Comprehensive Psychiatry 90 (2019) 37–42

Contents lists available at ScienceDirect

Comprehensive Psychiatry

journal homepage: www.elsevier.com/locate/comppsych

Reliability and validity of a self-report scale for daily assessments of the severity of anxiety symptoms

Mark Zimmerman *, Reina Kiefer, Sophie Kerr, Caroline Balling

Department of Psychiatry and Human Behavior, Brown Medical School, Rhode Island Hospital, Providence, RI, United States

ARTICLE INFO

ABSTRACT

Background: To evaluate the efficacy of rapidly effective treatments it is necessary to use measures that are designed to assess symptom severity over short intervals. In the present report from the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project, we modified our previously published anxiety scale and examined the reliability and validity of a daily version of the Clinically Useful Anxiety Outcome Scale (CUXOS-D, D indicates daily version).

Methods: Two thousand four hundred and ninety-one patients presenting for treatment to a partial hospital program completed the CUXOS-D as part of their initial paperwork and on a daily basis thereafter. Test-retest reliability was examined in 50 patients who completed the CUXOS-D twice on the same day. A subset of 73 patients was interviewed by a trained rater who administered the Hamilton Anxiety Rating Scale (HAMA) at baseline and on the day of discharge.

Results: The CUXOS-D had high internal consistency and test-retest reliability and was more highly correlated with another measure of anxiety symptoms than depressive symptoms. CUXOS-D scores progressively declined during the course of treatment, and scores on each successive day were significantly lower than the preceding day. The change in CUXOS-D scores was significantly correlated with a change in HAMA scores ($r = 0.61$, $p < .001$). A large effect size was found for both measures (CUXOS-D: $d = 1.22$; HAMA: $d = 0.93$).

Discussion: In a large sample of partial hospital patients mostly diagnosed with mood and anxiety disorders, we demonstrated some aspects of the reliability and validity of the CUXOS-D.

1. Introduction

The psychiatric field has only recently begun to ask fundamental questions regarding the effectiveness of our currently available treatments in real-world clinical practice. How well do these treatments work? For whom do they work best, and for whom are they ineffective? How many patients are and are not receiving evidence-based care, and is the provision of evidence-based care associated with better outcome? Are some clinicians more effective than others, and, assuming there are differences in patient outcomes between clinicians, is it possible to improve outcomes in patients treated by those clinicians who perform below average? Each of these, and other, questions can only be addressed if outcome is measured and measurement is incorporated into routine clinical practice.

The term "measurement-based care" has been coined in reference to the use of standardized scales to measure the outcome of psychiatric treatment [1,2]. In describing a measurement-based care approach towards treatment, Trivedi and colleagues [1] limited their focus to depressed patients. Other studies have likewise examined the benefit of measurement-based care with depressed patients [3,4]. However, most clinicians treat heterogeneous groups of patients with a variety of diagnoses. Diagnostic heterogeneity poses a challenge to adopt a measurement-based care approach towards outcome evaluation in clinical practice. To address the problem of diagnostic heterogeneity, we suggested that a couple of core and frequently occurring psychiatric constructs be measured for all patients regardless of their diagnoses [5]. Specifically, we recommended that depression and anxiety be measured in all patients and be considered psychiatric vital signs, analogous to the routine measurement of medical vital signs during medical evaluations.

The measurement of anxiety and depression would not preclude the use of additional disorder-specific assessments. However, considering the low frequency with which standardized scales are currently used by psychiatrists to measure outcome, we suggested that a focus on 2 measures of outcome would simplify and therefore advance efforts towards adopting measurement-based care approaches. In support of our recommendation to routinely assess anxiety and depression in clinical practice, we found that nearly 90% of 3000 psychiatric outpatients

* Corresponding author at: Rhode Island Hospital, Department of Psychiatry, 146 West River Street Suite 11b, Providence, RI 02904, United States.
E-mail address: mzimmerman@lifespan.org (M. Zimmerman).

https://doi.org/10.1016/j.comppsych.2018.12.014

who were interviewed with semi-structured diagnostic interviews reported clinically significant anxiety and/or depression [5]. In fact, the majority of patients in all 10 diagnostic categories examined had clinically significant anxiety or depression.

Four years ago, our clinical research group was asked to run the Rhode Island Hospital partial hospital program. In the partial hospital program (PHP) patients are treated 5 days per week. We thought it necessary to assess symptoms on a daily basis because we believe that quantitative, standardized measurement should be as routine a component of psychiatric care as the measurement of body temperature and blood pressure in patients who are hospitalized on medical and surgical units. We therefore modified the measures we had developed for outpatient treatment, which assess symptoms over the past week, to measure symptom intensity over the past day. Such a measure would also be useful in evaluating the efficacy of rapidly effective treatments such as ketamine in the treatment of depression and anxiety disorders.

In the present report from the MIDAS project, we examined the reliability and validity of a daily version of our anxiety measure—the Clinically Useful Anxiety Outcome Scale (CUXOS-D, the D indicates daily version). We examined the psychometric properties of the daily scale, its test-retest reliability, discriminant and convergent validity, and its sensitivity to change.

2. Method

A consecutive series of 2491 patients presenting for treatment to the Rhode Island Hospital Department of Psychiatry PHP completed the CUXOS-D as part of their initial paperwork and on a daily basis thereafter. The participation rate in the study was high. Only 2.1% patients refused consent, and 2.4% were unable to complete the scales. The Rhode Island Hospital institutional review committee approved the research protocol, and all patients provided informed, written consent.

The wording of the CUXOS-D items is the same as the original version of the scale, though the rating guidelines differ. Similar to the original CUXOS, the CUXOS-D contains 20 items assessing psychic and somatic symptoms of anxiety. It is not a disorder specific scale. We developed the scale as a general measure of anxiety symptoms so that it would be useful in the management of depressed patients, who often report high levels of anxiety in the absence of a specific anxiety disorder [6,7], in the monitoring of patients with different anxiety disorders, and patients with nondepressive and nonanxiety disorders. On the CUXOS-D the respondent is instructed to rate the symptom items on a 5-point ordinal scale indicating "how well the item describes you during the past day" (0 = not at all; 1 = a little bit; 2 = a moderate amount; 3 = quite a bit; 4 = extremely).

The patients also completed the Remission from Depression Questionnaire (RDQ) [8]. The domains covered on the RDQ were based on a literature review, our previous study of depressed patients' ratings of the relative importance of 16 factors in determining remission [9], and two focus groups with depressed patients. In the present study we used an expanded version of the nondepressive symptom subscale of the RDQ which included 8 items on an anxiety subscale. The items refer to the prior week and are rated on a 3-point rating scale (not at all or rarely true; sometimes true; often or almost always true).

A subset of 73 patients were interviewed by a trained rater who administered the 14-item Hamilton Anxiety Rating Scale (HAMA) [10] at baseline and on the day of discharge from the PHP.

The test-retest reliability of the CUXOS-D was examined in 50 patients, 28 of whom completed the measure twice on the same day of the initial evaluation and 22 of whom completed the scale in the middle or towards the end of their course of treatment. The retest interval was at least 2 h after the first completion of the scale. The patients were told that the purpose of the second administration was to test the performance of the scale, not to question the truthfulness or accuracy of their responses. When completing the CUXOS-D the first time, the patients were not aware that they would be asked to complete it a second time, thereby reducing recall bias.

2.1. Data analysis

We examined two types of reliability of the CUXOS-D—test-retest reliability and internal consistency. We examined convergent and discriminant validity [11] by comparing the correlation between the CUXOS-D and the RDQ subscales. We also compared the correlation between the CUXOS-D and RDQ depression and anxiety symptoms subscales by calculating the difference between the Fisher z transformations of the correlation coefficients and dividing the difference by the standard error [12].

We used paired *t*-tests to examine the significance of day-to-day change in CUXOS-D scores. We computed the effect size (Cohen's *d*) for each day of treatment compared to the first day of treatment. Because the treatment was naturalistic, the length of stay in the program depended on the patients' clinical status. That is, patients who improved quickly were discharged sooner than patients who improved more slowly or who did not improve. In order to examine the ability of the CUXOS-D to assess day-to-day change, it is necessary to control for length of stay. We therefore examined the day-to-day change for the first 5 days of treatment in patients who were in the program at least 5 days, and similarly examined the day-to-day change over the first 10 days of treatment in patients who were treated for at least 10 days.

Finally, we compared the change in CUXOS-D and HAMA scores from baseline to discharge for patients who completed at least one week of treatment. We also computed the effect size (Cohen's *d*) for both measures.

3. Results

3.1. Demographic and diagnostic characteristics

The sample of 2491 patients included 728 (29.2%) men, 1712 (68.7%) women, 32 (1.3%) transgender patients, 16 patients who identified as gender queer/androgynous (0.6%), and 3 patient who identified as "Other" (0.1%). The patients ranged in age from 18 to 88 years (M = 37.3, SD = 13.9). Approximately one-fourth of the subjects were married (22.5%, $n = 561$); the remainder were single (43.5%, $n = 1083$), divorced (14.6%, $n = 363$), separated (4.3%, $n = 107$), widowed (2.0%, $n = 50$), or living with someone as if in a marital relationship (13.1%, $n = 327$). The educational level achieved by the subjects was: 7.6% ($n = 188$) did not graduate high school, 53.2% ($n = 1325$) graduated high school or achieved equivalency, 27.5% ($n = 685$) graduated college, and 9.4% ($n = 233$) completed graduate/professional school. Sixty patients (2.4%) did not specify their level of education. The racial composition of the sample was 75.2% ($n = 1875$) white, 6.5% ($n = 163$) black, 11.0% ($n = 274$) Hispanic, 2.3% ($n = 53$) Asian, 4.9% ($n = 122$) from another or a combination of the above racial backgrounds, and data was missing for 4 patients.

The data in Table 1 shows the diagnostic characteristics of the 2491 patients who completed the CUXOS-D at admission to the partial program. The most frequent DSM-IV diagnoses were major depressive disorder (56.4%), generalized anxiety disorder (40.2%), posttraumatic stress disorder (28.2%), and panic disorder (19.7%).

3.2. Internal consistency and test-retest reliability of the CUXOS-D

Internal consistency coefficients were computed separately for the 2491 patients who completed the scale at intake and 1336 patients who completed the scale at discharge after being in treatment for at least a week. Patients with more than two items missing were excluded from the analyses. The CUXOS-D demonstrated excellent internal consistency at both time points (Cronbach's alpha at intake = 0.93;

Table 1
DSM-IV Axis I diagnoses of 2491 partial hospital patients.

DSM-IV diagnosis	*n*	%
Major depressive disorder	1405	56.4
Bipolar disorder	313	12.6
Dysthymic disorder	111	4.5
Generalized anxiety disorder	1002	40.2
Panic disorder	491	19.7
Social phobia	447	17.9
Specific phobia	107	4.3
Obsessive-compulsive disorder	130	5.2
Posttraumatic stress disorder	702	28.2
Adjustment disorder	128	5.1
Schizophrenia/schizoaffective	49	2.0
Eating disorder	82	3.3
Alcohol abuse/dependence	295	11.8
Drug abuse/dependence	307	12.3
Somatoform disorder	113	4.5
Attention deficit disorder	307	12.3
Impulse control disorder	133	5.3

Individuals could be given more than one diagnosis.

Cronbach's alpha at follow-up = 0.95). The data in Table 2 shows the correlation between each item and the total scale score (minus the contribution of the item to the total score). All item-scale correlations at baseline and follow-up were significant (Mean at intake = 0.62; Mean at follow-up = 0.68).

The aforementioned analyses were conducted in the diagnostically heterogeneous sample. We conducted post hoc analyses to examine internal consistency and item-scale correlations in diagnostically homogeneous groups. We conducted these analyses for the 7 disorders that were the principal diagnosis in at least 40 patients. The data in Table 3 shows that Cronbach's alpha was greater than 0.90 in all 7 diagnostic groups (range 0.91 to 0.95). All item-scale correlations were significant in all diagnostic groups, and the mean of the item-scale correlations ranged from 0.56 to 0.66.

The test-retest reliability of the CUXOS-D was examined in 28 patients at baseline and a separate group of 22 patients during follow-up. At both time points the test-retest reliability of the total symptom score was high ($r = 0.95$ and 0.96, respectively), and the test-retest reliability of every item was significant (Mean at intake = 0.81; Mean at follow-up = 0.85) (Table 4).

Table 2
Item-Total correlations of the Clinically Useful Anxiety Outcome Scale-Daily Version (CUXOS-D) items at intake ($n = 2491$) and discharge ($n = 1336$)[a].

CUXOS-D item	Intake	Discharge
Anxious mood	.63	.73
Worry about future	.63	.74
Excessive worry	.60	.73
Exaggerated startle response	.64	.74
"Keyed up" feeling	.67	.77
Fearfulness	.63	.74
Muscle tension	.61	.67
Jittery feeling	.69	.75
Shortness of breath	.70	.74
Pounding heart	.72	.78
Clammy palms	.61	.63
Dry mouth	.60	.63
Dizziness	.66	.72
Nausea	.64	.64
Diarrhea	.45	.41
Hot flashes/chills	.62	.65
Frequent urination	.48	.50
Lump in throat feeling	.58	.62
Sweating	.62	.68
Paresthesia	.57	.66

All correlation coefficients are significant at $p < .001$.
[a] The sample of patients at discharge were treated in the program for at least one week.

3.3. Discriminant and convergent validity of the CUXOS-D

The CUXOS-D was more highly correlated with the RDQ anxiety symptoms subscale than with the RDQ depression subscale (0.74 vs. 0.55, $z = 10.6$, $p < .01$). The data in Table 3 shows that for all 7 diagnostic subgroups the CUXOS-D was more highly correlated with the RDQ anxiety symptoms subscale than with the RDQ depression subscale. The difference between the correlations was significant for major depressive disorder ($z = -9.00$, $p < .001$), bipolar disorder ($z = -3.56$, $p < .001$), posttraumatic stress disorder ($z = -2.25$, $p < .05$), and adjustment disorder ($z = -2.29$, $p < .05$), but not for generalized anxiety disorder ($z = -0.97$, $p = .33$), panic disorder ($z = -1.02$, $p = .31$), and schizophrenia/schizoaffective disorder ($z = -0.31$, $p = .76$).

3.4. Sensitivity to change

The average length of stay in the PHP was 7.9 days (SD = 5.2). For the 1336 patients treated for 5 or more days, the mean CUXOS-D score decreased by 41.6% from admission to discharge (42.1 ± 18.2 vs. 24.6 ± 19.5, paired $t = 36.9$, $p < .001$). The CUXOS-D scores progressively declined throughout the course of treatment, and the scores on each successive day were significantly lower than the preceding day (Table 5). The results were similar for the 529 patients who were treated for 10 or more days though the degree of change from day to day was less because the rate of improvement was slower in the patients who were treated for a longer amount of time in the program (Table 6).

Seventy-three patients were reevaluated on the HAMA at discharge, at least one week after the initial evaluation (mean = 10.0 days, SD = 4.3, range 5–25 days). At discharge, CUXOS-D scores were significantly lower than baseline (48.0 ± 14.2 vs. 26.8 ± 20.2, paired $t = 11.1$, $p < .001$). Likewise, scores on the HAMA were significantly lower at discharge (25.0 ± 8.3 vs. 16.2 ± 10.6, paired $t = 8.1$, $p < .001$). The change in CUXOS-D scores was significantly correlated with a change in HAMA scores ($r = 0.61$, $p < .001$). A large effect size of treatment was found with both measures (CUXOS-D: $d = 1.22$; HAMA: $d = 0.93$).

4. Discussion

It is feasible to have patients in a PHP complete a self-administered anxiety scale on a daily basis. This is consistent with our previous study that demonstrated that brief self-report scales are feasible to incorporate into routine clinical outpatient practice [13]. On average the scale takes less than 2 min to complete, and more than 95% of patients are able to complete it in less than 3 min [14].

The results of this large validation study of the CUXOS-D show that it is a reliable and valid measure of anxiety severity. The CUXOS-D achieved high levels of internal consistency and test-retest reliability and was more highly correlated with another measure which assessed symptoms of anxiety than with a measure of depressive symptoms thereby supporting the convergent and discriminant validity of the scale. Finally, the CUXOS-D was a valid measure of symptom change.

The present study was of a heterogeneous patient sample. Earlier research from our clinical-research group found that most psychiatric patients, regardless of diagnosis, report clinically significant levels of anxiety, and we therefore recommended that clinicians consider anxiety to be a psychiatric vital sign to be measured at every clinical encounter. Medical vital signs such as blood pressure and body temperature provide important pieces of clinical information that are not diagnostic specific. We conceptualize anxiety in the same manner. That is, levels of anxiety are particularly elevated in patients with anxiety disorders, but also are elevated in many patients with other psychiatric disorders. This transdiagnostic approach is consistent with dimensional approaches towards psychiatric classification [15]. We also examined internal consistency, item scale correlations, convergent and discriminant validity

Table 3
Item-Total correlations, internal consistency, convergent and discriminant validity of the Clinically Useful Anxiety Outcome Scale-Daily Version (CUXOS-D) at intake in patients with various principal diagnoses.

CUXOS-D item	MDD ($n = 1099$)	Bipolar disorder ($n = 253$)	GAD ($n = 142$)	Panic disorder ($n = 77$)	PTSD ($n = 222$)	Adjustment ($n = 94$)	Schizophrenia ($n = 44$)
Anxious mood	.61	.61	.52	.64	.60	.62	.60
Worry about future	.61	.60	.48	.68	.55	.63	.58
Excessive worry	.56	.61	.57	.54	.62	.58	.64
Exaggerated startle response	.62	.61	.59	.69	.61	.77	.60
"Keyed up" feeling	.65	.67	.59	.73	.64	.61	.75
Fearfulness	.62	.63	.53	.58	.61	.61	.67
Muscle tension	.59	.54	.53	.74	.57	.65	.62
Jittery feeling	.68	.66	.67	.64	.64	.72	.57
Shortness of breath	.67	.71	.66	.67	.67	.78	.70
Pounding heart	.68	.72	.71	.78	.70	.78	.70
Clammy palms	.62	.59	.59	.67	.52	.66	.59
Dry mouth	.61	.56	.53	.63	.54	.78	.50
Dizziness	.64	.63	.58	.66	.62	.78	.66
Nausea	.61	.60	.58	.48	.65	.78	.79
Diarrhea	.41	.40	.42	.52	.42	.43	.40
Hot flashes/chills	.60	.60	.55	.52	.62	.63	.62
Frequent urination	.45	.47	.41	.60	.44	.63	.50
Lump in throat feeling	.57	.61	.53	.55	.51	.63	.64
Sweating	.60	.61	.57	.65	.57	.60	.64
Paresthesia	.57	.53	.56	.59	.51	.62	.57
Mean of item-scale correlations	.60	.60	.56	.63	.58	.66	.62
Internal consistency	.93	.93	.91	.93	.92	.95	.94
RDQ depression	.48	.59	.52	.53	.56	.69	.62
RDQ anxiety	.72	.76	.60	.64	.69	.83	.66

MDD indication major depressive disorder; GAD, generalized anxiety disorder; PTSD, posttraumatic stress disorder; and RDQ, Remission from Depression Questionnaire.
All correlation coefficients are significant at $p < .001$.

in 7 diagnosis specific groups and found that the reliability and validity of the scale generally held up across diagnoses. Nonetheless, while the overall sample size was large, the vast majority of patients had mood or anxiety disorders. We examined some of the psychometric properties of the scale in patients with schizophrenia, however, the sample size was small. Moreover, we did not examine the performance of the measure in other diagnostically homogeneous groups such as substance use disorders, somatoform disorders, or obsessive compulsive and related disorders.

While there are several self-report questionnaires that assess anxiety, few have been designed to assess anxiety severity on a daily basis. There are several potential uses of a daily measure of anxiety severity. As mentioned in the introduction, daily scale administration can be used to monitor outcome in inpatient settings or ambulatory treatment settings in which patients are seen every day or several times per week such as a partial hospital or an intensive outpatient program. In short-term ambulatory intensive treatment programs such as a partial hospital, some patients terminate treatment without warning. Outcome assessment based on admission and discharge evaluations will have missing discharge data in the patients who drop out of treatment or who do not return for their scheduled last day of treatment. Daily assessment provides data for the last day of treatment for the drop outs and obviates the need to impute missing information.

Early response during the first couple of weeks of outpatient treatment has prognostic significance [16,17]. If symptoms are assessed daily from the onset of treatment, clinicians and researchers might be able to identify trajectories even earlier in the course of outpatient treatment that predict outcome. In outpatient settings it might be

Table 4
Test-retest reliability of the Clinically Useful Anxiety Outcome Scale-Daily Version (CUXOS-D) items at baseline and follow-up.

CUXOS-D Item	Baseline ($n = 28$)	Follow-up ($n = 22$)
Anxious mood	.80	.88
Worry about future	.85	.94
Excessive worry	.77	.91
Exaggerated startle response	.94	1.0
"Keyed up" feeling	.74	.96
Fearfulness	.80	.72
Muscle tension	.74	1.0
Jittery feeling	.80	.92
Shortness of breath	.77	.89
Pounding heart	.88	.82
Clammy palms	.95	.46**
Dry mouth	.79	.65
Dizziness	.85	.64
Nausea	.69	.91
Diarrhea	.88	.63
Hot flashes/chills	.91	.78
Frequent urination	.93	.90
Lump in throat feeling	.52*	.85
Sweating	.54	.93
Paresthesia	.80	.53*

All correlations are significant at $p < .001$ except as noted.
* Correlations are significant at $p < .01$.
** Correlation is significant at $p < .05$.

Table 5
Mean (SD) scores on the Clinically Useful Anxiety Outcome Scale-Daily Version (CUXOS-D) during the first 5 days of treatment and at endpoint for patients treated for at least 5 days ($n = 1336$).

				Effect size	Paired *t*-test	
Treatment day	*n*	Mean	SD	Compared to baseline	Compared to baseline[a]	Compared to prior day[b]
Day 1 (baseline)	1336	42.1	18.2	–	–	–
Day 2	1230	32.4	18.8	0.52	27.9**	20.2**
Day 3	1203	30.4	18.9	0.63	29.9**	7.7**
Day 4	1202	29.1	19.1	0.70	30.1**	4.0**
Day 5	1213	28.6	19.7	0.74	29.7**	2.1*
Endpoint[c]	1336	24.6	19.5	0.93	36.9**	6.7**

[a] Each day is compared to Day 1. The sample consists of patients who had a value at baseline and on that day of treatment. For example, the Day 2 analysis includes the 1230 patients who completed the CUXOS-D on Day 1 and Day 2.
[b] Each day is compared to the previous day. The sample consists of patients who had a value on both of the days of treatment.
[c] The endpoint CUXOS-D represents the last CUXOS-D collected. For 84.9% of the patients this was their planned discharge day.
** $p < .001$.
* $p < .05$.

Table 6
Mean (SD) scores on the Clinically Useful Anxiety Outcome Scale-Daily Version (CUXOS-D) during the first 10 days of treatment and at endpoint for patients treated for at least 10 days ($n = 529$).

				Effect size	Paired *t*-test	
Treatment day	*n*	Mean	SD	Compared to baseline	Compared to baseline[a]	Compared to prior day[b]
Day 1 (baseline)	529	44.9	17.3	–	–	–
Day 2	484	37.0	18.3	0.44	17.1**	11.8**
Day 3	482	36.0	18.4	0.50	16.7**	2.1*
Day 4	470	34.8	18.7	0.56	16.3**	2.0
Day 5	474	35.4	19.3	0.52	15.7**	−0.1
Day 6	481	36.0	19.6	0.48	14.6**	−1.4
Day 7	469	35.5	19.5	0.51	13.8**	1.0
Day 8	475	35.0	19.9	0.53	14.8**	0.9
Day 9	470	33.9	20.0	0.58	15.7**	1.9
Day 10	491	32.9	20.3	0.64	16.6**	2.0*
Endpoint[c]	529	29.8	20.3	0.80	20.8**	5.13**

[a] Each day is compared to Day 1. The sample consists of patients who had a value at baseline and on that day of treatment. For example, the Day 2 analysis includes the 484 patients who completed the CUXOS-D on Day 1 and Day 2.
[b] Each day is compared to the previous day. The sample consists of patients who had a value on both of the days of treatment.
[c] The endpoint CUXOS-D represents the last CUXOS-D collected. For 89.2% of the patients this was their planned discharge day.
** $p < .001$.
* $p < .05$.

possible to use such information to more quickly modify treatment that is predicted to fail. In future analyses of our partial hospital data base, we will examine whether response during the first couple of days of treatment predicts outcome. To be sure, before the information from this or any other measure is considered when making changes to treatment regimens, replicated research across diverse clinical and demographic variables should demonstrate the superiority of such an approach over usual clinical practice.

A limitation of the study is that we did not ask patients to complete the CUXOS-D every day, but only asked them to complete the scale on the days they attended the program. The patients missed approximately 20% of the treatment days because of medical illness, child care responsibilities, family obligations, other appointments, or feeling too overwhelmed to attend. This accounts for the varying sample sizes in Tables 5 and 6. Also, we did not collect information on the weekends or holidays when the program was not open.

The present study was conducted in a partial hospital setting, and it is unknown how the findings would generalize to an outpatient setting. While we believe that the psychometric properties of the CUXOS-D would hold up in an outpatient setting, this needs to be demonstrated. Of greater uncertainty is whether outpatients would complete the measure every day. Outpatients are seen much less frequently than partial hospital patients, and it would be more difficult to get outpatients to complete the scale at home on a daily basis. Aside from the likely difficulty of having outpatients comply with a request for daily completion of a paper and pencil questionnaire, it would be difficult for clinicians or support staff to quickly score and assimilate the information from a stack of questionnaires completed over an extended interval of 1 or 2 weeks, or even longer. The use of modern technology could facilitate such efforts, though it is unknown how compliant outpatients would be with a daily web-based administration of the scale. One approach to enhance compliance with daily scale completion is to explain to patients that daily administration will only be necessary during the acute phase of outpatient treatment, and the frequency of scale administration will drop when clinical status has stabilized.

We did not administer an interviewer-based anxiety scale on a daily basis, and thus we were unable to compare daily self-reported and clinician ratings. However, in the subsample of patients evaluated with the HAMA at admission and discharge from the program, the effect size based on the CUXOS-D was similar to the HAMA, and the change in scores on both measures was significantly correlated. Though daily ratings on scales such as the HAMA are unlikely to be conducted in clinical settings, a future study should compare daily self-report and clinician-rated evaluations to further validate daily self-administered assessments of anxiety.

While the results supported some aspects of the validity of the measure in several diagnostic subgroups, we examined a limited number of diagnoses. Moreover, the convergent and discriminant validity of the scale was not supported in some diagnostic subgroups perhaps because sample sizes were relatively small thereby limiting statistical power.

We did not examine the impact of time of day of scale completion on scores on the measure. Our partial hospital treatment program begins at 0800 h and while patients are expected to complete the scale upon arrival, some patients complete the measure later in the day. It is possible that patients who complete the measure in the late morning or in the early afternoon use a different time frame when rating the items. That is, the patients who complete the measure in the early morning consider how they felt the preceding day whereas the patients who complete the measure later in the day may consider how they were feeling that day as well as the preceding day. The high test-retest reliability of CUXOS-D completed twice several hours apart suggests that time of completion did not have an effect on scale completion; nonetheless, this issue warrants additional study.

The interplay between depression and anxiety should be further elucidated in future research. While we found that the CUXOS-D was more highly correlated with a measure of anxiety than with a measure of depression, it should not be overlooked that the correlation with the depression measure was significant in the whole sample as well as each of the 7 diagnostic subgroups. It will be important to examine to what degree changes over time in depression and anxiety are concordant or discordant, and if discrepancy exists, what factors are associated with such discrepancies.

Experience sampling methodology (ESM) is another method of assessing psychopathology in a more rapid and fine-grained manner than that assessed with weekly or biweekly questionnaires or clinician administered scales. There is some evidence that ESM approaches are more sensitive to treatment effects than traditional assessment approaches [18]. The advantage of ESM over a measure such as the CUXOS-D is the opportunity to better understand the interaction between changes in environmental context and clinical symptoms. Future research should examine the association between the 2 approaches and whether they are equally sensitive to measuring change and predicting outcome.

In conclusion, to our knowledge the CUXOS-D is the first self-administered daily measure of anxiety severity that assesses the psychological and somatic symptoms of anxiety. The results of the present study provide some support for the scale's reliability, validity, and feasibility for use in an acute care partial hospital setting. While the results of this large validation study are encouraging, they require replication in samples with different demographic and clinical characteristics.

References

[1] Trivedi M, Rush A, Wisniewski S, Nierenberg A, Warden D, Ritz L, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in star*d implications for clinical practice. Am J Psychiatry 2006;163:28–40.
[2] Trivedi MH, Daly EJ. Measurement-based care for refractory depression: a clinical decision support model for clinical research and practice. Drug Alcohol Depend 2007;88S:S61–71.
[3] Guo T, Xiang YT, Xiao L, Hu CQ, Chiu HF, Ungvari GS, et al. Measurement-based care versus standard care for major depression: a randomized controlled trial with blind raters. Am J Psychiatry 2015;172:1004–13.
[4] Yeung AS, Jing Y, Brenneman SK, Chang TE, Baer L, Hebden T, et al. Clinical outcomes in measurement-based treatment (COMET): a trial of depression monitoring and feedback to primary care physicians. Depress Anxiety 2012;29:865–73.
[5] Zimmerman M, Young D, Chelminski I, Dalrymple K, Galione JN. Overcoming the problem of diagnostic heterogeneity in applying measurement-based care in clinical practice: the concept of psychiatric vital signs. Compr Psychiatry 2012;53:117–24.
[6] Fawcett J. The detection and consequences of anxiety in clinical depression. J Clin Psychiatry 1997;58(Suppl. 8):35–40.

[7] Fava M, Rankin MA, Wright E, Alpert JE, Nierenberg AA, Pava J, et al. Anxiety disorders in major depression. Compr Psychiatry 2000;41:97–102.
[8] Zimmerman M, Martinez J, Attiullah N, Friedman M, Toba C, Boerescu D, et al. A new type of scale for determining remission from depression: the Remission From Depression Questionnaire. J Psychiatr Res 2013;47:78–82.
[9] Zimmerman M, McGlinchey J, Posternak M, Friedman M, Attiullah N, Boerescu D. How should remission from depression be defined? The depressed patient's perspective. Am J Psychiatry 2006;163:148–50.
[10] Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56–62.
[11] Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait multi-method matrix. Psychol Bull 1959;56:81–105.
[12] Steiger JH. Tests for comparing elements of a correlation matrix. Psychol Bull 1980;87:245–51.
[13] Zimmerman M, McGlinchey JB. Depressed patients' acceptability of the use of self-administered scales to measure outcome in clinical practice. Ann Psychiatry 2008;20:125–9.
[14] Zimmerman M, Chelminski I, Young D, Dalrymple KL. A clinically useful anxiety outcome scale. J Clin Psychiatry 2010;71:534–42.
[15] Kotov R, Krueger RF, Watson D, Achenbach TM, Althoff RR, Bagby RM, et al. The hierarchical taxonomy of psychopathology (HITOP): a dimensional alternative to traditional nosologies. J Abnorm Psychol 2017;126:454–77.
[16] Ciudad A, Alvarez E, Roca M, Baca E, Caballero L, Garcia de Polavieja P, et al. Early response and remission as predictors of a good outcome of a major depressive episode at 12-month follow-up: a prospective, longitudinal, observational study. J Clin Psychiatry 2012;73:185–91.
[17] Schlagert HS, Hiller W. The predictive value of early response in patients with depressive disorders. Psychother Res 2017;27:488–500.
[18] Moore RC, Depp CA, Wetherell JL, Lenze EJ. Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults. J Psychiatr Res 2016;75:116–23.