

Feasibility and Diagnostic Validity of the M-3 Checklist: A Brief, Self-Rated Screen for Depressive, Bipolar, Anxiety, and Post-Traumatic Stress Disorders in Primary Care

Bradley N. Gaynes, MD, MPH¹

Joanne DeVeaugh-Geiss, MA, LPA¹

Sam Weir, MD²

Hongbin Gu, PhD¹

Cora MacPherson, PhD³

Herbert C. Schulberg, PhD, MSHyg⁴

Larry Culpepper, MD, MPH⁵

David R. Rubinow, MD¹

¹Department of Psychiatry, University of North Carolina School of Medicine, Chapel Hill, North Carolina

²Department of Family Medicine, University of North Carolina School of Medicine, Chapel Hill, North Carolina

³Social & Scientific Systems, Inc, Silver Spring, Maryland

⁴Department of Psychiatry, Weill Medical College, Cornell University, White Plains, New York

⁵Department of Family Medicine, Boston University School of Medicine, Boston, Massachusetts



MORE ONLINE
www.annfammed.org

Conflicts of interest: listed at the end of this article

CORRESPONDING AUTHOR

Bradley N. Gaynes, MD, MPH
Department of Psychiatry
CB #7160, University of North Carolina
School of Medicine
Chapel Hill, NC 27599-7160
bgaynes@med.unc.edu

ABSTRACT

PURPOSE Mood and anxiety disorders are the most common psychiatric conditions seen in primary care, yet they remain undetected and undertreated. Screening tools can improve detection, but available instruments are limited by the number of disorders assessed. We wanted to assess the feasibility and diagnostic validity of the My Mood Monitor (M-3) checklist, a new, 1-page, patient-rated, 27-item tool developed to screen for multiple psychiatric disorders in primary care.

METHODS We enrolled a sample of 647 consecutive participants aged 18 years and older who were seeking primary care at an academic family medicine clinic between July 2007 and February 2008. We used a 2-step scoring procedure to make screening more efficient. The main outcomes measured were the sensitivity and specificity of the M-3 for major depression, bipolar disorder, any anxiety disorder, and post-traumatic stress disorder (PTSD), a specific type of anxiety disorder. Using a split sample technique, analysis proceeded from determination of optimal screening thresholds to assessment of the psychometric properties of the self-report instrument using the determined thresholds. We used the Mini International Neuropsychiatric Interview as the diagnostic standard. Feasibility was assessed with patient and physician exit questionnaires.

RESULTS The depression module had a sensitivity of 0.84 and a specificity of 0.80. The bipolar module had a sensitivity of 0.88, and a specificity of 0.70. The anxiety module had a sensitivity of 0.82 and a specificity of 0.78, and the PTSD module had a sensitivity of 0.88 and a specificity of 0.76. As a screen for any psychiatric disorder, sensitivity was 0.83 and specificity was 0.76. Patients took less than 5 minutes to complete the M-3 in the waiting room, and less than 1% reported not having time to complete it. Eighty-three percent of clinicians reviewed the checklist in 30 or fewer seconds, and 80% thought it was helpful in reviewing patients' emotional health.

CONCLUSIONS The M-3 demonstrates utility as a valid, efficient, and feasible tool for screening multiple common psychiatric illnesses, including bipolar disorder and PTSD, in primary care. Its diagnostic accuracy equals that of currently used single-disorder screens and has the additional benefit of being combined into a 1-page tool. The M-3 potentially can reduce missed psychiatric diagnoses and facilitate proper treatment of identified cases.

Ann Fam Med 2010;8:160-169. doi: 10.1370/afm.1092.

INTRODUCTION

Psychiatric illness is common in primary care settings, where mood and anxiety disorders are the 2 psychiatric disorders most frequently encountered.^{1,2} Although primary care physicians account for most mental health visits and write the bulk of antidepressant and anti-anxiety

prescriptions for mood and anxiety disorders in the United States,³⁻⁵ underrecognition and inadequate treatment of these disorders in primary care practices remain substantial concerns.⁵⁻⁸ Tools to improve the identification and management of these disorders are therefore being developed to address this problem.⁹⁻¹¹

Despite these efforts, available tools narrowly focus on identifying either depressive or anxiety disorders and provide little guidance for management. Most screening tools target unipolar depression (16.1% prevalence),¹² and they do not help to differentiate unipolar from bipolar spectrum illness. The latter has a 3.9% prevalence in community settings,¹³ but it can be as high as 9.8% in primary care.¹⁴ A diagnostic error can lead to improper treatment of patients with bipolar depression; prescribing an antidepressant without a mood stabilizer potentially destabilizes the illness and increases the risk of a hypomanic, manic, or mixed episode.¹⁵⁻¹⁷ Indeed, bipolar patients seen in primary care appear to be at a particularly increased risk of inappropriate treatment.¹⁸

Also of concern is that depression-screening instruments often do not address anxiety syndromes (28% prevalence),¹² including generalized anxiety disorder, social anxiety disorder, panic disorder, obsessive-compulsive disorder, and post-traumatic stress disorder (PTSD). PTSD, in particular, is not only highly prevalent among those returning from service in the US Armed Forces,¹⁹ but it is also quite common in the general primary care population, where it is often underrecognized.²⁰ Coexisting anxiety disorders are associated with a more treatment-resistant depressive course, suggesting the need for more aggressive treatment or earlier referral to a mental health professional. Finally, available anxiety screens²¹ singularly address anxiety rather than the full spectrum of mood and anxiety disorders.

Given the limitations of currently available assessment instruments, we now report on the feasibility and diagnostic validity of the My Mood Monitor (M-3) checklist, a new, 1-page, patient-rated, 27-item tool that screens for 4 types of psychiatric disorders encountered in primary care. Our experience in developing and testing this instrument, and in determining the most appropriate means of scoring it, may provide useful information on the methods surrounding such instrument development. Specifically, this report addresses 3 questions:

1. What are the preferred M-3 screening thresholds for identifying psychiatric illness in primary care?
2. In a prospective testing of these thresholds, what are the psychometric characteristics of the M-3 checklist?
3. What is the feasibility of completing the M-3 in

a primary care setting from the patient and the clinician perspectives?

METHODS

In this study, we used a cross-sectional design enrolling a convenience sample of consecutive adult patients visiting a primary care clinic. After giving informed consent, participants completed a self-report symptom checklist in the waiting room. At the end of the clinic visits, participants and their physicians completed questionnaires assessing the feasibility of the symptom self-report. Within 30 days of the index visit, a research assistant administered the Mini International Neuropsychiatric Interview²² (MINI) to participating patients by telephone. Using a split sample technique, analysis proceeded from determining optimal screening thresholds to assessing psychometric properties of the self-report instrument using the determined thresholds and the MINI as the diagnostic standard.

This study was approved by the University of North Carolina Institutional Review Board.

Study Population and Sample

The study population comprised all patients visiting the Family Medicine Clinic at the University of North Carolina between July 2007 and February 2008 who were aged at least 18 years, English speaking, and mentally competent to provide informed consent. This clinic, staffed by 55 clinicians, saw approximately 18,000 patients per year, with a mean age of 45.7 years, 60% of whom were female. Nearly two-thirds of clinic patients were white (63%); the remainder identified themselves either as African American (30%) or as Native American, Asian, or other (7%).

Instruments

Study Questionnaire

Questions were generated by a group of experienced mental health clinicians and researchers and were specifically intended for use in primary care settings. The M-3 (Supplemental Figure, <http://www.annfam.org/cgi/content/full/8/2/160/DC1>) is a 23-item self-report symptom checklist that inquires whether during the past 2 weeks the patient experienced symptoms of major depressive disorder (7 questions), generalized anxiety disorder (2 questions), panic disorder (2 questions), social anxiety disorder (1 question), PTSD (4 questions), and obsessive compulsive disorder (3 questions). The M-3 also inquires about a lifetime history of symptoms of bipolar spectrum disorder (4 questions). At the end of the symptom checklist, the M-3 poses 4 functional impairment questions. Patient responses to each of the 27 questions can



range from 0 ("not at all") to 4 ("most of the time"). The wording and grammar of the M-3 place it at a sixth-grade reading level. The interactive Web site for the M-3 can be found at <http://www.mymoodmonitor.com/>.

Reference Standard

The Mini International Neuropsychiatric Interview (MINI), a reliable and valid diagnostic instrument,²²⁻²⁴ served as the reference standard to evaluate the performance of the M-3. The MINI is widely used,^{25,26} well accepted,²⁷ and validated in general medical settings.^{28,29} It has good concordance with other diagnostic measures²⁴ and can be administered by telephone when in-person interviews are not feasible or practical.³⁰⁻³⁷ We used the MINI to identify depressive disorders, bipolar spectrum disorders, and anxiety disorders, as categorized in the Analysis section below.

Study Procedures

A research assistant approached consecutive patients entering the Family Medicine Center's waiting area. To avoid sampling bias, the assistant approached a maximum of 3 consecutive patients of each attending clinician. To avoid overrepresentation of participants with mental health concerns, potential participants were invited to complete a general health survey for their clinician's use in guiding their health care management. Before the clinician visit, participants completed the M-3 checklist and returned it to the practice nurse, who attached the checklist to the top of the chart for review by the clinician before entering the examination room. The clinician was instructed to review the checklist with the patient and to use the information however he or she wished.

After the appointment ended, the clinician answered a brief exit questionnaire regarding the ease and usefulness of the M-3. Participants completed a similar exit questionnaire surveying its feasibility. We collected demographic data using a separate form and used record abstraction to obtain missing data.

Experienced master's level diagnostic interviewers blinded to M-3 results administered the MINI, either in person immediately after the clinic visit or as soon as possible by telephone but within 30 days after the index visit. The interviewers assigned final diagnoses after reviewing each interview with a psychiatrist (B.N.G.) blinded to M-3 results.

Analysis

Categorizing the Diagnoses

MINI diagnoses were sorted into 3 categories: depressive disorder, which includes major depressive disorder or dysthymia in the absence of major depression; bipolar disorder (type 1 or type 2), which could include a

current or past episodes; and anxiety disorders, which include generalized anxiety disorder, panic disorder, social anxiety disorder, obsessive compulsive disorder, agoraphobia, and PTSD. Participants experiencing an anxiety disorder were further assessed specifically for PTSD. Participants lacking an M-3 checklist or MINI diagnosis were excluded from the analysis. Each unanswered M-3 question was assigned a value of 0.

M-3 Scoring

Appreciating that false-positive findings can increase the workload of primary care physicians and lead to improper treatment of the patient, and that the *Diagnostic and Statistical Manual of Mental Disorders*³⁸ requires functional impairment for diagnosing psychiatric disorders, we used the functional impairment questions of the M-3 as a first-stage screen. The remaining checklist symptoms were then scored for only those patients whose screen was positive for functional impairment. This gateway method provided the best balance of increasing sensitivity and specificity while permitting a quick, visually intuitive method for scoring by hand.

The M-3 gateway method is a 2-step scoring procedure. First, the 4 lifestyle or functional impairment questions (questions 24 through 27) and the suicide question (question 5) were scored in the following manner. Participants responding negatively to the suicide question and who indicated their symptoms did not affect their lifestyle (ie, no single lifestyle question scored as "often" or "most of the time," and no more than 1 question scored "sometimes") were given a score of 0, and no further scoring was conducted. For all other participants passing through the "gate" and at risk for experiencing a psychiatric episode, subscores of the 4 diagnostic categories were summed as follows: responses of "not at all" and "rarely" were scored as 0; "sometimes" as 1; and "often" and "most of the time" as 2. After collapsing response categories from 5 to 3, a sensitivity analysis showed that there was no loss in sensitivity or specificity.

To determine optimal screening threshold cut-points for each diagnosis, we calculated the sensitivity and specificity of each score using data obtained from the first 80% (n = 525) of the cohort. We then calculated the threshold identified by both the Youden index³⁹ and the (0, 1) method by Holmes.⁴⁰ The Youden index, a commonly used measure of diagnostic effectiveness, optimizes and gives equal weight to sensitivity and specificity. The (0, 1) method by Holmes also weighs sensitivity and specificity equally but minimizes the distance between the receiver operating characteristics curve and the point (0, 1). These 2 methods are especially suitable for determining primary care cut-points of mental health measures, because they are least

dependent on population prevalence.⁴¹ For depression, anxiety, and PTSD, the Youden index³⁹ and (0, 1) method⁴⁰ resulted in the same optimal cut-point. For bipolar disorder, the 2 methods resulted in different cut-points, and their midpoint was used for the M-3 screening threshold (rather than 1 or 3, we chose 2).

After determining the optimal screening thresholds for sensitivity and specificity based on the initial 80% of cohort participants, we cross-validated these results with data from the remaining 20% ($n = 122$) of the cohort. For each diagnosis, the sensitivity and specificity resembled or improved upon those established with the original cohort. We therefore report results of this gateway method for the combined cohort.

Feasibility

We conducted descriptive analyses of the patient and physician exit questionnaire data. These analyses included the proportion of patients who reported discussing the M-3 checklist with their physicians, the duration of the discussion, and the proportion of patients perceiving the checklist as helpful to the clinical discussion. From physician responses to the exit questionnaire, we report the proportion who reviewed the M-3 form and the average time spent doing so, reasons for not reviewing the M-3, and its usefulness in clarifying the patient's emotional state.

RESULTS

A total of 723 patients (54% of all patients approached) consented to study participation. Although not systematically recorded, the most common reasons for nonparticipation were disinterest in filling out a checklist and anticipated difficulty in scheduling a follow-up interview. Of the 723 study participants, 15 lacked the M-3 form, 59 did not complete the MINI, and 2 were missing both, leaving analyzable data for 647. Compared with this cohort, patients with a missing M-3 screen or MINI were similar with regard to age, race, sex, and income level.

MINI interviews were successfully completed within 14 days of the index visit with 81% (523) of participants and within 30 days with 99% (639) of the participants. The mean number of days between administration of the M-3 and MINI for the full group was 8.8 days (SD 8.2 days). Because an analysis stratified by whether the MINI was administered within 2 weeks or 3 to 4 weeks of completion of the M-3 found no difference in study results between the 2 periods, we report analyses of the results of both periods combined. Excluding the 8 patients whose MINI interviews were conducted after 30 days did not affect study results, so the following analyses include these 8 patients, also.

Sample Characteristics

Table 1 displays baseline characteristics by type of psychiatric disorder. The study sample appeared representative of patients seen in this family medicine practice. Relative to the clinic's general population, participants were similar in age, race/ethnicity, and general income level; however, they were more likely to be female (71% vs 60%, $P < .001$). Fifty-four of the 55 practicing clinic physicians participated in the study, and they ranged in experience from second-year family practice residents to senior faculty.

MINI Mental Illness Prevalence

With regard to psychiatric diagnoses assigned to study participants by the MINI, 22% had a depressive disorder (16% had major depressive disorder, while 6% had bipolar depression), 9.3% had bipolar spectrum illness, 28.1% had an anxiety disorder, and 6.3% had PTSD (Table 2). Overall, 35% of study participants met MINI criteria for at least 1 psychiatric diagnosis. Because all 105 participants whose scores were positive for the MINI depressive disorder diagnoses met MINI criteria for a major depressive disorder and not dysthymia alone, these participants are reported as major depressive disorder. Of the 60 participants with a lifetime diagnosis of bipolar disorder, 8 currently met criteria for mania or hypomania (13%), and 37 were currently depressed (62% of participants with bipolar disorder and 5.7% of the sample, Figure 1).

Anxiety disorders were the most common psychiatric diagnosis among study participants. The most common subtype was generalized anxiety disorder ($n = 117$, 18.1%), followed by agoraphobia ($n = 96$, 14.8%), panic disorder ($n = 73$, 11.3%) and social phobia ($n = 52$, 8.0%). Forty-one patients (6.3%) had PTSD diagnosed. In this sample, obsessive-compulsive disorder was rare ($n = 21$, 3.3%).

Psychiatric comorbidities were common in this population (Figure 1). More participants had comorbid anxiety and depression (12.1%) than either major depression (4.2%) or anxiety (9.1%) alone.

M-3 and the 2-Step Scoring Method

The gateway method served as an efficient scoring strategy. The first-step screen for functional impairment eliminated 349 (53.9%) of 647 participants from the scoring process, 38 (10.9%) of whom nevertheless met MINI criteria for a psychiatric diagnosis. Twenty-three (6.6%) of the 349 participants who stopped at the functionality gate scored all 0s on the M-3; none had a psychiatric diagnosis. Of the 298 participants passing through the gate and assessed with the M-3, 186 (62.4%) had a psychiatric diagnosis. Participants who passed through the gate were nearly 6 times more

Table 1. Baseline Characteristics of the Study Sample

| Characteristic | Diagnosis by MINI | | | | | Total N = 647 |
|------------------------------------------|-----------------------------------|---------------------------------------------------|-------------------------------------------------|-----------------------------------------|---------------------------------------------------|------------------------|
| | No Disorder n = 423 (65.4%) | Depression Without Anxiety n = 27 (4.2%) | Depression With Anxiety n = 78 (12.1%) | Bipolar Disorder n = 60 (9.3%) | Anxiety Without Depression n = 59 (9.1%) | |
| Mean age (SD) [range], y | 46.4 (16.6) [18-92] | 44.9 (14.6) [19-70] | 45.0 (12.4) [19-70] | 41.2 (10.7) [18-63] | 41.0 (13.8) [21-72] | 45.2 (15.4) [18-92] |
| Women, % | 69.5 | 85.2 | 71.8 | 78.3 | 66.1 | 70.9 |
| Race, % | | | | | | |
| White | 67.4 | 59.3 | 65.4 | 60.0 | 76.3 | 66.9 |
| Black | 27.4 | 40.7 | 32.1 | 35.0 | 18.6 | 28.4 |
| Other | 5.2 | 0 | 2.5 | 5.0 | 5.1 | 4.6 |
| Marital status, % ^a | | | | | | |
| Married | 53.8 | 29.2 | 40.0 | 26.0 | 46.0 | 47.8 |
| Single | 26.2 | 29.2 | 24.6 | 38.0 | 30.0 | 27.6 |
| Divorced | 9.6 | 12.5 | 21.5 | 14.0 | 12.0 | 11.8 |
| Separated | 1.4 | 12.5 | 10.8 | 2.0 | 2.0 | 3.1 |
| Widowed | 4.2 | 8.3 | 1.5 | 4.0 | 2.0 | 3.9 |
| Living with partner | 4.8 | 8.3 | 1.5 | 16.0 | 8.0 | 5.9 |
| Education, % ^a | | | | | | |
| Without high school | 2.9 | 4.4 | 6.1 | 6.3 | 4.1 | 3.8 |
| Diploma or GED | 30.6 | 47.8 | 48.5 | 54.2 | 36.7 | 36.2 |
| High school diploma/GED | 13.5 | 26.1 | 18.2 | 6.3 | 20.4 | 14.6 |
| Associate/technical degree | 29.4 | 13.0 | 18.2 | 27.1 | 26.5 | 26.8 |
| College diploma | 23.6 | 8.7 | 9.1 | 6.3 | 12.2 | 18.6 |
| Graduate degree | | | | | | |
| Household's gross income, % ^a | | | | | | |
| ≤\$14,999 | 10.9 | 17.4 | 31.8 | 50.0 | 14.3 | 17.6 |
| \$15,000-39,999 | 24.6 | 39.1 | 31.8 | 27.1 | 26.5 | 26.5 |
| \$40,000-59,999 | 11.8 | 8.7 | 1.5 | 8.3 | 20.4 | 10.8 |
| ≥60,000 | 40.4 | 8.7 | 19.7 | 4.2 | 28.6 | 32.2 |
| Not know/refused | 12.3 | 26.1 | 15.2 | 10.4 | 10.2 | 12.9 |
| Employment status, % ^a | | | | | | |
| Unemployed | 13.8 | 43.5 | 35.4 | 50.0 | 18.4 | 21.4 |
| Employed | 69.3 | 47.8 | 47.7 | 45.8 | 71.4 | 63.8 |
| Retired, not working | 17.0 | 8.7 | 16.9 | 4.2 | 10.2 | 14.8 |

GED = general equivalency diploma; MINI = Mini International Neuropsychiatric Interview.

^a Employment status, marital status, education, and income were not collected on the first 99 participants. For all variables, percentages were calculated for participants with available data.

likely to have a psychiatric diagnosis than those who did not (62.4% vs 10.9%, $P < .001$).

Psychometrics of M-3, Specific Diagnoses

Table 2 specifies the screening cut-off score pertinent to each diagnosis (eg, a score of 2 or more indicates a positive screen for bipolar disorder), and the M-3 questions specific to the diagnosis (eg, bipolar disorder questions 20 through 23). When compared with the MINI-generated diagnosis, the M-3 depression module had a sensitivity of 0.84 (95% confidence interval [CI], 0.77-0.89) and a specificity of 0.80 (95% CI, 0.76-0.83). For this module, the positive likelihood ratio was 4.19, indicating that a positive screen was more than 4 times as likely to come from a patient with a depres-

sive disorder than from one without it. Further, given a 16% prevalence of depression in our population (or an odds of about 1 in 6), a patient with a positive screen had a post-test odds for depression of approximately 4:6, or 40%.

The M-3 bipolar module had a somewhat higher sensitivity (0.88; 95% CI, 0.77-0.95) but a lower specificity (0.70; 95% CI, 0.66-0.74). The anxiety module had a sensitivity of 0.82 (95% CI, 0.75-0.87) and a specificity of 0.78 (95% CI, 0.74-0.81), whereas the PTSD module had a sensitivity of 0.88 (95% CI, 0.74-0.96) and a specificity of 0.76 (95% CI, 0.73-0.80). A sensitivity analysis was performed to determine whether missing data influenced the results. When only patients with complete data were included in the

Table 2. M-3 Psychometrics for Specific Diagnoses and for Any Diagnosis by MINI (n = 647)

| Test Result | Diagnosis by MINI | | | | |
|-------------------------------------------------------|-------------------------------|-------------------------------|-------------------------------|------------------------------|----------------------------------|
| | Depression (n = 142) | Bipolar (n = 60) | Anxiety (n = 182) | PTSD (n = 41) | Any Diagnosis (n = 224) |
| M-3 subscore cutoff (≥) | 5 | 2 | 3 | 2 | Any positive screen ^a |
| Question No. | 1-7 | 20-23 | 8-19 | 13-16 | 1-23 |
| Sensitivity (95% CI) | 0.84 (0.77-0.89) | 0.88 (0.77-0.95) | 0.82 (0.75-0.87) | 0.88 (0.74-0.96) | 0.83 (0.77-0.88) |
| Specificity (95% CI) | 0.80 (0.76-0.83) | 0.70 (0.66-0.74) | 0.78 (0.74-0.81) | 0.76 (0.73-0.80) | 0.76 (0.72-0.80) |
| Positive LR (95% CI) | 4.19 (3.47-5.06) | 2.94 (2.53-3.44) | 3.65 (3.05-4.39) | 3.69 (3.08-4.44) | 3.48 (2.90-4.16) |
| Negative LR (95% CI) | 0.20 (0.14-0.29) | 0.17 (0.08-0.33) | 0.23 (0.17-0.32) | 0.16 (0.07-0.36) | 0.22 (0.17-0.30) |
| Positive M-3 screen, % | 34 | 35 | 39 | 28 | 44 |
| Diagnosed by MINI, % | 22 | 9 | 28 | 6 | 35 |
| Positive predictive value (95% CI) [n/n] ^b | 0.54 (0.47-0.61) [119/220] | 0.23 (0.18-0.29) [53/229] | 0.59 (0.53-0.65) [149/253] | 0.20 (0.15-0.27) [36/180] | 0.65 (0.59-0.70) [186/287] |
| Negative predictive value (95% CI) [n/n] ^c | 0.95 (0.92-0.96) [404/427] | 0.98 (0.96-0.99) [411/418] | 0.92 (0.88-0.94) [361/394] | 0.99 (0.97-1.0) [462/467] | 0.89 (0.86-0.92) [322/360] |

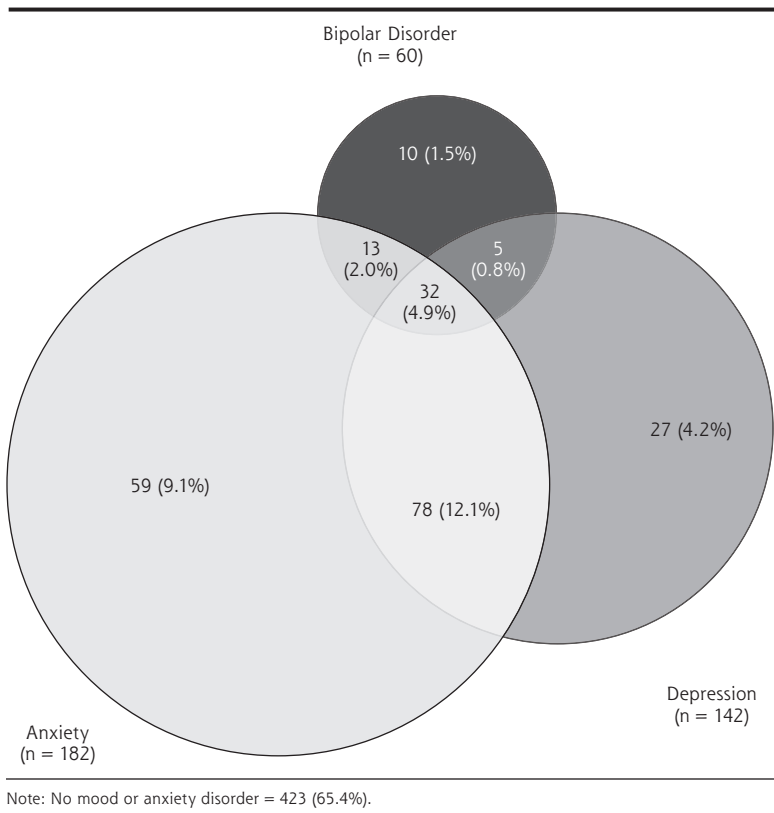
CI = confidence interval; LR = likelihood ratio; M-3 = My Mood Monitor checklist; MINI = Mini International Neuropsychiatric Interview; PTSD = post-traumatic stress disorder.

^a If any diagnosis cut-off score was met, the screen was considered positive.

^b Number with a positive screen and who have that diagnosis by MINI/total number with a positive screen by M-3.

^c Number with a negative screen and who do not have that diagnosis by MINI/total number with a negative screen by M-3.

Figure 1. Summary of 224 participants with a diagnosis by MINI.



screening instruments for depression, bipolar illness, PTSD, and anxiety (Table 3).

M-3 as a Screening Instrument for Any Mood or Anxiety Disorder

Given that the M-3 screens for both mood and anxiety disorders, we wondered whether a positive screen for any of the diagnostic categories could help identify any mood or anxiety disorder (even if different from what the positive screen would suggest). Of the 647 participants, 287 were positive (44%) on the M-3 checklist (Table 2). For all participants enrolled, sensitivity was 0.83 and specificity 0.76, with a positive likelihood ratio of 3.48 and a negative likelihood ratio of 0.22. Thus, as a general screen, the M-3 had a positive predictive value of 0.65 and a negative predictive value of 0.89 for any mood or anxiety disorder.

Feasibility

Participant Perspective

The M-3 checklist took less than 5 minutes to complete, and less than 1% of participants reported lacking sufficient time to complete it. Approximately 70% of participants reported talking to their clinician about mood or feelings, among those

analyses, the results were similar to the results obtained with the entire study population.

The sensitivity and specificity of the M-3 compared favorably with those of existing single-disorder

analyses, the results were similar to the results obtained with the entire study population.

Table 3. Comparison of Sensitivity and Specificity of M-3 With Reference Standard Instruments

| Outcome | PHQ-9 ⁴² Depression | Mood Disorder Questionnaire ⁴³ Bipolar | CAPS ⁴⁴ PTSD | GAD-7 ²¹ Anxiety |
|---------------------|-----------------------------------|---------------------------------------------------------|----------------------------|--------------------------------|
| Existing instrument | | | | |
| Sensitivity | 0.88 | 0.73 | 0.74 | 0.77 |
| Specificity | 0.88 | 0.90 | 0.84 | 0.82 |
| M-3 | | | | |
| Sensitivity | 0.84 | 0.88 | 0.88 | 0.82 |
| Specificity | 0.80 | 0.70 | 0.76 | 0.78 |

CAPS = Clinician-Administered PTSD Scale; GAD-7 = 7-item Generalized Anxiety Disorder scale; M-3 = My Mood Monitor checklist; PHQ-9 = 9-item Patient Health Questionnaire; PTSD = post-traumatic stress disorder.

Note: The sensitivity and specificity for any diagnosis by the M-3 (depression, bipolar, anxiety, or PTSD) was 0.83 and 0.76, respectively.

who did, 70% did so for at least 1 minute. Sixty-three percent of all participants reported that the M-3 helped them talk to their doctors about their mood or feelings. Among participants assigned a MINI diagnosis, 75% stated that the M-3 facilitated talking to their clinician about mood or feelings.

Clinician Perspective

Eighty-three percent of clinicians reviewed the checklist in 30 or fewer seconds. None found the M-3 too complicated, and 80% thought it was helpful in reviewing participants' emotional health.

DISCUSSION

The M-3 is a valid, efficient, and feasible 1-page tool for screening multiple common psychiatric illnesses in primary care and other settings. The psychometric properties of the M-3 compare favorably with those of currently used single-disorder screening tools (Table 3), thus permitting the M-3 to function both as a screen for specific anxiety and mood disorder diagnoses, as well as a screen for the presence of any mood or anxiety disorder (Table 2). Compared with other multi-condition screens, the M-3 has the additional benefit of integrating screening for bipolar disorder and PTSD while screening for other anxiety and mood disorders.

The extensive psychiatric comorbidity found in the primary care population underscores the need for clinicians to consider multiple psychiatric disorders rather than just depression or anxiety alone. In a study similar to ours with 100 patients, a primary care physician administering only the 9-item Patient Health Questionnaire⁴² (PHQ-9), which has a sensitivity of 0.88, would correctly identify 14 of the 16 depressed patients. The PHQ-9 would not identify the 9 patients experiencing anxiety alone, however, and would

misidentify 5 bipolar depressed patients as having a unipolar major depressive disorder. Similarly, physicians administering the 7-item Generalized Anxiety Disorder scale⁴⁵ (GAD-7) alone would capture 7 of 9 patients with an anxiety disorder alone but would miss approximately 20 patients with bipolar disorder or major depressive disorder. Given these diagnostic complexities, the M-3 potentially can reduce missed and misidentified psychiatric episodes and facilitate proper treatment of accurately identified disorders. Further,

the M-3 improves upon existing multiple-disorder instruments used in primary care practice by offering greater diagnostic specificity and by explicitly identifying those at risk for bipolar disorder and PTSD rather than identifying general diagnostic clusters or general levels of distress.⁴⁶⁻⁵⁰

In this initial study of the psychometric properties and utility of the M-3, we established preferred screening thresholds for specific psychiatric diagnoses using receiver operating characteristic curve analysis. We confirmed that the M-3 accurately screens for both specific psychiatric diagnoses (eg, major depressive disorder) and the presence of a psychiatric diagnosis in general (eg, either a mood or an anxiety disorder). Furthermore, both participants and clinicians found it easy to use, quick, and clinically helpful.

The 2-step method of scoring efficiently truncates the scoring process at the first step for most patients who report no functional impairment. For those reporting some degree of impairment, the second scoring step requires little time, and the clinician benefits from knowing that the total M-3 score has high predictive power for identifying psychiatric morbidity. Furthermore, if the clinician suspects that the patient is experiencing a psychiatric disorder, even though he or she denies such impairment, the M-3 data can guide a more detailed clinical interview.

As with all screening instruments, the M-3 seeks to identify efficiently those patients at high risk for 1 or more specific psychiatric conditions. Although the M-3 increases the likelihood of identifying a patient experiencing a psychiatric illness, the M-3 by itself is not a definitive diagnostic instrument. Indeed, as a screening tool, the M-3 screen was more likely to identify a risk of psychiatric illness than was confirmed by diagnostic interview. The clinician must fully investigate symptoms acknowledged by the patient to

confirm their diagnostic implications, including ruling out nonpsychiatric causes, such as physical illness, bereavement, or substance abuse. Nevertheless, the M-3 can facilitate the clinical assessment by identifying symptoms requiring fuller exploration and highlighting the patient's level of impairment.

Our study has several limitations. First, our sampling was conducted in a single family practice with a patient population and prevalence of psychiatric diagnoses that may not be representative of other primary care practices. Even so, the prevalences obtained in this study for anxiety disorders,²¹ PTSD,⁵¹ and bipolar disorder¹⁴ resemble those reported in the literature. The rate of major depression (16.2%) in our study, while higher than that reported for primary care settings in general,⁵² is consistent with the rate obtained in other studies of primary care settings,⁵³ including those conducted in low-income primary care settings.⁵⁴ Furthermore, the demographic characteristics of our participants correspond to those reported for US adult outpatients in the nationally representative samples assessed in the National Ambulatory Medical Care Survey.^{55,56}

As a second study limitation, we note that only 54% of the patients invited to participate did so, thereby raising the possibility of bias in that the participants could be either more or less likely to experience psychiatric symptoms. This participation rate exceeds the 50% criterion used in prior analyses of case-finding instruments⁵⁷ and is greater than some,^{43,44} but not all,^{45,46} recruitment rates reported in comparable studies validating other screening instruments. Further, our number of completed MINI reference standard interviews (647) is within the range of those performed in similar validation studies (eg, 585 with the PHQ-9,⁴⁶ and 965 reported with the GAD-7 scale⁴⁵).

A final study limitation is that faced by all studies of screening tools for psychiatric diagnoses, ie, the absence of a clear reference standard diagnostic test and the related evolving conceptualization of whether psychiatric research and clinical practice are best served by the current categorical approach (eg, normal vs disordered) or a dimensional approach (mild-severe) toward diagnoses.^{58,59} Although the current study cannot directly address these issues, the use of functional impairment by the M-3 as a necessary gateway toward positive screening, as well as its consideration of multiple symptoms, supports its use with either approach.

The M-3 demonstrates utility as a valid, efficient, and feasible tool for screening multiple common psychiatric illnesses, including bipolar disorder and PTSD, in primary care. Its screening accuracy equals that of currently used single-disorder screening instru-

ments and has the additional benefit of combining them into a 1-page tool. The M-3 potentially can reduce missed psychiatric diagnoses and facilitate proper treatment of identified cases. Our subsequent primary care research will seek to make M-3 self-rating and scoring increasingly efficient, assess its generalizability to other medical populations with varying sociodemographic profiles, and investigate whether the tool can promote collaborative discussion of mental health issues and more evidence-based management of psychiatric illness.

To read or post commentaries in response to this article, see it online at <http://www.annfammed.org/cgi/content/full/8/2/160>.

Key words: Mental health; health promotion; disease prevention; mass screening; depression; anxiety disorder; bipolar disorder; stress disorders, post-traumatic; primary health care

Submitted April 3, 2009; submitted, revised, July 29, 2009; Accepted August 18, 2009.

Funding support: Data collection was supported by a grant from M-3 Information. M-3 Information focuses on providing patient-rated tools to facilitate identifying, monitoring, and managing mood and anxiety problems in primary care.

Disclaimer: M-3 Information is a limited liability corporation funded by Michael Byer, who does not have relationships with members of the medical or pharmaceutical communities. M-3 Information provides this checklist to anyone without charge. All investigators had complete access to the data. The funding source had no role in the design, conduct, or reporting of the study or in the decision to submit this manuscript for publication.

Acknowledgments: We thank the patients and clinicians at the UNC Family Medicine clinic for participating in this study. We also acknowledge Aimee Reilly, MA, Linda S. Kalka-Juhl, CMA, and Brenda Pearson, LCSW, for their help in recruitment and assessment; Abby Scheer, MSIS, and Clayton Hainline for their help in data support; Collin Perry and Jennifer B. Tuttle for their assistance with administrative duties; and Amy Zoltick for organizational and technical support.

Conflicts of interest: Bradley N. Gaynes has received grants and research support from the National Institute of Mental Health; Agency for Healthcare Research and Quality; Robert Wood Johnson Foundation; the M-3 Corporation; Bristol-Myers Squibb Company; Novartis; Pfizer, Inc; and Ovation Pharmaceuticals. He has performed as an advisor or consultant for Bristol-Myers Squibb Company; Pfizer, Inc; Shire Pharmaceuticals; and Wyeth-Ayerst. He has also received a speaker's honorarium from GlaxoSmithKline.

Joanne DeVeugh-Geiss has received research support from the National Institute of Mental Health and the M-3 Corporation. She and members of her family are shareholders of GlaxoSmithKline.

Sam Weir has no conflicts to report.

Hongbin Gu has received grants and research support from the National Institute of Mental Health; National Institute of Child Health and Development; National Institute of Biomedical Imaging and Bioengineering; Autism Speaks; the M-3 Corporation; Eli Lilly and company and AstraZeneca Pharmaceuticals. She is also currently serving on Data Safety Monitoring Boards for Medical University of South Carolina and Pfizer, Inc.

Cora MacPherson has received support from AmpliMed Corporation, Atherogenics Inc, Kinexum LLC, M-3 Corporation, and the US Department of Defense.

Herbert C. Schulberg has received research support from the National Institute of Mental Health and has performed as a paid advisor or consultant to the MacArthur Foundation and the M-3 Corporation.

Larry Culpepper has served on advisory boards or speakers bureaus for AstraZeneca, Forest Labs, Lilly, Pfizer, Somaxon, Takeda, and Wyeth. He has a 1% interest in M-3 Corporation.

David R. Rubinow has received research support from the National Institute of Mental Health.

References

- Jackson JL, Passamonti M, Kroenke K. Outcome and impact of mental disorders in primary care at 5 years. *Psychosom Med*. 2007;69(3):270-276.
- Olfson M, Shea S, Feder A, et al. Prevalence of anxiety, depression, and substance use disorders in an urban general medicine practice. *Arch Fam Med*. 2000;9(9):876-883.
- Pincus HA, Tanielian TL, Marcus SC, et al. Prescribing trends in psychotropic medications: primary care, psychiatry, and other medical specialties. *JAMA*. 1998;279(7):526-531.
- Regier DA, Narrow WE, Rae DS, Manderscheid RW, Locke BZ, Goodwin FK. The de facto US mental and addictive disorders service system. Epidemiologic catchment area prospective 1-year prevalence rates of disorders and services. *Arch Gen Psychiatry*. 1993;50(2):85-94.
- Stafford RS, Ausiello JC, Misra B, Saglam D. National patterns of depression treatment in primary care. *Prim Care Companion J Clin Psychiatry*. 2000;2(6):211-216.
- Kessler RC, Berglund P, Demler O, et al.; National Comorbidity Survey Replication. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA*. 2003;289(23):3095-3105.
- Leclercq Y. Is depression under-recognised and undertreated? *Int Clin Psychopharmacol*. 1998;13(Suppl 5):S3-S6.
- Stein MB, Sherbourne CD, Craske MG, et al. Quality of care for primary care patients with anxiety disorders. *Am J Psychiatry*. 2004;161(12):2230-2237.
- Katon W, Roy-Byrne P. Anxiety disorders: efficient screening is the first step in improving outcomes. *Ann Intern Med*. 2007;146(5):390-392. [Comment on: Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Ann Intern Med*. 2007 Mar 6;146(5):317-325.]
- Pignone MP, Gaynes BN, Rushton JL, et al. Screening for depression in adults: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2002;136(10):765-776.
- US Preventive Services Task Force. Screening for depression: recommendations and rationale. *Ann Intern Med*. 2002;136(10):760-764.
- Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):617-627. [Erratum in: *Arch Gen Psychiatry*. 2005;62(7):709. Merikangas, Kathleen R [added]].
- Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):593-602. [Erratum in: *Arch Gen Psychiatry*. 2005 Jul;62(7):768. Merikangas, Kathleen R [added]].
- Das AK, Olfson M, Gameraff MJ, et al. Screening for bipolar disorder in a primary care practice. *JAMA*. February 23, 2005 2005; 293(8):956-963.
- Bond DJ, Noronha MM, Kauer-Sant'Anna M, Lam RW, Yatham LN. Antidepressant-associated mood elevations in bipolar II disorder compared with bipolar I disorder and major depressive disorder: a systematic review and meta-analysis. *J Clin Psychiatry*. 2008;69(10):1589-1601.
- Leverich GS, Altshuler LL, Frye MA, et al. Risk of switch in mood polarity to hypomania or mania in patients with bipolar depression during acute and continuation trials of venlafaxine, sertraline, and bupropion as adjuncts to mood stabilizers. *Am J Psychiatry*. 2006;163(2):232-239.
- Amsterdam JD, Wang C-H, Shwarz M, Shults J. Venlafaxine versus lithium monotherapy of rapid and non-rapid cycling patients with bipolar II major depressive episode: a randomized, parallel group, open-label trial. *J Affect Disord*. 2009;112(1-3):219-230.
- Merikangas KR, Akiskal HS, Angst J, et al. Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2007;64(5):543-552. Erratum in: *Arch Gen Psychiatry*. 2007;64(9):1039.
- Sareen J, Cox BJ, Afifi TO, et al. Combat and peacekeeping operations in relation to prevalence of mental disorders and perceived need for mental health care: findings from a large representative sample of military personnel. *Arch Gen Psychiatry*. 2007;64(7):843-852.
- Liebschutz J, Saitz R, Brower V, et al. PTSD in urban primary care: high prevalence and low physician recognition. *J Gen Intern Med*. 2007;22(6):719-726.
- Kroenke K, Spitzer RL, Williams JBW, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Ann Intern Med*. 2007;146(5):317-325.
- Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*. 1998;59(Suppl 20):22-33, quiz 34-57.
- Sheehan DV, Lecrubier Y, Harnett-Sheehan K, et al. The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *Eur Psychiatry*. 1997;12(5):232-241.
- Lecrubier Y, Sheehan D, Weiller E, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. *Eur Psychiatry*. 1997;12(5):224-231.
- Lamers F, Jonkers CCM, Bosma H, Penninx BWJH, Knottnerus JA, van Eijk JTM. Summed score of the Patient Health Questionnaire-9 was a reliable and valid method for depression screening in chronically ill elderly patients. *J Clin Epidemiol*. 2008;61(7):679-687.
- Stafford L, Berk M, Jackson HJ. Validity of the Hospital Anxiety and Depression Scale and Patient Health Questionnaire-9 to screen for depression in patients with coronary artery disease. *Gen Hosp Psychiatry*. 2007;29(5):417-424.
- Pinninti NR, Madison H, Musser E, Rissmiller D. MINI International Neuropsychiatric Schedule: clinical utility and patient acceptance. *Eur Psychiatry*. 2003;18(7):361-364.
- Bunevicius R, Velickiene D, Prange AJ Jr. Mood and anxiety disorders in women with treated hyperthyroidism and ophthalmopathy caused by Graves' disease. *Gen Hosp Psychiatry*. 2005;27(2):133-139.
- Friedman B, Conwell Y, Delavan RR, Wamsley BR, Eggert GM. Depression and suicidal behaviors in Medicare primary care patients under age 65. *J Gen Intern Med*. 2005;20(5):397-403.
- Simon GE, Revicki D, VonKorff M. Telephone assessment of depression severity. *J Psychiatr Res*. 1993;27(3):247-252.
- Potts MK, Daniels M, Burnam MA, Wells KB. A structured interview version of the Hamilton Depression Rating Scale: evidence of reliability and versatility of administration. *J Psychiatr Res*. 1990;24(4):335-350.

32. Wells KB, Burnam MA, Leake B, Robins LN. Agreement between face-to-face and telephone-administered versions of the depression section of the NIMH Diagnostic Interview Schedule. *J Psychiatr Res.* 1988;22(3):207-220.
33. Conejo-Galindo J, Medina O, Fraguas D, Terán S, Sainz-Cortón E, Arango C. Psychopathological sequelae of the 11 March terrorist attacks in Madrid : an epidemiological study of victims treated in a hospital. *Eur Arch Psychiatry Clin Neurosci.* 2008;258(1):28-34.
34. Esposito E, Wang JL, Adair CE, et al. Frequency and adequacy of depression treatment in a Canadian population sample. *Can J Psychiatry.* 2007;52(12):780-789.
35. Esposito E, Wang JL, Williams JV, Patten SB. Mood and anxiety disorders, the association with presenteeism in employed members of a general population sample. *Epidemiol Psychiatr Soc.* 2007;16(3):231-237.
36. Patten SB, Williams JV, Mitton C. Costs associated with mood and anxiety disorders, as evaluated by telephone survey. *Chronic Dis Can.* 2008;28(4):155-162.
37. Wang J, Adair CE, Patten SB. Mental health and related disability among workers: a population-based study. *Am J Ind Med.* 2006;49(7):514-522.
38. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, 4th edition, Text Revision.* Washington, DC: American Psychiatric Association; 2000.
39. Youden WJ. Index for rating diagnostic tests. *Cancer.* 1950;3(1):32-35.
40. Holmes WC. A short, psychiatric, case-finding measure for HIV seropositive outpatients: performance characteristics of the 5-item mental health subscale of the SF-20 in a male, seropositive sample. *Med Care.* 1998;36(2):237-243.
41. Kelly MJ, Dunstan FD, Lloyd K, Fone DL. Evaluating cutpoints for the MHI-5 and MCS using the GHQ-12: a comparison of five different methods. *BMC Psychiatry.* 2008;8:10.
42. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16(9):606-613.
43. Hirschfeld RM, Williams JB, Spitzer RL, et al. Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. *Am J Psychiatry.* 2000;157(11):1873-1875.
44. Hovens JE, van der Ploeg HM, Bramsen I, Klaarenbeek MT, Schreuder JN, Rivero VV. The development of the Self-Rating Inventory for Post-traumatic Stress Disorder. *Acta Psychiatr Scand.* 1994;90(3):172-183.
45. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med.* 2006;166(10):1092-1097.
46. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA.* 1999;282(18):1737-1744.
47. Veit CT, Ware JE Jr. The structure of psychological distress and well-being in general populations. *J Consult Clin Psychol.* 1983;51(5):730-742.
48. Goldberg DP. *The Detection of Psychiatric Illness by Questionnaire.* Oxford, England: Oxford University Press; 1972.
49. Goldberg D, Williams P. *A User's Guide to the General Health Questionnaire.* Berkshire, England: Nfer-Nelson; 1991.
50. Weinstein MC, Berwick DM, Goldman PA, Murphy JM, Barsky AJ. A comparison of three psychiatric screening tests using receiver operating characteristic (ROC) analysis. *Med Care.* 1989;27(6):593-607.
51. Stein MB, McQuaid JR, Pedrelli P, Lenox R, McCahill ME. Post-traumatic stress disorder in the primary care medical setting. *Gen Hosp Psychiatry.* 2000;22(4):261-269.
52. Depression Guideline Panel. *Depression in Primary Care: Volume 1, Diagnosis and Detection.* Vol. AHCPR, Publication No. 93-0550. Rockville, MD: US Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research; 1993.
53. Leon AC, Olfson M, Broadhead WE, et al. Prevalence of mental disorders in primary care. Implications for screening. *Arch Fam Med.* 1995;4(10):857-861.
54. Mauksch LB, Tucker SM, Katon WJ, et al. Mental illness, functional impairment, and patient preferences for collaborative care in an uninsured, primary care population. *J Fam Pract.* 2001;50(1):41-47.
55. Hing E, Cherry DK, Woodwell DA. National Ambulatory Medical Care Survey: 2004 summary. *Adv Data.* 2006;23(374):1-33.
56. Center for Disease Control and Prevention. National Center for Health Statistics. *Health, United States, 2006, With Chartbook on Trends in the Health of Americans.* Washington, DC: US Government Printing Office; 2006.
57. Mulrow CD, Williams JW Jr, Gerety MB, Ramirez G, Montiel OM, Kerber C. Case-finding instruments for depression in primary care settings. *Ann Intern Med.* 1995;122(12):913-921.
58. Phillips ML, Frank E. Redefining bipolar disorder: toward DSM-V. *Am J Psychiatry.* 2006;163(7):1135-1136.
59. Prisciandaro JJ, Roberts JE. A comparison of the predictive abilities of dimensional and categorical models of unipolar depression in the National Comorbidity Survey. *Psychol Med.* 2009;39(7):1087-1096.