It is illegal to post this copyrighted PDF on any website.

Direct Comparison of the Psychometric Properties of Multiple Interview and Patient-Rated Assessments of Suicidal Ideation and Behavior in an Adult Psychiatric Inpatient Sample

Eric A. Youngstrom, PhD[a,*]; Ahmad Hameed, MD[b]; Michael A. Mitchell, MA[c]; Anna R. Van Meter, PhD[d]; Andrew J. Freeman, PhD[e]; Guillermo Perez Algorta, PhD[f]; Amanda M. White, BS[b]; Paula J. Clayton, MD[g]; Alan J. Gelenberg, MD[b]; and Roger E. Meyer, MD[b]

ABSTRACT

Objective: Compare the accuracy, agreement, internal consistency, and interrater reliability of 3 interviews to assess suicidal ideation and behavior in accordance with US Food and Drug Administration guidance about reporting categories.

Method: Adults admitted to a psychiatric inpatient unit (N = 199) completed 3 assessments of past month and lifetime suicidal ideation and behavior—the Columbia Suicide Severity Rating Scale (C-SSRS), the Suicide Tracking Scale (STS), and the Sheehan Suicidality Tracking Scale (S-STS)—in randomized, counterbalanced order. "Missing gold standard" latent class analyses defined categories for ideation and behavior. Analyses also evaluated the S-STS mapping to C-SSRS categories. Three trained judges re-rated 89 randomly selected interview videotapes. Cohen κ, the primary outcome measure, quantified agreement above chance. Data were collected between November 2011 and June 2013.

Results: All 3 assessments showed excellent accuracy for suicidal ideation (κ = 0.72 to 1.00) and attempts (κ = 0.82 to 0.95) calibrated against latent classes. Interrater agreement ranged from κ = 0.52 to 1.00. Interrater agreement about more granular C-SSRS categories varied more widely (κ = 0.48 to 1.00), and the C-SSRS and S-STS assigned significantly different numbers of cases to many categories. Cronbach α was < 0.55 for the C-SSRS ideation and between 0.78 and 0.92 for the other scales.

Conclusions: All 3 assessments showed good accuracy for broad categories of suicidal ideation and behavior. More granular, specific categories usually were rated reliably, but the C-SSRS and S-STS differed significantly in regard to which patients were assigned to these subcategories. Using any of these interviews would improve reliability over unstructured assessment in evaluating suicidal ideation and behavior. Clinical predictive validity of these interviews, and particularly the more granular categories, remains to be shown.

J Clin Psychiatry 2015;76(12):1676–1682
dx.doi.org/10.4088/JCP.14m09353
© Copyright 2015 Physicians Postgraduate Press, Inc.

[a]Department of Psychology, University of North Carolina at Chapel Hill
[b]Department of Psychiatry, Pennsylvania State University College of Medicine and Milton S. Hershey Medical Center, Hershey
[c]VA Pittsburgh Healthcare System, Pittsburgh, Pennsylvania
[d]Ferkauf Graduate School, Yeshiva University, Bronx, New York
[e]Department of Psychology, University of Nevada at Las Vegas
[f]Division of Health Research, Lancaster University, Lancaster, United Kingdom
[g]Emeritus, University of Minnesota, Minneapolis, and University of New Mexico, Albuquerque
**Corresponding author:* Eric A. Youngstrom, PhD, Department of Psychology, University of North Carolina at Chapel Hill, CB #3270, Davie Hall, Chapel Hill, NC 27599-3270 (eay@unc.edu).

Most clinicians do informal, unstructured interviews to assess suicidal ideation and behavior as well as diagnoses. With regard to suicidal ideation and behavior, usual practice results in terminological inconsistency and suboptimal diagnostic accuracy.[1] There has been a surge in both suicide assessment tools and interest in their use. In the past several years, research and anecdotal reports that pharmacologic agents may contribute to an increase in suicidal thoughts and behaviors[2–10] led to a requirement from the US Food and Drug Administration (FDA) Division of Psychiatry Products (and an accompanying Draft Guidance Document in 2009[11]) that all participants in clinical trials of central nervous system–active drugs be evaluated for treatment-emergent suicidal ideation and behavior using a scale that mapped to the Columbia Classification Algorithm of Suicide Assessment (C-CASA).[2,12] The C-CASA was developed as a tool for coding information from record reviews in the context of the FDA-supported analysis of data from antidepressant clinical trials in adolescents and children. In August 2012, the FDA went beyond its initial recommendation to specifically endorse more fine-grained subcategories of suicidal ideation and behavior in the Columbia Suicide Severity Rating Scale (C-SSRS)[1] as the standard against which other scales must be measured (Table 1). Between the publication of the original FDA Draft Guidance Document in 2009[11] and the 2012 Revision,[13] several trials proceeded using alternative assessment instruments of suicidal ideation/behavior. Competition between instrument developers became part of the commerce of clinical trials, as well as within clinical settings such as emergency departments and suicide hotlines. In order to provide guidance about the acceptability of the different methods, it is crucial to evaluate the comparative psychometrics of different assessment instruments.

There are 2 technical challenges in evaluating how the different assessment formats of suicidal ideation and behavior map onto the C-CASA. First, the C-CASA was designed for archival record review and classification of surveillance data, whereas the C-SSRS and other such instruments are designed as interviews or patient inventories. The C-CASA includes a set of categories that are "indeterminate" or "insufficient information," because archival data often lack sufficient detail for clear evaluation of intent. Clinical interviews should rarely or never lead to these classifications, as the clinician would probe and gather

It is illegal to post this copyrighted PDF on any website.

For reprints or permissions, contact permissions@psychiatrist.com. ♦ © 2015 Copyright Physicians Postgraduate Press, Inc.

It is illegal to post this copyrighted PDF on any website.

more data until able to arrive at a final judgment. The second technical challenge is the definition of a criterion for measuring accuracy. In spite of the 2012 FDA endorsement of the C-SSRS, there is at present no "gold standard" with perfect reliability and validity as a criterion for establishing suicidal ideation and behavior. Any single interview is fallible. All candidate instruments need investigation of interrater reliability, especially for fine-grained subcategories of suicidal ideation and behavior defined by the C-SSRS. The challenge of having a missing or imperfect gold standard is common in medicine, and there is a well-established statistical framework for addressing the problem. Latent class analysis (LCA) can evaluate the performance of multiple methods when the gold standard is uncertain, missing, or biased.[14,15] Latent class analysis has been used successfully in radiology, oncology, and pathology to evaluate test performance. In psychiatry, LCA can identify clusters of patients with similar clinical presentations.[16–18]

Our study compared 2 versions of an alternative suicidality scale, the Suicide Tracking Scale (STS)[19] and the Sheehan Suicidality Tracking Scale (S-STS),[20] to the C-SSRS. While the S-STS and the C-SSRS approach the information-gathering effort in somewhat different ways, the S-STS and the C-SSRS both explicitly matched the C-CASA (2009) domains. The key question is how these assessment instruments compare in evaluating the C-CASA and C-SSRS subcategories of interest per the FDA guidance documents of 2009 and 2012. Table 1 lists the key categories for prospective assessment recommended in the 2012 FDA Guidance Document. To meet these criteria, instruments need to cover the C-CASA algorithm domains, as well as ideation and behavior subcategories, nonsuicidal self-injurious behavior, and accidental injuries that the C-SSRS enumerates. Our study had 3 objectives:

1. Calibrate clinical interview–based versions of the STS, S-STS, and C-SSRS against LCA categories of suicide ideation and behavior, corresponding to the C-CASA categories.
2. Quantify agreement between the C-SSRS and S-STS on C-SSRS categories.
3. Examine interrater reliability of clinician-administered formats to see whether finer granularity of the ideation and/or behavior domains is associated with lower rates of interrater reliability.[21,22]

Forthcoming publications will contrast the operational aspects, patient comprehension of items, and patient and rater satisfaction of self-report and interview-formatted versions of the STS and the S-STS with the interview version of the C-SSRS, as well as suicidal behavior in the past month in the context of systematic self-reported risk and protective factors.

Clinical Points

- Using standardized tools could greatly improve consistency of terminology and accuracy of assessment of suicidal ideation and behavior.
- Three tools show excellent accuracy about suicidal ideation and behavior, although agreement is less good about more fine-grained subcategories, such as interrupted attempts.
- Adopting any of these tools would improve accuracy of assessment of suicidal ideation and behavior.

Table 1. Subcategories Adopted in the Revised FDA Guidance (2012) and Corresponding Order in the Interview Sequence During C-SSRS and S-STS Administration

	Order During Interview	
Revised FDA Guidance (2012) Subcategory	C-SSRS	S-STS
Passive suicidal ideation	1[a]	1
Active suicidal ideation: nonspecific (no method, intent, plan)	2[a]	2
Active suicidal ideation: method, but no intent or plan	3	3
Active suicidal ideation: method and intent, but no plan	4	5, 6
Active suicidal ideation: method, intent, and plan	5	4
Completed suicide	11[b]	12
Suicide attempt	6	9, 10[c]
Interrupted attempt	8	7, 11[d]
Aborted attempt	9	7, 11[d]
Preparatory acts	10	7, 11[d]
Self-injurious behavior, no suicidal intent	7	8

[a]Screening items. If both #1 and #2 are negative, the interview proceeds to suicidal behavior questions.
[b]On the lifetime/most recent clinical interview, there is no item for "completed suicide." On the "since last visit" version, there is a "completed suicide" item.
[c]If positive (>0) on suicide attempt, there are immediate follow-up questions.
[d]If positive (>0) on "active steps to prepare for a suicide attempt," these questions are asked after the suicide attempt follow-up questions.
Abbreviations: C-SSRS = Columbia Suicide Severity Rating Scale, FDA = US Food and Drug Administration, S-STS = Sheehan Suicidality Tracking Scale.

METHOD

The design was a cross-sectional, randomized study with adult psychiatric inpatients. Methodology included stratifying patients by age (<25 vs 25+ years) and presence/absence of psychosis using urn randomization[23] to counterbalance administration order (Table 2). Procedures included preplanned breaks (~90 minutes) between assessments, interviews completed on the same day with the same interviewer, video recording for reliability, and standardized training by the scale developers for all raters and interraters. The protocol also included patient-reported and interviewer-rated measures of satisfaction, along with a comprehensive interview about established suicide risk and protective factors. Participants received a $50 store gift card after completing study procedures. Study data were collected and managed using REDCap electronic data capture tools[24] hosted at the Penn State Milton S. Hershey Medical Center and College of Medicine, Hershey, Pennsylvania. Because it is crucial to use a standard time interval when inquiring about suicidal ideation and behavior,[25] this study inquired about both the past 30 days and lifetime history.

Participants

Staff queried all patients within a few days of admission to an urban, university-affiliated hospital between November 2011 and

It is illegal to post this copyrighted PDF on any website.

For reprints or permissions, contact permissions@psychiatrist.com. ♦ © 2015 Copyright Physicians Postgraduate Press, Inc.

It is illegal to post this copyrighted PDF on any website.

Table 2. Randomization Schedule and Patient Flow (N = 199)[a]

Sequence	N Completing	Assessment 1	Assessment 2	Assessment 3
Group 1: C-SSRS, S-STS Self-Report, STS Interview				
A	17	C-SSRS	S-STS self-report	STS interview
B	18	C-SSRS	STS interview	S-STS self-report
C	17	S-STS self-report	STS interview	C-SSRS
D	17	S-STS self-report	C-SSRS	STS interview
E	18	STS interview	S-STS self-report	C-SSRS
F	20	STS interview	C-SSRS	S-STS self-report
Group 2: C-SSRS, STS Self-Report, S-STS Interview				
A	16	C-SSRS	STS self-report	S-STS interview
B	16	C-SSRS	S-STS interview	STS self-report
C	16	STS self-report	S-STS interview	C-SSRS
D	14	STS self-report	C-SSRS	S-STS interview
E	14	S-STS interview	STS self-report	C-SSRS
F	16	S-STS interview	C-SSRS	STS self-report

[a]Unequal n due to randomization; n = 4 patients withdrew, and 1 provided incomplete data. Self-reported versions not included in analyses comparing interview formats.
Abbreviations: C-SSRS = Columbia Suicide Severity Rating Scale, STS = Suicide Tracking Scale, S-STS = Sheehan Suicidality Tracking Scale.

June 2013 regarding interest in participating in a research project. The Penn State Hershey Medical School institutional review board approved the study. The enrollment goal was to evaluate the range of suicidal ideation and behavior in a case series. We approached 274 patients; 70 were either not interested (most often due to the video recording) or not eligible. Altogether, 204 patients provided informed consent and were randomized. Four patients (2%) withdrew consent either immediately following randomization or at some point during the suicide assessments, and 1 participant declined to answer some items. Data were analyzed for the 199 completed interview sets.

Analytic Plan

Latent class analysis[26,27] evaluated the accuracy of the STS, S-STS, and the C-SSRS by comparing the scores across all 199 patients (Objective 1). Latent class analysis estimates the "true" status of each patient with regard to each C-CASA category based on the observed ratings from each of the 3 measures, following established procedures for when there is a missing or imperfect gold standard,[14] focusing on a 2-class solution.[15,26] Latent class analysis estimated κ (weighted percent accuracy), sensitivity (of cases with suicidal ideation and/or behavior, how many does the test identify correctly?), and specificity (how many cases without suicidal ideation and/or behavior does the test identify correctly?). Analyses used bootstrapping with 1,000 resamples to provide estimates robust to violations of assumptions.[28]

Because the FDA defined the C-SSRS subcategories of suicidal ideation and behavior as the "acceptable standard" in its 2012 guidance document, we also compared the newest version of the clinician-administered S-STS and the C-SSRS to each other to produce κ estimates for all categories with sufficient numbers of events in the newer C-CASA categories (Objective 2). Interrater reliability for each instrument was quantified via coding of video recordings by independent judges (Objective 3). Cronbach α quantified internal consistency reliability. Sensitivity analyses checked for administration order effects.

Measures

Columbia Suicide Severity Rating Scale. The Columbia Suicide Severity Rating Scale (C-SSRS)[1] is a semistructured interview widely used in clinical trials to document neuropsychiatric adverse events and treatment-emergent suicidal ideation severity and intensity, suicidal behavior, and nonsuicidal self-injurious behavior,[29,30] with good reported reliability and validity.[1] Two screening questions elicit whether there is any type of suicidal ideation. If yes, the interviewer asks about the most severe type of ideation. The remaining 3 ideation categories focus on the level of devising a specific plan and intent on carrying out such a plan. The ideation intensity section assesses several evidence-based risk factors such as frequency, duration, controllability, deterrents, and reasons for ideation. Suicidal behavior is grouped into 4 categories including actual suicide attempt, aborted attempt, interrupted attempt, and preparatory behaviors. Another item inquires about nonsuicidal self-injurious behavior. Questions have yes/no responses, except the intensity of ideation questions, which are on a 5-point Likert scale. Lethality uses a 5-point Likert rating scale ranging from "No physical damage or very minor physical damage" to "Death." Another item ranks potential lethality for any attempts with no actual medical damage on a 3-point Likert scale ranging from "Behavior not likely to result in injury" to "Behavior likely to result in injury." Kelly Posner, PhD, trained the interviewers in C-SSRS administration and scoring. The 3 interrater reliability judges read the same training materials and watched video of the training.

Suicide Tracking Scale. The Suicide Tracking Scale (STS) is an 8-item instrument assessing suicidal ideation and behavior initially derived from the Mini-International Neuropsychiatric Interview (MINI),[19] a diagnostic interview assessment. The interviewer is supposed to educate the patient about the different suicide-related definitions and terms before proceeding with the interview. Patients are asked to consider time, intensity, and severity when answering. All questions inquire about the seriousness of the thought(s) and/or behavior(s), eg, "How seriously did you think that you would be better off dead or wish you were dead?" Standardized response options include "Not at all," "A little," "Moderately," "Very," and "Extremely." The STS has shown good psychometric properties in prior clinical trials.[20,31]

Sheehan Suicidality Tracking Scale. The Sheehan Suicidality Tracking Scale (S-STS)[20] is a revised version of the Suicide Tracking Scale (STS) designed to more closely map the C-CASA algorithm. New components include capturing data points related to missed visits ("Fatal completed suicide," "Fatal, but not enough information to code as a completed suicide," "Known death from causes other than suicide," "Subject alive, but not available because of a suicide attempt," "Subject alive, but not available, for reasons other than suicide, or for uncertain

It is illegal to post this copyrighted PDF on any website.

For reprints or permissions, contact permissions@psychiatrist.com. ♦ © 2015 Copyright Physicians Postgraduate Press, Inc.

It is illegal to post this copyrighted PDF on any website.

Table 3. Objective 1: Accuracy of Interviews for Detecting Past Month and Lifetime Suicidal Ideation and Attempts Calibrated Against Latent Class Analysis Categories (N = 199)[a]

Measure	κ	(95% CI)	Sensitivity	Specificity	NPV	PPV
Suicidal ideation—past month						
C-SSRS (N = 199; 90% base rate)	0.78	(0.64–0.90)	0.95	1.00	0.68	1.00
S-STS (n = 92; 94% base rate)	0.75	(0.37–1.00)	0.98	0.83	0.71	0.99
STS (n = 107; 88% base rate)	0.91	(0.74–1.00)	1.00	0.85	1.00	0.98
Suicidal ideation—lifetime						
C-SSRS (N = 199; 98% base rate)	0.72	(0.22–1.00)	0.99	0.80	0.67	1.00
S-STS (n = 92; 99% base rate)	1.00	Undefined	1.00	1.00	1.00	1.00
STS (n = 107; 96% base rate)	1.00	Undefined	1.00	1.00	1.00	1.00
Suicide attempt—past month						
C-SSRS (N = 199; 37% base rate)	0.90	(0.78–0.96)	0.90	0.98	0.95	0.97
S-STS (n = 92; 37% base rate)	0.95	(0.88–1.00)	1.00	0.97	1.00	0.94
STS (n = 107; 36% base rate)	0.75	(0.61–0.87)	0.97	0.82	0.98	0.76
Suicide attempt—lifetime						
C-SSRS (N=199; 76% base rate)	0.82	(0.72–0.91)	0.93	0.94	0.80	0.98
S-STS (n = 92; 78% base rate)	0.87	(0.73–0.97)	0.97	0.90	0.90	0.97
STS (n = 107; 74% base rate)	0.88	(0.76–0.97)	0.99	0.86	0.96	0.95

[a]Confidence intervals based on 1,000 bootstrapping re-samples. All κ values and diagnostic efficiency parameters significant, $P < .0005$, 2-tailed. Statistics for nonsuicidal self-injury and preparatory acts available upon request as supplemental tables.

Abbreviations: C-SSRS = Columbia Suicide Severity Rating Scale, NPV = negative predictive value, PPV = positive predictive value, STS = Suicide Tracking Scale, S-STS = Sheehan Suicidality Tracking Scale.

reasons, or lost to follow up"), detailed suicide-related events history, and number of suicidal thoughts for both active and passive ideation. The S-STS, unlike the C-SSRS, measures "self-injurious behavior, intent unknown," fatal events with insufficient information to decide whether it was suicide, nonfatal ambiguous events, or "other (accidental, psychiatric medical), no deliberate self-harm." We did not report categories that only the S-STS measured, as they do not map to the FDA 2012 subcategories, but they were rare in our inpatient setting (0% to 2% endorsement). Preti et al[32] examined the psychometric properties of the scale in a sample of young Italian adults. David V. Sheehan, MD, trained the interviewers, and interrater reliability judges read the same training materials and watched video of the training.

RESULTS

Participants were 199 adult inpatients (mean ± SD age 38.5 ± 12.4 years; 57% female; 79% white, 18% black, 8% other, 13% Hispanic; 20% married; 87% earning a high school diploma or equivalent). Demography was similar to the catchment region. There was less than 2% attrition, and no significant associations with demographic or clinical characteristics were found. Contrary to concerns about patients "learning" from 1 interview and changing their responses to later interviews based on the "coaching," administration order did not account for significant variance (largest $\eta^2 = 0.02$).

The primary outcome measure was Cohen κ, using the established criteria from multiple authorities[33–35]: $\kappa \geq 0.75$ is considered "excellent," 0.6 to < 0.75 "substantial agreement," and < 0.4 "fair" or "poor." Rates of suicidal ideation were extremely high, with 98% of the sample reporting lifetime ideation and 90% reporting past month suicidal ideation. Similarly, 76% reported at least 1 lifetime attempt, and 37% reported a past month attempt. The κ value is affected by the base rate of the condition; it will be highest with a 50% base rate and often drops sharply with few misclassifications when the rates are extremely low or high.[36] Despite this, agreement between assessment methods and the latent class for each C-CASA category ranged from "good" (κ values = 0.60 to 0.74) to "excellent" (0.75 or higher) (Table 3). All instruments showed excellent sensitivity, detecting cases with suicidal ideation or behavior from 90% (C-SSRS for past month attempt) to 100% of the time (achieved by multiple scales, see Table 3). Specificity, or avoiding "false alarms" in cases that did not have suicidal ideation or behavior, ranged from 66% (STS past month attempt, patient-rated format) to 100% (again, attained by multiple scales). Table 3 also reports the positive and negative predictive values, illustrating how performance also changes depending on the base rate of the target.

Supplemental analyses compared agreement between the C-SSRS and S-STS on both the original C-CASA categories and the C-SSRS subcategories of suicidal ideation and behavior. Table 4 presents results for the C-SSRS and S-STS; pairwise agreement between the C-SSRS and S-STS ranged from κ = 0.51 to 0.86 about ideation and attempts on the original C-CASA categories. Agreement about the C-SSRS subcategories was poor, ranging from κ = −0.03 to 0.78, with more than a dozen falling in the "poor" range (κ < 0.4). These findings are largely driven by different administration orders and scoring algorithms: The S-STS attempts to classify each patient into a single category of ideation based on severity, whereas the C-SSRS may assign the same patient more than 1 category of ideation, depending on presentation (K. Posner, PhD, personal communication, May 7, 2014). For example, a participant who reported episodes with active ideation with intent, plan, and method might be counted in all 4 C-SSRS active ideation categories, whereas she/he would be counted only in the most severe ideation category on the S-STS (as well as in the severity score for the C-SSRS). As a result, the C-SSRS assigned significantly more cases to multiple intermediate categories, whereas the S-STS tended to classify cases at the highest level of reported ideation rather than listing each of the intermediate levels (Table 4).

We selected a random set of 90 tapes to recode. We then compared interrater reliability for 49 usable STS tapes versus 49 tapes from the same participants on the C-SSRS, and 39 usable S-STS and C-SSRS tapes available for interrater reliability; randomization resulted in slightly different *N*s (Table 5). Interrater agreement ranged from κ = 0.52 (C-SSRS about past month attempts) to 1.00 (STS about ideation, lifetime and past month). Agreement about the

It is illegal to post this copyrighted PDF on any website.

For reprints or permissions, contact permissions@psychiatrist.com. ♦ © 2015 Copyright Physicians Postgraduate Press, Inc.

It is illegal to post this copyrighted PDF on any website.

Table 4. Objective 2: Agreement Between C-SSRS and S-STS Interview Format on C-CASA and C-SSRS Categories (N = 92)[a]

	Past Month					Lifetime				
			Event Identification Rate[b,c]					Event Identification Rate[b,c]		
Category	% Agreement	κ	Both	C-SSRS Only	S-STS Only	% Agreement	κ	Both	C-SSRS Only	S-STS Only
C-CASA categories (2009)										
2) Attempt	94%	0.86***	35%	2%	4%	91%	0.72***	73%	4%	5%
3) Preparatory acts	81%	0.60***	30%	14%	5%	85%	0.67***	58%	11%	4%
4) Ideation	92%	0.51***	86%	2%	6%	99%	Undefined	99%	1%	0%
7) Nonsuicidal self-injury	87%	0.67***	21%	2%	**11%***	75%	0.51***	39%	5%	**20%***
C-SSRS subcategories (2012)										
1) Passive suicidal ideation	90%	0.14†	89%	8%	2%	98%	0.49***	97%	0%	2%
2) Active suicidal ideation—nonspecific (no method, intent, plan)	95%	−0.03†	0%	3%	2%	98%	0.49***	1%	2%	0%
3) Active suicidal ideation—method without intent or plan	85%	−0.02†	0%	**14%****	1%	85%	−0.02†	0%	**14%****	1%
4) Active suicidal ideation—method and intent, but no plan	71%	0.15†	6%	**26%*****	3%	69%	0.09†	3%	**29%*****	2%
5) Active suicidal ideation—method, intent, and plan	68%	0.37***	39%	10%	22%	67%	0.35***	46%	2%	**31%*****
7) Suicide attempt (cf C-CASA #2)	58%	0.28***	36%	**42%*****	0%	93%	0.78***	74%	4%	3%
8) Interrupted attempt	75%	0.33***	10%	**24%*****	1%	82%	0.59***	22%	12%	6%
9) Aborted attempt	66%	0.16†	8%	**30%*****	4%	74%	0.42***	22%	15%	11%
10) Preparatory acts (cf C-CASA #3)	61%	0.19*	11%	**37%*****	2%	68%	0.35***	21%	**28%*****	4%
11) Nonsuicidal self-injury (cf C-CASA #7)	71%	0.40***	24%	20%	9%	75%	0.52***	36%	8%	17%

[a]Subcategory 6, "completed suicide," has been deleted from the lifetime and past month subcategories because there were no completed suicides during the study.
[b]McNemar test determined whether one of the interviews identified significantly more test positives than the other. Significantly higher rates denoted with **boldface** and asterisks to show level of significance.
[c]Note that the C-SSRS asks about method, then intent, then plan, whereas the S-STS goes from method to plan, then intent.
†Not statistically significant.
*$P < .05$.
**$P < .005$.
***$P < .0005$, 2-tailed. All *** survive Holm stepdown Bonferroni correction.
Abbreviations: C-CASA = Columbia Classification Algorithm of Suicide Assessment, C-SSRS = Columbia Suicide Severity Rating Scale, S-STS = Sheehan Suicidality Tracking Scale.

more fine-grained C-SSRS categories ranged from 0.52 to 0.93 in the cases that also had the S-STS, whereas the S-STS was much more variable, κ = −0.07 for interrupted attempts (albeit with 85% classification accuracy) to 1.00 for multiple other categories. Although several of the S-STS κ values look lower than the corresponding C-SSRS values, there was no significant difference between them overall ($t_{27} = 1.17$, $P = .253$). One source of disagreement in the interrater reliability was that the patient sometimes changed her/his story over the course of the interview, often after clarifying definitions, but sometimes recanting. If the patient expressed 2 or more statements that could be scored differently, the interviewer and the re-rater sometimes based their scores on different statements or portions of the interview. Final scores usually represented the most severe interpretation reported.

Regarding internal consistency, from the C-SSRS, only the ideation score might be amenable to estimating internal consistency. Cronbach α was 0.53 for the lifetime and 0.50 for past month. On the S-STS, the total score (including both ideation and behavior) had α = 0.85 for lifetime and 0.87 for past month; the ideation score had α values = 0.90 and 0.92. The STS had α values = 0.80 and 0.78.

DISCUSSION

Assessment of suicidal ideation and behavior is 1 of the most high-stakes activities for patients, practicing clinicians, and researchers.[22,25] Concern about adverse events in clinical trials[4,5,10] led the FDA to issue guidance about acceptable strategies for measuring suicidal ideation, attempts, and related behaviors. The present study compared the C-SSRS, a semistructured interview endorsed by the FDA, with both the STS and the S-STS. All 3 interviews showed good interrater reliability and strong convergence measuring suicidal ideation, preparatory acts, suicidal behavior, and nonsuicidal self-injury. Reliability estimates were higher benchmarked against the latent class categories than head-to-head, because head-to-head comparisons are penalized for error in either interview.

The revised FDA guidance (2012) increased the granularity required for tracking suicidal ideation and behavior, following the C-SSRS structure. More granularity increases both interview complexity and burden on the interviewer and patient[37] and raises concerns about balancing false-positive and -negative findings both clinically and statistically.[38,39] The more granular categories (eg, ideation with method and intent, but no plan) increased disagreement between measures and often had lower interrater reliability within the same instrument. Some disagreement results from the C-SSRS algorithm assigning some patients to multiple ideation categories, whereas the S-STS algorithm assigns patients exclusively to the most severe reported category. It is not clear that greater granularity offers more clinical value than simpler indication of the most serious form of ideation.

It is illegal to post this copyrighted PDF on any website.

For reprints or permissions, contact permissions@psychiatrist.com. ♦ © 2015 Copyright Physicians Postgraduate Press, Inc.

It is illegal to post this copyrighted PDF on any website.

Table 5. Objective 3: Interrater Reliability Based on Independent Re-Rating of Videotape[a]

C-CASA (2009) Category	C-SSRS (n = 39)	S-STS (n = 39)	C-SSRS (n = 49)	STS[b] (n = 49)
Lifetime				
2) Attempt	0.75	0.62***	0.96***	0.91***
3) Preparatory acts	0.84	0.35*	0.67***	0.86***
4) Ideation	(100% ideation)	(100% ideation)	0.81***	1.00***
7) Nonsuicidal self-injury	0.77	0.69***	0.88***	0.76***
Past Month				
2) Attempt	0.52	0.67***	0.94***	0.96***
3) Preparatory acts	0.79	0.59***	0.82***	0.92***
4) Ideation	0.84	0.89***	1.00***	1.00***
7) Nonsuicidal self-injury	0.77	0.77***	0.91***	0.81***
2012 Subcategories, Lifetime				
1) Passive suicidal ideation	0.66	1.00***	0.83***	...
2) Active suicidal ideation—nonspecific (no method, intent, plan)	0.66 (97% accuracy)	0.00† (97% accuracy)	0.86***	...
3) Active suicidal ideation—method without intent or plan	0.72	1.00***	0.86***	...
4) Active suicidal ideation—method and intent, but no plan	0.77	1.00***	0.83***	...
5) Active suicidal ideation—method, intent, and plan	0.84	0.93***	0.77***	...
7) Suicide attempt (cf C-CASA #2)	0.75	0.69***	0.96***	...
8) Interrupted attempt	0.59	0.57***	0.90***	...
9) Aborted attempt	0.63	0.48**	0.76***	...
10) Preparatory acts (cf C-CASA #3)	0.89	0.56**	0.70***	...
11) Nonsuicidal self-injury (cf C-CASA#7)	0.73	1.00***	0.88***	...
2012 Subcategories, Past Month				
1) Passive suicidal ideation	0.79	0.86***	1.00***	...
2) Active suicidal ideation—nonspecific (no method, intent, plan)	0.93	1.00***	1.00***	...
3) Active suicidal ideation—method without intent or plan	0.83	1.00***	1.00***	...
4) Active suicidal ideation—method and intent, but no plan	0.89	1.00***	0.96***	...
5) Active suicidal ideation—method, intent, and plan	0.79	1.00***	0.75***	...
7) Suicide attempt (cf C-CASA #2)	0.52	0.39† (75% accuracy)	0.95***	...
8) Interrupted attempt	0.69	−0.07† (85% accuracy)	0.55***	...
9) Aborted attempt	0.60	0.00† (85% accuracy)	1.00***	...
10) Preparatory acts (cf C-CASA #3)	0.85	0.48* (85% accuracy)	0.69***	...
11) Nonsuicidal self-injury (cf C-CASA #7)	0.77	0.90***	0.92***	...

[a]Subcategory 6, "completed suicide," has been deleted from the lifetime and past month subcategories because there were no completed suicides during the study.
[b]The STS does not cover the 2012 subcategories, so they are omitted from the table.
†Not statistically significant.
*$P < .05$.
**$P < .005$.
***$P < .0005$, 2-tailed. All *** survive Holm stepdown Bonferroni correction.
Abbreviations: C-CASA = Columbia Classification Algorithm of Suicide Assessment, C-SSRS = Columbia Suicide Severity Rating Scale, STS = Suicide Tracking Scale, S-STS = Sheehan Suicidality Tracking Scale.

Prior studies suggest that assessing preparatory acts or aborted or interrupted attempts may have clinical value,[40,41] but future work needs to show whether the specific C-SSRS or S-STS algorithms show clinical predictive validity.[42,43]

Limitations include the high ideation rate in an inpatient sample, increasing the standard error of estimates (though all results still achieved high degrees of statistical significance) and lowering the κ estimates (often quite low compared to raw agreement rates). We could not examine associations between psychiatric diagnoses and suicidal ideation and behavior, nor prospective predictive validity. Strengths include large sample size and use of statistical methods designed for situations with imperfect gold standards (ie, LCA)[15] and robust to violations of assumptions (eg, bootstrapping).[28]

Overall, the C-SSRS, STS, and S-STS all demonstrated good detection of suicidal ideation and attempts in a psychiatric inpatient sample and good interrater reliability. Cross-instrument agreement varied widely depending on the subcategory. The instruments and scoring algorithms are not interchangeable for the subcategories. Future research needs to investigate the predictive validity, clinical utility, and patient acceptability of the more fine-grained categories, to determine if the extra complexity is warranted. All 3 are an improvement over the unstructured assessment methods that still are typical of clinical practice.[44–46] However, it is unclear whether the information codified by these instruments is sufficient by itself for assessment of risk for imminent suicidal behavior or suicide.

Submitted: June 29, 2014; accepted February 27, 2015.
Online first: November 10, 2015.
Potential conflicts of interest: **Dr Youngstrom** has consulted with Lundbeck and Otsuka. **Dr Gelenberg** has consulted with Zynx Health and owns stock in Healthcare Technology Systems, Inc. **Dr Meyer** has received grants or research support from Alkermes, NIDA, and grand rounds at the University of Connecticut. The other authors have no potential conflicts of interest to disclose.
Funding/support: This research was supported in part by an investigator initiated grant from Pfizer (PI: Dr Gelenberg) and partially supported by the Department of Veterans Affairs Office of Academic Affiliations Advanced Fellowship Program in Addiction Treatment (Mr Mitchell) and with resources and facilities at the VA Pittsburgh Healthcare System, Pittsburgh, Pennsylvania.

It is illegal to post this copyrighted PDF on any website.

For reprints or permissions, contact permissions@psychiatrist.com. ♦ © 2015 Copyright Physicians Postgraduate Press, Inc.

It is illegal to post this copyrighted PDF on any website.

Disclaimers: Dr Gelenberg, *JCP* Editor in Chief, was not involved in the editorial evaluation or decision to publish this article. The views expressed in this article are those of the authors and do not necessarily reflect the position of the Department of Veterans Affairs or the United States government.

REFERENCES

1. Posner K, Brown GK, Stanley B, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *Am J Psychiatry*. 2011;168(12):1266–1277.
2. Hammad TA, Laughren TP, Racoosin JA. Suicide rates in short-term randomized controlled trials of newer antidepressants. *J Clin Psychopharmacol*. 2006;26(2):203–207.
3. Christman DS, Faubion MD. Suicide attempt following initiation of topiramate. *Am J Psychiatry*. 2007;164(4):682–683.
4. Hammad TA, Laughren T, Racoosin J. Suicidality in pediatric patients treated with antidepressant drugs. *Arch Gen Psychiatry*. 2006;63(3):332–339.
5. Stone M, Laughren T, Jones ML, et al. Risk of suicidality in clinical trials of antidepressants in adults: analysis of proprietary data submitted to US Food and Drug Administration. *BMJ*. 2009;339:b2880.
6. Kuehn BM. FDA warns of adverse events linked to smoking cessation drug and antiepileptics. *JAMA*. 2008;299(10):1121–1122.
7. Kalinin VV. Suicidality and antiepileptic drugs: is there a link? *Drug Saf*. 2007;30(2):123–142.
8. Jepsen P, Johnsen SP, Sørensen HT. Risk of suicide in users of cardiovascular drugs: a review of the epidemiological evidence. *Am J Cardiovasc Drugs*. 2003;3(3):163–167.
9. Sørensen HT, Mellemkjaer L, Olsen JH. Risk of suicide in users of beta-adrenoceptor blockers, calcium channel blockers and angiotensin converting enzyme inhibitors. *Br J Clin Pharmacol*. 2001;52(3):313–318.
10. Miller M, Swanson SA, Azrael D, et al. Antidepressant dose, age, and the risk of deliberate self-harm. *JAMA Intern Med*. 2014;174(6):899–909.
11. US Department of Health and Human Services, Food and Drug Administration Center for Drug Evaluation and Research (CDER), Division of Psychiatry Products. Guidance for industry: suicidality: prospective assessment of occurrence in clinical trials. Silver Spring, MD: Department of Health and Human Services; August 2010.
12. Posner K, Oquendo MA, Gould M, et al. Columbia Classification Algorithm of Suicide Assessment (C-CASA): classification of suicidal events in the FDA's pediatric suicidal risk analysis of antidepressants. *Am J Psychiatry*. 2007;164(7):1035–1043.
13. US Department of Health and Human Services, Food and Drug Administration Center for Drug Evaluation and Research (CDER), Division of Psychiatry Products. Guidance for industry: suicidal ideation and behavior: prospective assessment of occurrence in clinical trials. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM225130.pdf. Published August 6, 2012. Accessed October 20, 2015.
14. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York, NY: Wiley; 2003.
15. Zhou X-H, Obuchowski NA, McClish DK. *Statistical Methods in Diagnostic Medicine*. New York, NY: Wiley; 2002.
16. Althoff RR, Rettew DC, Faraone SV, et al. Latent class analysis shows strong heritability of the child behavior checklist-juvenile bipolar phenotype. *Biol Psychiatry*. 2006;60(9):903–911.
17. Mahoney WJ, Szatmari P, MacLean JE, et al. Reliability and accuracy of differentiating pervasive developmental disorder subtypes. *J Am Acad Child Adolesc Psychiatry*. 1998;37(3):278–285.
18. Szatmari P, Volkmar F, Walter S. Evaluation of diagnostic criteria for autism using latent class models. *J Am Acad Child Adolesc Psychiatry*. 1995;34(2):216–222.
19. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (MINI.): 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.
20. Coric V, Stock EG, Pultz J, et al. Sheehan Suicidality Tracking Scale (Sheehan-STS): preliminary results from a multicenter clinical trial in generalized anxiety disorder. *Psychiatry (Edgmont)*. 2009;6(1):26–31.
21. Silverman MM, Berman AL, Sanddal ND, et al. Rebuilding the tower of Babel: a revised nomenclature for the study of suicide and suicidal behaviors, part 2: suicide-related ideations, communications, and behaviors. *Suicide Life Threat Behav*. 2007;37(3):264–277.
22. Silverman MM, Berman AL, Sanddal ND, et al. Rebuilding the tower of Babel: a revised nomenclature for the study of suicide and suicidal behaviors, part 1: background, rationale, and methodology. *Suicide Life Threat Behav*. 2007;37(3):248–263.
23. Project MATCH Research Group. Matching Alcoholism Treatments to Client Heterogeneity: Project MATCH posttreatment drinking outcomes. *J Stud Alcohol*. 1997;58(1):7–29.
24. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–381.
25. Meyer RE, Salzman C, Youngstrom EA, et al. Suicidality and risk of suicide—definition, drug safety concerns, and a necessary target for drug development: a consensus statement. *J Clin Psychiatry*. 2010;71(8):e1–e21.
26. McCutcheon AL. *Latent Class Analysis*. Thousand Oaks, CA: Sage; 1987.
27. Vermunt JK, Magidson J. Latent class cluster analysis. In: Hagenaars JA, McCutcheon AL, eds. *Applied Latent Class Analysis*. Cambridge, UK: Cambridge Univ Press; 2002:89–106.
28. Efron B. *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadelphia, PA: Society for Industrial and Applied Mathematics; 1982.
29. Hesdorffer DC, French JA, Posner K, et al. Suicidal ideation and behavior screening in intractable focal epilepsy eligible for drug trials. *Epilepsia*. 2013;54(5):879–887.
30. Williams JM, Anthenelli RM, Morris CD, et al. A randomized, double-blind, placebo-controlled study evaluating the safety and efficacy of varenicline for smoking cessation in patients with schizophrenia or schizoaffective disorder. *J Clin Psychiatry*. 2012;73(5):654–660.
31. Khan A, Khan SR, Hobus J, et al. Differential pattern of response in mood symptoms and suicide risk measures in severely ill depressed patients assigned to citalopram with placebo or citalopram combined with lithium: role of lithium levels. *J Psychiatr Res*. 2011;45(11):1489–1496.
32. Preti A, Sheehan DV, Coric V, et al. Sheehan Suicidality Tracking Scale (S-STS): reliability, convergent and discriminative validity in young Italian adults. *Compr Psychiatry*. 2013;54(7):842–849.
33. Fleiss JL. Measuring agreement between two judges on the presence or absence of a trait. *Biometrics*. 1975;31(3):651–659.
34. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
35. Cicchetti D, Bronen R, Spencer S, et al. Rating scales, scales of measurement, issues of reliability: resolving some critical issues for clinicians and researchers. *J Nerv Ment Dis*. 2006;194(8):557–564.
36. Kraemer HC. *Evaluating Medical Tests: Objective and Quantitative Guidelines*. Newbury Park, CA: Sage; 1992.
37. Silverman MM. Challenges to classifying suicidal ideations, communications, and behaviours. In: O'Connor RC, Platt S, Gordon J, eds. *International Handbook of Suicide Prevention: Research, Policy and Practice*. New York, NY: Wiley; 2011:9–25.
38. Kraemer HC, Lowe KK, Kupfer DJ. *To Your Health: How to Understand What Research Tells Us about Risk*. New York, NY: Oxford University Press; 2005.
39. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, MI: Lawrence Erlbaum; 1988.
40. Steer RA, Beck AT, Garrison B, et al. Eventual suicide in interrupted and uninterrupted attempters: a challenge to the cry-for-help hypothesis. *Suicide Life Threat Behav*. 1988;18(2):119–128.
41. Barber ME, Marzuk PM, Leon AC, et al. Aborted suicide attempts: a new classification of suicidal behavior. *Am J Psychiatry*. 1998;155(3):385–389.
42. Potkin SG, Alphs L, Hsu C, et al; InterSePT Study Group. Predicting suicidal risk in schizophrenic and schizoaffective patients in a prospective two-year trial. *Biol Psychiatry*. 2003;54(4):444–452.
43. Mundt JC, Greist JH, Jefferson JW, et al. Prediction of suicidal behavior in clinical research by lifetime suicidal ideation and behavior ascertained by the electronic Columbia-Suicide Severity Rating Scale. *J Clin Psychiatry*. 2013;74(9):887–893.
44. Jensen-Doss A, Youngstrom EA, Youngstrom JK, et al. Predictors and moderators of agreement between clinical and research diagnoses for children and adolescents. *J Consult Clin Psychol*. 2014;82(6):1151–1162.
45. Garb HN. Clinical judgment, clinical training, and professional experience. *Psychol Bull*. 1989;105(3):387–396.
46. Hughes CW. Objective assessment of suicide risk: significant improvements in assessment, classification, and prediction. *Am J Psychiatry*. 2011;168(12):1233–1234.

Editor's Note: We encourage authors to submit papers for consideration as a part of our Focus on Suicide section. Please contact Maria A. Oquendo, MD, at moquendo@psychiatrist.com.

It is illegal to post this copyrighted PDF on any website.

For reprints or permissions, contact permissions@psychiatrist.com. ♦ © 2015 Copyright Physicians Postgraduate Press, Inc.