Web-Based Assessment of Depression in Patients Treated in Clinical Practice: Reliability, Validity, and Patient Acceptance

Mark Zimmerman, MD, and Jennifer H. Martinez, BA

ABSTRACT

Objective: Calls for the use of standardized assessments in clinical practice have been increasing. A Web-based administration of outcome assessments offers several potential advantages over paper-and-pencil assessments, such as patient convenience, reduced missing data, reduced costs, automatic scoring, and generation of large databases. The present study from the Rhode Island Methods to Improve Diagnostic Assessments and Services (MIDAS) project evaluated the acceptability, reliability, and validity of a Web-based administration of a depression scale in patients receiving ongoing care for depression.

Method: From June 2009 to July 2010, fifty-three depressed outpatients completed a Web-based and a paper version of the Clinically Useful Depression Outcome Scale (CUDOS). The vast majority of patients met *DSM-IV* criteria for either major depressive disorder (n = 36) or bipolar disorder (n = 9). Patients were also asked to complete a brief 6-question survey of the acceptability of the 2 modes of scale administration. At the time of the visit, the patients' psychiatrist completed the Montgomery-Asberg Depression Rating Scale (MADRS) and rated patients on the Clinical Global Impressions-Severity of Illness (CGI-S) scale and Global Assessment of Functioning (GAF).

Results: The correlation between the Web-administered and paper versions of the CUDOS was high ($P < .001$). The mean scores were similar on the paper and Internet administrations. The internal consistency of the paper and Internet administrations of the CUDOS was high (both values, Cronbach $\alpha = .93$), and all item-scale correlations for the paper and Internet versions were significant (median for paper administration = 0.76; median for Internet administration = 0.74). The paper and Internet versions of the CUDOS were equally correlated with clinicians' ratings on the MADRS, CGI-S, and GAF (all P values < .001). Patients reported high levels of satisfaction with Internet administration and preferred this method of monitoring outcome to paper administration in the office (all P values < .001).

Conclusions: The results of this first study of the use of a Web-based system of monitoring outcome in routine clinical practice supported the reliability and validity of Internet administration of a depression scale, and patients clearly preferred Internet administration to completion of a paper-and-pencil questionnaire in the office.

J Clin Psychiatry 2012;73(3):333–338

Submitted: *August 23, 2010; accepted February 18, 2011.*
Online ahead of print: *November 1, 2011 (doi:10.4088/JCP.10m06519).*
Corresponding author: *Mark Zimmerman, MD, Bayside Medical Center, 235 Plain St, Providence, RI 02905 (mzimmerman@lifespan.org).*

To determine the impact of treatment, one needs to evaluate outcome. Calls for the use of standardized assessments in clinical practice have been increasing. One of the principal lessons of the National Institute of Mental Health (NIMH)–funded Sequenced Treatment Alternatives to Relieve Depression (STAR*D) effectiveness trial was the potential benefit of practicing measurement-based care.[1] The Centers for Medicare and Medicaid Services Physician Quality Reporting Initiative provides financial incentive to clinicians who monitor outcome.[2] Financial incentives are likely to increase clinicians' motivation to systematically evaluate outcome depending on how much reimbursement is provided and how burdensome it is to collect the information.

Recent studies of psychiatrists in the United Kingdom and the United States found that only a minority of clinicians regularly used standardized scales to measure outcome in the treatment of depression.[3,4] It will most likely be easier to get clinicians to change their behavior and utilize quantitative assessments of outcome if these assessments are not intrusive, burdensome, or costly. If clinicians find that such assessments have practical value, such as improving the efficiency of the clinical encounter, this is also likely to increase the likelihood of clinical utilization.

More than 10 years ago, we established the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project—a clinical research program in which we have integrated the assessment methods characteristic of research protocols into a routine psychiatric outpatient clinical practice.[5] One of the goals of the MIDAS project has been to develop instruments for use in routine clinical practice. One such measure is the Clinically Useful Depression Outcome Scale (CUDOS),[6–8] a brief (completed in less than 3 minutes), quickly scored (in less than 15 seconds), clinically useful (fully covering the *DSM-IV* symptoms of major depressive disorder as well as global psychosocial functioning and quality of life) measure that we found to be reliable, valid, and feasible to incorporate into clinical practice.

Coincident to the increasing emphasis on measuring outcome in clinical practice, there has been a rapid expansion during the past decade in the provision of mental health information and services over the Internet.[9] Web sites provide information about nearly every disorder in the *DSM*. Tests of intelligence, personality, and various major psychiatric disorders are increasingly available. As the role of the Internet in daily life continues to expand, it is likely clinicians will be increasingly interested in conducting Internet-based assessments. In discussing the delivery of therapeutic services over the Internet, Cartreine et al[10] recommended that it be paired with routine symptom monitoring and outcome assessment.

Web-based administration of outcome assessments offers several potential advantages over paper-and-pencil assessments, such as patient convenience, reduced missing data, reduced costs, automatic scoring, and aggregation of data across patients. Web-based scales can be completed by patients at their convenience in their home rather

Clinical Points

- Standardized assessments of outcome should be used in clinical practice.
- A Web-based administration of outcome measurement offers several potential advantages over paper-and-pencil assessments, such as patient convenience, reduced missing data, reduced costs, automatic scoring, and generation of large databases.
- The result of this first study of a Web-based system of monitoring outcome in routine clinical practice supported the reliability, validity, and patient acceptability of the system.

than arriving early or staying after their clinical appointment to complete the measure. A computer-administered survey can prompt respondents to ensure all questions are answered, thereby reducing missing data. The administrative costs associated with the copying, handing out, and scoring of paper questionnaires are reduced with a Web-based system. Similarly, the high costs of establishing and maintaining a database to evaluate treatment outcome for a large sample of patients based on the administration, scoring, and data entry of paper questionnaires could be markedly reduced with a Web-based system. Because the data collected via the Internet are automatically entered into a database, data entry errors are reduced.[11]

It cannot be assumed that paper-and-pencil and computer-based administrations of the same test will produce equivalent results.[12–14] Despite identical content, differences in format can influence results.[15,16] For example, some computer-administered tests present items one at a time on a screen, whereas paper tests usually present multiple questions on a page. Increased attention to items presented singly could result in differential item-scale correlations. Some computer-administered tests do not allow a return to previous items, whereas paper tests allow the respondent to change prior responses. This could produce respondent frustration with computer administration thereby influencing responses to remaining items on the scale. Also, respondents might reconsider prior answers after prompting from answering additional questions. Some computer-administered scales do not allow missing answers. The "candor hypothesis" suggests that respondents are more truthful when responding to computer-administered tests. Each of these factors could influence the psychometric properties of a test, as well as the cutoff values used to determine caseness (or remission status), thereby warranting the demonstration of equivalence between computerized and paper versions of a scale.[12,13]

Internet administration of scales adds another potential source of error variance because patients might complete questionnaires while watching television, talking on the telephone, eating, or participating in another distracting activity.[17] The nonstandardized conditions under which scales might be completed can increase error variance, thereby requiring demonstration of equivalence to in-office administration.[11]

With the growing interest in developing electronic medical records, together with the increased pressure to adopt measurement-based care in clinical practice, the time is right to develop a Web-based system to monitor depression in clinical practice. We recently developed the CUDOS Web version (CUDOS-W) to monitor the course of depression for patients in ongoing treatment. Most prior studies of computer and Web-based depression assessment have been of college students and research subjects in clinical trials.[18–22] We are not aware of any such studies of patients receiving ongoing treatment for depression in usual clinical practice. The goal of the present study from the MIDAS project was to evaluate the acceptability, reliability, and validity of the CUDOS-W. Specifically, the aims of the present study were to (1) determine whether the psychometric properties of the CUDOS-W were equivalent to the paper-and-pencil version of the scale; (2) determine whether the validity of the CUDOS-W was comparable to the paper-and-pencil version by examining the correlation of each version with an independent clinician rating measure of depression severity; and (3) compare patients' ratings of acceptability of Internet and paper-and-pencil versions of the scale.

METHOD

The study was conducted from June 2009 to June 2010. The participants were 53 psychiatric outpatients who were in ongoing treatment for depression. All patients had been interviewed with the Structured Clinical Interview for *DSM-IV*.[23] Measurement-based care was the standard in the practice, and all patients had prior experience with completing the paper version of the CUDOS at the time of their visits. Patients were told that a Web-based system had been developed that enabled them to complete the measure at home. The patients were given the Web site address, and they were required to register on the site and indicate the date and time of their next appointment. Registration was done at home, not in the office. The outcome tracking system (www.outcometracker.org) includes an appointment reminder function; thus, 48 hours before their appointment, the patients automatically received an e-mail reminding them of their forthcoming appointment and directing them to complete the online version of the CUDOS. The completed form is e-mailed to the treating clinician.

At the end of their appointment with their psychiatrist, the patients were asked if they would complete the paper version of the CUDOS. It was explained to them that the circumstances and setting of scale completion sometimes influences responses to a scale; therefore, it was important to examine the comparability of computer and paper administrations of a scale. Patients were also asked to complete a brief 6-question survey of the acceptability of the 2 modes of scale administration; the survey asks which of the 2 approaches took less time to complete, was easier to understand, less burdensome to complete, and more acceptable

to complete at every follow-up appointment. Two additional, open-ended questions asked the patient to describe the advantages and disadvantages of each method. During the index visit, the clinician completed the Montgomery-Asberg Depression Rating Scale (MADRS)[24] and rated the patients on the Global Assessment of Functioning (GAF) and Clinical Global Impressions-Severity of Illness (CGI-S) scale.[25] The raters were blind to the scores on both versions of the CUDOS. The Rhode Island Hospital institutional review committee approved the research protocol, and all patients provided informed, written consent.

The paper and Internet versions of the CUDOS are identical. The CUDOS-W presents the entire scale at once, rather than 1 item at a time. Patients are able to change their responses after answering a question. The questionnaire cannot be submitted unless all items are completed. Both versions of the CUDOS contain 18 items assessing all of the *DSM-IV* inclusion criteria for MDD as well as psychosocial impairment and quality of life. According to the Flesch-Kincaid scoring system, the symptom items of the CUDOS are written at the 6th-grade level.

On the CUDOS the respondent is instructed to rate the symptom items on a 5-point Likert scale indicating "how well the item describes you during the past week, including today" (0 = not at all true/0 days, 1 = rarely true/1–2 days, 2 = sometimes true/3–4 days, 3 = usually true/5–6 days, 4 = almost always true/every day). A Likert rating of the symptom statements was preferred in order to keep the scale brief. A study[7] comparing the acceptability of the CUDOS and the Beck Depression Inventory to monitor treatment outcome found that the CUDOS was perceived as less burdensome and was the preferred measure to be completed on a regular basis to monitor outcome.

The initial studies of the reliability and validity of the CUDOS indicated that the scale has strong psychometric properties.[6] Briefly, the scale's internal consistency reliability coefficient was 0.90, and test-retest reliability was 0.92. The CUDOS was more highly correlated with other measures of depression than with measures of the other symptom domains, thereby supporting the scale's convergent and discriminant validity. Moreover, the CUDOS was nearly as highly correlated with clinician ratings of the severity of depressive symptoms as with another self-report measure of depression severity. In a separate study[8] of depressed outpatients, the CUDOS was found to be a valid measure of remission.

Data Analysis

We computed Cronbach α to determine the internal consistency of both versions of the scale. Intraclass correlation coefficients were used to determine the association between the total scores and item scores of the 2 versions, and Cohen κ was used to determine agreement in remission status. A paired *t* test was used to determine if the mean scores of the paper and web versions of the CUDOS were significantly different. To compare the differences between the Internet and paper versions of the CUDOS in terms of perceived burden, understandability, and preference for future completion, we conducted separate tests of approximate inference for a single proportion on those patients who made a clear selection. When using this test, a *z* score is computed from the observed proportions of favorability for the 2 versions of the scale and compared against a null hypothesis of equivalent selection.

Table 1. Correlations Between the Paper and Internet Versions of the Clinically Useful Depression Outcome Scale (CUDOS) and Clinician Ratings (N = 53)[a]

Clinician Rating Scale	Paper	Internet	Difference Between Correlations, *P* Value[b]
Montgomery-Asberg Depression Rating Scale	0.94	0.91	< .05
Clinical Global Impressions-Severity of Illness scale	0.92	0.92	NS
Global Assessment of Functioning	−0.91	−0.91	NS

[a]All correlations are significant at $P < .001$.
[b]This column indicates whether the difference between the correlations of the paper and Internet versions of the CUDOS with each of the clinician rating scales was significant.
Abbreviation: NS = nonsignificant.

RESULTS

The sample included 13 men (24.5%) and 40 women (75.5%) who ranged in age from 19 to 85 years (mean = 45.1, SD = 12.3). The vast majority of patients were diagnosed with either major depressive disorder (n = 36) or bipolar disorder (n = 9). The mean scores on the MADRS (13.3 [SD = 12.0]) and CGI-S (1.4 [SD = 1.2]) indicated a mild level of depression severity. The mean score on the GAF was 64.7 (SD = 10.2).

The mean interval between the completion of the paper and Internet versions of the scale was 1.2 days (SD = 0.9). The correlation between the CUDOS and CUDOS-W was high (ICC = 0.96, $P < .001$). The mean ± SD scores were nearly identical on the paper and Internet administrations (20.0 ± 14.6 vs 20.6 ± 13.9, paired $t = 1.0$, NS). In our previous validation studies[6–8] of the paper version of the CUDOS, we found that a cutoff score of 20 identified patients who were in remission. Given this cutoff score, there was 96.7% agreement between the paper and Internet administrations in determining patients' remission status (κ = 0.85).

The internal consistency of the paper and Internet administrations of the CUDOS was high (Cronbach α = .93 and .93, respectively). All item-scale correlations for the paper and Internet versions were significant (median for paper administration = 0.76; median for Internet administration = 0.74). For each item the correlation between the CUDOS and CUDOS-W was significant (median = 0.86).

Both the paper and Internet versions of the CUDOS were significantly correlated with the MADRS, CGI-S, and GAF (Table 1). The paper version of the CUDOS was significantly more highly correlated with the MADRS than the Web version.

© Copyright 2011 Physicians Postgraduate Press, Inc.

Table 2. Depressed Patients' Perceptions of the Paper and Internet Administration of the Clinically Useful Depression Outcome Scale (N = 53)[a]

Question	Paper, % (n)	Web, % (n)	About the Same, % (n)	*z* Value	*P* Value
Which scale took less time to complete?	5.9 (3)	58.8 (30)	35.3 (18)	4.6	<.001
Which scale was more of a burden to complete?	76.5 (39)	2.0 (1)	21.6 (11)	−6.0	<.001
Which scale would you prefer to complete to monitor your progress in treatment?	0.0 (0)	100.0 (51)	0.0 (0)	7.1	<.001
Which method of collecting this information do you feel is safer and more secure?	3.9 (2)	54.9 (28)	41.2 (21)	4.8	<.001
Which method of collecting information do you think is more accurate and valid?	0.0 (0)	41.5 (17)	58.5 (24)	4.2	<.001

[a]Due to missing data, the sample size responding to each question ranged from 41 to 51.

The patients preferred to complete the scale on the Internet (Table 2). Internet administration was viewed as less burdensome, less time consuming, more secure, and even more accurate and valid.

We examined the possible impact of age, gender, and number of times the self-administered scale had been previously completed prior to the study. A median split was used to divide the patients by age (above and below 46 years). Patients who previously completed the scale fewer than 5 times (n = 13) were compared to patients who completed the scale 5 or more times. None of these variables were significantly associated with the perception or preference of mode of administration, item-scale correlations, scale reliability, or the correlation between different versions of the CUDOS and the validity scales.

DISCUSSION

The results of this study supported the reliability and validity of Internet administration of an outcome measure for depression. Internal consistency, item-scale correlations, and correlations with external validators were as high, or nearly as high, with Internet administration as with paper administration of the scale. This is not surprising because the correlation between the 2 modes of administering the CUDOS was high. The MADRS was significantly more highly correlated with the paper than with the Web-based version of CUDOS. This difference was likely because the MADRS was administered on the same day as the completion of the paper version of the scale, whereas the Web-based version was completed 1 or 2 days earlier. Thus, the design was, in small measure, biased against Web-based administration.

The patients clearly preferred Internet administration of the scale. Internet administration was reported to be less burdensome and time consuming. Importantly, the patients indicated that they perceived Internet administration to be as secure and private as paper administration. Surprisingly, Web-based administration was also perceived as more valid, though this opinion was not supported by the analyses comparing the validity of the respective approaches. We believe that the patients generally had a more favorable view of Web-based administration, and this perspective influenced all of their ratings comparing the 2 approaches.

Other studies have demonstrated the equivalence of computer and paper administration of measures of depression, though many of these studies have been of college students.[19–22] The usual methodology of studies comparing computer and paper versions of a scale has been based on administration of the alternate versions at the same testing session.[18,21,26–28] Only a small number of studies have compared paper and Internet administrations of depression scales,[19,29] and we are not aware of any studies of patients in ongoing treatment in clinical practice.

Two types of equivalence need to be demonstrated before concluding that different modes of scale administration are interchangeable.[14] The first type of equivalence relates to the absolute metric or score of the test and is based on a demonstration of comparability in measures of central tendency. If one mode of administration results in scores that are 20% higher than the other, this disparity creates problems when cutoff scores are used to determine status. Buchanan[13] reviewed several studies suggesting that online score distributions tend to be higher than offline scores, and the author cautioned against extrapolating norms and cutoff scores based on paper administration. In the case of depression, rating scale cutoff scores have been used to determine caseness, levels of severity, and remission status. Before concluding that paper- and computer-administered depression scales are equivalent, it is necessary to demonstrate that the same individuals are identified as depressed cases, or as being in remission, on the different versions of the test. The cutoff score might well be the same, or scores could be transformed to achieve equivalence. The second type of equivalence relates to the underlying construct of the test. If the mode of administering a depression test is influenced by differences in the environment or such factors as computer experience or computer anxiety, then the underlying psychometrics of the test may change, as might the rank order of test takers. Examination of the psychometric properties and reliability coefficients are relevant to this concern. The results of the present study indicated that the modes of administering the CUDOS achieved both types of equivalence. The psychometric properties and validity correlations with clinician measures of depression severity of the Web-based and paper administration of the CUDOS were nearly identical. High agreement was found in determining which patients were in remission.

There are several differences between the current study and most other studies comparing paper and computer assessments of symptoms. Other studies typically counterbalance the order of presentation.[11,18,20,26,29] That is, some subjects complete the paper version first, and others complete the computer version prior to the paper version. In general, these studies have not found evidence of an order effect. In the present study, all patients completed the Internet version prior to the paper administration. We chose this methodology because we wanted to reduce the impact of

study participation on reliability. At the time of the Internet scale completion, the patients were not aware that they were going to be recruited into a study comparing Internet and paper completion of the scale. In contrast, in most other studies the paper and computer administrations are completed at the same testing session, and subjects are aware, when completing the first administration of the scale, that they will be completing it again, albeit in a different format. Thus, the design of the current study reduced the demand characteristics that might artificially enhance agreement between the 2 administrations of the scale.

The results of the present study are limited to patients who had access to the Internet and e-mail accounts. Also, the participants were sufficiently motivated to register on the Web site in order to receive the e-mail requesting scale completion. It is possible that, despite reassurance that the information was encrypted and being sent to a secure server, patients who were concerned about privacy issues chose not to register. Any Web-based system of outcome assessment will be limited to patients with Internet access who are not so concerned about privacy issues that they are unwilling to complete measures about their health status using this medium. Thus, while the results of the study may not generalize to all patients being treated for depression, they are pertinent to those patients to whom a Web-based outcome system would apply.

In conclusion, the results of the present study indicate that an Internet version of a depression scale was equivalent to the paper version. Internet assessment was reliable and valid, and patients preferred Web-based scale administration to paper-and-pencil administration in the office. The availability of a reliable, valid, and acceptable Web-based system for measuring outcome can enhance capabilities for tracking longitudinal course of this often chronic disorder. Another advantage of Internet scale administration is that it provides an inexpensive method for data aggregation across a sample of patients. With paper-and-pencil–administered scales, data aggregation could be labor intensive and, thus, costly. While data aggregation is not necessary to realize the clinical benefit of measurement-based care in the treatment of individual patients, data aggregation is necessary for quality improvement efforts in which outcome across a case load is compared before and after a change in service delivery. The Web site for the CUDOS can be found at www.outcometracker.org and is available for clinicians to use at no cost.

A limitation of the study is that it was conducted in a single practice where most of the patients have health insurance and only patients with Internet access participated. Moreover, using scales to measure outcome is commonplace in the practice, which is at variance with the practice of most psychiatrists.[3,4] However, there was no association between experience with completing the self-report scale and the results. We did not systematically record compliance rates with scale completion at home and in the office. Anecdotal observation was that, when patients were asked to complete the scale in the office, almost everyone did so. When patients were asked to complete the scale on the Web site, the percentage was lower because some patients forgot to do so. While we studied only patients with Internet access at home, it is possible to establish computer access for patients to complete the Web-based scale in the office. The sample size was relatively small. However, the problem with a small sample size is low power and subsequent type I error, and this does not account for the high, and significant, correlation between the 2 versions of the scale, the significant correlations with the clinician ratings, and the significant results indicating that patients preferred Web-based scale completion.

Author affiliations: Department of Psychiatry and Human Behavior, Brown University School of Medicine, Rhode Island Hospital, Providence, Rhode Island.
Potential conflicts of interest: None reported.
Funding/support: None reported.

REFERENCES

1. Trivedi MH, Rush AJ, Wisniewski SR, et al; STAR*D Study Team. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry*. 2006;163(1):28–40.
2. 2008 Physician Quality Reporting Initiative. Center for Medicare and Medicaid Services. US Department of Health and Human Services Web site. https://www.cms.gov/PQRS/2008/list.asp#TopOfPage. Accessed September 29, 2011.
3. Gilbody SM, House AO, Sheldon TA. Psychiatrists in the UK do not use outcomes measures: national survey. *Br J Psychiatry*. 2002;180(2): 101–103.
4. Zimmerman M, McGlinchey JB. Why don't psychiatrists use scales to measure outcome when treating depressed patients? *J Clin Psychiatry*. 2008;69(12):1916–1919.
5. Zimmerman M. Integrating the assessment methods of researchers in routine clinical practice: the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project. In: First MB, ed. *Annual Review of Psychiatry*. Washington, DC: American Psychiatric Press, Inc; 2003:29–34.
6. Zimmerman M, Chelminski I, McGlinchey JB, et al. A Clinically Useful Depression Outcome Scale. *Compr Psychiatry*. 2008;49(2):131–140.
7. Zimmerman M, McGlinchey JB. Depressed patients' acceptability of the use of self-administered scales to measure outcome in clinical practice. *Ann Clin Psychiatry*. 2008;20(3):125–129.
8. Zimmerman M, Posternak MA, Chelminski I. Using a self-report depression scale to identify remission in depressed outpatients. *Am J Psychiatry*. 2004;161(10):1911–1913.
9. Emmelkamp PM. Technological innovations in clinical assessment and psychotherapy. *Psychother Psychosom*. 2005;74(6):336–343.
10. Cartreine JA, Ahern DK, Locke SE. A roadmap to computer-based psychotherapy in the United States. *Harv Rev Psychiatry*. 2010;18(2):80–95.
11. Coles ME, Cook LM, Blake TR. Assessing obsessive compulsive symptoms and cognitions on the internet: evidence for the comparability of paper and Internet administration. *Behav Res Ther*. 2007;45(9): 2232–2240.
12. Buchanan T. Online assessment: desirable or dangerous? *Prof Psychol Res Pr*. 2002;33(2):148–154.
13. Buchanan T. Internet-based questionnaire assessment: appropriate use in clinical contexts. *Cogn Behav Ther*. 2003;32(3):100–109.
14. McDonald A. The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Comput Educ*. 2002;39(3):299–312.
15. American Psychological Association. *Guidelines for Computer-Based Tests and Interpretations*. Washington, DC: American Psychological Association; 1986.
16. Webster J, Compeau D. Computer-assisted versus paper-and-pencil administration of questionnaires. *Behav Res Methods Instrum Comput*. 1996;28(4):567–576.
17. Barbeite F, Weiss E. Computer self-efficacy and anxiety scales for an internet sample: testing measurement equivalence of existing measures and development of new scales. *Comput Human Behav*. 2004;20(1):1–15.
18. Cook IA, Balasubramani GK, Eng H, et al. Electronic source materials in

clinical research: acceptability and validity of symptom self-rating in major depressive disorder. *J Psychiatr Res*. 2007;41(9):737–743.
19. Herrero J, Meneses J. Short web-based versions of the perceived stress (PSS) and Center for Epidemiological Studies-Depression (CESD) Scales: a comparison to pencil and paper responses among internet users. *Comput Human Behav*. 2006;22(5):830–846.
20. Merten T, Ruch W. A comparison of computerized and conventional administration of the German versions of the Eysenck Personality Questionnaire and the Carroll Rating Scale for Depression. *Pers Individ Dif*. 1996;20(3):281–291.
21. Schulenberg S, Yutrzenka B. Equivalence of computerized and conventional versions of the Beck Depression Inventory-II (BDI-II). *Curr Psychol*. 2001;20(3):216–230.
22. Vallejo MA, Jordán CM, Díaz MI, et al. Psychological assessment via the internet: a reliability and validity study of online (vs paper-and-pencil) versions of the General Health Questionnaire-28 (GHQ-28) and the Symptoms Check-List-90-Revised (SCL-90-R). *J Med Internet Res*. 2007;9(1):e2.
23. First MB, Spitzer RL, Williams JBW, et al. *Structured Clinical Interview for DSM-IV (SCID)*. Washington, DC: American Psychiatric Association; 1997.
24. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134(4):382–389.
25. Guy W. *ECDEU Assessment Manual for Psychopharmacology*. Rockville, MD: National Institute of Mental Health; 1976.
26. Glaze R, Cox JL. Validation of a computerised version of the 10-item (self-rating) Edinburgh Postnatal Depression Scale. *J Affect Disord*. 1991; 22(1–2):73–77.
27. Kurt R, Bogner HR, Straton JB, et al. Computer-assisted assessment of depression and function in older primary care patients. *Comput Methods Programs Biomed*. 2004;73(2):165–171.
28. Ogles BM, France CR, Lunnen KM, et al. Computerized depression screening and awareness. *Community Ment Health J*. 1998;34(1):27–38.
29. Carlbring P, Brunt S, Bohman S, et al. Internet vs paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. *Comput Human Behav*. 2007;23(3):1421–1434.