Computerized Screening for Psychiatric Disorders in an Outpatient Community Mental Health Clinic

Kenneth A. Kobak, Ph.D.
Leslie vH. Taylor, M.D.
Susan L. Dottl, Ph.D.
John H. Greist, M.D.
James W. Jefferson, M.D.
Diane Burroughs, B.A.
David J. Katzelnick, M.D.
Monica Mandell, Ph.D.

Objective: **This study examined the validity and utility of two types of computer-administered versions of a screening interview, PRIME-MD (Primary Care Evaluation of Mental Disorders), in a mental health setting: one administered by desktop computer and one by computer using a touch-tone telephone and interactive voice response (IVR) technology.** *Methods:* **Fifty-one outpatients at a community mental health clinic were given both IVR and desktop PRIME-MD and the Structured Clinical Interview for DSM-IV (SCID-IV), which was administered by a clinician, in a counterbalanced order. Diagnoses were also obtained from charts.** *Results:* **Prevalence rates found by both computer interviews were similar to those obtained by the SCID-IV for the presence of any diagnosis, any affective disorder, and any anxiety disorder. Prevalence rates for specific diagnoses were also similar to those found by the SCID-IV except for dysthymia, obsessive-compulsive disorder, and panic disorder; the first two conditions were found to be more prevalent by the computer, and panic disorder was more prevalent by the SCID. Compared with the prevalence rates in the charts, the rates found by the computer were higher for anxiety disorders, particularly for obsessive-compulsive disorder and social phobia. Using the SCID-IV as the criterion, both computer-administered versions of PRIME-MD had high sensitivity, specificity, and positive predictive value for most diagnoses. No significant difference was found in how well patients liked each form of interview.** *Conclusions:* **Results support the validity and utility of both desktop and IVR PRIME-MD for gathering information from mental health patients about certain diagnoses. (*Psychiatric Services* 48:1048–1057, 1997)**

The first step in the development of a successful treatment plan for patients with psychiatric disorders is a comprehensive and accurate diagnostic evaluation. Although the clinician's theoretical framework determines to a large extent the domains of interest for evaluation, current diagnostic practice focuses largely on description of symptoms and signs and is atheoretical in terms of etiology. Thus an accurate and comprehensive evaluation of all psychiatric morbidity provides a basis from which clinicians can then turn to theoretical orientations for an appropriate intervention strategy.

Patients typically come for treatment with a salient presenting problem, and clinicians often focus primarily on evaluating the dimensions of this problem. However, successful treatment is often hampered by covert comorbidity. For example, a patient who presents complaining of depression but also has an underlying problem with alcohol abuse may not respond to typical interventions with proven efficacy, such as antidepressant medications or cognitive therapy, and may in fact be more appropriately treated with an intervention focused primarily on the alcohol abuse.

In other cases, treatment efficacy may be maximized by the awareness of underlying comorbidity. For example, a patient who presents complaining of symptoms of depression might not reveal comorbid symptoms of obsessive-compulsive disorder. A thorough diagnostic interview at intake could identify this comorbidity, permitting a better treatment plan—for instance, prescription of a potent selective serotonin reuptake inhibitor

Except for Dr. Mandell, the authors are affiliated with the Dean Foundation for Health, Research, and Education, 2711 Allen Boulevard, Middleton, Wisconsin 53562. Dr. Mandell is with the Hennepin County Medical Center in Minneapolis.

that is effective for both disorders rather than a potent norepinephrine reuptake inhibitor helpful only for depression (1).

Although ideally such thorough diagnostic evaluation should precede treatment, in real life this ideal is not fully realized for several reasons. One issue is clinicians' unfamiliarity with diagnostic criteria, which often results in mistakes of commission (an incorrect diagnosis) as well as omission (failure to detect comorbidity). For example, Skodol and others (2) reported a study of *DSM-III* diagnoses made for 200 outpatients by a group of ten psychiatric residents in their third postgraduate year and five clinical psychology interns; 37 percent of axis I diagnoses were found to be incorrect on review during supervision. Errors in applying *DSM-III* conventions or criteria accounted for 78 percent of the incorrectly diagnosed cases. These errors occurred even though the residents and interns took a didactic course on *DSM-III* and the residents took an additional three-month course on diagnostic interviewing.

In a related study, 30 percent of diagnoses made by a group of 18 psychiatric residents and ten clinical psychology interns, which were reviewed and approved in absentia by their supervisor, were found to differ from diagnoses made subsequently by consensus between the resident or intern and an experienced clinician who sat in on the initial interview (3). The major reason for error was inadequate information resulting from focusing too narrowly on the initial complaint and thus failing to attend to information that suggested an additional or alternate diagnosis.

In a third study, knowledge of the diagnostic criteria for major depression was evaluated among the mental health staff of a large training hospital using a 15-item true-false questionnaire (4). Staff included psychiatrists, psychiatric residents, psychologists, social workers, and master's-level nurses. The rate of incorrect responses ranged from 13 to 48 percent: 13 percent incorrectly thought that depression brought on by psychosocial factors cannot be melancholic, and 48 percent incorrectly thought that at least one vegetative sign was required to make the diagnosis.

Because *DSM-IV* lists more than 300 disorders, it is understandable that clinicians often make diagnostic errors. One solution to this problem has been the development of structured diagnostic interviews. However, few clinicians are trained in conducting formal diagnostic evaluations using structured or semistructured clinical interviews, such as the Structured Clinical Interview for DSM-IV (5) or its predecessor, the Schedule for Affective Disorders and Schizophrenia (6). Even when structured interviews are conducted, clinicians often forget to ask up to 5 percent of the required questions (7) and fail to process diagnostic logic accurately (Kobak KA, Taylor LvH, Dottl SL, et al., unpublished manuscript, 1996). A thorough diagnostic evaluation takes substantial clinician time and results in additional costs to the patient or insurer. These economic concerns, as well as clinicians' time constraints, often result in a narrow focus on the presenting problem and its remediation.

Use of computers frees clinicians to focus on what they do best: forming relationships with patients, weighing the available information, making mutually informed treatment decisions, and conducting the appropriate therapy.

One solution to these problems is the use of computer-administered diagnostic interviews. Information obtained by computer-administered interviews is thorough and complete. Computers never fail to ask the required questions as clinicians often do. For example, in one study physicians missed or failed to record 35 percent of the information obtained by a computer-administered medical history (8). Computers always make perfect branching decisions if programmed to do so, asking the appropriate follow-up questions based on the patient's response and the diagnostic algorithm. Scoring errors are eliminated, as diagnoses are computed automatically according to the patient's responses and the diagnostic criteria. Standardization of administration minimizes the problem of interrater reliability, as computers ask the same questions in the same way of all patients.

Computers allow patients to work at their own pace and are available wherever a computer terminal (or in the case of interactive voice response, a touch-tone telephone) is available. Results can be scored and presented to the clinician or patient immediately, enabling clinicians to make more informed treatment decisions. Computer-administered rating scales are cost-effective and time-efficient, eliminating the clinician's time involved in learning and administering structured diagnostic interviews. They can be available at all times, and they free clinicians to focus on what they do best: forming relationships with patients, weighing the available information, making mutually informed treatment decisions, and conducting the appropriate therapy.

Patients' reaction to computer interviews has generally been positive. Studies have found that patients are more likely to disclose information of a sensitive nature to a computer than to a clinician in areas such as suicide (9), alcohol and drug abuse (10), and high-risk sexual behavior (11). Psychiatric inpatients reported an average of five and one-half more items of information on a computerized psy-

chiatric history than on a clinician interview, including having a criminal record (reported on a computer but not to the clinician by 26 percent of patients), blackouts from drinking (23 percent), impotence (20 percent), being fired (17 percent), and suicide attempts (17 percent) (12). Subjects feel less embarrassed giving information to the computer (13), and some patients, such as socially phobic individuals (14) and suicide attempters (15), prefer being interviewed by a computer rather than by a clinician.

Direct patient-computer interviews have been available for more than 30 years, beginning with the automated medical history of Slack and associates (16). Since that time, empirical evidence has supported the reliability, validity, and equivalence of computer-administered versions of clinician-administered symptom rating scales for disorders such as depression (17), obsessive-compulsive disorder (18), generalized anxiety disorder (19), and social phobia (20). Several structured diagnostic interviews have also been adapted for computer administration, such as the Structured Clinical Interview for DSM-IV (SCID-IV) (21), the Symptom-Driven Diagnostic System for Primary Care (SDDS-PC) (22), and the Composite International Diagnostic Interview (CIDI-Auto) (23).

As of this writing, no published data are available on the reliability and validity of these computer-administered diagnostic interviews except for the CIDI-Auto, which was studied in the diagnosis of major depression and six of the anxiety disorders (agoraphobia, panic disorder with and without agoraphobia, social phobia, simple phobia, obsessive-compulsive disorder, and generalized anxiety disorder) (23). Results were mixed, with levels of agreement with clinician diagnoses ranging from .02 (kappa statistic) for generalized anxiety disorder to .81 for obsessive-compulsive disorder, and with an overall level of agreement of .40.

The study reported here examined the validity of a computer-administered version of a new diagnostic instrument, PRIME-MD (Primary Care Evaluation of Mental Disorders) (24). PRIME-MD was originally designed for use by primary care physicians in detecting the mental disorders most commonly found in primary care settings—mood, anxiety, somatoform, alcohol, and eating disorders.

PRIME-MD consists of two components: a brief, 26-item patient questionnaire in a yes-no format that the patient completes before seeing the primary care physician and a structured clinical interview, or clinician evaluation guide, containing five modules in a yes-no format that correspond to the five classes of disorders covered. Clinicians administer only the modules that are indicated, based on patients' answers on the patient questionnaire.

The validity of the clinician-administered PRIME-MD was demonstrated by a study in which the same patients were diagnosed by primary care physicians using PRIME-MD and by experienced mental health professionals (24). A high level of agreement was found for the presence of any diagnosis (kappa=.71; overall accuracy rate=88 percent).

Recently, a computer-administered version of PRIME-MD was developed using interactive voice response (IVR) technology (25). With this method, patients dial a central phone number using a touch-tome phone, listen to questions read by the computer over the telephone, and respond by pressing numbers on the telephone. Touch-tone IVR technology offers the advantage of not requiring patients to be physically present to complete the interview. It is more familiar to most patients than desktop computers and obviates the need for hardware and software at all locations. Compared with other computer-administered diagnostic interviews, the telephone-administered computer-assisted PRIME-MD—referred to in this paper as IVR PRIME-MD—has the advantage of being relatively brief, taking about 10 minutes to complete. The computerized SCID-IV takes about 25 minutes, the computerized SDDS-PC about 15 minutes, and the CIDI-Auto about 65 minutes.

A validation study of IVR PRIME-MD was done with 200 outpatients from four primary care clinics (N=80), an eating disorders clinic (N=10), an alcohol treatment facility (N=10), and a psychiatric research foundation (N=70); community control subjects (N=30) also participated (Kobak KA, Taylor LvH, Dottl SL, et al., unpublished manuscript, 1996). Over the telephone, the subjects were administered both the SCID-IV by a trained clinician and IVR PRIME-MD. A subsample also was given the clinician-administered version of PRIME-MD in person by a primary care physician.

Similar prevalence rates for any psychiatric disorder among the primary care patients were obtained by the computer and by a mental health professional using the SCID-IV (38.8 percent and 36.3 percent, respectively). Primary care physicians using the clinician-administered version of PRIME-MD identified significantly less psychopathology in this sample (12.5 percent) than did either IVR PRIME-MD or the SCID-IV (χ^2= 9.23, df=2, $p<.01$).

Prevalence rates for individual diagnoses were generally similar for IVR PRIME-MD and PRIME-MD. However, primary care patients reported twice as much alcohol abuse on the computer (15 percent) than on either the SCID-IV (7.5 percent) or the clinician-administered PRIME-MD (7.5 percent). Using the SCID-IV as the criterion (the standard against which both interviews were compared), both the computer- and clinician-administered versions of PRIME-MD demonstrated high and roughly equivalent levels of sensitivity and specificity. Overall agreement for any diagnosis was .67 (kappa statistic) for the computer-administered PRIME-MD and .70 for the clinician-administered PRIME-MD. Overall accuracy was also high at 84 percent and 86 percent, respectively.

The study reported here examined the validity and utility of computer-administered versions of PRIME-MD in a community mental health population. Use of a computer-administered version of PRIME-MD in this population could greatly increase the quality of patient care by increasing clinicians' recognition of both overt and covert disorders, resulting in more effective treatment planning.

In this study, two types of computer administration of PRIME-MD were examined—one was administered by a desktop computer and one by a computer over the telephone using IVR technology (IVR PRIME-MD). Diagnoses obtained by both methods were compared with diagnoses obtained by an expert clinician conducting the SCID-IV by telephone.

Methods

Subjects and setting

Participants were 51 outpatients who were currently receiving psychotherapy at the Hennepin County Medical Center's evaluation and therapy clinic in Minneapolis. The center is a large urban teaching hospital, and its psychiatric clinics primarily serve a low-income population. The majority of patients are supported by general assistance, Social Security Disability Insurance, or Aid to Families With Dependent Children. Although Medicaid and Medicare are the primary insurance providers, a smaller percentage of patients have private insurance, such as Blue Cross/Blue Shield and Medica.

At the time of the study (May to September 1996) the professional staff consisted of two full-time and five part-time doctoral-level psychologists, four part-time psychology fellows, two full-time and two part-time doctoral-level psychology interns, two full-time master's-level psychologists, and two full-time master's-level psychiatric nurses. Patients requiring psychotropic medication were followed in the adjoining but separate medication clinic, staffed by two full-time and 14 part-time psychiatrists and four psychiatric nurses. In 1995 a total of 28,574 visits were made to the two clinics, 17,929 to the therapy clinic and 10,645 to the medication clinic.

The age of participants ranged from 23 to 66 years (mean±SD= 44.12± 9.01). Twenty-nine participants (57 percent) were women). Forty-three participants (84 percent) were Caucasian, four (8 percent) were African American, two (4 percent) were Hispanic, and two (4 percent) were Native American.

Twenty-eight participants (55 percent) were on disability, 15 (29 percent) were unemployed, five (10 percent) were employed, one (2 percent) was retired, and data were unavailable for two participants. Seventeen (33 percent) listed their current or former occupation as clerical, skilled, or semiskilled manual labor; six (12 percent) were in unskilled occupations; five (9 percent) were in administration- or management-level occupations; and data were unavailable for 23 participants. Five participants (10 percent) did not finish high school, 14 (28 percent) had a high school education or its equivalent, 19 (37 percent) had completed some college, five (10 percent) graduated from a two-year college, four (8 percent) finished four years of college, three (6 percent) had some graduate or professional school training, and data were unavailable for one participant (2 percent).

The versions of PRIME-MD that were administered by computer successfully identified the majority of psychiatric disorders present in this sample of mental health patients.

Although a decision was made before the study to exclude psychotic patients, no patient had to be excluded for this reason.

Procedure

Subjects were recruited through flyers posted in the patient waiting area of the evaluation and therapy clinic or through word of mouth from their psychotherapist. Persons interested in participating spoke with a research coordinator who explained the purpose of the study, answered the subject's questions, and obtained informed consent. Subjects then called the Dean Foundation in Madison, Wisconsin, using an 800 number, and interview times were arranged.

All subjects were administered three interviews: IVR PRIME-MD; a desktop-computer-administered version of PRIME-MD (hereafter referred to as desktop PRIME-MD); and the mood, anxiety, eating disorders, and psychoactive substance use modules from the SCID-IV (5). A three-way, counterbalanced order of administration was used. Each interview was scheduled 24 to 72 hours apart to minimize memory and other order effects.

Subjects called an 800 number to be administered IVR PRIME-MD. Desktop PRIME-MD was administered at the clinic on a desktop computer located in the patients' waiting area. Subjects who did not have a home telephone (N=3) were invited to come to the clinic to use the telephone.

The SCID-IV was administered over the telephone by an advanced doctoral student in psychology who was proficient in diagnostic interviewing and trained in administration and scoring of the SCID-IV. The SCID-IV rater also administered the Hamilton Depression Rating Scale (26) to examine the relationship between the diagnosis of a mood disorder and the severity of depressive symptoms. The SCID rater was blind to the results of both computer-administered PRIME-MD interviews for the duration of the study protocol.

Assessment via telephone has been shown to yield scores similar to face-to-face interviews in the areas of depression (27–29), health status (30), and psychiatric diagnosis (31). The SCID-IV was used as the criterion against which the accuracy of both desktop and IVR PRIME-MD were compared. Spitzer (32) has described a semistructured interview (such as the SCID-IV) that uses expert clinical judgment and open-ended questions as the criterion for assessing the validity of a fully structured interview with closed-ended questions that does not require clinician judgment.

Table 1

Patients screening positive for at least one symptom on the interactive voice response (IVR) and desktop-computer-administered versions of PRIME-MD, by diagnosis

Diagnosis	IVR (N=51)		Desktop (N=51)	
	N	%	N	%
Any psychiatric disorder	49	96	51	100
Any mood disorder[1]	41	80	45	88
Any anxiety disorder	48	94	48	94
Generalized anxiety disorder	47	92	47	92
Panic disorder	36	71	47	92
Social phobia	33	65	34	67
Obsessive-compulsive disorder	30	55	32	63
Alcohol abuse	20	39	16	31
Any eating disorder[1]	18	35	19	37

[1] The screening items for all diagnoses within this module were identical.

Both desktop PRIME-MD and IVR PRIME-MD screened subjects for several *DSM-IV* disorders. Mood disorders included major depressive disorder, partial remission or recurrence of major depressive disorder, dysthymia, minor depressive disorder (as defined by PRIME-MD), and bipolar disorder. Anxiety disorders screened for were panic disorder, social phobia, obsessive-compulsive disorder, and generalized anxiety disorder. Alcohol abuse was also assessed. Eating disorders screened for were binge eating disorder (a *DSM-IV* disorder proposed for further study) and bulimia nervosa.

Data from the somatization module were not included in the analyses (the module was administered by desktop PRIME-MD but not IVR PRIME-MD) because of the low prevalence of somatoform disorder and because the module requires a physician's evaluation to determine whether each somatic symptom has a physical explanation that adequately accounts for its severity and associated disability. The diagnoses of social phobia and obsessive-compulsive disorder were not part of the original clinician-administered PRIME-MD because these disorders are rarely treated in the primary care setting. However, these modules were added to the IVR and desktop versions of the PRIME-MD because of their substantial prevalence and often hidden nature. In addition, the increasing gatekeeper role of primary care clinicians makes comprehensive screening vital.

A chart review was conducted to obtain diagnoses made by the treating clinicians. These diagnoses were compared with diagnoses obtained by both computer interviews and the SCID-IV. Chart diagnoses included those that were made by the current clinician and the psychiatrist managing medication, as well as discharge diagnoses if the subject was recently hospitalized.

Subjects were paid $50 for their participation. The study was reviewed and approved by both the Dean Foundation institutional review board and the human subjects research committee of Hennepin County Medical Center.

IVR PRIME-MD

Subjects self-administered IVR PRIME-MD by telephoning an 800 number and entering a patient identification number assigned by the study coordinator and an individual password chosen by the subject. They then listened to prerecorded questions played by the computer over the telephone and answered by pressing number keys on the touch-tone telephone. The interview began with a brief introduction, followed by an optional offer of instructions on how to use the IVR system.

The computer then asked the subject four distractor questions (for example, "Have you been bothered a lot by back pain during the past two weeks?") before proceeding to the patient questionnaire. This questionnaire contained 13 screening questions for the four PRIME-MD modules being evaluated: two for the mood module, six for anxiety, four for alcohol abuse, and one for eating disorders. A positive response to any screening question in a module resulted in the computer's branching to that module for further evaluation and possible confirmation of the diagnosis; a negative response to all the screening questions in a module resulted in the computer's skipping further evaluation in that module.

Desktop PRIME-MD

The desktop PRIME-MD was given at the clinic on a IBM-compatible computer located in the patient waiting area. Subjects checked in with the receptionist who oriented them to the computer and then left them to complete the interview alone. At the beginning of the interview the subject was asked to enter his or her name, gender, date of birth, and identification number. On the screen the computer then asked questions that were almost identical to those in IVR PRIME-MD, following the same branching logic.

Subjects answered questions on the screen by using a mouse to point and click on a box containing "yes" or "no." Subjects could skip a question they chose not to answer, which was not an option of IVR PRIME-MD. After the interview, the results were automatically stored on the computer hard disk. A hard copy was printed out, and it was filed by the receptionist.

After subjects completed all three interviews, they were mailed a seven-item questionnaire asking them to evaluate their experience with each of the three interviews. Subjects mailed the results back to the study coordinator in Wisconsin. All questionnaires were completed anonymously.

Statistical analyses

The analyses examined several features of the two versions of PRIME-MD. They included sensitivity, which is the proportion of cases given a SCID-IV diagnosis that were also correctly given the diagnosis by the computer, and specificity, or the proportion of cases not given a SCID-IV diagnosis that were also correctly not given the diagnosis by the computer. Positive predictive value, which is the proportion of cases given the diagnosis by the computer that were also given the diagnosis by the SCID-IV, was also examined. The overall accuracy rate, or the proportion of total patients correctly identified by the computer as having or not having the diagnosis, was also calculated, using the SCID-IV as the standard.

Positive predictive value (that is,

the "hit rate") is particularly important in examining the clinical utility of a screening instrument, as it tells the probability that a person who screens positive on the test actually has the disorder. Agreement between SCID-IV diagnoses and both IVR PRIME-MD and desktop PRIME-MD diagnoses were computed using kappa coefficients. The prevalence rates of diagnoses found by the three methods of assessment were examined. A separate analysis was conducted for the presence of any disorder within a diagnostic category (for example, within the category of mood disorders or anxiety disorders) and for the presence of specific diagnoses. Prevalence rates for chart diagnoses were compared with rates obtained by the SCID-IV and both computer versions of PRIME-MD.

Table 2

Diagnoses of 51 outpatients obtained by PRIME-MD administered by interactive voice technology (IVR) and a desktop computer, by a clinician using the Structured Clinical Interview for DSM-IV (SCID-IV), and by chart review

Diagnosis	IVR		Desktop		SCID-IV		Chart review	
	N	%	N	%	N	%	N	%
Any diagnosis	48	94	47	92	51	100	50	98
Any mood disorder	43	84	44	86	50	98	48	94
Major depression	31	61	35	69	35	67	39	77
Major depression, partial remission	9	18	6	12	26	51	1	2
Dysthymia	34	67	31	61	17	33	20	39
Minor depression	1	2	2	4	4	8	1	2
Any anxiety disorder	40	78	42	82	44	86	28	55
Panic disorder	23	45	22	43	33	65	14	28
Generalized anxiety disorder	22	43	26	51	29	57	6	12
Obsessive-compulsive disorder	22	43	27	53	13	26	7	14
Social phobia	25	49	26	51	27	53	1	2
Alcohol use disorder	15	29	9	18	11	22	16	31
Any eating disorder	7	14	9	18	5	10	3	6

Results

Patient questionnaire

As shown in Table 1, of the 51 patients screened, 49 (96 percent) screened positive on the patient questionnaire for the presence of a symptom of any psychiatric disorder on IVR PRIME-MD, and 100 percent screened positive on desktop PRIME-MD. Screening positive for a symptom means the person answered "yes" to a screening item (a symptom) for a particular disorder, which resulted in the computer's branching to the appropriate diagnostic module for further evaluation and possible confirmation of the diagnosis. Patients screened positive for a symptom of mood and eating disorders slightly more often on desktop PRIME-MD than on the IVR version, slightly less often for alcohol abuse, and equally as often for anxiety disorders.

Diagnostic rates

Table 2 presents a summary of prevalence rates of psychiatric disorders identified by IVR PRIME-MD, desktop PRIME-MD, and SCID-IV. The percentage of patients receiving any diagnosis was similar for all three interview methods (94 percent, 92 percent, and 100 percent, respectively). Rates were also similar for any mood disorder and for any anxiety disorder. In order to receive a diagnosis on PRIME-MD, a person screened positive for a symptom on the patient questionnaire, and the diagnosis was confirmed by the appropriate follow-up diagnostic module.

The prevalence rates found by IVR and desktop PRIME-MD were similar to those found by the SCID-IV for major depressive disorder, generalized anxiety disorder, social phobia, and any eating disorder. The computer interviews found significantly greater rates of dysthymia ($\chi^2=12.98$, df=2, $p<.01$) and obsessive-compulsive disorder ($\chi^2=8.2$, df=2, $p<.025$) than the SCID-IV, while the SCID-IV found higher rates of panic disorder than either IVR PRIME-MD ($\chi^2=3.96$, df=1, $p<.05$) or the desktop version ($\chi^2=4.77$, df=1, $p<.05$).

The mean±SD number of diagnoses per patient was 4.33±2.18 using IVR PRIME-MD (range, 0 to 8 diagnoses), 4.41±2.18 on the desktop version (range, 0 to 8), and 4.37±1.92 on the SCID-IV (range, 0 to 7). It took a mean of 7.36±4.48 minutes for patients to complete desktop PRIME-MD (range, 1.18 to 12.86 minutes), 8.61±2.80 minutes to complete IVR PRIME-MD (range, 2.40 to 14.50 minutes), and 76.17±17.98 minutes to complete the SCID-IV, including the Hamilton Depression Rating Scale (range, 40 to 150 minutes).

Chart diagnoses

Chart diagnoses were examined to compare diagnostic rates found by the clinicians working with the patients and those obtained by both the computer and the structured clinical interviews. Results are presented in Table 2. Prevalence rates of major depressive disorder, alcohol abuse, and eating disorders were similar for all four diagnostic methods. The rate of chart diagnosis was significantly lower than the rates found by either computer interview or the SCID-IV for panic disorder, ($\chi^2=14.41$, df=3, $p<.005$), generalized anxiety disorder ($\chi^2=25.57$, df=3, $p<.001$), and social phobia ($\chi^2=38.9$, df=3, $p<.001$). Only one subject was given a chart diagnosis of social phobia, compared with 25, 26, and 27 subjects given such a diagnosis by IVR PRIME-MD, the desktop version, and the SCID-IV, respectively. The rate of chart diagnoses for obsessive-compulsive disorder was significantly lower than for either IVR PRIME-MD ($\chi^2=10.84$, df=1, $p<.001$) or desktop PRIME-MD ($\chi^2=17.65$, df=1, $p<.001$) and was almost half the rate found by the SCID-IV, although this difference was not statistically significant.

Table 3

Indexes of agreement between PRIME-MD administered by interactive voice technology (IVR) and a desktop computer and the Structured Clinical Interview for DSM-IV (SCID-IV) administered by a clinician to diagnose 51 outpatients, in percentages

Diagnosis	Sensitivity		Specificity[1]		Positive predictive value		Overall accuracy		Kappa[1]	
	IVR	Desktop	IVR	Desktop	IVR	Desktop	IVR	Desktop	IVR	Desktop
Any diagnosis	94	92	—	—	100	100	94	92	—	—
Any mood disorder	84	88	—[2]	100	98	100	82	88	-.04	.22*
Major depression	77	77	75	50	87	77	77	69	.49***	.27
Major depression, partial remission	23	12	88	88	67	50	55	49	.11	-.01
Dysthymia	77	65	38	41	38	36	52	49	.12	.05
Minor depression	0	0	98	96	0	0	90	88	-.03	-.06
Any anxiety disorder	86	93	71	86	95	98	84	92	.47***	.70***
Panic disorder	61	58	83	83	87	86	69	67	.39**	.36**
General anxiety disorder	55	76	73	82	73	85	63	78	.27*	.57***
Obsessive-compulsive disorder	54	77	61	55	32	37	59	61	.12	.24*
Social phobia	74	78	79	79	80	81	77	78	.53***	.57***
Alcohol abuse and dependence	82	64	85	95	60	78	84	88	.59***	.63***
Any eating disorder	100	60	96	87	71	33	96	84	.81***	.35**

[1] Cannot be computed when the SCID-IV rate of diagnosis is 100 percent
[2] Cannot be computed due to small cell size
*p<.05
**p<.01
***p<.001

Diagnostic accuracy

Table 3 presents the indexes of agreement between the IVR and desktop versions of PRIME-MD and the SCID-IV. Sensitivity of both computer interviews was similar and was high for most diagnoses. For both computer interviews, specificity was also high (except for dysthymia, for which both values were low) and was similar (except for major depression, for which the IVR version had greater specificity). Positive predictive value for the presence of any diagnosis was 100 percent for both the desktop and IVR versions, meaning that in all cases in which the computer detected a mental disorder, at least one disorder was present.

The high rate of positive predictive value for the presence of any diagnosis is mainly an artifact of the population studied. However, positive predictive values for specific diagnoses were generally quite high as well, with the exception of dysthymia and obsessive-compulsive disorder. Previous research on clinicians' errors in using diagnostic criteria found that underdiagnosis of dysthymia is one of the most common problems, in that clinicians often fail to give the diagnosis of dysthymia when it is warranted if there is a comorbid major depressive episode.

The overall accuracy rate was quite high for both computer interviews. For IVR PRIME-MD, overall accuracy ranged from 96 percent for any eating disorder to 51 percent for dysthymia. For the desktop version, rates ranged from 88 percent for alcohol abuse and dependence to 49 percent for dysthymia. The kappa coefficients, representing agreement of both computer interviews corrected for chance, ranged from good to fair, and most comparisons were statistically significant. Overall, the level of diagnostic accuracy of both computer versions was quite good.

The diagnostic accuracy of the mood disorder module of IVR PRIME-MD was further examined by calculating a partial correlation coefficient between scores on the clinician-administered Hamilton Depression Rating Scale and the presence or absence of a diagnosis of any mood disorder. Scores were adjusted for the proportion of the variance attributed to age, gender, and ethnicity. A strong partial correlation of .5033 ($p<.001$) was obtained for desktop PRIME-MD, and a strong partial correlation of .5046 ($p<.001$) was found for IVR PRIME-MD. These results provide support for the diagnostic accuracy of the mood modules of both computer versions of PRIME-MD.

Patients' reactions

A survey was mailed to all participants after they completed all three interviews to evaluate their reaction to each interview form. A total of 37 patients (72 percent) returned the anonymous survey. Results are presented in Table 4. In response to the question "How comfortable were you being interviewed?" subjects rated the clinician higher on both ends of the extremes—that is, more were either very comfortable or very uncomfortable with the clinician compared with the computer. Ratings for the computer on this dimen-

sion fell more in the middle. A significantly greater percentage of subjects, 57 percent, rated themselves as being very comfortable with the clinician compared with 30 percent for the IVR PRIME-MD ($\chi^2=5.51$, df=1, $p<.05$); no significant difference in being very comfortable was found between the clinician and the desktop version.

In response to the question "How well were you able to describe your feelings?" a significantly greater percentage of patients, 92 percent, rated the clinician "OK" or higher compared with 62 percent for IVR PRIME-MD ($\chi^2=9.24$, df=1, $p<.006$) and 68 percent for desktop PRIME-MD ($\chi^2=6.77$, df=1, $p<.01$). A significantly greater percentage of patients, 51 percent, felt a little, somewhat, or very embarrassed with the clinician compared with 14 percent for the IVR version ($\chi^2=12.09$, df=2, $p<.001$) and 22 percent for the desktop version ($\chi^2=7.06$, df=2, $p<.01$).

No significant difference was found between the percentage of patients who did not like being interviewed by each interview form—that is, those who responded "didn't like at all" or "didn't like" (clinician, 8 percent; IVR PRIME-MD, 16 percent; desktop PRIME-MD, 14 percent). Similarly, no significant difference was found between the percentage of patients who felt they were able to understand the questions "OK," "well," or "very well" (clinician, 100 percent; IVR, 97 percent; desktop, 100 percent).

Finally, when asked which interview they preferred, a significantly greater percentage of patients, 73 percent, chose the clinician than either IVR PRIME-MD, 3 percent ($\chi^2=38.84$, df=1, $p<.001$), or desktop PRIME-MD, 16 percent ($\chi^2=24.12$, df=1, $p<.001$); 8 percent of patients had no preference. The difference between the percentage of patients who preferred desktop PRIME-MD (16 percent) and IVR PRIME-MD (3 percent) was also significant ($\chi^2=4.1$, df=1, $p<.05$).

Table 4

Responses of 37 patients to a survey about their reactions to IVR PRIME-MD, desktop PRIME-MD, and the clinician-administered SCID-IV, in percentages

Survey item	Rating 1	2	3	4	5
Comfortable[1]					
IVR	8	8	27	27	30
Desktop	3	5	27	24	41
SCID-IV	11	8	5	19	57
Able to describe feelings[2]					
IVR	11	27	30	19	14
Desktop	8	24	27	22	19
SCID-IV	3	5	16	11	65
Embarrassed[3]					
IVR	0	0	5	8	87
Desktop	3	0	8	11	78
SCID-IV	8	11	19	14	49
Understood questions[2]					
IVR	0	3	32	22	43
Desktop	0	0	22	35	43
SCID-IV	0	0	5	14	81
Liked being interviewed[4]					
IVR	11	5	51	19	14
Desktop	3	11	38	11	38
SCID-IV	5	3	17	14	61

[1] On a scale from 1, very uncomfortable, to 5, very comfortable
[2] On a scale from 1, very poorly, to 5, very well
[3] On a scale from 1, very embarrassed, to 5, not at all embarrassed
[4] On a scale from 1, didn't like at all, to 5, liked a lot

Discussion and conclusions

Diagnosis is the cornerstone of effective treatment. Direct computer-administered diagnostic interviews provide clinicians with useful additional information for obtaining a complete and accurate diagnostic evaluation of patients with mental disorders. In this study the computer-administered versions of PRIME-MD successfully identified the majority of psychiatric disorders present in this sample of mental health patients. Compared with chart diagnoses made by clinicians, the computer-administered interviews found higher rates of anxiety disorders, with strikingly higher rates of obsessive-compulsive disorder and social phobia. Social phobia was virtually undocumented in the charts.

Whether these findings accurately reflect clinicians' lack of awareness of these disorders, a simple failure to record the diagnosis in the chart, or patients' overendorsement of symptoms resulting from misunderstanding the computer questions cannot be determined from our data. However, low rates of detection of obsessive-compulsive disorder and social phobia have traditionally been a problem (33,34). Both disorders cause substantial functional impairment (35,36), and detection of their comorbidity with other diagnosed disorders would result in more efficacious treatment planning. Patients generally feel ashamed of these disorders and often make substantial efforts to hide their symptoms from others. In our study, patients seemed to be more willing to disclose symptoms of these disorders to the computer than to a clinician, although it may be that clinicians failed to ask about symptoms of these disorders.

Overall, the computer versions of PRIME-MD worked well. Prevalence rates for any diagnosis, for any affective disorder, and for any anxiety disorder obtained by the computer interviews were similar to those obtained by the SCID-IV. Prevalence rates for individual diagnoses were also similar, with the exception of dysthymia and obsessive-compulsive disorder, both of which were found to be more prevalent by the computer, and panic disorder, which was found to be more prevalent by the SCID-IV.

Although lower clinician-detected prevalence rates (chart diagnoses) for obsessive-compulsive disorder may be expected without a struc-

tured interview, it is unclear why the rates found by the SCID-IV were still lower than those obtained by the computer interviews. Helzer and colleagues (37) found that psychiatrists diagnosed lower rates of this disorder than did lay interviewers who used the Diagnostic Interview Schedule, a structured clinical interview. The computer interviews may be overdiagnosing obsessive-compulsive disorder, or patients may feel uncomfortable discussing this disorder with a clinician. Both Rasmussen and Tsuang (33) and Pollitt (38) found that an average of 7.5 years elapsed between onset of obsessive-compulsive disorder and first psychiatric attention. Another possibility is that PRIME-MD overdiagnoses obsessive-compulsive disorder because it has a new module for this disorder that has not been previously validated.

As noted, underdiagnosis of dysthymia was found to be a common error in clinical practice (2), often as a result of failure to make the diagnosis when major depression is superimposed. In a previous study clinicians failed to give a diagnosis of dysthymia when it was warranted on the clinician-administered PRIME-MD because of incorrect application of the scoring algorithm in 4 percent of cases (Kobak KA, Taylor LvH, Dottl SL, et al., unpublished manuscript, 1996).

Diagnostic accuracy using the SCID-IV as the "gold standard" was high for both the IVR and the desktop versions of PRIME-MD. Sensitivity, specificity, positive predictive value, and overall accuracy rates were high for most diagnoses, and the accuracy rates were similar to those obtained by clinician-administered primary care screening interviews, such as the Symptom-Driven Diagnostic System for Primary Care (22), the Composite International Diagnostic Interview (23), and the clinician-administered PRIME-MD (24). Problems with poor positive predictive value (patients screening positive who did not have the disorder) that have plagued past screeners (39) have been largely overcome with the computer PRIME-MD.

The computer-administered versions of PRIME-MD are also time-efficient—for example, even though patients had a mean of 4.4 diagnoses, the mean interview time was only 7.4 minutes on the desktop version and 8.6 minutes on the IVR version. Furthermore, IVR administration makes the evaluation available even in sites without access to computer hardware and permits evaluation 24 hours a day from any location with a touch-tone phone.

Although patients were free to ask for help at any time, all patients were able to complete the desktop interview unassisted once the program was started. Four patients had difficulty using the IVR system; two had to hang up before the interview was completed and were assigned a new identification number to re-enter the system; one person called the system before her scheduled time and thus was unable to enter the system at the appropriate time; and one person had technical difficulties and was unable to complete the interview.

Both computer-administered interviews were well accepted by patients, with no difference between the computer and the clinician in how well patients liked being interviewed by each interview form or in how well they were able to understand the questions. Although patients said that the clinician was better able to understand their feelings, they also felt more embarrassed with the clinician interview than with either computer interview. As in a previous study (13), when asked to choose, psychiatric patients preferred being interviewed by a clinician, although no significant difference was found in how well patients liked each interview, and no patient refused to do any interview. Patients' preferences may vary by diagnosis; those with certain disorders such as social phobia may prefer to be interviewed by computer (14).

A significantly greater percentage of the patients preferred desktop PRIME-MD to IVR PRIME-MD; only 2.7 percent of the patients preferred the IVR version to the other interviews. Several factors may have influenced this preference. Auditory presentation of the information places a heavier reliance on short-term memory during response selection. The desktop interview permits patients to quickly refer to the monitor for clarification before entering their responses on the keyboard. Although patients can choose to have IVR-administered items repeated, this process is time consuming and may be tedious for some patients.

Engineering psychologists have found that auditory presentation of information, such as by a clinician and in an IVR interview, are more compatible with verbal responses, while visual presentation of information, such as in the desktop version, are more compatible with manual responses (40). Thus the current IVR PRIME-MD requires a less compatible manual response to the auditorally presented questions. Speaker-independent speech recognition technology would allow patients to respond verbally to IVR-administered questions using a limited vocabulary and might enhance the acceptability of IVR assessments. In addition, greater familiarity with the interview method through use over time and changes in the program allowing greater control over the timing of questions and responses may increase preference for this form of administration in the future (41).

One limitation of the study reported here is the lack of patients with a diagnosis of psychosis. It is unclear whether psychotic patients would be incorrectly diagnosed as having the disorders included in the computer-administered PRIME-MD because the program's algorithm does not contain exclusionary criteria for psychosis. This limitation may make its use with psychotic patients problematic.

Overall, the high levels of sensitivity, specificity, and accuracy found in this study support the use of both desktop and IVR PRIME-MD as a valid and reliable instrument for clinicians gathering information from mental health patients. Although the interview was originally designed for use with a primary care population, results of the study support the utility of the computer-administered versions of the PRIME-MD in a mental health population.

The computer-administered ver-

sions of PRIME-MD provide a standardized, cost-effective, and time-efficient source of useful diagnostic information to clinicians. By increasing the thoroughness and accuracy of the diagnostic evaluation, PRIME-MD offers better treatment planning to patients, which in turn increases the quality of patient care. ♦

Acknowledgments

The authors acknowledge the contributions of Amy James, who conducted the SCID-IV interviews; Bill Gaber, who wrote the computer data extract programs; Randal Norton, B.A., who was system administrator for the PRIME-MD program; Toni Lundeen, for assisting with study patients; Jim Mundt, Ph.D., for his helpful suggestions and input; Keith Wenzel, B.S., and Chris Spierings, B.S., who developed the software used to deliver the IVR PRIME-MD; and Robert L. Spitzer, M.D., Janet B. Williams, D.S.W., and Jean Endicott, Ph.D., who were consultants to the project. This project was partly funded by a grant from Pfizer, Inc.

References

1. Kobak KA, Greist JH, Jefferson JW, et al: Decision support for patient care: computer-administered rating scales. Behavioral Healthcare Tomorrow 5(1):25–29, 1996
2. Skodol AE, Williams JBW, Spitzer RL, et al: Identifying common errors in the use of DSM-III through supervision. Hospital and Community Psychiatry 35:251–255, 1984
3. Spitzer RL, Skodol AE, Williams JBW, et al: Supervising intake diagnosis: a psychiatric "Rashomon." Archives of General Psychiatry 39:1299–1305, 1982
4. Rubinson EP, Asnis GM, Friedman JM: Knowledge of the diagnostic criteria for major depression: a survey of mental health professionals. Journal of Nervous and Mental Disease 176:480–484, 1988
5. First MB, Spitzer RL, Gibbon M, et al: Structured Clinical Interview for DSM-IV Axis I Disorders—Patient Edition (SCID-I/P, Version 2.0). New York, New York State Psychiatric Institute, 1995
6. Endicott J, Spitzer RL: A diagnostic interview: the Schedule for Affective Disorders and Schizophrenia. Archives of General Psychiatry 35:837–844, 1978
7. Fairbairn AS, Wood CH, Fletcher CM: Variability in answers to a questionnaire on respiratory symptoms. British Journal of Preventive Social Medicine 13:175–193, 1959
8. Simmons EM, Miller OW: Automated patient history-taking. Hospitals 45:56–59, 1971
9. Greist JH, Gustafson DH, Stauss FF, et al: Suicide risk prediction: a new approach. Life Threatening Behavior 4:212–223, 1974
10. Lucas RW, Mullins PJ, Luna CB, et al: Psychiatrists and a computer as interrogators of patients with alcohol-related illnesses: a comparison. British Journal of Psychiatry 131:160–167, 1977
11. Locke SE, Kowaloff HB, Hoff RG, et al: Computer-based interview for screening blood donors for risk of HIV transmission. JAMA 268:1301–1305, 1992
12. Carr AC, Ghosh A, Ancill RJ: Can a computer take a psychiatric history? Psychological Medicine 13:151–158, 1983
13. Kobak KA, Reynolds WM, Greist JH: Computerized and clinician assessment of depression and anxiety: respondent evaluation and satisfaction. Journal of Personality Assessment 63:173–180, 1994
14. Katzelnick DJ, Kobak KA, Greist JH, et al: Sertraline in social phobia: a double-blind, placebo-controlled cross-over study. American Journal of Psychiatry 152:1368–1371, 1995
15. Petrie K, Abell W: Responses of parasuicides to a computerized interview. Computers in Human Behavior 10:415–418, 1994
16. Slack WV, Hicks GP, Reed CE, et al: A computer-based medical history system. New England Journal of Medicine 274:194–198, 1966
17. Kobak KA, Reynolds WM, Rosenfeld R, et al: Development and validation of a computer-administered version of the Hamilton Depression Rating Scale. Psychological Assessment 2:56–63, 1990
18. Rosenfeld R, Dar R, Anderson D, et al: A computer-administered version of the Yale-Brown Obsessive Compulsive Scale. Psychological Assessment 4:329–332, 1992
19. Kobak KA, Reynolds WM, Greist JH: Development and validation of a computer-administered version of the Hamilton Anxiety Scale. Psychological Assessment 5:487–492, 1993
20. Kobak KA, Greist JH, Jefferson JW, et al: Computerized assessment in clinical drug trials. Psychopharmacology Bulletin 32: 464, 1996
21. First MB, Gibbon M, Williams JBW, et al: Mini-SCID, Version 2. North Tonawanda, NY, Multi-Health Systems, 1995
22. Weissman MM, Olfson M, Leon AC, et al: Brief diagnostic interview (SDDS-PC) for multiple mental disorders in primary care: a pilot study. Archives of Family Medicine 4:220–227, 1995
23. Peters L, Andrews G: Procedural validity of the computerized version of the Composite International Diagnostic Interview (CIDI-Auto) in the anxiety disorders. Psychological Medicine 25:1269–1280, 1995
24. Spitzer RL, Williams JBW, Kroenke K, et al: Utility of a new procedure for diagnosing mental disorders in primary care: the PRIME-MD 1000 study. JAMA 272:1749–1756, 1994
25. Greist JH, Jefferson JW, Wenzel K, et al: Telephone-Administered Computer Assisted PRIME-MD (computer program). Madison, Wis, Dean Foundation, 1995
26. Hamilton M: A rating scale for depression. Journal of Neurology, Neurosurgery, and Psychiatry 23:56–62, 1960
27. Potts MK, Daniels M, Burnam MA, et al: A structured interview version of the Hamilton Depression Rating Scale: evidence of reliability and versatility of administration. Journal of Psychiatric Research 24:335–350, 1990
28. Simon GE, Revicki D, VonKorf M: Telephone assessment of depression severity. Journal of Psychiatric Research 27:247–252, 1993
29. Aneshensel CS, Frerichs RR, Clark VA, et al: Measuring depression in the community: a comparison of telephone and personal interviews. Public Opinion Quarterly 46: 110–121, 1982
30. Aneshensel CS, Frerichs RR, Clark VA, et al: Telephone versus in-person surveys of community health status. American Journal of Public Health 72:1017–1021, 1982
31. Wells KB, Burnam MA, Leake B, et al: Agreement between face-to-face and telephone-administered versions of the depression section of the NIMH Diagnostic Interview Schedule. Journal of Psychiatric Research 22:207–220, 1988
32. Spitzer RL: Psychiatric diagnosis: are clinicians still necessary? in Psychotherapy Research: Where Are We and Where Should We Go? Edited by Williams JBW, Spitzer RL. New York, Guilford, 1984
33. Rasmussen SA, Tsuang MT: Epidemiology of obsessive compulsive disorder: a review. Journal of Clinical Psychiatry 45:450–457, 1984
34. Liebowitz MR, Gorman JM, Fyer AJ, et al: Social phobia: review of a neglected anxiety disorder. Archives of General Psychiatry 42: 726–729, 1985
35. Rachman SJ, Hodgson RJ: Obsessions and Compulsions. Englewood Cliffs, NJ, Prentice-Hall, 1980
36. Schneier FR, Heckelman LR, Garfinkel R, et al: Functional impairment in social phobia. Journal of Clinical Psychiatry 55:322–331, 1994
37. Helzer JE, Robins LN, McEvoy LT, et al: A comparison of clinical and diagnostic interview schedule diagnoses: physician reexamination of lay-interviewed cases in the general population. Archives of General Psychiatry 42:657–666, 1985
38. Pollitt J: Natural history of obsessional states. British Journal of Medicine 1:194–198, 1957
39. Campbell TL: Is screening for mental health problems worthwhile in family practice? An opposing view. Journal of Family Practice 25:184–187, 1987
40. Wicken CD: Engineering Psychology and Human Performance. Columbus, Ohio, Merrill, 1984
41. Baer L, Brown-Beasley MW, Sorce J, et al: Computer-assisted telephone administration of a structured interview for obsessive-compulsive disorder. American Journal of Psychiatry 150:1737–1738, 1993