ORIGINAL RESEARCH

Development and Psychometric Evaluation of a Clinical Global Impression for Schizoaffective Disorder Scale

by **MICHAEL H. ALLEN, MD, DFAPA; DAVID G. DANIEL, MD; DENNIS A. REVICKI, PhD; CARLA M. CANUSO, MD; IBRAHIM TURKOZ, MS; DONG-JING FU, MD, PhD; LARRY ALPHS, MD, PhD; K. JACK ISHAK, PhD; JOHN J. BARTKO, PhD; and JEAN-PIERRE LINDENMAYER, MD**

Dr. Allen is from the Colorado Depression Center at the University of Colorado School of Medicine, Aurora, Colorado, and VISN 19 Suicide MIRECC, Denver, Colorado; Dr. Daniel is from United BioSource Corporation, McLean, Virginia; Dr. Revicki is from United BioSource Corporation, Bethesda, Maryland; Drs. Canuso and Turkoz are from Janssen Pharmaceutical Research and Development, LLC, Titusville, New Jersey; Drs. Fu and Alphs are from Janssen Scientific Affairs, LLC, Titusville, New Jersey; Dr. Ishak is from United BioSource Corporation, Dorval, Quebec, Canada; Dr. Bartko is an independent consultant, Newville, Pennsylvania; and Dr. Lindenmayer is from the New York University School of Medicine, New York, New York.

ABSTRACT

Objective: The Clinical Global Impression for Schizoaffective Disorder scale is a new rating scale adapted from the Clinical Global Impression scale for use in patients with schizoaffective disorder. The psychometric characteristics of the Clinical Global Impression for Schizoaffective Disorder are described.

Design: Content validity was assessed using an investigator questionnaire. Inter-rater reliability was determined with 12 sets of videotaped interviews rated independently by two trained individuals. Test-retest reliability was assessed using 30 randomly selected raters from clinical trials who evaluated the same videos on separate occasions two weeks apart. Convergent and divergent validity and effect size were evaluated by comparing scores between the Clinical Global Impression for Schizoaffective Disorder and the Positive and Negative Syndrome Scale, 21-item Hamilton Rating Scale for Depression, and Young Mania Rating Scale scales using pooled patient data from two clinical trials. Clinical Global Impression for Schizoaffective Disorder scores were then linked to corresponding Positive and Negative Syndrome Scale scores.

Results: Content validity was strong. Inter-rater agreement was

good to excellent for most scales and subscales (intra-class correlation coefficient ≥0.50). Test-retest showed good reproducibility, with intraclass correlation coefficients ranging from 0.444 to 0.898. Spearman correlations between Clinical Global Impression for Schizoaffective Disorder domains and corresponding symptom scales were 0.60 or greater, and effect sizes for Clinical Global Impression for Schizoaffective Disorder overall and domain scores were similar to Positive and Negative Syndrome Scale, Young Mania Rating Scale, and 21-item Hamilton Rating Scale for Depression scores. Raters anticipated that the scale might be less effective in distinguishing negative from depressive symptoms, and, in fact, the results here may reflect that clinical reality.

Conclusion: Multiple lines of evidence support the reliability and validity of the Clinical Global Impression for Schizoaffective Disorder for studies in schizoaffective disorder.

INTRODUCTION

Schizoaffective disorder is a complex psychiatric condition characterized by concurrent psychotic and mood symptoms. A recent review examining schizoaffective disorder compared with schizophrenia and mood disorders showed that patients with schizoaffective disorder have higher hospitalization rates and greater frequencies of suicidal behavior than patients with schizophrenia and mood disorders.[1] The prognosis for patients with schizoaffective disorder lies between that of patients with schizophrenia and mood disorders in terms of overall outcomes, medication use, work functioning, and rehospitalization.[2]

The *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)*, recognizes schizoaffective disorder as a distinct entity whose symptoms encompass criteria for schizophrenia, as well as for a major

depressive, manic, or mixed episode.[3] Diagnosis with schizoaffective disorder requires that symptoms of a mood disorder be present for a substantial proportion of the illness and that psychotic symptoms be present for at least two weeks in the absence of prominent mood symptoms. As a result of these criteria, challenges exist regarding diagnostic reliability and stability of the diagnosis.[4,5] Treatment can differentially affect the various symptom domains in schizoaffective disorder, including positive, negative, cognitive, excited, manic, and depressive symptoms. Thus, for a comprehensive measurement of changes in the severity of the condition, a wide variety of phenomena must be encompassed, and redundancy must be avoided.

Studies of schizoaffective disorder have traditionally used a combination of scales designed to assess schizophrenia or mood disorders. However, there has been no scale to assess the global severity and change specific to schizoaffective disorder. The Positive and Negative Syndrome Scale (PANSS)[6] has been found to differentiate between schizophrenia and schizoaffective patients[7] but has not been specifically validated for schizoaffective disorder. Scales that evaluate mood symptoms, such as the Hamilton Rating Scale for Depression (HAM-D),[8] the Montgomery-Asberg Depression Rating Scale (MADRS),[9] and the Young Mania Rating Scale (YMRS),[10] have been validated only for patients with mood disorders. Although the Calgary Depression Scale for Schizophrenia has been validated for schizophrenia,[11] it has not been validated for schizoaffective disorder.

Global scales are valuable to research and clinical practice because they take into consideration all available evidence about the patient's situation, capture data that may be missed in a particular rating scale, and are sensitive to the clinical and functional impact of

severe symptoms.[12-15] They tend to be easy for raters to learn and have strong face validity that is broadly accepted within the clinical field. The original Clinical Global Impression (CGI) scale has been widely used to measure symptom severity and treatment efficacy in patients with psychoses, some of whom had schizoaffective disorder.[16] More recently, many disorder-specific CGI scales have been introduced, among them the CGI for schizophrenia (CGI-SCH)[17] and for bipolar disorder (CGI-BP)[12]; these scales have been useful in capturing the full extent of the disorders' manifestations. Development of a CGI scale specific for schizoaffective disorder and its four syndromal domains (positive, negative, manic, and depressive) would be useful for researchers and clinicians.

A panel of experts created a new scale based on the original CGI,[16] the CGI-BP,[12] and the CGI-SCH[17] and adapted it using the CGI-Severity (CGI-S) and CGI-Improvement (CGI-I) scales. The new Clinical Global Impression for Schizoaffective Disorder (CGI-SCA) scale measures severity (CGI-S-SCA) and change (CGI-C-SCA) with each measurement including ratings of the four domains of schizoaffective disorder, as well as overall ratings. A CGI-SCA instruction manual also has been developed to define the four domains and to help clinicians differentiate manic from positive symptoms and depressive from negative symptoms.

The objective of this study is to describe the psychometric evaluation of the CGI-SCA. The psychometric evaluation plan included tests for 1) content validity, 2) inter-rater reliability, and 3) test-retest reliability. Convergent and divergent validity and effect size also were determined after actual use of the CGI-SCA in two large international trials of paliperidone extended-release (ER) that were part of the first registration study for the treatment of schizoaffective disorder.[18,19]

TABLE 1. Clinical Global Impression Scale for Schizoaffective Disorder

I. SYMPTOM SEVERITY RATING

First rate the severity of the four main symptom domains.

Domain	Normal, not at all ill	Minimally ill	Mildly ill	Moderately ill	Markedly ill	Severely ill	Among most severely ill
1. POSITIVE symptoms	1	2	3	4	5	6	7
2. NEGATIVE symptoms	1	2	3	4	5	6	7
3. DEPRESSIVE symptoms	1	2	3	4	5	6	7
4. MANIC symptoms	1	2	3	4	5	6	7

Now, taking into account your symptomatic ratings, the patient's level of functioning, all other ratings you have completed, and any other relevant information, assign an overall global severity rating.

	Normal, not at all ill	Minimally ill	Mildly ill	Moderately ill	Markedly ill	Severely ill	Among most severely ill
OVERALL illness severity	1	2	3	4	5	6	7

II. SYMPTOM CHANGE RATING

First rate the degree of change in the four main symptom domains.

Domain	Very much improved	Much Improved	Minimally Improved	No Change	Minimally Worse	Much Worse	Very Much Worse
1. POSITIVE symptoms	1	2	3	4	5	6	7
2. NEGATIVE symptoms	1	2	3	4	5	6	7
3. DEPRESSIVE symptoms	1	2	3	4	5	6	7
4. MANIC symptoms	1	2	3	4	5	6	7

Now, taking into account your symptomatic ratings, the patient's level of functioning, all other ratings you have completed, and any other relevant information, assign an overall global change rating.

	Very much improved	Much Improved	Minimally Improved	No Change	Minimally Worse	Much Worse	Very Much Worse
OVERALL illness change	1	2	3	4	5	6	7

METHODS

Description of the CGI-SCA. The CGI-SCA consists of two subscales: severity of illness (CGI-S-SCA) and degree of change (CGI-C-SCA) (Table 1). The severity of illness subscale evaluates the overall severity of the disorder during the week before the assessment, whereas the degree of change subscale evaluates the change during the week before assessment compared with the week before beginning treatment. Each scale includes four domains (positive, negative, manic, and depressive), which are evaluated using a 7-point ordinal scale. A short definition of each domain is included in the instrument itself, and an instruction manual contains a detailed description of each domain. Overall ratings were performed as a separate and distinct gestalt that considered total severity and impact on functioning rather than a mathematical operation performed using the domain scores.

Content validity. *Sample.* The raters for the two clinical trials that used the CGI-SCA[18,19] were trained at three investigator meetings in the United States, India, or Eastern Europe by two of the authors (Allen, Daniel). These meetings used similar training materials, and distinctions were made between positive, negative, manic, and depressive symptoms. Raters were required to have clinical experience with at least one patient with schizophrenia, schizoaffective disorder, and bipolar disorder and to have rated at least 10 patients with the PANSS, at least 10 patients with the YMRS, and at least 10 patients with any version of the HAM-D (or ≥5 patients with any version of the 21-item HAM-D [HAM-D-21] plus ≥5 patients with the MADRS), and at least 10 patients with CGI-S or CGI-C. The CGI raters also were required to have an MD/DO, PhD, or PsyD degree. As part of their training, raters scored two sets of videos that included PANSS, HAM-D-21, and YMRS interviews of one subject, played by a professional actor, at two time points. These videos were rated on severity at each time point and change between the two time points. A passing grade for raters training on the PANSS, HAM-D-21, YMRS, and CGI was defined as scoring within 1 point of an expert panel consensus score on 80 percent of the items in each scale of this first set of videos. This procedure yielded a total of 132 raters (41 from the United States, 35 from India, and 56 from Eastern Europe).

Analysis. To obtain an impression of the content validity of the CGI-SCA, raters completed a survey instrument as part of their initial training exercises. They were queried about the content of the items, the general clinical and research utility of the instrument, the clinical relevance of the overall measure of symptom severity, and functional impairment in evaluating schizoaffective disorder, whether or not the items were sensitive to changes in clinical status, ability of the CGI-SCA to distinguish between positive and manic and negative and depressive symptoms, ability to measure variance of disease severity and change, instrument's ease of use, and potential changes that might be made to the proposed scale. The first 14 items in the survey were assessed with a 6-point Likert scale ranging from 1 (strongly disagree) to 6 (strongly agree). A neutral rating was not included. Item 15 was an open-ended invitation: "If you feel it would be helpful to add an additional component(s), please specify what that would be." Descriptive statistics were used. Agreement was defined as a score of 4 or greater on the 6-point Likert scale. A threshold of 80-percent agreement was considered good agreement.[20]

Inter-rater reliability. *Sample.* To power a formal test of inter-rater reliability, ratings of the two previously described sets of training videos were explored using intra-class correlation coefficients (ICCs). It was determined that two raters scoring 12 subjects with sufficient variation in their symptoms would be adequate to achieve 80-percent power to demonstrate an ICC of 0.7 versus a value of zero using an F test at a 0.05 level of significance. To create this sample, two experienced clinicians not involved in the schizoaffective disorder trials interviewed an independent sample of 12 individuals diagnosed with schizoaffective disorder and administered the CGI-S-SCA, PANSS, YMRS, and HAM-D-21. Subjects were selected to exhibit a defined range of mild, moderate, or severe symptoms including excitement, depressive, and manic features. These interviews were videotaped for later scoring using the same procedures as in the clinical trials. Information covering the preceding one-week period was derived from patient self-report and direct patient observation and was included for rating. The two raters for this study were trained in the same way as the clinical trial investigators. An institutional review board approved the protocol for this portion of the study.

Analysis. Inter-rater reliability was measured by ICC. Inter-rater reliability was calculated separately for the CGI-SCA global score and each of its subscales (positive, negative, manic, and depressive) and for the PANSS, YMRS, and HAM-D-21. An ICC of 0.5 to 0.7 was considered moderate to good agreement, and >0.7 was considered excellent agreement.[20]

Test-retest reliability. *Sample and analysis.* The agreement of measures of the same subject by the same rater at different times was evaluated. Thirty trained and certified clinical raters participating in the schizoaffective disorder trials were chosen at random, 10 each from the United States, India, and Eastern Europe. The raters viewed and scored two sets of interviews of a single actor at two time points, designed to represent change in a single patient's condition over time. The CGI-S-SCA, PANSS, YMRS, and HAM-D-21 scales were scored for each set, and the CGI-C-SCA was scored for the change in clinical

status between the two sets of interviews. The same interviews were then viewed and scored again two weeks later. The correlation of pairs of scores from each rater for the first and second viewings was measured with the ICC using the same threshold previously described.

Validation in clinical trials.
Study design and sample. To compare the performance of the CGI-SCA with that of other concurrently measured established instruments in a large patient sample, pooled data were used from two international randomized, double-blind, placebo-controlled, six-week trials of paliperidone ER in the treatment of schizoaffective disorder.[18,19] Subjects had to be 18 to 65 years of age, inclusive, with a diagnosis of schizoaffective disorder based on the Structured Clinical Interview for *DSM-IV* Disorders. Entry criteria required that all subjects be experiencing an acute exacerbation lasting fewer than four weeks but occurring more than four days before screening and that all subjects have a PANSS total score 60 or greater and a score 4 or greater on at least two of the following: hostility, excitement, tension, uncooperativeness, or poor impulse control; and prominent mood symptoms (YMRS score of ≥16 and/or HAM-D-21 score of ≥16). A total of 614 subjects were available for analysis. Details of the subject population are reported elsewhere.[18,19]

Psychometric analyses.
Convergent and divergent validity. Convergent validity was evaluated by examining the correlation between baseline measurements of the CGI-S-SCA (overall and four domains) with other instruments measuring conceptually related symptom domains: PANSS positive, PANSS negative, HAM-D-21 (in patients with depression), and YMRS (in patients with mania). Correlations between CGI-S-SCA overall and individual CGI-S-SCA domains also were determined. Divergent validity was tested by evaluating the correlations between CGI-S-SCA domain scores

and conceptually unrelated symptom scales. Spearman rank correlations were used for these analyses. Correlation <0.5 indicated a weak association, 0.5 to 0.7 a strong association, and >0.7 a very strong association.

Effect size for change. Effect sizes for change for the CGI-S-SCA overall scale, positive domain, and negative domain, and for the PANSS total, positive factor, and negative factor scales were calculated for the total trial population. Effect sizes for change for the CGI-S-SCA manic and depressive domains, YMRS and HAM-D-21, were calculated for subpopulations according to their mood state at entry, as reflected by a baseline YMRS ≥16 or HAM-D-21 ≥16. Effect size was calculated using Cohen's d.[21] Scores for all doses of paliperidone ER were combined and compared with placebo for each time point (Day 4 and Weeks 1, 2, 3, 4, and 6) using observed cases. Effect sizes <0.3 were considered small, 0.3 to 0.5 moderate, and >0.5 large.

Linking analysis of CGI-SCA and PANSS. Linking analyses were performed to determine the correspondence of CGI-S-SCA severity ratings to the PANSS score and of CGI-C-SCA improvement ratings to percentage change from baseline in PANSS score by the method of Leucht.[22] The observed value of the CGI-S-SCA overall score at each time point was linked to the observed mean PANSS total score for subjects with that CGI score; for example, at Week 6, there were 20 subjects with a CGI-S-SCA score of 4, and the mean PANSS total for those subjects was 83.5. Likewise, the observed change from baseline as reflected in the CGI-C-SCA score was associated with the percent difference from baseline in PANSS total score at that time point. The EQUIPERCENTILE macro was used for these analyses.[23]

RESULTS

Content validity. Overall, 84 percent or greater of raters in each location concurred with each item

SCALE	INTRA-CLASS CORRELATION COEFFICIENT[†]
PANSS positive	0.813
PANSS negative	0.740
PANSS total	0.700
YMRS total	0.798
HAM-D-21 total	0.926
CGI-S-SCA positive	0.619
CGI-S-SCA negative	0.655
CGI-S-SCA depressive	0.883
CGI-S-SCA manic	0.754
CGI-S-SCA overall	0.500

TABLE 2. Inter-rater reliability*

* Based on videotaped interviews of 12 patients with schizoaffective disorder independently rated by two trained raters

† *p*<0.05 for all values

CGI-S-SCA: Clinical Global Impression of Severity for Schizoaffective Disorder; HAM-D-21: 21-item Hamilton Rating Scale for Depression; PANSS: Positive and Negative Syndrome Scale; YMRS, Young Mania Rating Scale

supporting the statement that the CGI-SCA was a useful measure of symptom severity and functional impairment in schizoaffective disorder (Likert scale rating ≥4), and 81 percent or greater concurred with each item supporting the statement that the instrument captured a large fraction of the variance. More than 90 percent of raters in each location concurred that the scale could distinguish between positive and manic symptoms (United States, 96.7%; India, 92.2%; Eastern Europe, 91.4%). The raters also agreed, although with greater regional

TABLE 3. Test-retest reliability*

SCALE	INTRA-CLASS CORRELATION COEFFICIENT		
	TIME 1	TIME 2	CHANGE
PANSS positive	0.755	0.855	–
PANSS negative	0.747	0.873	–
PANSS total	0.884	0.898	–
YMRS total	0.844	0.737	–
HAM-D-21 total	0.816	0.759	–
CGI-S-SCA positive	0.573	0.632	–
CGI-S-SCA negative	0.720	0.559	–
CGI-S-SCA depressive	0.667	0.444	–
CGI-S-SCA manic	0.692	0.756	–
CGI-S-SCA overall	0.888	0.517	–
CGI-C-SCA positive	–	–	0.526
CGI-C-SCA negative	–	–	0.475
CGI-S-SCA depressive	–	–	0.514
CGI-C-SCA manic	–	–	0.634
CGI-C-SCA overall	–	–	0.576

* Based on two sets of videotaped interviews of a single actor representing two time points rated by 10 randomly selected clinical trial raters from the US, India, and Eastern Europe for a total of 30 raters. First and second ratings by each rater were separated by 2 weeks.

† $p<0.05$ for all values

CGI-S-SCA: Clinical Global Impression of Severity for Schizoaffective Disorder; HAM-D-21: 21-item Hamilton Rating Scale for Depression; PANSS: Positive and Negative Syndrome Scale; YMRS, Young Mania Rating Scale

differences (United States, 80.6%; India, 74.8%; Eastern Europe, 86.2%), that the CGI-SCA could distinguish between negative and depressive symptoms.

Inter-rater and test-retest reliability. Good to excellent agreement among raters for most scales (ICC) was observed, especially those scales measuring affective symptoms (Table 2). The weakest correlation among raters was for the CGI-S-SCA overall rating (ICC=0.5). For test-retest reliability, the ICC for rating pairs on the CGI-S-SCA ranged from 0.573 to 0.888 for Time 1 and from 0.444 to 0.756 for Time 2 (Table 3). Pairs of test and retest ratings of the CGI-C-SCA (reflecting the ratings of change in the subjects' condition between Time 1 and Time 2) were less well correlated and

ranged from 0.475 to 0.634. A higher correlation between rating pairs (ICC>0.7) was observed for all traditional measures (PANSS positive and negative factor scores and PANSS, YMRS, and HAM-D-21 total scores).

Convergent and divergent validity. In general, Spearman coefficients were strong (>0.6; $p<0.0001$) between CGI-S-SCA domains and rating instruments that measured similar constructs. The strength of these correlations indicated good convergence of the severity ratings of various symptom clusters and their respective CGI-S-SCA domains (Table 4). Analysis of divergent validity showed that coefficients were strongest (>0.4) between CGI-S-SCA negative and CGI-S-SCA depressive (0.400) scores and between CGI-S-SCA depressive and PANSS negative (0.318) scores.

The CGI-S-SCA overall score was most strongly correlated with the CGI-S-SCA positive (0.719) and CGI-S-SCA manic (0.402) scores and less correlated with CGI-S-SCA negative (0.244) and CGI-S-SCA depressive (0.204) scores. Correlations between CGI-S-SCA domains were generally weak (<0.50).

Sensitivity to change. Effect sizes for change from baseline in CGI-S-SCA overall and domain scores were moderate and similar to those of the PANSS, YMRS, and HAM-D-21 at Week 6 and time points in between (Table 5). Among CGI-S-SCA scores at week 6, the CGI-S-SCA mania rating had the largest effect size (0.37), as did the YMRS among the other measures (0.43). At Week 6, the CGI-S-SCA negative and the PANSS negative performed least well (both 0.21).

Linking analysis of CGI-SCA and PANSS. The relationships between CGI-S-SCA overall and the PANSS total score (Figure 1A) and between CGI-C-SCA overall and PANSS percentage change (Figure 1B) both appear to be approximately linear. A change of 1 severity point on the CGI-S-SCA corresponded to an approximately 20-point increase

in PANSS total score in the mid-range of the CGI-S-SCA scale, and a CGI-S-SCA overall score of 5 (markedly ill) was equivalent to a PANSS total score of 97 to 101 at various time points. A CGI-C-SCA score of 4 (no change) was associated with 2 to 4 percent improvement in PANSS total score; a rating of 2 (much improved) corresponded to 33 to 38 percent improvement in PANSS total score.

DISCUSSION

The CGI-SCA was developed to provide global clinical ratings in the first large international clinical trials in schizoaffective disorder. The scale assesses multiple symptom domains as well as overall impression of severity and change. This is critical for trials in schizoaffective disorder because what may appear as improvement in one symptom domain may be associated with worsening in others. This study provides multiple lines of evidence to warrant the use of the CGI-SCA in clinical trials.

Raters polled about the instrument's content validity agreed that the CGI-SCA was a useful measure of symptom severity and functional impairment for patients with schizoaffective disorder. In particular, they found the scale to be effective in distinguishing between manic and positive symptoms. Investigators in some regions believed that this scale would be less effective in differentiating depressive from negative symptoms. This belief may reflect the difficulty of this distinction in theory, cultural differences between the raters, and the possibilities of different conceptions of the symptoms. Studies on additional subjects and further examination by region may be warranted.

Initial correlations of ratings among investigators were used to power formal testing of inter-rater reliability. This test showed good to excellent agreement for most scales. The overall rating was the least strongly correlated (ICC=0.5, Table 1). The relatively modest level of agreement for the overall rating may reflect differences in rating method: one rater used the

TABLE 4. Convergent and divergent validity correlations at baseline for 614 schizoaffective subjects in two pooled international trials

CORRELATION	SPEARMAN RANK CORRELATION N=614	P-VALUE
Convergence		
CGI-S-SCA positive and PANSS positive	0.609	<0.0001
CGI-S-SCA negative and PANSS negative	0.651	<0.0001
CGI-S-SCA depressive and total HAM-D-21*	0.653	<0.0001
CGI-S-SCA manic and total YMRS†	0.611	<0.0001
CGI-S-SCA overall and CGI-S-SCA positive	0.719	<0.0001
CGI-S-SCA overall and CGI-S-SCA negative	0.244	<0.0001
CGI-S-SCA overall and CGI-S-SCA depressive*	0.204	<0.0001
CGI-S-SCA overall and CGI-S-SCA manic†	0.402	<0.0001
Divergence		
CGI-S-SCA positive and PANSS negative	0.075	0.062
CGI-S-SCA negative and PANSS positive	−0.042	0.302
CGI-S-SCA negative and total HAM-D-21*	0.267	<0.0001
CGI-S-SCA negative and total YMRS†	−0.256	<0.0001
CGI-S-SCA depressive and PANSS negative	0.318	<0.0001
CGI-S-SCA positive and CGI-S-SCA negative	0.171	<0.0001
CGI-S-SCA positive and CGI-S-SCA depressive	0.089	0.073
CGI-S-SCA positive and CGI-S-SCA manic	0.255	<0.0001
CGI-S-SCA negative and CGI-S-SCA depressive	0.400	<0.0001
CGI-S-SCA negative and CGI-S-SCA manic	−0.127	0.005
CGI-S-SCA depressive and CGI-S-SCA manic	−0.304	<0.0001

*Baseline HAM-D-21 ≥16 (n=411)

† Baseline YMRS ≥16 (n=488)

CGI-S-SCA: Clinical Global Impression of Severity for Schizoaffective Disorder; HAM-D-21: 21-item Hamilton Rating Scale for Depression; PANSS: Positive and Negative Syndrome Scale; YMRS, Young Mania Rating Scale

TABLE 5. Effect sizes for change based on pooled clinical trial data including all doses of paliperidone extended release combined compared with placebo (N=614)

SCALE	DAY 4	WEEK 1	WEEK 2	WEEK 3	WEEK 4	WEEK 6
CGI-S-SCA						
Overall	0.16‡	0.19‡	0.29‡	0.22‡	0.24‡	0.32‡
Positive	0.19‡	0.20‡	0.27‡	0.14	0.13	0.30‡
Negative	0.02	0.14	0.21‡	0.08	0.20‡	0.21‡
Depression*	0.09	0.13	0.29‡	0.26‡	0.20	0.27‡
Manic†	0.19‡	0.23‡	0.19	0.24‡	0.12	0.37‡
PANSS						
Total	0.19‡	0.25‡	0.34‡	0.20‡	0.20‡	0.33‡
Positive	0.16‡	0.21‡	0.32‡	0.20‡	0.18	0.37‡
Negative	0.11	0.21‡	0.22‡	0.12	0.17	0.21‡
YMRS†	0.27‡	0.27‡	0.33‡	0.31‡	0.2	0.43‡
HAM-D-21*	0.22‡	0.23‡	0.41‡	0.19	0.14	0.26‡

* Baseline HAM-D-21 $\geq$16 (n=411)

† Baseline YMRS $\geq$16 (n=488)

‡ p<0.05 vs. placebo

CGI-S-SCA: Clinical Global Impression of Severity for Schizoaffective Disorder; HAM-D-21: 21-item Hamilton Rating Scale for Depression; PANSS: Positive and Negative Syndrome Scale; YMRS, Young Mania Rating Scale

recommended gestalt method, whereas the other used a more arithmetic averaging method. Revision of the CGI-SCA user manual will emphasize this point. Additionally, although there was some variation among subjects in the various domains, there was less variation in the overall scores, with two-thirds of subjects being rated 5 or 6. ICC calculation depends on the differences within and between subjects; hence, lack of variation mathematically determines the bounds of the ICC, and identical subjects will produce an ICC of 0.

Test-retest reliability performed with randomly selected raters using the CGI-SCA in clinical trials confirmed that measurements made by the scale were reproducible by raters over time. ICCs were higher for traditional instruments, likely reflecting the more extensive experience with these measures and the greater specificity of their individual items and anchor points. The finding of higher ICCs for severity ratings than for change ratings was consistent with the same finding for the CGI-BP.[12] The CGI-C-SCA ratings also were restricted because extreme ratings are rare in clinical trials.[24] In most instances, subjects were minimally improved (score of 3) or much improved (score of 2). This reflects a psychometric and statistical limitation of the bidirectional and categorical nature of the CGI-C-SCA.

Analyses of clinical trial data using the new CGI-SCA and traditional instruments suggested good convergent validity between CGI-S-SCA and scales measuring related symptom domains, with Spearman correlations generally >0.6, indicating strong correlation. The lower correlations between the overall CGI-S-SCA score and CGI-S-SCA negative and depressive scores suggested that the psychopathology in these domains contributed less to the overall perception of illness in this population. The divergent validity analysis found moderate correlation between the CGI-S-SCA depressive and negative symptoms domain scores and the PANSS negative factor scores. These results were anticipated by the investigators due to the difficulty in distinguishing between depressive and negative symptoms. Future analyses may need to use this scale in conjunction with additional depressive and negative symptoms assessment scales to confirm improvement in these symptoms.

Effect sizes for change from baseline obtained with the CGI-S-SCA were moderate and comparable to those obtained with established symptom measures, including the PANSS, YMRS, and HAM-D-21, at all time points. This finding demonstrates the sensitivity of the CGI-S-SCA to

change and its ability to differentiate between active treatment and placebo.

The relationship between scores on the CGI-SCA and PANSS total and between changes on these two scales is approximately linear, suggesting the possibility of extrapolating between CGI-SCA and PANSS scores for both severity and change, as has previously been demonstrated for the CGI.[24] A change of 1 point on the CGA-S-SCA overall scale corresponded to an approximately 20-point change in the PANSS total score in the mid-range of the CGI-SCA scale. The relationship between the CGI-S-SCA and PANSS total score is similar to that described for the CGI-S-SCH, in which a score of 5 (markedly ill on the CGI-S) corresponded to a PANSS total score of 93 to 96 at various time points, and a CGI-I score of 2 (much improved) was associated with a 40- to 53-percent reduction in PANSS total score at various time points.[22] These associations may help clinicians relate a change in a traditional instrument score to a change based on broader clinical judgment and may provide a framework for interpreting clinical trial results. Moreover, there is potentially great clinical benefit for mental health practitioners in having a succinct tool to assess and track changes in the multiple dimensions of symptoms associated with schizoaffective disorder.

Some limitations of the study should be considered when interpreting the findings. The lower inter-rater correlation for the CGI-SCA overall score reveals the disparity among raters in the method for making the global overall ratings and reinforces the importance of clarity in instructions and investigator training. Also, because the results are based on pooled data from two trials, some differences between the studies may be obscured. The ability of the scale to perform across the full range of severity could not be tested because the data were derived from clinical trials with few extreme cases. As a result, these data may not be generalizable to other populations of subjects with schizoaffective disorder,
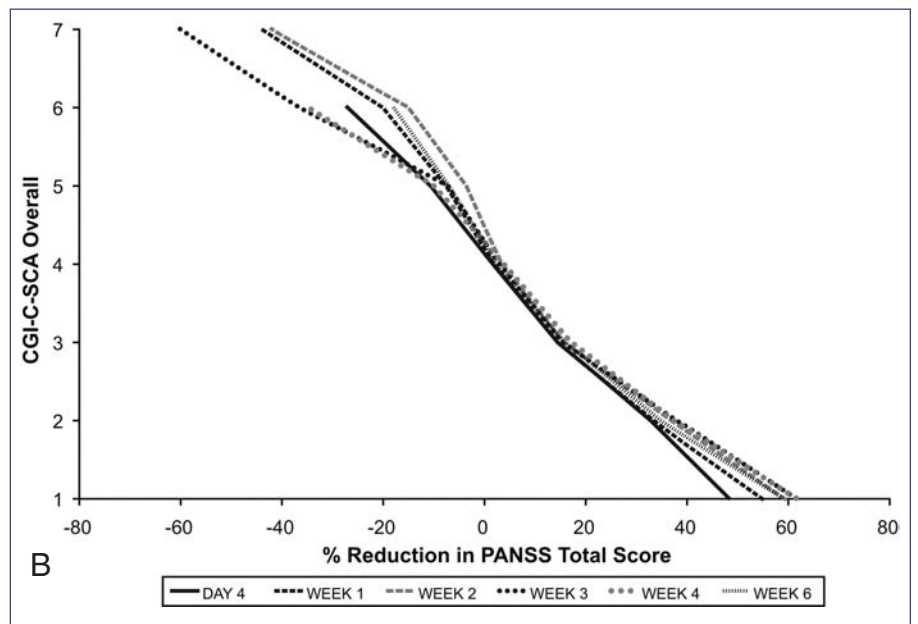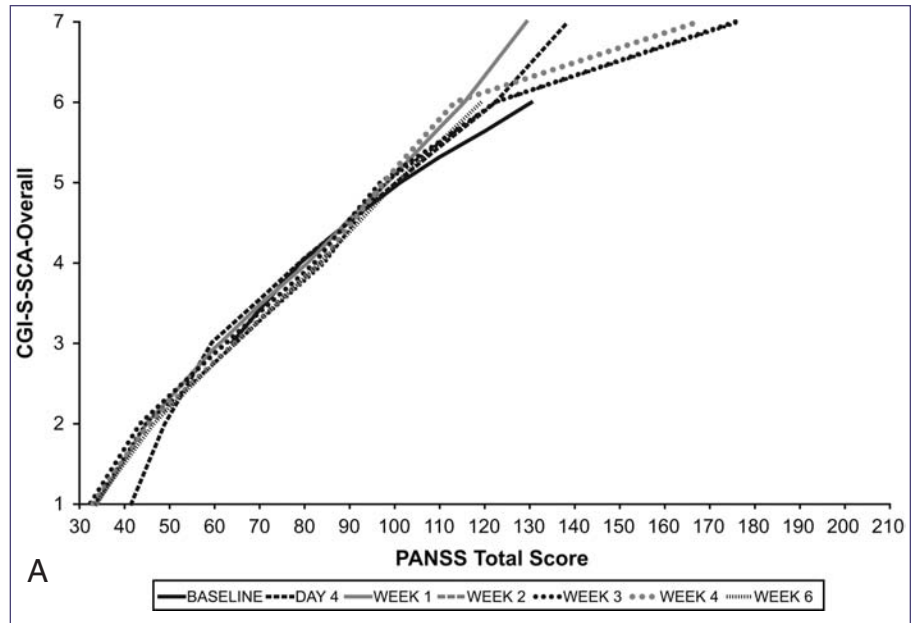

FIGURE 1. Relationship of (A) CGI-S-SCA overall and PANSS total scores and (B) CGI-C-SCA overall and percentage reduction from baseline in PANSS total scores.

particularly community samples. Additional analyses in these groups of subjects are needed to confirm these findings.

In conclusion, the CGI-SCA was developed to evaluate global impressions across the four relevant domains of schizoaffective disorder and to provide a global overall rating of patients' clinical status. These results provide multiple lines of evidence supporting the reliability and validity of the CGI-SCA for use in clinical trials

of patients with schizoaffective disorder. The CGI-SCA was able to measure severity and change in individual symptom domains, as well as in the overall status of patients with schizoaffective disorder in a manner comparable to established rating scales.

ACKNOWLEDGMENTS

REFERENCES

1. Cheniaux E, Landeira-Fernandez J, Lessa TL, et al. Does schizoaffective disorder really exist? A systematic review of the studies that compared schizoaffective disorder with schizophrenia or mood disorders. *J Affect Discord.* 2008;106:209–217.

2. Harrow M, Grossman LS, Herbener ES, Davies EW. Ten-year outcome: patients with schizoaffective disorders, schizophrenia, affective disorders and mood-incongruent psychotic symptoms. *Br J Psychiatry.* 2000;177:421–426.

3. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision.* Washington, DC: American Psychiatric Association; 2000.

4. Abrams DJ, Rojas DC, Arciniegas DB. Is schizoaffective disorder a distinct categorical diagnosis? A critical review of the literature. *Neuropsychiatric Diseases and Treatment.* 2008;4:1089–1109.

5. Craddock N, O'Donovan MC, Owen MJ. Psychosis genetics: modeling the relationship between schizophrenia, bipolar disorder, and mixed (or "schizoaffective") psychoses. *Schizophr Bull.* 2009;35:482–490.

6. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull.* 1987;13:261–276.

7. Lindenmayer JP, Kay SR, Van Praag H. Distinction of schizoaffective from schizophrenic profile: independent validation. *Schizophr Res.* 1989;2:423–424.

8. Hamilton M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol.* 1967;6:278–296.

9. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry.* 1979;134:382–389.

10. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry.* 1978;133:429–435.

11. Addington D, Addington J, Maticka-Tyndale E. Specificity of the Calgary Depression Scale for schizophrenics. *Schizophr Res.* 1994;11:239–244.

12. Spearing MK, Post RM, Leverich GS, et al. Modification of the Clinical Global Impressions (CGI) scale for use in bipolar illness (BP): the CGI-BP. *Psychiatry Res.* 1997;73:159–171.

13. Lehmann E. Practicable and valid approach to evaluate the efficacy of nootropic drugs by means of rating scales. *Pharmacopsychiatry.* 1984;17:71–75.

14. Leucht S, Engel RR. The relative sensitivity of the Clinical Global Impressions Scale and the Brief Psychiatric Rating Scale in antipsychotic drug trials. *Neuropsychopharmacology.* 2006;31:406–412.

15. Leucht S, Kane JM, Kissling W, et al. Clinical implications of Brief Psychiatric Rating Scale scores. *Br J Psychiatry.* 2005;187:366–371.

16. Guy W. *ECDEU Assessment Manual for Psychopharmacology (028 Clinical Global Impressions [CGI]).* Rockville, MD: National Institutes of Health, 1976:218–222.

17. Haro JM, Kamath SA, Ochoa S, et al. The Clinical Global Impression-Schizophrenia scale: a simple instrument to measure the diversity of symptoms present in schizophrenia. *Acta Psychiatr Scand Suppl.* 2003:16–23.

18. Canuso CM, Lindenmayer JP, Kosik-Gonzalez C, et al. A randomized, double-blind, placebo-controlled study of 2 dose ranges of paliperidone extended-release in the treatment of subjects with schizoaffective disorder. *J Clin Psychiatry.* 2010;71:587–598.

19. Canuso CM, Schooler N, Carothers J, et al. Paliperidone extended-release in schizoaffective disorder: a randomized, controlled study comparing a flexible dose with placebo in patients treated with and without antidepressants and/or mood stabilizers. *J Clin Psychopharmacol.* 2010;30:487–495.

20. Hays R, Revicki DA. Reliability and validity (including responsiveness). In: Fayers P, Hays R (eds). *Assessing Quality of Life in Clinical Trials, Second Edition.* New York, NY: Oxford University Press; 2005:25–40.

21. Cohen J. *Statistical Power Analysis for the Behavioral Sciences, Second Edition.* Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

22. Leucht S, Kane JM, Kissling W, et al. What does the PANSS mean? *Schizophr Res.* 2005;79:231–238.

23. Price LR, Lurie A, Wilkins C. Computer program exchange: EQUIPERCENT: a SAS program for calculating equivalent scores using the equipercentile method. *Appl Psychological Meas.* 2001;25:332. DOI 10.1177/01466210122032172.

24. Levine SZ, Rabinowitz J, Engel R, et al. Extrapolation between measures of symptom severity and change: an examination of the PANSS and CGI. *Schizophr Res.* 2008;98:318–322. ■