Published in final edited form as:

Bipolar Disord. 2010 August ; 12(5): 521–527. doi:10.1111/j.1399-5618.2010.00844.x.

The interactive computer interview for mania

Noreen A Reilly-Harrington^{a,b}, Daniel DeBonis^a, Andrew C Leon^c, Louisa Sylvia^{a,b}, Roy Perlis^{a,b}, Daniel Lewis^a, and Gary S Sachs^{a,b}

^aConcordant Rater Systems, Boston, MA

^bMassachusetts General Hospital and Harvard Medical School, Boston, MA

^cWeill Medical College of Cornell University, New York, NY, USA

Abstract

Objectives—The Interactive Computer Interview for Mania (ICI-M) is a computer-administered interview that presents probes to assess symptom severity and utilizes a scoring algorithm to select follow-up questions and rate subject responses in accordance with rating scale anchor points. The current study examines the acceptability, feasibility, and reliability of the ICI-M as a potential method for evaluating the performance of human raters.

Methods—Participants with a diagnosis of bipolar I or II disorder completed both a live interview of the Young Mania Rating Scale with a human rater (LR) and the ICI-M. A panel of three expert raters reviewed each videotaped LR and assigned a consensus rating (CR). Participants completed a modified version of the Client Satisfaction Questionnaire to assess each method.

Results—Intraclass correlation coefficients were 0.91 between the ICI-M and CR and 0.97 between the LR and CR ($n = 100$), providing empirical support for the inter-rater reliability of each approach. Coefficient alphas indicated comparable internal consistency reliability: ICI-M = 0.82, LR = 0.83, and CR = 0.84. The ICI-M was significantly more sensitive in detecting symptomatology than the LR ($p < 0.001$) and the CR ($p < 0.001$), and resulted in significantly higher ratings than CR on mood, speech, psychotic content, and disruptive-aggressive behavior. While participants endorsed significantly higher overall satisfaction with LR, no significant differences emerged between ICI-M and LR regarding willingness to participate again or ability to understand the questions.

Conclusions—The ICI-M is a well-accepted and reliable method for assessing manic symptoms. The ICI-M is a tool with adequate sensitivity to elicit symptoms and rate severity and

Corresponding author: Noreen A. Reilly-Harrington, Ph.D., Bipolar Clinic and Research Program, Massachusetts General Hospital, 50 Stanford Street, Suite 580, Boston, MA 02114, USA, Fax: (617) 726-6768, nhreilly@partners.org.

Disclosures NAR-H is a shareholder and consultant for Concordant Rater Systems and receives royalties from Oxford University Press and the American Psychological Association. DD is president and shareholder of Concordant Rater Systems. ACL serves on the Data and Safety Monitoring Boards for AstraZeneca, Daiippon Sumitomo Pharma America, and Pfizer; has served as a consultant/advisor to the FDA, NIMH, Cyberonics, MedAvante, Schering Plough, and Takeda; and has equity in MedAvante. LS is a consultant for Concordant Rater Systems. RP has received speaker's fees/honoraria from AstraZeneca, Eli Lilly & Co., Pfizer, GlaxoSmithKline, and Bristol-Myers Squibb; has served on the consulting/advisory boards for AstraZeneca, Eli Lilly & Co., Pfizer, Bristol-Myers Squibb, and Proteus LLC; and has received royalties from Concordant Rater Systems. DL is a shareholder in Concordant Rater Systems. GSS serves on the speakers bureaus of Abbott, AstraZeneca, Bristol-Myers Squibb, Eli Lilly & Co., GlaxoSmithKline, Janssen, Memory Pharmaceuticals, Novartis, Pfizer, Sanofi-aventis, and Wyeth; serves as an advisory board member or consultant for Abbott, AstraZeneca, Bristol-Myers Squibb, Cephalon, CNS Response, Concordant Rater Systems, Elan Pharmaceuticals, Eli Lilly & Co., GlaxoSmithKline, Janssen, Memory Pharmaceuticals, Merck, Novartis, Organon, Otsuka, Pfizer, Repligen, Sanofi-aventis, Schering Plough, Sepracor, Shire, Sigma-Tau, Solvay Pharmaceuticals, and Wyeth; and his spouse holds shares in Concordant Rater Systems.

is recommended as a tool to monitor and improve rater performance, not as a replacement to a human rater.

Keywords

bipolar disorder; computerized interviews; rater reliability

The failure of clinical trials to detect significant differences between groups is a well-recognized and increasingly costly impediment to clinical drug development (1,2). This difficulty is particularly acute in clinical trials of psychiatric drugs, where placebo response rates of above 30% are not uncommon (3-5), making discrimination of active drug effect especially challenging. Some studies of major depression have reported placebo response rates as high as 70% (6). Moreover, placebo-controlled trials are increasingly difficult to justify on ethical grounds when an effective treatment is known (7).

To enhance the power of a given clinical trial, investigators can simply include greater numbers of patients. However, this approach has several major drawbacks. First, it adds substantially to the cost of performing clinical trials, as the per-subject costs associated with additional subjects often represent the majority of the total costs of a trial. More importantly, it requires that larger numbers of patients be exposed to experimental drugs, or drugs that may not yet have shown clear benefit for their illness. In order to assure assay sensitivity, or in some countries based on regulatory requirements, more trials are incorporating active comparators. Unfortunately, increasing the number of arms in a trial not only increases the sample size but is also associated with an increased placebo response rate (8-10). Mounting ethical objections to placebo-controlled studies further slows recruitment and thereby inhibits development of new therapies.

An alternative mechanism to enhance the power of a clinical trial involves improving the reliability of outcome measurements (11,12). In psychiatric trials, where objective biological outcome measures may be lacking, reliability may be particularly poor, as investigators typically rely on rating scales completed by clinicians. This reliance on clinician-administered rating scales likely contributes to the high failure rate of large, multicenter, randomized clinical trials in psychiatry. Gains in rater reliability can result in substantial reduction in the sample size requirement, time, cost, and risk of failure that can thwart development of promising therapeutic agents. For example, Perkins et al. (13) calculated that an improvement in reliability from $R = 0.7$ to $R = 0.9$ could reduce sample size requirements by 22%, with an associated cost savings.

Results reported by Calabrese et al. (14) are consistent with the expectation of rater drift over the course of a multicenter trial. In this multicenter, double-blind trial, a preliminary analysis found a significant antimanic advantage over placebo for both 256 mg and 512 mg doses of topiramate with only 36 patients. However, one year later, when the study was completed, results for the full sample ($n = 96$) showed no significant difference.

Interview quality has also been shown to affect signal detection in clinical trials (15). In a study of active drug (paroxetine) versus placebo, Kobak and colleagues (15) found that only subjects whose ratings which were classified as 'good' or 'excellent' achieved drug-placebo separation.

These risks to rater reliability suggest the need for improved procedures for the training, certification, recertification, and monitoring of raters. Such procedures can decrease rater variance, thereby increasing statistical power and enhancing the likelihood of detecting a true difference between active and inactive treatments. However, such quality improvements

are difficult to achieve. Numerous obstacles and conflicting ethical, economic, and scientific considerations require solutions to be both sophisticated and pragmatic.

Concordant Rater Systems (CRS) developed and patented an automated system that can administer and score rating scales in a manner similar to that done by an experienced human rater. Specifically, the Interactive Computer Interview (ICI) is a computer-administered interview that both presents probes designed to elicit information about the presence and severity of symptoms and utilizes a scoring algorithm to select follow-up questions and rate subject responses in accordance with rating scale anchor points. CRS has developed versions of the ICI for psychiatric rating scales commonly used to assess depression, mania, and anxiety. The goal of the ICI is not to replace the human rater, but to provide a standard comparator that can be used to enhance the sensitivity and consistency of human raters by providing ongoing feedback on the concordance of their ratings with the ICI ratings and identify those in need of specific remediation during the course of study operations. The ICI is based on the premise that computer administration and scoring not only provides a consistent measure against which the human rater can be measured, but also that its use in parallel with human ratings will tend to enhance the consistency of human raters. While the current study does not test the impact of the ICI on human raters, its hypothesized effect on human raters is explained here in order to provide the context and rationale for its eventual use.

The current study focuses on establishing the psychometric properties of the ICI for Mania (ICI-M) which is a self-rated, computerized interview that simulates a clinician's administration of the Young Mania Rating Scale (YMRS) (16). The specific objective of the current study was to determine the feasibility and reliability of the ICI-M in a sample of 100 patients with bipolar mood disorder. We hypothesize that the ICI-M and a live rating (LR) of the YMRS will have comparable inter-rater reliability with an expert consensus rating (CR). The primary analyses will focus on the inter-rater reliability of the ICI and LRs by comparing the agreement of each method with an expert CR. In addition, the internal consistency reliability of the ICI and LRs will be compared. It is hypothesized that the internal consistency reliability is not sacrificed with the ICI methodology.

The current study also examined the feasibility and tolerability of administration of the ICI-M across 100 patients with a range of severity from euthymic, i.e., non-symptomatic, to severely manic. We hypothesize that the ICI would be time-efficient for patient completion (< 35 minutes) and should be rated as *acceptable or better* by most patients on a modified version of the Client Satisfaction Questionnaire (CSQ) (17).

Materials and methods

Subjects

Participants were recruited from sites in the greater Boston area that were engaged in clinical care and research through the use of posters, electronic announcements, and newspaper advertisements offering compensation for participation. The principal investigator also recruited subjects from local chapters of the Depression and Bipolar Support Alliance by speaking at meetings and describing the study. Subjects were offered \$100.00 for their participation.

Procedure

Subjects first verbally consented to a phone screen in which they were questioned about their lifetime history of moods, symptoms, and behaviors. As part of this consent, subjects were informed that if current severe symptoms were reported, they would be referred for psychiatric care. A modified version of the Structured Clinical Interview for DSM-IV Axis I

disorders (SCID) lifetime mood disorder modules was then used to screen patients by phone. Eligibility requirements for the study included a DSM-IV diagnosis of bipolar I or II disorder, a history of treatment for a bipolar mood disorder, a minimum age of 18 years, and the capacity to read and understand spoken English and provide written informed consent. The study purpose and design were described, and those willing to consent to the two study assessments (videotaped live interview and ICI) were scheduled for a single 1.5-2 hour study visit.

Upon arrival at the study site, the study purpose and procedures were again explained to the potential subject. After obtaining written consent, study staff administered the SCID Mood Disorder modules to confirm lifetime diagnosis. Participants also completed a brief questionnaire inquiring about previous and current treatment. After eligibility was confirmed, participants were issued a study ID number and were assessed both by the ICI-M and by a live experienced rater using a scripted version of the YMRS. The live interview was videotaped. The order of administration of the ICI-M and live interview were counterbalanced to control against order effects using random assignment, with a 1:1 ratio of the two orders.

Assessments

Client Satisfaction Questionnaire—A modified version of the CSQ was developed and administered for this study to assess patients' satisfaction and acceptability of both the clinician-administered YMRS (LR) and the ICI-M. Eight self-report items asked about the overall quality of the interview, ability to understand questions, the ability of the interview to help a clinician understand patients' symptoms, emotional difficulty, comfort level, willingness to participate in an ongoing study with this type of assessment, overall satisfaction, and willingness to participate again.

Young Mania Rating Scale—The YMRS is an 11-item, clinician-administered scale that provides a continuous measure of the symptoms associated with mania. Items assess mood elevation, increased motor activity/energy, sexual interest, sleep, irritability, speech, language-thought disorder, content (hyper-religiosity, grandiosity, delusions, etc.), disruptive-aggressive behavior, appearance, and insight. The traditional anchor points of 0-4 (on items 1, 2, 3, 4, 7, 10, and 11) and 0-8 (on items 5, 6, 8, and 9) were utilized.

Interactive Computer Interview for Mania—The ICI-M is a self-rated, computerized interview that simulates a clinician's administration of the YMRS by using a computer screen to deliver scripted probes similar to those used by human raters to elicit information and score items. The interview proceeds based on a subject's response to each probe until sufficient information is available for the scoring algorithm to map responses into a unique anchor point. The ICI-M concludes when scores are obtained for all YMRS items.

The ICI-M software presents a series of 'yes' or 'no' and multiple-choice questions, with an option for the questions to be read aloud. Questions appear on the screen one at a time. Depending on the hardware capabilities of the computer, subjects will either use a touch screen (similar to an automated teller machine screen) or a mouse click to select an answer. After a question is answered, the next question in the sequence appears. The ICI-M selects questions depending on answers to previous questions, similar to probing or follow-up questions. When the database determines that enough data have been collected for a specific item, the ICI-M will record a score for the subject and move to the next item. The CRS implementation methods and scoring algorithms have received U.S. Patent 7,315,725.

A research assistant not involved in the LR interview oriented subjects to the computer interface for the ICI-M. Once the patient was oriented to the computer and the procedure,

the research assistant left the room until the ICI-M was completed. The research assistant was available nearby to answer any technical questions. The results for each item were entered directly into a password-protected study database and transmitted via a secure-socket Internet connection to a central database. Immediately following completion of the ICI-M, the modified version of the CSQ was administered to elicit patients' comments and ratings of satisfaction and acceptability regarding the ICI-M. The time to complete the ICI-M was also measured.

An experienced live rater (with at least one year of experience training and certifying other raters) administered and rated the live YMRS (LR), and the interviews were videotaped. The rater and technician controlling the camera were blind to the ICI-M results. Directly following the live interview, the modified version of the CSQ was also used to elicit patients' comments and ratings of satisfaction and acceptability regarding the live, videotaped interview. Immediately after the LR, its results were double-entered into the study database and only accepted when scores for all items were verified as accurate.

The videotaped interviews were then reviewed by a consensus panel consisting of two of the investigators (NAR-H and GSS) and at least one other experienced rater in order to establish *gold-standard* CRs for each item on the rating scale. These gold-standard CRs were then compared to the scores generated by the ICI-M scoring algorithm and the LRs.

Analyses

The primary analyses focused on the inter-rater reliability of the ICI-M and the LR by comparing each method with the expert CR. The intraclass correlation coefficient (ICC) of ICI-M ratings with CRs was compared with the ICC of LRs with CRs. The ICCs were also compared based on the counterbalanced order of administration, and homogeneity of variance was examined using Levene's test. In addition, the internal consistency reliability among scale items of the ICI-M and the LRs was compared using Cronbach's coefficient alpha. The feasibility and tolerability of each method were compared by examining levels of acceptability on the modified CSQ.

Results

The sample consisted of 100 participants meeting DSM-IV criteria for bipolar I or II disorder (see Table 1):

- 60 women, 40 men
- Age range 19-71 (mean age 37.6)
- 17% minority (10 African American, 4 Asian American, 2 Latino, 1 Native American)
- 94% currently receiving treatment for bipolar disorder

Overall mean (SD) total scores were: ICI-M = 24.7 (12.4), LR = 21.3 (12.1), CR = 20.9 (12.1). The ICI-M score was significantly higher than the LR ($t = 6.52$, $df = 1,99$, $p < 0.001$, $d = 0.28$) and CR scores ($t = 3.741$, $df = 1,99$, $p < 0.001$, $d = 0.31$). ICI-M items had significantly higher ratings than CR on mood, speech, psychotic content, and disruptive-aggressive behavior. The mean scores, SDs, and corrected item total correlations for each method are included in Table 2. The ICI-M mean corresponds to the 61st and 62nd percentile of the LR and CR distributions, respectively.

The inter-rater reliability of the ICI-M and the LR was examined by comparing each method with the expert CR. Using a random effects model, the ICC of ICI-M with CR was

compared with the ICC of LR with CR, using a Z-test of the Fisher-transformed ICCs, as described by Donner (18). As continuous rather than categorical variables were being assessed, the ICC was utilized in lieu of kappa as the test statistic. ICCs were 0.91 between the ICI-M and CR and 0.97 between the LR and CR, providing empirical support for the inter-rater reliability of the ICI-M.

The above analyses were also repeated to examine the potential order effects of administration, as the ICI-M and LR were administered in a 1:1 counterbalanced order. No significant differences were found. The ICCs between the ICI-M and CR were 0.88 when the ICI-M was administered first and 0.93 when the live interview was administered first. The ICCs between the LR and the CR were 0.97 when the ICI-M was administered first and 0.98 when the live interview was administered first. Homogeneity of variance was also examined using Levene's test. There were no significant differences in variances of the LR due to the order of administration (LR administered first: SD = 12.36; LR administered second: SD = 11.54; $t = 0.87$, ns).

In addition, the internal consistency reliability of the ICI-M and LR was compared. The internal consistency reliability among scale items of the ICI-M and the LR was examined using Cronbach's coefficient alpha. As hypothesized, internal consistency reliability was not sacrificed with the ICI methodology. Coefficient alphas indicated comparable internal consistency reliability across all three methods: ICI-M = 0.82, LR = 0.83, and CR = 0.84.

Finally, within-subject analyses of acceptability of the ICI methodology were conducted by comparing the CSQ subject satisfaction ratings of the ICI-M and LR using Wilcoxon signed rank tests. Participants endorsed significantly higher overall satisfaction ($Z = -3.054$, $p < 0.002$), significantly higher quality ($Z = -5.734$, $p < 0.001$), and a significantly better ability for a clinician to understand their symptoms ($Z = -4.010$, $p < 0.001$) with the LR as compared to the ICI-M. However, participants reported that they were significantly more comfortable ($Z = 3.808$, $p < 0.001$) and had significantly less difficulty emotionally ($t = 4.328$, $p < 0.001$) participating in the ICI-M as compared to the LR. No significant differences emerged between ICI-M and LR on items assessing willingness to participate again or ability to understand the questions. The mean scores and SDs by item for each method are included in Table 3. The average length of time to complete the ICI-M was 13.7 minutes.

Discussion

The results of the current study suggest that the ICI-M is a well-accepted and reliable method for assessing manic symptoms. This study also illustrated the feasibility of the ICI-M in a population of bipolar disorder outpatients ranging from euthymic to severely manic. All of the participants were able to complete the ICI-M, and the levels of acceptability were high. These findings are consistent with previous research suggesting that psychiatric patients exhibit positive attitudes toward computerized assessments (19-21).

One issue in using a self-report computerized measure to assess manic symptoms in bipolar disorder patients involves the concern that patients will lack the insight to recognize and report their symptoms without the careful probing of a live interviewer. However, the algorithmic probe questions characteristic of the ICI-M resulted in a significantly more sensitive ratings in detecting symptomatology than both the LR and the CR. We must also acknowledge the possibility that the ICI-M may simply result in higher scores and we plan to further investigate the discriminant validity of the ICI-M by utilizing samples of normal controls in addition to participants with bipolar disorder.

Future studies are also planned to pilot the ICI-M in a population of inpatients with bipolar disorder. It will be of interest to explore the impact of symptom severity on reliability as well as to explore the effects of depressive symptomatology and comorbidity on ratings. It will also be of interest to compare the ICI-M with clinical ratings of mania, as clinical evaluations have been shown to differ often from standardized interviews such as the ICI-M (22). The ICI-M does have the capacity to gather high-quality information in a short timeframe (mean = 13.7 minutes) and may have potential applications in clinical as well as research settings.

This demonstration of the acceptability and psychometric properties of the ICI-M is a crucial step in the eventual implementation of the ICI-M as a tool to monitor and improve human rater performance. In the field of psychiatry, we routinely rely on the quantification of subjective feelings and behaviors to demonstrate the efficacy of interventions. Even minor improvements in the reliability of such ratings can dramatically reduce sample size requirements, time, cost, and risks of failed trials. The ICI method has the potential to improve the reliability of LR's through providing corrective feedback and remediation to raters when scores significantly deviate from the ICI. CRS is currently engaged in studies implementing the ICI-M and it is hypothesized that this system will enhance the detection of drug versus placebo differences for promising therapeutic agents.

Acknowledgments

This study was funded by NIMH Grant 1 R43 MH65796-01A1 to NAR-H, Principal Investigator.

References

1. Kobak KA, Kane JM, Thase ME, Nierenberg AA. Why do clinical trials fail? The problem of measurement error in clinical trials: time to test new paradigms? *J Clin Psychopharmacol.* 2007; 27:1–5. [PubMed: 17224705]
2. Robinson DS, Rickels K. Concerns about clinical drug trials. *J Clin Psychopharmacol.* 2000; 20:593–596. [PubMed: 11106129]
3. Thase M. How should efficacy be evaluated in randomized clinical trials of treatments for depression? *J Clin Psychiatry.* 1999; 60:23–31. [PubMed: 10086480]
4. Trivedi MH, Rush J. Does a placebo run-in or a placebo treatment cell affect the efficacy of antidepressant medications? *Neuropsychopharmacology.* 1994; 11:33–43. [PubMed: 7945742]
5. Quitkin FM, Rabkin JG, Gerald J, Davis J, Klein DF. Validity of clinical trials of antidepressants. *Am J Psychiatry.* 2000; 157:327–337. [PubMed: 10698806]
6. Brown WA, Dornseif IS, Wernicke JF. Placebo response in depression: a search for predictors. *Psychiatry Res.* 1988; 26:259–264. [PubMed: 3065807]
7. Quitkin FM. Placebos, drug effects, and study design: A clinician's guide. *Am J Psychiatry.* 1999; 156:829–836. [PubMed: 10360119]
8. Keck PE, Welge JA, McElroy SL, Arnold LM, Strakowski SM. Placebo effect in randomized, controlled studies of acute bipolar mania and depression. *Biol Psychiatry.* 2000; 47:748–755. [PubMed: 10773184]
9. Keck PE, Welge JA, Strakowski SM, Arnold LM, McElroy SL. Placebo effect in randomized, controlled maintenance studies of patients with bipolar disorder. *Biol Psychiatry.* 2000; 47:756–761. [PubMed: 10773185]
10. Schatzberg AF, Kraemer HC. Use of placebo control groups in evaluating efficacy of treatment of unipolar major depression. *Biol Psychiatry.* 2000; 47:736–744. [PubMed: 10773182]
11. Leon AC. Implications of clinical trial design on sample size requirements. *Schizophr Bull.* 2008; 34:664–669. [PubMed: 18469326]
12. Leon AC, Marzuk PM, Portera L. More reliable outcome measures can reduce sample size requirements. *Arch Gen Psychiatry.* 1995; 52:867–871. [PubMed: 7575107]

13. Perkins D, Wyatt RJ, Bartko JJ. Penny-wise and pound-foolish: the impact of measurement error on sample size requirements in clinical trials. *Biol Psychiatry*. 2000; 47:762–766. [PubMed: 10773186]
14. Calabrese J, Shelton M, Rapport D, Kimmel S. Current state of the art in pharmacologic treatment of bipolar disorder. *Int J Neuropsychopharmacol*. 2000; 3:S50.
15. Kobak KA, Feiger AD, Lipsitz JD. Interview quality and signal detection in clinical trials. *Am J Psychiatry*. 2005; 162:628. [PubMed: 15741493]
16. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry*. 1978; 133:429–435. [PubMed: 728692]
17. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfactions: development of general scale. *Eval Program Plann*. 1979; 2:197–207. [PubMed: 10245370]
18. Donner A. Sample size requirements for the comparison of two or more coefficients of inter-observer agreement. *Stat Med*. 1998; 17:1157–1168. [PubMed: 9618775]
19. Carr AC, Ghosh A. Response of phobic patients to direct computer assessment. *Br J Psychiatry*. 1983; 142:60–65. [PubMed: 6831131]
20. Erdman HP, Klein MH, Geist JH. Direct patient computer interviewing. *J Consult Clin Psychol*. 1985; 53:760–773. [PubMed: 3841354]
21. Lucas RW, Mullin PJ, Luna CBX, McInroy DC. Psychiatrists and a computer as interrogators of patients with alcohol-related illness: a comparison. *Brit J Psychiatry*. 1977; 131:160–167. [PubMed: 334310]
22. Rettew DC, Lynch AD, Achenbach TM, Dumenci L, Ivanova MY. Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *Int J Methods Psychiatr Res*. 2009; 18:169–184. [PubMed: 19701924]

Table 1

Demographic characteristics and treatment status of sample

Variable	n
Gender	
Female	60
Male	40
Age	
18-29	10
30-39	30
40-49	26
Over 50	34
Ethnicity	
Caucasian	83
African American	10
Asian American	4
Latino	2
Native American	1
Treatment status	
In therapy or on medication	94
Currently not medicated	9
Taking only one medication	22
Most commonly reported medications	
Depakote	24
Lithium	23
Wellbutrin	13
Zoloft	13

Table 2

Means, standard deviations, and corrected item total correlations (CITC) for each method (n = 100)

YMRS item	ICIM		Live rating		Consensus rating	
	Mean (SD)	CITC	Mean (SD)	CITC	Mean (SD)	CITC
1	3.1 (1.3)	0.60	2.3 (1.4)	0.53	2.3 (1.3)	0.53
2	1.7 (1.5)	0.73	1.7 (1.5)	0.67	1.8 (1.5)	0.64
3	0.9 (1.2)	0.38	0.7 (1.1)	0.33	0.8 (1.1)	0.34
4	1.1 (0.9)	0.38	1.5 (1.2)	0.60	1.4 (1.1)	0.56
5	2.5 (1.8)	0.56	2.9 (1.7)	0.65	2.9 (1.7)	0.63
6	3.7 (3.1)	0.57	3.1 (3.0)	0.57	2.7 (2.8)	0.65
7	2.1 (1.6)	0.58	2.0 (1.3)	0.73	1.9 (1.2)	0.76
8	4.9 (3.2)	0.51	3.6 (3.0)	0.53	3.6 (3.0)	0.53
9	3.5 (2.3)	0.49	2.5 (2.3)	0.36	2.4 (2.3)	0.51
10	1.1 (1.4)	0.44	1.0 (1.0)	0.57	0.9 (0.9)	0.65
11	0.1 (0.5)	0.22	0.1 (0.4)	0.24	0.2 (0.7)	0.16
Total	24.7 (12.4)		21.3 (12.1)		20.9 (12.1)	

YMRS = Young Mania Rating Scale; ICI-M = Interactive Computer Interview for Mania.

Table 3

Means and standard deviations of Client Satisfaction Questionnaire (CSQ) items for each method (n = 100)

CSQ Items	ICI-M Mean (SD)	Live rating Mean (SD)
Overall quality	1.88 (0.74)	1.38 (0.55)
Ability to understand questions	1.32 (0.51)	1.27 (0.47)
Ability to help clinician understand symptoms	2.22 (0.95)	1.77 (0.84)
Emotional difficulty	1.07 (0.26)	1.33 (0.53)
Comfort level	1.28 (0.47)	1.56 (0.64)
Willingness to be in an ongoing study with this assessment	1.46 (0.73)	1.44 (0.64)
Overall satisfaction	1.78 (0.73)	1.52 (0.59)
Willingness to participate again	1.25 (0.50)	1.24 (0.43)
Total	12.26 (3.31)	11.51 (3.04)

Lower scores indicate higher satisfaction. ICI-M = Interactive Computer Interview for Mania.