ELSEVIER

Journal of Affective Disorders 71 (2002) 85–96

JOURNAL OF AFFECTIVE DISORDERS

www.elsevier.com/locate/jad

Research report

Rating dysthymia: an assessment of the construct and content validity of the Cornell Dysthymia Rating Scale☆

David J. Hellerstein[a,b,*], Sarai T. Batchelder[b], Agnes Lee[c], Margarita Borisovskaya[c]

[a]*New York State Psychiatric Institute, New York, NY, USA*
[b]*Department of Psychiatry, Beth Israel Medical Center, New York, NY, USA*
[c]*Albert Einstein College of Medicine, Bronx, NY, USA*

Received 19 October 1999; accepted 7 May 2001

Abstract

Background: Mason et al. developed the Cornell Dysthymia Rating Scale (CDRS), a 20-item clinician-rated inventory, and hypothesized that it may be superior to the commonly-used Hamilton Depression Rating Scale (HDRS) in assessing the symptoms of dysthymia, a form of chronic depression. The purpose of this study was to compare these instruments in an outpatient sample of dysthymic patients. *Method*: The CDRS and the HDRS and other inventories (including the Hopkins Symptom Check List (SCL)) were administered to 110 patients meeting DSM-IIIR diagnosis of dysthymia. *Results:* There was a significant correlation between the CDRS and the HDRS at baseline and termination, indicating concurrent validity. Distributional statistics were compared for baseline and termination severity scores, showing that the CDRS has greater severity range scores than the HDRS. Furthermore, results of the DSM-IV Mood Disorders Field Trial suggest that the CDRS has better content validity than the HDRS when it comes to the dysthymic population. *Limitations:* The results are limited by the use of a homogeneous sample, the absence of observer ratings of divergent symptoms, and less than excellent validity of self-report divergent symptoms. *Conclusions:* Our results support the value of the CDRS in assessing symptoms of dysthymia.

Keywords: Dysthymic disorder; Rating scales; Test validation

1. Introduction

In recent years, data from a number of sources (Robins and Regier, 1991; Barrett et al., 1988; Goldberg, 1985; Wells et al., 1989) have demonstrated that dysthymia (American Psychiatric Association, 1994), a form of chronic depression, is a significant public health problem, affecting approxi-

☆Previously presented at the 1999 American Psychiatric Association Annual Meeting, Washington, DC.

*Corresponding author. New York State Psychiatric Institute, Unit no. 101, 1051 Riverside Drive, New York, NY 10032, USA. Tel.: +1-212-543-5790; fax: +1-212-543-5674.

E-mail address: hellers@pi.cpmc.columbia.edu (D.J. Hellerstein).

0165-0327/02/$ – see front matter
PII: S0165-0327(01)00371-8

mately 3% of the adult population (Weissman et al., 1988). Social impairment in dysthymia is significant (Friedman, 1993), with a notable impact on family life, intimate relationships and social activities.

Despite the prevalence of this disorder, there are few scales designed specifically to assess the symptoms of chronic depression. The instrument most commonly used to assess depressive symptoms, the Hamilton Depression Rating Scale (HDRS) (Hamilton, 1960), was developed for the assessment of the acute and episodic disorder of major depression, and includes questions for symptoms uncommon in dysthymia (e.g. psychomotor retardation/agitation, decreased appetite, paranoia). Mason et al. (1993) developed the Cornell Dysthymia Scale, intended to be sensitive to the chronic and less severe symptomatology of dysthymia. Mason et al. (1995) found that the CDRS showed greater breadth in the range of individual item and sum scores than the HDRS when measuring symptoms for dysthymic disorder. An NIMH conference recently identified a need to assess the sensitivity of scales used to assess dysthymia (Gwirtsman et al., 1997). Gwirtsman et al. note that the most common symptoms of dysthymia include low self-esteem, pessimism and hopelessness, rather than neurovegetative signs such as insomnia or appetite disturbance. The accurate assessment of symptoms seems particularly important given the increasing recognition of dysthymia and the widespread use of SSRI and other new classes of medications for treatment of this disorder.

2. Method

2.1. Subjects

The sample consists of 110 outpatients from four clinical drug trials for dysthymic disorder that were conducted in the Beth Israel Medical Center Mood Disorders Research Unit. In all studies, subjects underwent standardized assessment batteries with trained raters and clinicians that had been trained on the SCID, the CDRS and the HDRS. The studies included here are three open-label studies of (1) fluoxetine vs. trazodone ($n = 17$); (2) fluoxetine plus group therapy vs. fluoxetine alone ($n = 39$), (3) venlafaxine ($n = 21$), as well as a double-blind trial of fluoxetine vs. placebo ($n = 33$). All protocols were approved by the Beth Israel Medical Center Institutional Review Board, and all subjects provided informed consent after study procedures had been explained to them.

All patients met DSM-IIIR criteria for dysthymia, based on a Structured Clinical Interview for DSM-IIIR (SCID; Spitzer et al., 1987) and were between 18 and 65 years old. Patients who also met criteria for Major Depressive Disorder were excluded from the study. Other exclusion criteria included DSM-IIIR or DSM-IV diagnoses of schizophrenia, psychotic disorders, substance abuse, bipolar disorder, delirium, dementia, or a primary diagnosis of panic disorder, generalized anxiety disorder, or obsessive–compulsive disorder. Patients with serious medical illness or strong suicidal feelings were also excluded. Only one study (open-label venlafaxine, $n = 21$) had an inclusion criteria of an HDRS score of at least 14 at baseline. The mean age of the subjects was 41 ± 10 years and 56% were female.

Fourteen (12.7%) of the subjects dropped out of their study prior to completion of the protocol, and thus the sample at termination contained 96 subjects. Subject were primarily Caucasian (87.8%), college educated (78.2%) and employed (83.6%). Over half the subjects had never married (59.6%), with 20.9% currently married and 19.3% separated or divorced.

2.2. Measures

The following tests were administered to all subjects at baseline and at the time of termination: the Cornell Dysthymia Rating Scale (CDRS; Mason et al., 1993); the Hamilton Rating Scale for Depression (HDRS: Hamilton, 1960); the Clinical Global Impression Scale (CGI; Guy, 1976); and the Hopkins Symptom Check List (SCL; Derogatis, 1974). Two versions of the HDRS were used: the 21-item version ($n = 18$) and the 24-item version ($n = 92$). To maximize the amount of data available, we used the 21-item HDRS for analyses using HDRS scores and used all items available for item-by-item analyses. We also used two versions of the SCL: the 90-item (Derogatis, 1983) and the 58-item (Derogatis, 1974) versions. Thus, again to maximize the available data, we transformed the raw SCL scores to *T* scores using the psychiatric outpatient norms for these scales.

This made the data obtained from the two measures comparable and *T* scores were used in all analyses.

2.3. Construct validity

Construct validity refers to the degree to which a measure is actually measuring what it purports to measure. There are four aspects to construct validity: content validity, concurrent validity (both convergent and divergent validity) and predictive validity

1. Content validity is the degree to which a measure is considered to cover a representative sample of the content in question. Specifically, adequate content validity of the CDRS would require that the range of symptoms present in dysthymic disorder were assessed by that measure. Internal consistency is the degree to which all items of a measure are associated with each other
2. Convergent validity refers to the degree to which other measures of the same or similar constructs are positively associated with the scale
3. Divergent validity is the degree to which the scale does not measure other, dissimilar constructs
4. Predictive Validity refers to the ability of the measure to predict outcomes related to the variable in question.

2.4. Statistical analyses

Internal consistency was assessed using Cronbach's α coefficient. For construct validity, the following analyses were used

1. Content validity was assessed by comparing means, standard deviations and frequency distributions of ratings on items of the CDRS and HDRS. Symptom ratings on these two scales were compared with data from the DSM-IV Field Trial. In addition, the frequency of symptoms in the DSM-IV Field Trial was compared with the frequency of symptoms in our sample as measured by both scales using Spearman rank order correlations
2. Convergent validity was assessed using Pearson product–moment correlations to assess the relationship of the CDRS, the HDRS-21 and the SCL depression subscale. Differences in the strength of associations were tested using *z* tests (Meng et al., 1992)
3. Divergent validity was assessed using Pearson product–moment correlations, to assess the relationships of the CDRS and the HDRS-21 to SCL subscales other than Depression. Again, differences in the strength of associations were tested using *z* tests
4. Predictive validity was assessed by first categorizing subjects as Responders or Non-Responders using cut-off scores (responders had HDRS-21 scores <8, or CDRS score <20). The resulting classifications of completed subjects were compared descriptively. Responders and non-responders (using the HDRS-21 and the CDRS cut-off criteria separately) were compared on severity of intake symptoms using independent sample *t*-tests. Responders and non-responders (again using both criteria) were compared on termination scores on the SCL-GSI (the overall mean score of all items on the SCL, a measure of general distress) and the four SCL subscales.

3. Results

3.1. Scores on instruments

The average score on the CDRS was 35.96 (S.D. = 9.02) at baseline and 17.29 (S.D. = 12.71) at termination. The average score on the HDRS-21 was 14.79 (S.D. = 4.55) at baseline and 6.93 (S.D. = 5.35) at termination. We found no significant association between the scores on the CDRS and the demographic variables of age, race, education level, employment or marital status.

3.2. Internal consistency

Estimates of internal consistency (Cronbach's α) of the 20-item CDRS scale were 0.72 at baseline and 0.90 at termination. These can be interpreted as the lower bound of reliability of the CDRS. In comparison, the HDRS had internal consistency estimates of 0.64 at intake and 0.83 at termination.

3.3. Inter-rater reliability

Inter-rater reliability was calculated for one of the studies ((fluoxetine) vs. placebo, $n = 33$) using a subsample of eight patients. Four raters were paired up and each patient was rated by a pair of raters (during the same interview). Average intraclass correlations (1,*k*) for these eight pairs were 0.92 for the CDRS and 0.81 for the HDRS, indicating that both scales were reliably rated.

3.4. Construct validity

3.4.1. Content validity

3.4.1.1. Assessment of specific symptoms

Table 1 and Fig. 1 list the most frequent to least frequent symptoms of dysthymia found in the DSM-IV Mood Disorders Field Trials dysthymic population (Keller et al., 1995). Table 1 and Fig. 1 also details the corresponding item on the HDRS and

Table 1
Comparison between the rates of depressive symptoms in subjects meeting criteria for dysthymia ($n = 193$) from the DSMIV Field Trial, and ratings of Beth Israel sample ($n = 110$) on representative items on CDRS and HDRS

Symptoms	DSMIV field trial sample (%)	BI sample with symptom rated mild or greater			
		CDRS (≥2)		HDRS (≥2)[d]	
		(%)	Item no.	(%)	Item no.
Low self-esteem[a,b,c]	84	81.0	5	54.3	[24]
Pessimism[b,c]	77	78.3	3	41.3	[23]
Feelings of inadequacy[b]	73	67.3	7	54.3	[24]
Social withdrawal[b]	71	71.8	8		
Loss of interest or pleasure[b]	70	71.8	2	70.0	7
Low energy or fatigue[a,b,c]	66	79.9	17	70.0	13
Hopelessness[a,b,c]	65	78.3	3	41.3	[23]
Irritability or excessive anger[b,c]	65	63.3	14		
Brooding[b]	65	73.4	13	60.9	10
Decreased effectiveness or productivity[b]	62	60.9	16		
Poor concentration[a,b,c]	60	64.5	10	20.0	8
Self-pity	59				
Difficulty making decisions[a,b]	59	52.7	9		
Less talkative	58				
Tearfulness or crying	54				
Insomnia[a]	50	48.2	19	71.8	4–6[e]
Feeling slowed down	50			2.7	8
Inability to respond to praise or rewards	47				
Overeating[a]	44				
Recurrent thoughts of death or suicide	43	18.1	4	15.4	3
Restlessness	41			17.2	9
Hypersomnia[a]	38				
Poor appetite[a]	32			17.6	12

Note: field trial data and table adapted from Keller et al., 1995.

[a] Core features for dysthymic disorder, using various criterion sets: DSMIV Diagnostic Criteria B for Dysthymic Disorder (note: all subjects had depressed mood (criterion A)).

[b] DSMIV Alternative Research Criterion B for Dysthymic Disorder.

[c] Essential Symptom Criteria List for Dysthymia (Gwirtsman et al., 1997).

[d] For HDRS items ranging from 0 to 2 (items 4–6,12–17), scores of 1 or 2 were considered as 'at least mild'.

[e] Score on any of these items rated mild or greater.

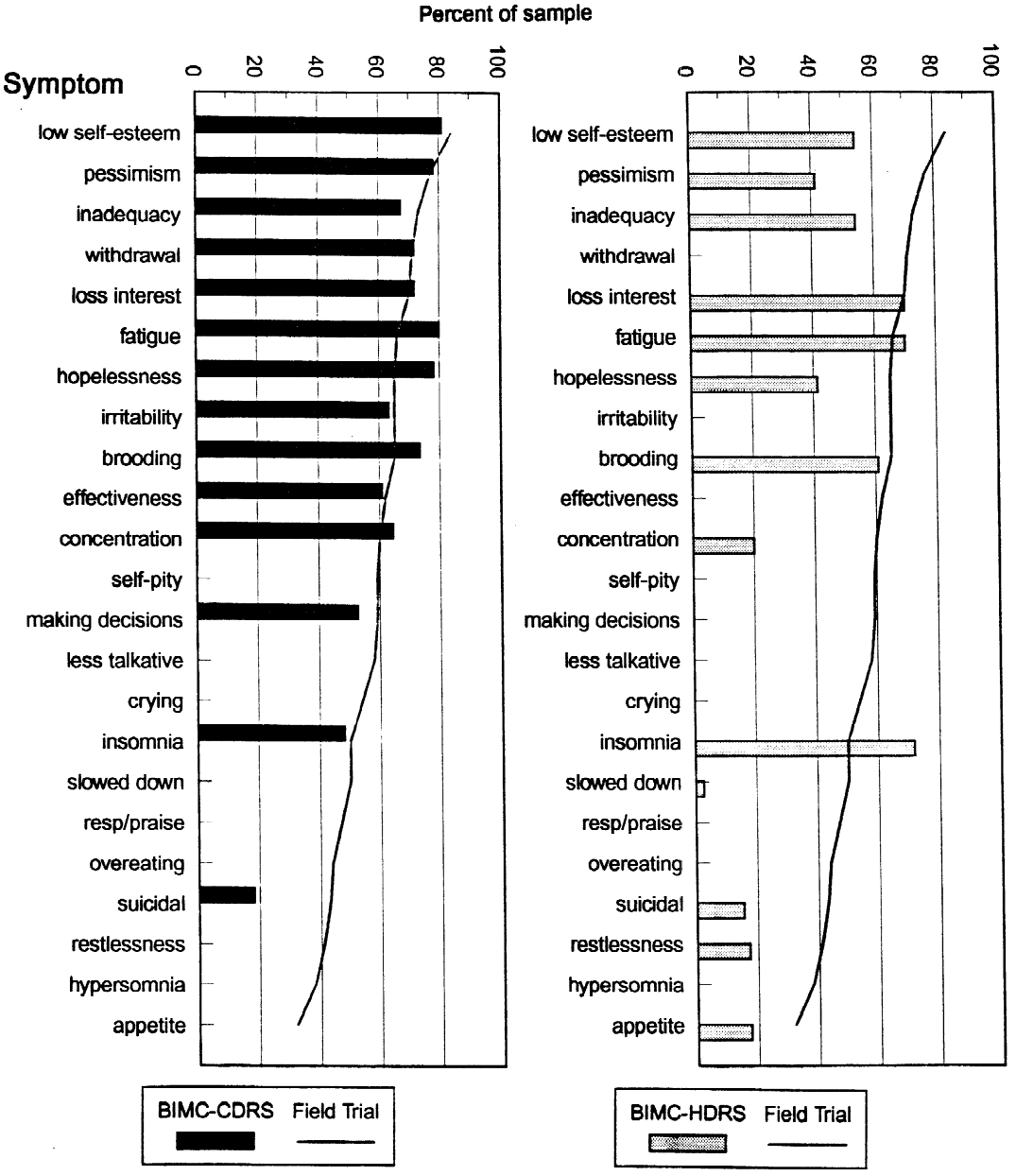

Fig. 1. Comparison of BIMC sample CDRS and HDRS ratings with the Field DSMIU Trial sample in rating common symptoms of dysthymic disorder.

CDRS and the percentage of subjects in our sample rated as having that symptom (mild or greater) using each scale. Table 1 and Fig. 1 clearly demonstrate that the CDRS assesses a greater number of the most frequent dysthymic symptoms than does the HDRS. For example, the CDRS assesses all eleven symptoms that were found to be present in ≥60% of the dysthymic population while the HDRS-21 covers

only four of these symptoms, with four more being assessed by two items in the HDRS-24. To statistically examine whether the CDRS and HDRS were picking up symptoms with the same frequency as the Field Trial data suggest are in the general population, Spearman rank correlations of the frequency of symptoms in the two samples were conducted. The CDRS correlated 0.87 ($P<0.001$) and the HDRS correlated 0.39 ($P=0.07$) with the Field Trial rankings. This indicates that the CDRS is more accurate and sensitive in assessing symptoms of Dysthymic Disorder than is the HDRS.

In addition, Table 1 and Fig. 1 demonstrate that severity ratings on the CDRS corresponded more closely to the Field Trials data (which we are using here as a normative sample of dysthymic patients) than did the HDRS ratings. Other sets of dysthymia criteria are noted in Table 1, namely (a) DSMIV Core Features of Dysthymic Disorder (b) Alternative Research Criteria B for Dysthymic Disorder, from DSMIV Appendix B and (c) Essential Symptom Criteria List for Dysthymia (Gwirtsman et al., 1997). Criteria sets (b) and (c) correspond well to the most common symptoms noted in the Field Trials and Beth Israel samples and all items of both these sets are assessed by the CDRS. In contrast, criteria set (a), DSMIV Core Features, which include neurovegetative symptoms, do not correspond well with the most frequent symptoms seen in these dysthymic samples.

3.4.1.2. Distribution of severity ratings

Distributional statistics for the severity ratings on the CDRS and the HDRS from our sample were computed and are presented in Tables 2 and 3. On the CDRS, a full range of severity ratings was found on 19 items (95%). Item 1, depressed mood, was the exception, ranging from 1 to 4, with no ratings of zero. The standard deviation exceeded 1.0 on 17 (85%) CDRS items. In contrast, on the HDRS, a full range of scores was reached only on 12 (50%) of the 24 items. The standard deviation did not exceeded 1.0 on any of the HDRS items, further evidence of a narrower range of symptom severity as rated by the HDRS. A mean score above 1 was found for only 3 of the 24 HDRS items, while 19 of the 20 CDRS items had a mean score above 1. The remaining item (suicidal ideation) was artificially low due to the exclusion criteria in our studies of significant suicidal ideation or plan.

Table 2
Distributional statistics for the severity ratings on the Cornell Dysthymia Rating Scale (CDRS)

CDRS item		Mean	S.D.	Severity rating (% of sample)				
No.	Description			None	Slight	Mild	Mod	Severe
1	Depressed mood	2.53	0.67	0.0	7.3	35.5	54.5	2.7
2	Lack of interest or pleasure	2.13	1.03	5.5	22.7	33.6	30.0	8.2
3	Pessimism	2.24	0.92	2.7	19.1	35.5	37.3	5.5
4	Suicidal ideation	0.68	0.93	56.4	25.5	12.7	4.5	0.9
5	Low self-esteem	2.44	1.11	7.3	11.8	26.4	39.1	15.5
6	Guilt	1.73	1.17	16.4	30.0	23.6	24.5	5.5
7	Lack of control	1.92	1.21	18.2	14.5	31.8	28.2	7.3
8	Social withdrawal	1.94	1.14	15.5	12.7	43.6	19.1	9.1
9	Indecisiveness	1.62	1.17	22.7	24.5	22.7	28.2	1.8
10	Low attention/concentration	1.89	1.30	20.9	14.5	30.9	21.8	11.8
11	Psychic anxiety	1.90	1.05	13.6	17.3	36.4	30.9	1.8
12	Somatic anxiety	1.17	1.07	32.7	33.6	18.2	14.5	0.9
13	Worry	2.11	1.10	6.4	20.2	31.2	30.3	11.9
14	Irritability or excessive anger	1.73	1.14	22.0	14.7	33.0	29.4	0.9
15	Somatic general	1.17	1.15	38.5	23.9	21.1	14.7	1.8
16	Low productivity	1.77	1.18	19.1	20.0	30.9	24.5	5.5
17	Low energy	2.30	1.12	9.1	10.9	34.5	31.8	13.6
18	Low sexual interest, activity	1.70	1.57	35.5	14.5	14.5	15.5	20.0
19	Insomnia	1.54	1.28	28.2	23.6	21.8	19.1	7.3
20	Diurnal variation	1.40	1.16	33.9	12.8	32.1	21.1	–

Table 3
Distributional statistics for the severity ratings on the HDRS

HDRS item		Mean	S.D.	Score (% of sample)				
No.	Description			0	1	2	3	4
1	Depressed mood	1.98	0.79	2.7	23.6	46.4	27.3	0.0
2	Feelings of guilt	1.13	0.78	24.5	38.2	37.3	0.0	0.0
3	Suicidality	0.59	0.82	59.1	25.5	12.7	2.7	0.0
4	Insomnia (early)	0.56	0.77	61.8	20.9	17.3	–	–
5	Insomnia (middle)	0.73	0.82	50.9	25.5	23.6	–	–
6	Insomnia (late)	0.60	0.77	57.3	25.5	17.3	–	–
7	Work and activities	1.89	0.94	10.0	19.1	43.6	26.4	0.9
8	Retardation	0.23	0.48	80.0	17.3	2.7	–	–
9	Agitation	0.60	0.84	60.0	22.7	14.5	2.7	0.0
10	Anxiety-psychic	1.64	0.93	13.6	25.5	45.5	14.5	0.9
11	Anxiety-somatic	0.91	0.81	37.3	34.5	28.2	0.0	0.0
12	Somatic (gastrointestinal)	0.18	0.38	82.4	17.6	0.0	–	–
13	Somatic (general)	1.23	0.76	20.0	37.3	42.7	–	–
14	Genital symptoms	0.87	0.84	42.2	28.4	29.4	–	–
15	Hypochondriasis	0.21	0.51	83.6	11.8	4.5	–	–
16	Loss of weight	0.06	0.28	94.5	4.5	0.9	–	–
17	Insight	0.03	0.16	97.3	2.7	0.0	–	–
18	Diurnal variation	0.90	0.78	35.5	39.1	25.5	–	–
19	Depersonalization	0.13	0.41	90.0	7.3	2.7	0.0	0.0
20	Paranoid symptoms	0.18	0.41	82.7	16.4	0.9	0.0	–
21	Obsessive and compulsive	0.16	0.40	84.5	14.5	0.9	–	–
22[a]	Helplessness	0.97	0.85	34.8	35.9	27.2	2.2	0.0
23[a]	Hopelessness	1.38	0.91	16.3	42.4	28.3	13.0	0.0
24[a]	Worthlessness	1.50	0.66	6.5	39.1	52.2	2.2	0.0

[a] Note: Items 22–24 (as part of the HDRS-24 scale) were only collected on a subsample ($n=92$). These items not included in the calculation of HDRS scores for the overall analyses.

3.4.2. Convergent validity

3.4.2.1. Measures of depressive symptoms

Pearson correlations revealed significant associations between the HDRS-21 and the CDRS at both intake ($r=0.70$, $P<0.001$) and at termination ($r=0.90$, $P<0.001$). The greater association found at termination is most likely due to the larger range of scores at this time point. Significant associations were also found between the CDRS and the SCL Depression subscale *T* scores at both time points (intake: $r=0.26$, $P=0.007$; termination: $r=0.40$, $P<0.001$).

Comparatively, the HDRS did not show a positive association at intake ($r=0.09$, n.s.) but was associated with SCL Depression *T* scores at termination ($r=0.38$, $P<0.001$). To compare the strengths of the pairs of correlations at each time point, *z* tests to compare dependent correlations (Meng et al., 1992) were conducted (see Table 4). Although the CDRS and the SCL Depression scale were significantly associated at intake while the same association with the HDRS was not, the two correlations were not statistically different from each other. In addition, none of the other parallel correlations were significantly different.

3.4.3. Measures of overall symptomatology

The CDRS was significantly associated with CGI-Severity scores at both baseline ($r=0.42$, $P=0.001$) and termination ($r=0.65$, $P<0.001$). A similar association between the HDRS and the CGI was found at baseline ($r=0.34$, $P=0.008$) and termination ($r=0.69$, $P<0.001$). The CDRS was also significantly and moderately associated with the SCL global severity index (GSI) *T* score [a measure of overall psychological distress due to symptomatology] at both baseline ($r=0.38$, $P<0.001$) and termination ($r=0.40$, $P<0.001$). When the HDRS was correlated with SCL GSI T scores, no association

Table 4
Measures of convergent and divergent validity, and significance tests of the differences in associations with the CDRS and the HDRS

	Intake ($n=110$)				Termination ($n=92$)			
	CDRS	HDRS	z	P	CDRS	HDRS	z	P
CDRS		0.70[a]***				0.90***		
CGI-S	0.42[b]***	0.34[b]**	0.72	0.47	0.65[c]***	0.69***	0.81	0.42
SCL-GSI T	0.38***	0.13 ns	2.18	0.03	0.40***	0.34***	1.37	0.17
SCL-Depr T	0.26**	0.09 ns	1.45	0.15	0.40***	0.38***	0.46	0.65
Somat T	0.38***	0.07 ns	2.67	0.008	0.22*	0.17 +	1.08	0.28
Anxiety T	0.37***	0.17+	1.75	0.08	0.35***	0.34***	0.23	0.82
ObsComp T	0.35***	0.03 ns	2.73	0.006	0.44***	0.35***	2.08	0.04
InterpSens T	0.25**	0.05 ns	1.69	0.09	0.44***	0.34***	2.30	0.02

*, $P<0.05$; **, $P<0.01$; ***, $P<0.001$; +, $P<0.10$.
[a] $n=45$.
[b] $n=60$.
[c] $n=46$.

was found at intake ($r=0.13$, $P=0.17$) but a significant association emerged at termination ($r=0.34$, $P=0.001$). The difference between the correlations at intake is statistically significant, z (44)=2.18, $P<0.03$. None of the other pairs of correlations (CDRS vs. HDRS) were significantly different from each other.

3.4.4. Divergent validity

The relationships between the CDRS and the SCL subscales that are distinct from depression (Somatization, Anxiety, Obsessive Compulsive and Interpersonal Sensitivity) were assessed using Pearson product–moment correlations (see Table 4). All four subscales were significantly associated with the CDRS at both Intake and Termination. We also calculated Pearson correlations between the HDRS and the SCL non-depression subscales and found no significant associations at Intake and positive associations with Anxiety, Obsessive Compulsive, and Interpersonal Sensitivity *T* scores at termination. In addition, a trend toward a significant association between the HDRS and the Somatization subscale *T* score was found at termination. The majority of the paired correlations were significantly different, with the CDRS showing greater association with divergent scales than the HDRS. The only scales which were equivalently associated with both the CDRS and the HDRS were termination scores on the Somatization and Anxiety subscales.

3.4.5. Predictive validity

3.4.5.1. Associations with treatment response

3.4.5.1.1. *Responder status* Using the CDRS responder criteria (i.e. CDRS<20), 58 (61.1%) of the 95 completed subjects were classified as responders. Of these 58, 51 were also classified as responders using the HDRS-21 cutoff criteria (i.e. HDRS<8), while seven subjects were 'misclassified' as nonresponders using the HDRS. Of 37 nonresponders by CDRS criteria, (38.9%), 31 were also classified as nonresponders using the HDRS-21 cut-off criteria.

We then tested for differences between responders and non-responders, using the CDRS criteria, on intake scores on the CDRS, the HDRS-21, and the subscales of the SCL. Significant differences were found for the HDRS-21, the CDRS, as well as the Interpersonal Sensitivity subscale of the SCL. In each case, non-responders had higher levels of symptoms at intake than did than non-responders (see Table 5). When responder status was measured by the HDRS criteria, similar findings emerged (see Table 6): significant differences on intake scores on the CDRS, HDRS-21, and Interpersonal Sensitivity. In addition, differences were found on intake SCL Anxiety scores. As with the CDRS, nonresponders

Table 5
Differences in intake symptom variables between responders and nonresponders, as measured by the CDRS

	Responder ($n=58$)	Non-responder ($n=37$)	t
HDRS-21	13.57 (4.59)	16.27 (4.33)	2.86**
CDRS	33.38 (8.51)	40.49 (7.96)	4.07***
SCL scales			
Somatization	57.00 (5.82)	58.89 (5.09)	1.62
Obsessive–Compulsive	55.41 (5.80)	56.24 (4.67)	0.73
Interpersonal Sensitivity	54.29 (5.54)	56.46 (4.61)	2.06*
Depression	53.53 (5.09)	53.51 (4.41)	−0.02
Anxiety	48.72 (5.78)	49.81 (4.82)	0.95

*, $P<0.05$; **, $P<0.01$; ***, $P<0.001$.

Table 6
Differences on intake symptom variables between responders and nonresponders, as measured by the HDRS

	Responder ($n=57$) M (S.D.)	Non-responder ($n=38$) M (S.D.)	t
HDRS-21	13.26 (4.54)	16.66 (4.11)	3.71***
CDRS	33.33 (8.11)	40.37 (8.59)	4.04***
SCL scales			
Somatization	56.86 (6.04)	59.05 (4.62)	1.90
Obsessive–Compulsive	55.60 (5.88)	55.95 (4.59)	0.31
Interpersonal Sensitivity	54.16 (5.76)	56.61 (4.10)	2.26*
Depression	53.18 (5.02)	54.05 (4.51)	0.87
Anxiety	48.12 (5.98)	50.68 (4.07)	2.31*

*, $P\leq0.05$; ***, $P<0.001$.

had higher levels of symptomatology on all these scales than did responders.

3.4.5.2. Sensitivity to change

We also compared the ability of the CDRS and the HDRS to detect change over time as a result of treatment. Both scales detected statistically significant decreases from baseline to termination (HDRS: t (94)=12.64, $P<0.001$, CDRS: t (94) 15.43, $P<0.001$). However, the effect size of the CDRS difference ($d=3.18$) was greater than the HDRS effect size ($d=2.61$). We also examined the ability of the two scales to detect drug–placebo differences, and again while both scales detected significant differences (see Table 7) the CDRS again showed a

Table 7
Changes from intake to termination as measured by the CDRS and HDRS, t-tests and effect sizes

	Intake		Termination		t-test (df 94)	Effect size (d)
	Mean	(S.D.)	Mean	(S.D.)		
CDRS	36.15	(8.96)	17.29	(12.71)	15.43***	3.18
HDRS	14.62	(4.66)	6.93	(5.35)	12.64***	2.61

***, $P<0.001$.

larger effect size ($d = 1.00$) than the HDRS ($d = 0.72$).

3.4.5.2.1. *Ratings of illness severity* Termination scores on the HDRS-21 and the CDRS were correlated with ratings on the CGI-Severity scale rated at termination. Both were positively associated with the CGI-S at termination (CDRS: $r = 0.65$, $P < 0.001$; HDRS-21: $r = 0.69$, $P < 0.001$). These two associations were not statistically different (see Table 4).

4. Discussion

This study provides independent confirmation of the validity and utility of the CDRS in the assessment of dysthymic patients. In a sample of 110 dysthymic patients, we found that distributions of severity ratings were similar to those found by Mason et al. (1993), and that the CDRS assessed most of the common symptoms found in the DSM-IV Field Trials sample. It also assessed the most common symptoms in the DSMIV Alternative Research Criterion B for Dysthymic Disorder, and the Essential Symptom Criteria List for Dysthymia of Gwirtsman et al. (1997). In our sample, the CDRS demonstrated good to excellent internal consistency, providing evidence that all items on the CDRS tap into a single construct. In addition, all four aspects of construct validity (content, convergent, divergent, and predictive) were examined and, as will be discussed below, the findings support the CDRS's construct validity.

4.1. Content validity

Excellent content validity was demonstrated, since the CDRS assesses the primary and most common symptoms of Dysthymic Disorder (see Table 1). In addition, our findings suggest that the CDRS is superior to the HDRS in this population as it covers a broader range of these symptoms and measures them more sensitively. More specifically, of the 12 Alternative Research Criteria ([b] in Table 1), the CDRS covers all 12 items, the HDRS-24 covers eight items and the HDRS-17 covers only four items. Of the Essential symptom criteria list ([c]) (six items), the CDRS covers all six, the HDRS-24 covers five and the HDRS-17 covers only two. Of the nine DSMIV criteria ([a]), the CDRS and HDRS-24 both cover six, and the HDRS-17 covers four. Clearly the CDRS is superior in this regard, especially for research criteria. (It is also clear that the 24-item HDRS is far superior to the 17-item version in assessing dysthymia. The 17-item HDRS covers less than half the criteria, no matter which set is referenced.)

4.2. Convergent validity

The presence of adequate convergent validity of the CDRS was also demonstrated in this study. The CDRS was positively associated with several standard measures of depression and with global measures of psychopathology, both doctor-rated and patient-rated. These findings are consistent with the Mason et al. (1993, 1995) studies. In addition, the majority of the associations between the CDRS and these measures are virtually identical in strength to the associations between these measures and the HDRS. The one difference however, was that at intake the CDRS was associated with the SCL GSI, a measure of overall psychological distress, while the HDRS was not. Thus, the CDRS shows very similar convergent validity as does the industry-standard measure of depressive symptomatology, the HDRS, with some evidence that the CDRS is more strongly associated with patient-perceived distress at intake than the HDRS.

4.3. Divergent validity

On the other hand, our study provides inconclusive evidence regarding the divergent validity of the CDRS. The CDRS showed significant associations with scales purported to measure divergent symptoms of Anxiety, Somatization, Obsessive Compulsivity, and Interpersonal Sensitivity at both intake and termination. In addition, the HDRS also showed significant associations with Anxiety, Obsessive Compulsive, and Interpersonal Sensitivity scales at termination, as well as a trend toward significant associations with Somatization.

The relationship between the CDRS and these SCL factors was unexpected. The association with Somatization scores at termination may reflect medication side effects, but this cannot be directly tested

with the current data and cannot account for the association at intake. The association with the Obsessive–Compulsive scale may be due to the considerable overlap between depression symptoms and the some of the symptoms assessed in the Obsessive–Compulsive subscale (e.g. difficulty making decisions, trouble concentrating, difficulty with memory, mind going blank, feeling blocked in getting things done). The presence of an association only at termination may be due to a greater range of scores at that assessment point than at intake. Use of a more specific measure of obsessive symptoms would provide a more definitive assessment of the association with CDRS scores. Finally, given the level of comorbidity of dysthymic disorder and social anxiety, in hindsight it seems evident that some relationship between the two scales would be expected. The scales used to test divergent validity here, in retrospect, are not as distinct as would be desired and thus may not adequately assess the divergent validity of the CDRS.

4.4. Predictive validity

4.4.1. Associations with treatment response

Significant associations were found between the CDRS and measures of treatment response, demonstrating predictive validity. The CDRS was found to be equally accurate as the HDRS in categorizing treatment responders. In addition, the level of initial symptomatology (as measured by the CDRS) was associated with treatment outcome, with greater symptomatology at baseline associated with poorer outcome. Finally, treatment non-responders (categorized by the CDRS and the HDRS separately) showed higher levels of symptomatology at intake as assessed by other measures as well. These results demonstrate not only the predictive validity of the CDRS, but also the equivalence of the CDRS's predictive validity to that of the HDRS.

4.5. Study strengths and limitations

This study involves a fairly large sample of dysthymics, evaluated with a standard methodology, including the SCID interview. It excluded patients with Major Depression or Double Depression, focusing on 'pure' dysthymics. Measures from multiple perspectives, both the patient's and the doctor's, support the value of the CDRS. However, the use of only patients with dysthymic disorder limits our discussion of the divergent validity of the CDRS, since there is no data on the ability of the CDRS to distinguish between dysthymic patients and either 'normals' or patients with other psychiatric diagnoses. Another weakness of the current study relates to our method of measuring divergent symptoms within our sample. We did not have access to observer ratings of divergent symptoms. In addition, the SCL subscales have shown less than robust divergent validity themselves, have overlap with common depression symptoms, and thus are not the optimum measure of divergent symptoms. Future studies would benefit by using more robust measures of other symptoms such as OCD, panic, anxiety, etc. and by using measures that also assess these symptoms from an observer's perspective. Finally, convincing evidence for the value of the CDRS or other rating scales in comparison to the HDRS could come from using independent raters to assess global improvement, and determining which scale was more closely correlated with overall response.

4.6. Future directions

Because of the introduction of more effective and more tolerable medications to treat chronic depression, there is a greater need for more specific and sensitive measures of the symptoms of this disorder. Several authors have described multidimensional systems for measuring outcome. Rush et al. (1998) used the HDRS and Clinical Global Impressions scales as primary measures of symptom outcome, with secondary measures including the Montgomery–Asberg Depression Rating Scale; the Cornell Dysthymia Rating Scale; and the Beck Depression Inventory; along with several measures of psychosocial function. Haykal and Akiskal (1999) describe clinical criteria for dysthymia treatment including symptomatology, the Global Assessment of Functioning scale, the patient's ability to cope with stress, suicidal ideation and behavior, temperament and psychosocial functioning. In another such effort with broad-based recommendations across the course of assessment and treatment, Gwirtsman et al. (1997) summarize what they believe to be 'essential' and

'recommended and optional measures'. Given the limitations of the HDRS, Gwirtsman recommends the use of a number of change scales on a supplementary basis, including the CDRS as well as others.

At present, our data do not suggest that the CDRS should replace the HDRS as a symptom or outcome measure; rather, its value appears to be supplemental. Because of its improved content validity and increased sensitivity to common dysthymic symptoms, it may be useful as a means of assessing residual symptoms and characterizing treatment response in both research and clinical settings.

5. Conclusions

We found the CDRS to have adequate internal consistency and construct validity (content, convergent, and predictive validity), and that its validity is equivalent or superior to the HDRS in the assessment of dysthymic disorder.

Acknowledgements

This study was supported by the Eli Lilly Company.

References

American Psychiatric Association, 1994. Diagnostic and Statistical Manual, 4th ed. American Psychiatric Association, Washington, DC.

Barrett, J.E., Barrett, J.A., Oxman, T.E., Gerber, P.D., 1988. Prevalence of psychiatric disorders on primary care practice. Arch. Gen. Psychiatry 45, 1100–1106.

Derogatis, L., 1974. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. Behav. Sci. 19, 1–13.

Derogatis, L., 1983. The Symptom Checklist-90 Revised: Administration, Scoring and Procedures Manual II. Clinical Psychometric Research, Baltimore, MD.

Friedman, R.A., 1993. Social impairment in dysthymia. Psychiatric Ann. 23, 623–637.

Goldberg, D., 1985. Identifying psychiatric illness among general medical patients. Br. Med. J. 291, 161–162.

Guy, W. (Ed.), 1976. SECDEU Assessment manual for psychopharmacology. US Department of Health. Education, and Welfare Publication (ADM 76-338). National Institute of Mental Health, Rockville, MD.

Gwirtsman, H.E., Blehar, M.C., McCullough, J.P., Kocsis, J.H., Prien, R.F., 1997. Standardized assessment of dysthymia: report of a National Institute of Mental Health Conference. Psychopharm. Bull. 33, 3–11.

Hamilton, M., 1960. A rating scale for depression. J. Neurol. Neurosurg. Psychiatry 25, 56–62.

Haykal, R.F., Akiskal, H.S., 1999. The long-term outcome of dysthymia in private practice: clinical features, temperament and the art of management. J. Clin. Psychiatry 60, 508–518.

Keller, M.B., Klein, D.N., Hirshfeld, R.M.A. et al., 1995. Results of the DSM-IV mood disorders field trial. Am. J. Psychiatry 152, 843–849.

Mason, B.J., Kocsis, J.H., Leon, A.C. et al., 1993. Measurement of severity and treatment response in dysthymia. Psychiatric Ann. 23, 625–631.

Mason, B.J., Kocsis, J.H., Leon, A.C. et al., 1995. Assessment of symptoms and change in dysthymic disorder. In: Kocsis, J.H., Klein, D.N. (Eds.), Diagnosis and Treatment of Chronic Depression. Guilford Press, New York, pp. 73–88.

Meng, X.L., Rosenthal, R., Rubin, D.B., 1992. Comparing correlated correlation coefficients. Psych. Bull. 111, 172–175.

Robins, L.N., Regier, D.A. (Eds.), 1991. Psychiatric Disorders in America: The Epidemiologic Catchment Area Study. The Free Press, New York.

Rush, A.J., Koran, L.M., Keller, M.B. et al., 1998. The treatment of chronic depression. Part I: study design and rationale for evaluating the comparative efficacy of sertraline and imipramein as acute, crossover, continuation, and maintenance phase therapies. J. Clin. Psychiatry 59, 589–597.

Spitzer, R.L., Williams, J.B.W., Gibbon, M., 1987. Structured Clinical Interview for DSM-IIIR (SCID-I). New York State Psychiatric Institute, Biometrics Research, NY.

Weissman, M.M., Leaf, P.J., Bruce, M.L., Florio, L., 1988. The epidemiology of dysthymia in five communities: rates, risks, comorbidity, and treatment. Am. J. Psychiatry 145, 815–819.

Wells, K.B., Stewart, A., Hays, R.D. et al., 1989. The functioning and well-being of depressed patients: results from the Medical Outcomes Study. J. Am. Med. Assoc. 262, 914–919.