Assessing the Reliability and Validity of the Sheehan Irritability Scale in Patients With Major Depressive Disorder

Sajjad A. Khan, PhD[a]; Dennis A. Revicki, PhD[b,*]; Mariam Hassan, PhD[c,i]; Julie C. Locklear, PharmD, MBA[c,j]; Lisa Aronson Friedman, ScM[d]; Sally Mannix, BA[b]; Raj Tummala, MD[e]; Geoffrey C. Dunbar, MD[f,j,k]; Hans Eriksson, MD, PhD, MBA[g]; and David V. Sheehan, MD, MBA[h]

ABSTRACT

Objective: Irritability is a significant component in the clinical manifestation of major depressive disorder (MDD). The Sheehan Irritability Scale (SIS) was developed to assess irritability-related symptoms in patients with psychiatric disorders. Data from a phase 2 clinical trial (June 2008–July 2009) was utilized to evaluate the psychometric properties of the SIS. The trial population included patients diagnosed with MDD, according to *DSM-IV* and confirmed via the MINI diagnostic scale, who had inadequate response to citalopram.

Method: The secondary analyses included 586 patients from the United States and India. Data from the SIS, depression severity measures (17-item Hamilton Depression Rating Scale [HDRS-17], Montgomery-Asberg Depression Rating Scale [MADRS], Quick Inventory of Depressive Symptomatology–Self-Report [QIDS-SR]), and other measures (Sheehan Disability Scale [SDS], Clinical Global Impressions-Severity of Illness scale [CGI-S]) were used in the psychometric evaluation. All statistical tests used a significance level of .05 unless otherwise noted.

Results: Internal consistency (0.92–0.99) and test-retest reliability (0.83 to 0.98) were excellent. Concurrent validity was demonstrated through strong correlations between the SIS total score and HDRS-17, QIDS-SR, SDS, CGI-S, and MADRS scores. SIS total scores were significantly different by clinical severity level ($P < .001$). Minimally important difference estimates suggest that a 7- to 8-point change in the SIS total score may be clinically meaningful.

Conclusions: The SIS has excellent reliability, acceptable validity, and good responsiveness, making the SIS appropriate for use in clinical research and practice.

Trial Registration: ClinicalTrials.gov identifier: NCT00692445

J Clin Psychiatry 2016;77(8):1080–1086
dx.doi.org/10.4088/JCP.14m09719

[a]NeuroScience, Inc, Herndon, Virginia
[b]Evidera, Bethesda, Maryland
[c]AstraZeneca Pharmaceuticals LP, Wilmington, Delaware
[d]Johns Hopkins University School of Medicine, Baltimore, Maryland
[e]AstraZeneca Pharmaceuticals LP, Wilmington, Delaware
[f]Targacept Inc, Winston-Salem, North Carolina
[g]AstraZeneca R&D, Södertälje, Sweden
[h]University of Southern Florida College of Medicine, Tampa, Florida
[i]Vertex Pharmaceuticals (current affiliation)
[j]EMD Serono (current affiliation)
[k]Lundbeck (current affiliation)
**Corresponding author:* Dennis A. Revicki, PhD, Evidera, Outcomes Research, 7101 Wisconsin Ave, Ste 1400, Bethesda, MD 20814 (dennis.revicki@evidera.com).

Irritability has been identified as a key component in many neuropsychiatric conditions that adversely impacts activities of daily living, mood regulation, interpersonal relationships, functional outcomes, and treatment outcomes.[1–9] Irritability has been described as the temporary subjective experience of impatience, intolerance, and poorly controlled anger.[9] Irritability is known to be a clinical component of bipolar disorder,[10,11] personality disorders,[12,13] anxiety disorders,[7] panic disorder,[14] posttraumatic stress disorder,[8,15] and major depressive disorder (MDD).[6,16–18] Irritability may also be an indicator of suboptimal treatment response and may be an indicator of persistent symptomatology that may require adjunctive pharmacotherapy.[6,19,20]

A key symptom associated with depression is the subjective feeling of irritability,[18] which has been estimated to occur in up to 60% of MDD patients.[6] Irritability has been associated with treatment nonadherence, violence, accidents, and suicidal ideation.[6] In acutely symptomatic untreated MDD patients, irritability was a clinically relevant symptom, along with feelings of anger, anxiety, and depression.[17] While most depression rating scales do not distinguish irritability symptoms from the core MDD criteria, irritability has been studied as an important concept in MDD. In a study of 1,500 MDD participants, irritability was shown to persist 50% of the time in 40% of the sample and was deemed a "broad indicator" of severe depression, thus underscoring the clinical importance of measuring irritability.[6]

The 7-item Sheehan Irritability Scale (SIS)* was developed to measure the frequency, severity, and impairment associated with irritability in psychiatric patients. The SIS includes items on irritability, frustration, edginess/impatience/overreaction, moodiness, anger with self, anger with others, and temper. Items are answered on an 11-point numeric rating scale where higher scores indicate greater severity (0 = not at all, 10 = extremely). A qualitative study has been conducted to assess the content validity of the SIS in 24 patients diagnosed with MDD (S. Mannix, BA, manuscript submitted). The results of the content validity study provided evidence supporting the content validity of the SIS in patients diagnosed with MDD, supporting the use of the SIS as a measure of irritability in patients with depression. Specifically, the majority of participants in the content validity study reported that the concepts asked about on the SIS were

*Readers can contact the developer (Dr Sheehan) for more information on the scale (dsheehan@health.usf.edu).

Clinical Points

- Irritability is a significant component in the clinical manifestation of major depressive disorder (MDD). The Sheehan Irritability Scale (SIS) was developed to assess irritability-related symptoms in patients with psychiatric disorders. The SIS has not previously undergone extensive psychometric evaluation in MDD populations. This study examined the psychometric properties of the SIS in patients with MDD.
- The SIS complements existing measures of depression symptom severity and extends assessment to irritability symptoms. The SIS was demonstrated to be reliable and valid in an MDD clinical trial sample. Most important for clinical trials comparing different antidepressant treatments, the SIS total scores demonstrated good responsiveness to changes in depression severity and discriminated between responders and nonresponders.

part of their experience and that they understood the SIS instructions, item contents, and response scales. The content validity study also further confirmed the utility of the 0–10 Discan design rating scale[21] that has been used in assessing other measures, such as the Sheehan Disability Scale.[21,22]

The SIS has not previously undergone extensive psychometric evaluation in MDD populations. The objective of the current study was to examine the psychometric properties of the SIS in patients with MDD.

METHOD

Data Source

This study involved secondary analyses of phase 2 clinical trial data (June 2008–July 2009) from MDD in subjects who were inadequate responders to citalopram therapy.[23] The study consisted of an open-label phase to identify the patient population of inadequate responders to citalopram (8 weeks) and a double-blind phase in which the inadequate responders were randomized to either TC-5214 (2–8 mg) or placebo plus citalopram for 8 weeks. Inadequate responders were subjects who had total score reduction of < 50% on the Montgomery-Asberg Depression Rating Scale (MADRS) but no lower than 17. Week 8 was the start of double-blinded randomization treatment, which lasted 8 weeks (weeks 8–16). The clinical trial included subjects from the United States and India. Subjects were required to be between 18 and 70 years of age, have a diagnosis of MDD according to *DSM-IV* and confirmed via the MINI diagnostic scale,[24] have had no more than 1 prior antidepressant course of treatment before trial entry for the current episode of depression, have a MADRS score greater than 27, have a Clinical Global Impressions-Severity of Illness scale (CGI-S)[25] score greater than or equal to 4, and provide written informed consent to participate. The scales administered during the trial were administered in English. The clinical trial received institutional review board approval. The study was registered on ClinicalTrials.gov (identifier: NCT00692445).

Measures

Clinician-reported measures. Hamilton Depression Rating Scale. This study used the 17-item version of the Hamilton Depression Rating Scale (HDRS-17).[26] The HDRS scores range from 0 to 52, with higher scores indicating greater depressive symptom severity. The HDRS was administered at screening, baseline, and weeks 8, 9, 10, 12, 14, 16, and 18/19.

MADRS. The MADRS was developed to be sensitive to change.[27] MADRS total scores range from 0 to 60, with higher scores indicating greater depression severity. The MADRS was administered at screening, baseline, and weeks 8, 16, and 18/19.

CGI. The CGI-S and CGI-Improvement (CGI-I) (global improvement) were included in this study to assess severity of mental illness.[25] The CGI-S scores range from 0 (not assessed) to 7 (among the most extremely ill), and CGI-I scores range from 0 (not assessed) to 7 (very much worse). The CGI-S was administered at screening, baseline, and weeks 2, 8, 16, and 18/19, and the CGI-I was administered at weeks 2, 4, 8, 9, 10, 12, 14, 16, and 18/19.

Patient-reported measures. SIS. The 7-item SIS assesses symptoms of irritability, frustration, edginess/impatience, moodiness, anger with self, anger with others, and temper during the previous week. Each item is assessed on an 11-point numeric rating scale ranging from 0 (not at all) to 10 (extremely). The SIS total score ranges from 0 to 70, with higher scores indicating greater symptomatology. The SIS was administered at screening, baseline, and weeks 8, 10, 12, 14, 16, and 18/19.

Sheehan Disability Scale (SDS). The 3-domain SDS assesses patient disability (work/school, social, family life)[21,22] using a 0- to 10-point numeric rating scale. SDS scores range from 0 to 30 and higher scores indicate greater disability. The SDS was administered at baseline and weeks 8, 16, and 18/19.

Quick Inventory of Depressive Symptomatology–Self-Report (QIDS-SR). The 16-item QIDS-SR is a self-administered scale that assesses the 9 *DSM-IV* domains of depression.[28] QIDS-SR scores range from 0 to 27, with higher scores indicating greater depression severity. The QIDS-SR was administered at baseline and weeks 8, 10, 12, 14, 16, and 18/19.

Statistical Analysis

The psychometric analyses evaluated the internal consistency, test-retest reliability, concurrent validity, longitudinal validity, known-groups validity, and minimally important difference of the SIS. The analyses were conducted while investigators were blinded to treatment group and used all available observed data at selected study visits (baseline and weeks 8 and 16). All analyses were conducted using SAS (version 9.1.3; SAS Institute; Research Triangle Park, North Carolina,) or MPlus (version 5.21; Muthen & Muthen; Los Angeles, California). All statistical tests used a significance level of .05 unless otherwise noted.

Confirmatory Factor Analysis

The unidimensionality of the SIS items was evaluated using confirmatory factor analysis (CFA)[29] at baseline and at week 8. Fit of the data to a single factor model was based on the comparative fit index (CFI; >0.9), standardized root mean residual (SRMR; <0.10), and root mean square error of approximation (RMSEA; <0.08).[30,31] Factor loadings of 0.40 or greater are considered acceptable.

Reliability

Internal consistency reliability was evaluated using Cronbach coefficient α (baseline and weeks 8 and 16).[32] Cronbach α estimates >0.70 are considered acceptable for aggregate data.[33]

Test-retest reliability was assessed by evaluating the reproducibility of mean SIS scores at baseline to week 8, weeks 8–10, weeks 12–14, and weeks 16–18/19. Test-retest reliability was evaluated in a subset of patients rated as having "no change" based on the CGI-I. Intraclass correlation coefficients (ICC) and change scores using paired *t* tests were calculated between the relevant mean and retest mean SIS scores. ICC values >0.70 are considered acceptable for establishing test-retest reliability.[33]

Table 1. Patient Demographic Characteristics at Baseline

Variable	Total (N=586)	India (n=525)	United States (n=61)
Age, y			
Mean (SD)	36.4 (11.1)	35.2 (10.5)	46.4 (11.0)
Range	18.1–68.8	18.1–63.0	22.2–68.8
Gender, n (%)			
Male	301 (51.4)	285 (54.3)	16 (26.2)
Female	285 (48.6)	240 (45.7)	45 (73.8)
Race, n (%)			
Asian	525 (89.6)	525 (100.0)	
Black or African American	7 (1.2)		7 (11.5)
Hispanic or Latino	29 (4.9)		29 (47.5)
White or Caucasian	25 (4.3)		25 (41.0)
Highest level of regular school education completed, n (%)			
Did not complete high school/secondary school	225 (38.4)	223 (42.5)	2 (3.3)
Completed high school/secondary school	242 (41.3)	204 (38.9)	38 (62.3)
Completed college	97 (16.6)	79 (15.0)	18 (29.5)
Completed postgraduate education	22 (3.8)	19 (3.6)	3 (4.9)
HDRS-17 total score, mean (SD)	27.9 (5.3)	28.5 (5.1)	22.8 (4.5)
MADRS total score, mean (SD)	35.2 (4.5)	35.5 (4.5)	32.8 (4.4)
QIDS-SR total score, mean (SD)	17.1 (3.9)	17.4 (3.9)	15.4 (4.2)
SDS total score, mean (SD)	16.7 (5.2)	16.6 (5.1)	17.8 (5.6)
CGI-S total score, mean (SD)	4.6 (0.6)	4.6 (0.6)	4.5 (0.6)
SIS total score, mean (SD)	34.3 (13.4)	33.6 (13.2)	40.2 (13.5)

Abbreviations: CGI-S = Clinical Global Impressions-Severity of Illness scale, HDRS-17 = 17-item Hamilton Depression Rating Scale, MADRS = Montgomery-Asberg Depression Rating Scale, QIDS-SR = Quick Inventory of Depressive Symptomatology–Self-Report, SDS = Sheehan Disability Scale, SIS = Sheehan Irritability Scale.

Validity

The concurrent validity of the SIS was examined using Spearman correlation coefficients with HDRS, MADRS, SDS, CGI-S, and QIDS-SR scores at baseline and weeks 8 and 16.

Longitudinal validity was evaluated by comparing correlations between changes from baseline to week 8 and weeks 8–16 between SIS scores and changes in HDRS, QIDS-SR, MADRS, SDS, and CGI-S scores.

Known-groups validity was evaluated by comparing mean SIS scores by clinically relevant groups based on the CGI-S, HDRS, and MADRS scores at baseline, week 8, and week 16. Analysis of covariance (ANCOVA) models, adjusting for age, sex, and country, were used to compare mean SIS scores by differing levels of clinical severity. CGI-S severity groups were defined as normal, borderline mentally ill, mildly ill, moderately ill, and markedly, severely, or extremely ill. Three severity categories were defined for the HDRS (mild, 0–15; moderate, 16–27; and severe, >28) and the MADRS (mild, 0–10; moderate, 11–30; and severe, >30). Scheffé test was used to conduct post hoc between-category comparisons.

Responsiveness and Minimally Important Difference

Responsiveness was evaluated by comparing mean changes in SIS scores for responders and nonresponders based on HDRS, MADRS, and CGI-I scores between baseline and week 8 and weeks 8–16. Responders were defined as those subjects who achieved a 50% reduction in HDRS scores, a 50% reduction in MADRS scores, or a CGI-I score <4. Remission of depression symptoms was defined as an HDRS score ≤7 or a MADRS score ≤10.[34–38] ANCOVA models were used to examine mean change in SIS scores by responder or remission status.

An anchor-based approach based on the CGI-I was implemented by examining baseline to week 8 mean changes in SIS scores and weeks 8–16 mean changes in SIS scores by CGI-I categories. The CGI-I was grouped into 3 categories: very much improved/much improved, minimally improved, and no change/worsened. ANCOVA, adjusting for age, sex, and country, was used to compare mean changes in SIS scores by these clinician-rated change groups. Scheffé test was used to make post hoc comparisons between groups. The minimally important difference was determined by examining the mean SIS change scores for the small improvement group (minimally improved), focusing on weeks 8–16 change.

RESULTS

The study recruited 61 US patients and 525 patients from India. Mean age was 36.4 years (range, 18.1–68.8). Most participants were male (51.4%) and Asian (89.6%); approximately 62% had at least a high school education (Table 1).

Measurement Descriptive Statistics

Mean (SD) SIS total scores were 34.3 (13.4) at baseline, 24.8 (15.3) at week 8, and 16.1 (15.1) at week 16. SIS total scores

showed > 50% reduction between baseline and week 16.

Confirmatory Factor Analysis

Confirmatory factor analyses at both time points indicated that the items of the SIS loaded to a single factor of "irritability." All individual factor loadings were higher than 0.60, indicating acceptable factor loading. The CFI at baseline was 0.96 and was 0.95 at week 8; the SRMR at baseline was 0.032 and was 0.026 at week 8; and the RMSEA at baseline was 0.125 and was 0.168 at week 8.

Reliability

The internal consistency reliability of the SIS was excellent, 0.92, 0.96, and 0.99 at baseline, week 8, and week 18/19, respectively. At baseline, item-to-total correlations ranged from 0.72 to 0.88.

Test-retest reliability was assessed by identifying clinical trial participants who were rated as not changed and comparing SIS scores between selected time intervals. ICC values ranged from 0.83 to 0.98, indicating excellent reproducibility.

Concurrent Validity

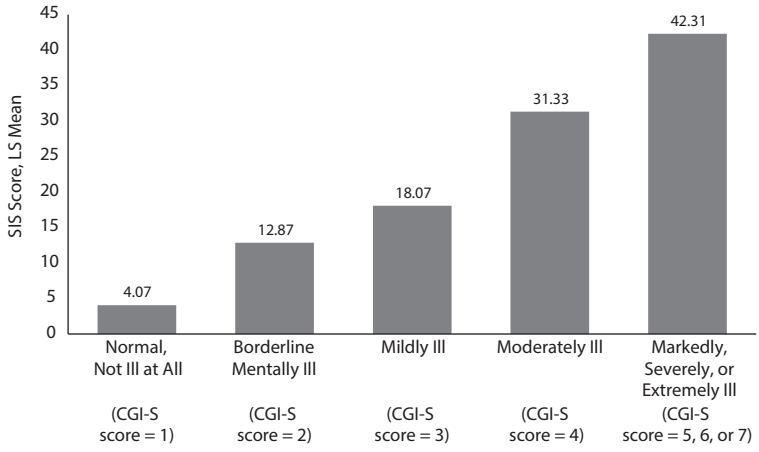
At baseline, SIS total scores were not strongly correlated (all r values < 0.10) with HDRS, MADRS, or QIDS-SR scores, although there were moderate to strong correlations observed at weeks 8 and 16 (Table 2). The nonsignificant correlations between HDRS, MADRS, and QIDS-SR scores at baseline suggest that irritability represents a separate component of MDD-related symptoms. Moderate to strong correlations were seen between the SIS and SDS and CGI-S scores, indicating that increased irritability was associated with greater disability and greater overall clinical severity.

Table 2. Concurrent and Longitudinal Validity: Correlations Between the Sheehan Irritability Scale (SIS) Total Score and Depression Severity and Disability Measures

	Concurrent Validity[a]			Longitudinal Validity[b]	
SIS Total Score	Baseline (n = 586)	Week 8 (n = 414)	Week 16 (n = 241)	Baseline to Week 8 (n = 414)	Week 8 to Week 16 (n = 241)
HDRS-17 total score	0.07	0.64***	0.81***	0.66***	0.66***
QIDS-SR total score	0.03	0.61***	0.63***	0.59***	0.57***
SDS total score	0.54***	0.78***	0.91***	0.67***	0.72***
CGI-S total score	0.16***	0.69***	0.75***	0.57***	0.67***
MADRS total score	0.09*	0.67***	0.84***	0.69***	0.68***
CGI-I total score				0.61***	0.59***

[a]Spearman rank correlation coefficient between measurements at the same visit.
[b]Spearman rank correlation coefficient between changes in scores.
*P < .05. ***P < .001.
Abbreviations: CGI-I = Clinical Global Impressions-Improvement scale, CGI-S = Clinical Global Impressions-Severity of Illness scale, HDRS-17 = 17-item Hamilton Depression Rating Scale, MADRS = Montgomery-Asberg Depression Rating Scale, QIDS-SR = Quick Inventory of Depressive Symptomatology–Self-Report, SDS = Sheehan Disability Scale.

Figure 1. Known-Groups Validity of the Sheehan Irritability Scale (SIS): Analysis of Covariance by CGI-S Score at Week 8[a]



[a]Overall F test = 56.68; P < .0001.
Abbreviations: CGI-S = Clinical Global Impression-Severity of Illness scale, LS = least squares.

Longitudinal Validity

Longitudinal validity of the SIS was examined by correlating baseline to week 8 and weeks 8–16 SIS change scores and changes in HDRS, QIDS-SR, MADRS, SDS, and CGI-S scores. Changes in the SIS scores were statistically significantly and moderately correlated with change scores on the depression severity and disability measures (P < .001) (Table 2).

Known-Groups Validity

Mean SIS total scores varied significantly by CGI-S severity groups at all visits (P < .0001), and a similar pattern of mean SIS scores was seen at each visit. Figure 1 summarizes mean SIS scores by CGI-S groups at week 8.

Mean SIS scores varied significantly by depression severity groups based on the HDRS and MADRS. Mean SIS scores differed significantly between the HDRS mild, moderate, and severe groups at baseline, week 8, and week 16 (all P values < .0001). Mean SIS scores differed significantly between the MADRS mild, moderate, and severe groups at baseline, week 8, and week 16 (all P values < .0001).

Responsiveness and Minimally Important Difference

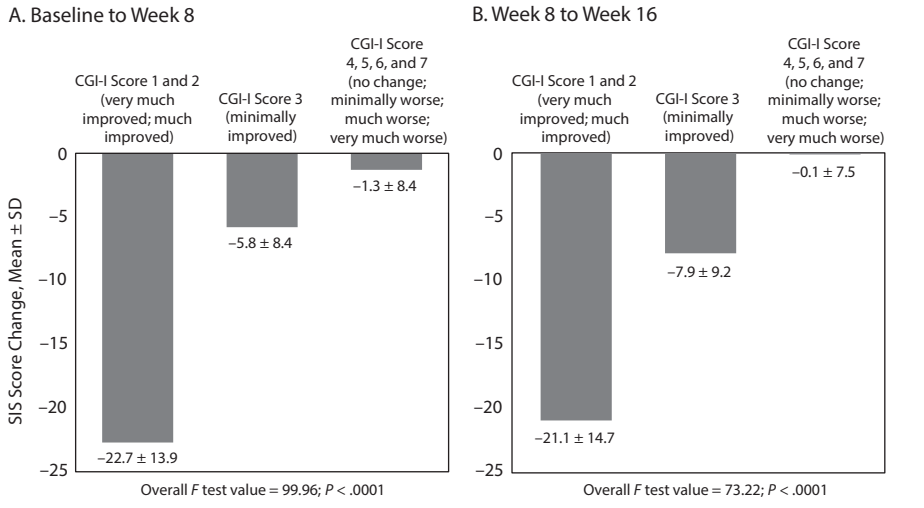
The responsiveness of the SIS was examined by evaluating mean change in SIS scores by responder status based on changes in HDRS, MADRS, and CGI-I from baseline to week 8 and weeks 8–16 (Table 3). Between baseline and week 8, mean changes in the SIS total scores were significantly different between responders and nonresponders defined by the HDRS (P < .0001), MADRS (P < .0001), and CGI-I (P < .0001). Comparable results were seen for weeks 8 and 16 (Table 3).

Table 3. Mean Change[a] in Sheehan Irritability Scale (SIS) Total Scores by HDRS-17, MADRS, and CGI-I Responder Groups

Responder Criteria	Responders, n	Responders, LS Mean (SE)	Nonresponders, n	Nonresponders, LS Mean (SE)	Overall *F* Test	*P* Value
Baseline to week 8 (randomization)						
HDRS-17 responder[b]	121	−24.8 (1.1)	293	−4.2 (0.9)	112.9	<.0001
HDRS-17 remission[c]	59	−29.0 (1.6)	355	−6.5 (1.0)	67.6	<.0001
MADRS responder[d]	120	−24.3 (.97)	294	−4.7 (0.62)	290.1	<.0001
MADRS remission[e]	64	−28.2 (1.5)	350	−6.0 (1.0)	71.5	<.0001
CGI-I[f]	311	−13.2 (1.1)	103	1.7 (1.4)	41.6	<.0001
Week 8 (randomization) to week 16						
HDRS-17 responder[b]	128	−26.5 (1.5)	113	−5.7 (1.5)	78.2	<.0001
HDRS-17 remission[c]	76	−26.9 (1.9)	165	−10.2 (1.7)	37.3	<.0001
MADRS responder[d]	135	−21.8 (1.1)	106	−5.5 (1.2)	97.3	<.0001
MADRS remission[e]	96	−27.8 (1.7)	145	−8.9 (1.5)	57.8	<.0001
CGI-I[f]	205	−18.2 (1.5)	36	6.4 (2.4)	45.9	<.0001

[a]Least square (LS) mean values and standard errors, adjusted by baseline total SIS score, age, sex, and country.
[b]Responder is defined as 50% reduction in HDRS-17 scores.
[c]Remission defined as HDRS-17 score ≤ 7.
[d]Responder is defined as 50% reduction in MADRS scores.
[e]Remission defined as MADRS score ≤ 10.
[f]CGI-I score at week 8 is < 4.
Abbreviations: CGI-I = Clinical Global Impressions-Improvement scale, HDRS-17 = 17-item Hamilton Depression Rating Scale, LS = least squares, MADRS = Montgomery-Asberg Depression Rating Scale, SE = standard error.

Figure 2. Anchor-Based Analysis of Mean Change in Sheehan Irritability Scale (SIS) Total Score by CGI-I Group From Baseline to Week 8 (A) and From Week 8 to Week 16 (B)



Abbreviation: CGI-I = Clinical Global Impressions-Improvement scale.

Mean changes in SIS total scores were compared by CGI-I categories from baseline to week 8 and weeks 8–16 for the minimally important difference anchor-based approach (Figure 2). Reductions in SIS total scores were statistically different ($P<.0001$), and the greatest reduction was experienced by those patients who were rated as "very much improved" or "much improved." The reduction in total SIS scores for the "minimally improved" group, estimating the minimally important difference using an anchor-based approach, was 7.9 for weeks 8–16.

DISCUSSION

The SIS was developed to assess the frequency, severity, and impairment associated with self-reported irritability in patients with psychiatric disorders. Irritability is an important clinical feature in severely depressed patients and those with suboptimal antidepressant treatment response.[6,17] Results from this assessment of the psychometric properties of the SIS in MDD clinical trial participants are supportive of the reliability, validity, and responsiveness of the SIS.

Internal consistency and test-retest reliability results indicated that the SIS total score has excellent reliability. The SIS has excellent internal consistency reliability (α values > 0.90); test-retest reliability for the SIS total scores ranged from 0.83 to 0.98. Generally, α values and intraclass correlations coefficients > 0.70 are considered acceptable for internal consistency reliability and test-retest reliability.[33]

The CFA results indicated acceptable model fit, and item factor loadings supported a single factor of "irritability." On the basis of the CFA findings and the excellent internal

consistency reliability, there is evidence supporting the unidimensionality of the SIS.

The SIS total scores varied by differing levels of MDD severity, supporting the known-groups validity of the SIS total score. Mean SIS total scores were highest for patients with clinician ratings of severe or moderate compared to mild. These findings suggest that patients with more severe depression are more likely to experience greater irritability-related symptoms.

Estimates of concurrent and longitudinal validity indicated that the SIS total scores correlated well with depression severity measures and clinical assessment of depression severity and change. At baseline, the SIS was not correlated with the HDRS or MADRS, suggesting that irritability is a somewhat different concept than depression symptoms. The observed correlations were moderate to strong at 8 and 16 weeks, indicating that the concept of irritability, as measured by the SIS, was related to measures of depression symptom severity and disability. Irritability-related symptoms may be more closely related to more severe or nonresponsive symptoms since moderate to strong correlations were seen at the follow-up assessments.

Longitudinal validity of the SIS was demonstrated with statistically significant correlations between changes in SIS total scores and changes in clinician-rated and self-reported depression symptom severity. Changes in SIS total scores were also significantly related to changes in SDS scores, demonstrating that greater irritability was associated with greater patient-reported disability.

The SIS total scores were responsive to changes in MDD symptoms based on MADRS or HDRS ratings. SIS total score changes discriminated between responders and nonresponders. Responders demonstrated significantly greater improvements in SIS total scores than nonresponders. Patients who reached criteria for remission of depressive symptoms also showed greater improvements in SIS total scores.

The minimally important difference estimates indicate that a 7- to 8-point change in the SIS total score may be indicative of a clinically important change in irritability symptoms. The estimated minimally important differences can be used to power clinical trials and aid in interpreting clinical studies findings where irritability, as assessed by the SIS, is included as an end point.

Sample size was a study limitation. Sample sizes varied over follow-up assessments due to the clinical trial design and subject discontinuation. Analyses at later time points were conducted on smaller samples. Additionally, while equivalency testing between the US and Indian populations was performed with no significant differences noted (data not shown), given small numbers of US participants (approximately 10%), equivalency estimates may not have been robust or stable. Furthermore, 43% of the Indian patients had less than a high school education, which may affect the generalizability. To our knowledge, the psychometric properties of the SIS have not been established in other studies, making this study the first to examine the psychometric properties of the SIS in MDD patient populations. Future research around the measurement properties of the SIS in other clinical studies and populations is needed.

In conclusion, the SIS was designed to assess irritability-related symptoms for clinical studies in MDD. The SIS complements the existing measures of depression symptom severity and extends assessment to irritability symptoms. On the basis of this study, the SIS has excellent reliability and evidence supporting concurrent and longitudinal validity. Most important for clinical trials comparing different antidepressant treatments, the SIS total scores demonstrated good responsiveness to changes in depression severity and discriminated between responders and nonresponders. Preliminary estimates of minimally important difference suggest that changes in SIS total scores of 7–8 points may represent a clinically significant effect. The SIS was demonstrated to be reliable, valid in a MDD clinical trial sample, and responsive to changes in depression severity.

Submitted: December 9, 2014; accepted June 1, 2015.

Online first: September 15, 2015.

Drug names: citalopram (Celexa and others).

Author contributions: Dr Khan contributed to this article by designing the analysis plan, overseeing the analyses, and taking the lead role in writing and revising this article. Dr Revicki contributed to this article by providing guidance and oversight to the analysis plan, overseeing the analyses, and by reviewing and contributing to the writing of this article. Ms Friedman contributed to this article in the design, implementation, and interpretation of statistical data and review of the manuscript. Ms Mannix contributed by reviewing this article. Dr Hassan contributed to study conception and design, acquisition and interpretation of clinical study data, oversight of the study plan and review of initial statistical plans and results, and assisted with interpretation and finalization of this article. Dr Locklear contributed to study conception and design, review of initial statistical plans and results, and assisted with interpretation and finalization of this article. Dr Tummala contributed to interpretation of clinical study data, clinical guidance and review of initial statistical plans and results, and assisted with interpretation and finalization of this article. Dr Dunbar contributed to acquisition and interpretation of clinical study data, clinical guidance, and review of this article. Dr Eriksson contributed to acquisition and interpretation of clinical study data, clinical guidance and review of initial statistical plans and results, and assisted with interpretation and finalization of this article. Dr Sheehan is the author and copyright holder of the Sheehan Irritability Scale, and advised on the design and interpretation of this study, and reviewed the manuscript.

Potential conflicts of interest: **Drs Khan** and **Revicki** and **Mss Friedman** and **Mannix** worked on this project as salaried employees of United BioSource Corporation (UBC), now Evidera, which performed the analysis work under contract with AstraZeneca. **Dr Dunbar** was a full-time employee of Targacept, Inc at the time of the study. **Drs Hassan, Locklear, Tummala** and **Eriksson** worked on this project as full-time employees of AstraZeneca. **Drs Tummala** and **Eriksson** are stock shareholders of AstraZeneca. **Dr Sheehan** has received grant funding support, or been affiliated or received honoraria and travel expenses related to lectures/presentations or consultant activities from the following organizations (key: 1–consultant; 2–grant/research support; 3–lectures/presentations; 4–stock holder): Abbott,[1,2,3] Actavis,[1] Ad Hoc Committee, Treatment Drug & Assessment Research Review,[1] Alexa,[1] Alza,[1] American Medical Association,[2] American Psychiatric Association Task Force on Benzodiazepine Dependency,[1] American Psychiatric Association Task Force on Treatments of Psychiatric Disorders,[1] American Psychiatric Association Working Group to revise *DSM-III* Anxiety Disorders Section,[1] Anclote Foundation,[2] Anxiety Disorders Resource Center,[1] Anxiety Drug Efficacy Case, US Food and Drug Administration,[1] Applied Health Outcomes/XCENDA,[1] AstraZeneca,[1,2,3] Avera,[1,2] Boehringer Ingelheim,[3] Boots Pharmaceuticals,[3] Bristol-Myers Squibb,[1,2,3] Burroughs Wellcome,[2,3] Cephalon,[1] Charter Hospitals,[3] Ciba Geigy,[3] Committee (RRC) of NIMH on Anxiety and Phobic Disorder Projects,[1] Connecticut & Ohio Academies of Family Physicians,[1] Cortex,[1] Council on Anxiety Disorders,[1] CPC Coliseum Medical Center,[1] Cypress Bioscience,[1] Daiichi Sankyo,[1] Dainippon Sumitomo,[2] Dista Products Company,[3] Division of Drugs & Technology, American Medical Association,[1] EISAI,[1,2] Eli Lilly,[2,3] Excerpta Medica Asia,[3] Faxmed,[1] Forest Laboratories,[1,2] Glaxo,[3]

GlaxoSmithKline,[1,2,3] Glaxo-Wellcome,[2] Hospital Corporation of America,[3] Humana,[3] ICI,[3] INC Research,[1] International Clinical Research (ICR),[2] International Society for CNS Drug Development,[1] Janssen,[1,2,3] Jazz,[1,2] Kali-Duphar,[2,3] Labopharm,[1,2,3] Layton Bioscience,[1] Lilly Research Laboratories,[1] Eli Lilly,[1,2,3] Lundbeck, Denmark,[1] Marion Merrill Dow,[3] McNeil,[3] Mead Johnson,[2,3] Medical Outcome Systems,[4] MediciNova,[1,2] Merck Sharp & Dohme,[2,3] National Anxiety Awareness Program,[1] National Anxiety Foundation,[1] National Depressive & Manic Depressive Association,[1] National Institute of Drug Abuse,[2] National Institutes of Health,[2] Neuronetics,[1] NovaDel,[1] Novartis,[2] Novo Nordisk,[3] Organon,[1,3] Orion,[1] Parexel International,[1] Parke-Davis,[2,3] Pfizer,[1,2,3] Pharmacia,[1] Pharmacia & Upjohn,[1,3] Philadelphia College of Pharmacy & Science,[1] Pierre Fabre, France,[1] Quintiles,[2] Rhone Laboratories,[3] Rhone-Poulenc Rorer,[3] Roche,[1] Roerig,[3] Sandoz,[2,3] Sanofi-Aventis,[1,2,3] Sanofi-Synthelabo Recherche,[1,2] Schering,[3] Sepracor,[1,2] Shire Laboratories,[1] SmithKlineBeecham,[1,2,3] Solvay,[1,3] Sunovion,[2,3] Takeda,[1,2,3] Tampa General Hosp,[1] University of South Florida Psychiatry Center,[2] University of South Florida College of Medicine, Takeda,[1,2,3] TAP,[2,3] Targacept,[1] TGH-University Psychiatry Center,[3] Tikvah Therapeutics,[1] Titan,[1] United Bioscience,[2] Upjohn Company,[1,2,3] US Congress-House of Representatives Committee,[1] USF Friends of Research in Psychiatry, Board of Trustees,[1] Warner Chilcott,[2,3] World Health Organization,[1] Worldwide Clinical Trials,[2] Wyeth-Ayerst,[1,2,3] ZARS,[1] Zeneca,[2] Neuronetics.[1]

Funding/support: This analysis of study TC-5214-23-CRD-001 was supported by AstraZeneca Pharmaceuticals and Targacept, Inc.

Role of the sponsor: The study sponsor was involved in the study conception and design, acquisition and interpretation of clinical study data, oversight of the study plan, clinical guidance, review of initial statistical plans and results, and interpretation of clinical study data; and assisted with interpretation and finalization of the manuscript.

Acknowledgment: The authors thank Christopher D. Thomas, BA, and Amara R. Tiebout, BA, United BioSource Corporation (UBC) now Evidera, for editorial assistance with formatting this article, funded by AstraZeneca.

Additional information: The Sheehan Irritability Scale is copyrighted by Dr Sheehan.

REFERENCES

1. Born L, Steiner M. Irritability: the forgotten dimension of female-specific mood disorders. *Arch Women Ment Health*. 1999;2(4):153–167.
2. Buss AH, Durkee A. An inventory for assessing different kinds of hostility. *J Consult Psychol*. 1957;21(4):343–349.
3. Dickstein DP, Towbin KE, Van Der Veen JW, et al. Randomized double-blind placebo-controlled trial of lithium in youths with severe mood dysregulation. *J Child Adolesc Psychopharmacol*. 2009;19(1):61–73.
4. Labiner DM, Ettinger AB, Fakhoury TA, et al. Effects of lamotrigine compared with levetiracetam on anger, hostility, and total mood in patients with partial epilepsy. *Epilepsia*. 2009;50(3):434–442.
5. Norton LE, Malloy PF, Salloway S. The impact of behavioral symptoms on activities of daily living in patients with dementia. *Am J Geriatr Psychiatry*. 2001;9(1):41–48.
6. Perlis RH, Fraguas R, Fava M, et al. Prevalence and clinical correlates of irritability in major depressive disorder: a preliminary report from the Sequenced Treatment Alternatives to Relieve Depression study. *J Clin Psychiatry*. 2005;66(2):159–166, quiz 147, 273–274.
7. Ressler KJ, Nemeroff CB. Role of serotonergic and noradrenergic systems in the pathophysiology of depression and anxiety disorders. *Depress Anxiety*. 2000;12(suppl 1):2–19.
8. Sims A, Sims D. The phenomenology of post-traumatic stress disorder: a symptomatic study of 70 victims of psychological trauma. *Psychopathology*. 1998;31(2):96–112.
9. Snaith RP, Constantopoulos AA, Jardine MY, et al. A clinical scale for the self-assessment of irritability. *Br J Psychiatry*. 1978;132(2):164–171.
10. Bowden CL. Spectrum of effectiveness of valproate in neuropsychiatry. *Expert Rev Neurother*. 2007;7(1):9–16.
11. Deckersbach T, Perlis RH, Frankle WG, et al. Presence of irritability during depressive episodes in bipolar disorder. *CNS Spectr*. 2004;9(3):227–231.
12. Kavoussi RJ, Coccaro EF. Divalproex sodium for impulsive aggressive behavior in patients with personality disorder. *J Clin Psychiatry*. 1998;59(12):676–680.
13. Tyrer P, Alexander J. Classification of personality disorder. *Br J Psychiatry*. 1979;135(2):163–167.
14. Nutt DJ, Glue P. Irritability in panic disorder: effects of imipramine treatment. *Hum Psychopharmacol: Clin*. 1991;6(1):49–52.
15. Adamou M, Puchalska S, Plummer W, et al. Valproate in the treatment of PTSD: systematic review and meta analysis. *Curr Med Res Opin*. 2007;23(6):1285–1291.
16. Brooks SJ, Kutcher S. Diagnosis and measurement of adolescent depression: a review of commonly utilized instruments. *J Child Adolesc Psychopharmacol*. 2001;11(4):341–376.
17. Pasquini M, Picardi A, Biondi M, et al. Relevance of anger and irritability in outpatients with major depressive disorder. *Psychopathology*. 2004;37(4):155–160.
18. Vigo DV, Baldessarini RJ. Anticonvulsants in the treatment of major depressive disorder: an overview. *Harv Rev Psychiatry*. 2009;17(4):231–241.
19. American Psychiatric Association. Practice guideline for the treatment of patients with major depressive disorder, 3rd ed. 2010. http://www.psychiatry.org/practice/clinical-practice-guidelines. Updated November 2010. Accessed July 15, 2015.
20. Bristow M, Bright J. Group cognitive therapy in chronic depression: results from two intervention studies. *Behav Cogn Psychother*. 1995;23(04):373–380.
21. Sheehan KH, Sheehan DV. Assessing treatment effects in clinical trials with the discan metric of the Sheehan Disability Scale. *Int Clin Psychopharmacol*. 2008;23(2):70–83.
22. Sheehan DV. *The Anxiety Disease*. New York, NY: Scribners; 1983.
23. Dunbar G, Hosford DP. 2.a.002 the potential of the nicotinic channel blocker tc-5214 as augmentation treatment in patients with major depression. *Eur Neuropsychopharmacol*. 2010;20:S334.
24. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for *DSM-IV* and *ICD-10*. *J Clin Psychiatry*. 1998;59(Suppl 20):22–33, quiz 34–57.
25. Guy W. *ECDEU Assessment Manual for Psychopharmacology*. US Dept Health, Education, and Welfare publication (ADM) 73-338. Rockville, MD: National Institute of Mental Health; 1976:218–222.
26. Hamilton M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol*. 1967;6(4):278–296.
27. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134(4):382–389.
28. Rush AJ, Trivedi MH, Ibrahim HM, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry*. 2003;54(5):573–583.
29. Muthen LK, Muthen BO. *Mplus User's Guide*. Los Angeles, CA: Muthen & Muthen; 1998.
30. Browne M, Cudeck R. Alternative ways of assessing model fit. In: Bollen K, Long J, eds. *Testing Structural Equation Models*. Newbury Park, CA: Sage; 1993:136–162.
31. Kline R. *Principles and Practice of Structural Equation Modeling*. 2nd ed. New York, NY: The Guilford Press; 2005.
32. Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16(3):297–334.
33. Hays RD, Revicki DA. Reliability and validity, including responsiveness. In: Fayers P, Hays RD, eds. *Assessing Quality of Life in Clinical Trials*. 2nd edition. New York, NY: Oxford University Press; 2005:25–39
34. Bandelow B, Baldwin DS, Dolberg OT, et al. What is the threshold for symptomatic response and remission for major depressive disorder, panic disorder, social anxiety disorder, and generalized anxiety disorder? *J Clin Psychiatry*. 2006;67(9):1428–1434.
35. Frank E, Prien RF, Jarrett RB, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: remission, recovery, relapse, and recurrence. *Arch Gen Psychiatry*. 1991;48(9):851–855.
36. Hirschfeld RM, Montgomery SA, Aguglia E, et al. Partial response and nonresponse to antidepressant therapy: current approaches and treatment options. *J Clin Psychiatry*. 2002;63(9):826–837.
37. Riedel M, Möller HJ, Obermeier M, et al. Response and remission criteria in major depression—a validation of current practice. *J Psychiatr Res*. 2010;44(15):1063–1068.
38. Zimmerman M, Posternak MA, Chelminski I. Derivation of a definition of remission on the Montgomery-Asberg depression rating scale corresponding to the definition of remission on the Hamilton rating scale for depression. *J Psychiatr Res*. 2004;38(6):577–582.