by about 4 points for every drop of 3 points in HRS-D score. The IDS-SR score dropped at about the same rate as the HRS-D. These results are consistent with the notion that the IDS-C may be the most sensitive of the three measures to symptomatic change.

Optimal cut-point

Finally, a ROC analysis was conducted on each of the four instruments to identify the optimal total score thresholds to distinguish between

symptomatic and euthymic subjects for each scale. Table 6 shows the cut-point with a combination of optimal sensitivity and optimal specificity for each of the four instruments. This information may be useful in determining whether or not a patient has remitted.

DISCUSSION

Both the 28- and 30-item versions of the IDS-C and IDS-SR reveal adequate face validity, a complete coverage of criterion diagnostic symptoms, reasonable internal consistency (e.g. Cronbach's α), interrater reliability, and concurrent and discriminant validity.

Both clinician and self-report forms of the IDS 28- and 30-item versions related moderately to the DSM-III-R fifth digit severity coding. These also discriminated highly between symptomatic and euthymic groups. However, because of recruitment procedures, few symptomatic patients scored below 14 on the HRS-D and most of the 448 subjects were rated on the fifth digit code as either moderate ($N = 304$) or severely ($N = 101$) ill, resulting in reduced covariation with instrument total score. These factors may also have artefactually increased the discrimination between out-patients with MDD and normal controls. For the same reason, the threshold values set in the ROC analysis may be somewhat higher than would have been found if the depressed sample had not been selected on minimum severity. The IDS was designed to match closely the current thinking regarding the symptomatology of MDD and its major subtypes (melancholic, atypical). This includes rating symptoms on amount of time experienced rather than intensity; this is also the approach used in diagnosis by DSM-III-R and DSM-IV criteria. Even though items on the IDS-C and IDS-SR differ from those on the HRS-D and BDI in what aspect of symptoms they measure, the concurrent validity index is above 0·90 for same-rater instruments.

Total scores on the 30-item versions were not associated with age, gender, education, or marital status, while only gender was very slightly related to the BDI and IDS-SR (28-item) total score. This association with gender has been previously reported for the BDI (Beck *et al.* 1988). The negative correlations of the HRS-D, BDI and IDS-C (28-item) with education is also consistent with other reports (Beck, 1967; Beck & Beamsderfer, 1974; Blumenthal, 1975; Dorus & Senay, 1980; Oliver & Simmons, 1984), which support the conclusion that these associations may be characteristics of the population studied rather than flaws in a specific instrument.

Two similar factors were substantiated for both forms in the 30-item versions of the IDS: a cognitive/mood factor and an anxiety/arousal factor. Varying items reflecting sleep, appetite and weight regulation, suggest a third factor that is not identical for the clinician and self-report versions.

Findings of a division between depressive and vegetative features common to both the IDS-C and IDS-SR are analogous to other factor analyses of the HRS-D (Rhoades & Overall, 1983) and the BDI (Gibbons *et al.* 1985; Young *et al.* 1986; Bouman & Kok, 1987), although differences exist among these studies and scales with regard to which items contribute to each particular factor across these reports.

Results of the principal components analysis suggest that the following items are candidates for creation of a shorter, single factor version of each rating scale: 'Feeling sad' (5); 'Reactivity of mood' (8); 'Quality of mood' (10); 'Concentration/decision making' (15); 'Self-criticism and blame' (16); 'Future pessimism' (17); 'Suicidal thoughts' (18); 'Interest in people/activities' (19); 'Energy/fatiguability' (20); 'Pleasure/enjoyment (not sex)' (21); 'Interest in Sex' (22); and 'Appetite disturbance' (11, decrease or 12, increase). It would also be entirely reasonable to use the items on the first two factors to compute two subscores, which would then provide somewhat more homogeneous measures of these target symptoms. However, the high internal consistency suggests that the loss of sensitivity resulting from combining all items in a total score is negligible. Further studies to evaluate whether a briefer version of the IDS, using selected items from both the clinician-rated and self-reported format, could be used to assess outcome in primary care practice are indicated, even though briefer versions would reduce item coverage, especially across subtypes.

There are, as well, limitations to the present study. First, the same rater completed both the

HRS-D and IDS-C based on the same interview. While, in general, the HRS-D was completed prior to the IDS-C, the interviewer was not blind to the results of one scale compared with the other. This procedure may have resulted in an increased correlation between the two measures that may not have been apparent if different blinded interviewers had completed each scale. Also, the inclusion of the euthymic group increases the variation on a general factor of severity and inflates the correlation; this effect is also apparent in Table 2 (Cronbach's α).

Secondly, the sample, while representative of depressed psychiatric out-patients, did not include many mildly depressed or psychotic patients neither did it include patients with general medical conditions and mood symptoms, those with bipolar, depressive disorders not otherwise specified, or those with concurrent mood and other major Axis I psychiatric disorders (e.g. panic disorder, eating disorder, obsessive–compulsive disorder). Thus, while the current sample was selected to optimally evaluate psychometric properties in MDD, generalization to other mood disordered samples requires further evaluation.

In summary, both the 28- and 30-item versions of the IDS-C and IDS-SR have psychometric properties that allow recommendation of their use in research, as well as in clinical practice with depressed patients. The self-report and clinician-rated perspectives each contribute related, but also to some degree distinct, information about the nature of the symptomatology. The broader item coverage of the IDS as compared to the HRS-D or BDI ensures an adequate assessment of all symptoms currently used to define MDD, as well as the symptoms of both the melancholic and atypical subtypes. The apparently greater lower range of symptom severity and the development of the IDS on out-patients, may also argue for a greater sensitivity to less severely ill patients than at least the standard 17-item HRS-D can provide. The more complete atypical symptom feature coverage by the 30-item versions of the IDS-C and IDS-SR would presently make them preferable to their 28-item counterparts for future research.

The authors wish to express their appreciation to David Savage for secretarial assistance, Mary Fitch and Andrea Simmons for data analytical support, Dr Kenneth Z. Altshuler, Stanton Sharp Distinguished Chair and Chairman, Department of Psychiatry, for administrative support, and to the many clinicians who assisted in the collection of these data including Drs John Battaglia, John W. Cain, John Debus, G. Gregory Eaves, Donna E. Giles, Rodger Kobes and Paul Silver. This work was supported in part by a grant from the National Institute of Mental Health (MH-41115; MH-35370; MH-38238; MH-45043) and by contracts for pharmaceutical development from Eli Lilly and Company, SmithKline Beecham, Bristol-Myers Squibb and Pfizer, Inc.

REFERENCES

American Psychiatric Association (1980). *Diagnostic and Statistical Manual of Mental Disorders, 3rd edn.* APA: Washington, DC.

American Psychiatric Association (1987). *Diagnostic and Statistical Manual of Mental Disorders, 3rd edn. Revised.* APA: Washington, DC.

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders, 4th edn.* APA: Washington, DC.

Beck, A. T. (1967). *Depression: Causes and Treatment.* University of Pennsylvania Press: Philadelphia.

Beck, A. T. & Beamesderfer, A. (1974). Assessment of depression: the depression inventory. In *Modern Problems in Pharmacopsychiatry, vol. 7* (ed. P. Pichot), pp. 151–169. Karger: Basel.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. E. & Erbaugh, J. K. (1961). An inventory for measuring depression. *Archives of General Psychiatry* **4**, 561–571.

Beck, A. T., Rush, A. J., Shaw, B. F. & Emery, G. (1979). *Cognitive Therapy of Depression.* Guilford Press: New York.

Beck, A. T., Steer, R. A. & Garbin, M. M. (1988). Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. *Clinical Psychology Review* **8**, 77–100.

Blumenthal, M. D. (1975). Measuring depressive symptomatology in a general population. *Archives of General Psychiatry* **32**, 971–978.

Bouman, T. K. & Kok, A. R. (1987). Homogeneity of Beck's depression inventory (BDI): Applying Rasch analysis in conceptual exploration. *Acta Psychiatrica Scandinavia* **76**, 568–573.

Carroll, B. J., Feinberg, M., Smouse, P. E., Rawson, S. G. & Greden, J. F. (1981). The Carroll Rating Scale for depression. I: Development, reliability and validity. *British Journal of Psychiatry* **138**, 194–209.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* **16**, 297–309.

Dorus, W. & Senay, E. C. (1980). Depression, demographic dimensions, and drug abuse. *American Journal of Psychiatry* **137**, 689–704.

Frank, E., Prien, R., Jarrett, R. B., Keller, M. B., Kupfer, D. J., Lavori, P., Rush, A. J. & Weissman, M. M. (1991). Conceptualization and rationale for consensus definitions of response, remission, recovery, relapse and recurrence in major depressive disorder. *Archives of General Psychiatry* **48**, 851–855.

Gibbons, R. D., Clark, D. C., Cavanaugh, S., Von Ammon Cavanaugh, S. & Davis, J. M. (1985). Application of modern psychometric theory in psychiatric research. *Journal of Psychiatric Research* **19**, 43–45.

Gibbons, R. D., Clark, D. C. & Kupfer, D. J. (1993). Exactly what does the Hamilton Depression Rating Scale measure? *Journal of Psychiatric Research* **27**, 259–273.

Grundy, C. T., Lunnen, K. M., Lambert, M. J., Ashton, J. E. & Tovey, D. R. (1994). The Hamilton Rating Scale for Depression: one scale or many? *Clinical Psychology: Science and Practice* **1**, 197–205.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry* **23**, 56–62.

Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology* **6**, 278–296.

Hedlund, J. L. & Vieweg, B. W. (1979). The Hamilton Rating Scale for Depression: a comprehensive review. *Journal of Operational Psychiatry* **10**, 149–165.

Kraemer, H. C. (1988). Assessment of 2 × 2 associations: generalizability of signal-detection methodology. *American Statistician* **42**, 37–49.

Lahey, R. A., Downey, R. B. & Saal, F. E. (1983). Intraclass correlations: there's more here than meets the eye. *Psychological Bulletin* **93**, 586–595.

Nelson, J. C. & Mazure, C. M. (1990). A scale for rating tricyclic response in major depression: The TRIM. *Journal of Clinical Psychopharmacology* **10**, 252–260.

Oliver, J. M. & Simmons, M. E. (1984). Depression as measured by the DSM-III and the Beck Depression Inventory in an unselected adult population. *Journal of Consulting and Clinical Psychology* **52**, 892–898.

Prien, R. F., Carpenter, L. L. & Kupfer, D. J. (1991). The definition and operational criteria for treatment outcome of major depressive disorder. A review of the current research literature. *Archives of General Psychiatry* **48**, 796–800.

Rehm, L. P. & O'Hara, M. W. (1985). Item characteristics of the Hamilton Rating Scale for Depression. *Journal of Psychiatric Research* **19**, 31–41.

Rhoades, H. M. & Overall, J. E. (1983). The Hamilton depression scale: factor scoring and profile classification. *Psychopharmacology Bulletin* **19**, 91–96.

Rush, A. J., Giles, D. E., Schlesser, M. A., Fulton, C. L., Weissenburger, J. E. & Burns, C. T. (1986). The Inventory for Depressive Symptomatology (IDS): preliminary findings. *Psychiatry Research* **18**, 65–87.

Spitzer, R. L., Endicott, J. & Robins, E. (1978). Research diagnostic criteria: rationale and reliability. *Archives of General Psychiatry* **36**, 773–782.

Spitzer, R. L., Williams, J. B. W., Gibbon, M. & First, M. B. (1992). The Structured Clinical Interview for *DSM-III-R* (SCID) I: History, rationale, and description. *Archives of General Psychiatry* **49**, 624–629.

Young, M. A., Scheftner, W. A., Klerman, G. L., Andreasen, N. C. & Hirschfeld, R. M. A. (1986). The endogenous sub-type of depression: a study of its internal construct validity. *British Journal of Psychiatry* **148**, 257–267.