

Validating the M-3 Checklist: a Novel, Broad Spectrum Screen for Mood and Anxiety Disorders in the Primary Care Setting

Overview: Several validated self-administered instruments are available in primary care for the assessment of psychiatric illness, including the PHQ-9 for depression, the GAD-7 for anxiety, the CAPS for PTSD, and the MDQ for Bipolar Disorder. Taken individually these instruments have demonstrated validity and feasibility for their respective diagnoses. Yet collectively, they fail to address feasibility: patients confronted with several pages of questionnaires and doctors with too much and uncoordinated data within the time constraints of a typical office visit. The M-3 was designed to redress these shortcomings. A recently concluded validation and feasibility study at the University of North Carolina has found:

- The M-3 performs as well as the existing instruments in each of the specific diagnostic categories.
- The M-3 covers a broader spectrum of disorders (depression, anxiety, bipolar, PTSD).
- The M-3 is quick and easy to use. Eighty percent of the primary care physicians read the checklist responses in 30 seconds or less and felt that the M-3 was helpful in reviewing their patients' emotional health.
- The M-3 scores well in terms of sensitivity and specificity as compared to the other instruments. However, the M-3 has much greater functionality, as it is a multi-diagnostic instrument that is user-friendly for both PCPs and their patients. The M-3 facilitates the dialogue between doctor and patient about their mood and stress in their daily lives. Doctors and patients in the UNC study found the M-3 to be useful and easy to use.

Objective: Available tools for screening and monitoring psychiatric illness in primary care are limited by the number of disorders they assess, the degree to which they involve patients in their ongoing management, and the degree to which they encourage communication between patient and physician. The objective of this study was to assess the feasibility and diagnostic validity of a new tool developed for use in primary care, the M-3 checklist.

Design: Six hundred and forty-seven patients were recruited from the Family Medicine Center at the University of North Carolina. Patients were enrolled between July 2007 and February 2008. Patients filled out the M-3 checklist in the waiting room prior to their visit with the doctor. The doctor reviewed the checklist and asked questions regarding their mood and anxiety if needed. Following completion of the visit, the patient was contacted by phone within average of 9 days and the MINI Neuropsychiatric Interview was administered by trained personnel. The interviewer was blinded to the patient's answers on the M-3.

Measurements: The M-3 checklist is a comprehensive screening tool that was evaluated for feasibility and validity. The MINI was designed as a brief structured interview for the major Axis I psychiatric disorders in the DSM-IV and ID-10. Validation and reliability studies have been done comparing the MINI to the SCID-P for DSM-III-R and the CIDI. The results of these studies show that the MINI has acceptable high validation and reliability scores, but can be administered in a much shorter period of time than the above referenced instruments. The MINI was used as the reference standard against which the accuracy of the M-3 was tested.

Demographics: Patients in the study ranged from 18 to 92 years old, with a mean age of 45. Eighteen percent of the participants were at least 60 years old and 71% were female. Two-thirds of patients were white and 28% were black. Nearly 50% percent were married and only 8% were unemployed and looking for work. Forty-three percent of the patients reported living in households with incomes of \$40,000 per year or more.

representative studies for our comparison (Table 1).

***Expanded Depression and Subclinical Conditions:** Expanded depression defines a group of patients who fall on the spectrum between unipolar depression and anxiety disorders. In the UNC study, 80% of the depressed patients had a comorbid anxiety disorder diagnosed by the MINI. These patients may be picked up by the PHQ-9 or

Table 1: Comparison of Sensitivity/Specificity					
	PHQ-9 ¹	GAD-7 ²	CAPS ³	Mood Disorders Questionnaire ⁴	M-3
Outcome	Depression	Anxiety	PTSD	Bipolar	Depression, Anxiety, PTSD, or Bipolar
Sensitivity	.75	.77	.74	.73	.83
Specificity	.90	.82	.84	.90	.76

- 1.Spitzer, 1999. "The PHQ primary care study"
- 2.Kroenke, 2007. "Anxiety Disorders in Primary Care"
- 3.Hovens, 1994. "The development of the Self-Rating Inventory for PTSD"
- 4.Hirschfeld, 2000. "Development and Validation of a Screening Instrument for Bipolar Spectrum Disorder: The Mood Disorder Questionnaire"

Table 2: M-3 Results for Individual Diagnoses (Using sub-scores for each diagnostic category)				
	Expanded Depression*	Anxiety	PTSD	Bipolar
Sensitivity	.84	.82	.88	.88
Specificity	.80	.78	.76	.70

Results: The M-3 Performs as Well as Other Instruments: The most straightforward way to compare the M-3 with other instruments in identifying cases is to look at the sensitivity and specificity of each test. The exact numbers for the other instruments vary depending on the reference used, but generally these values do fall "within the same ball park," and so we chose

by the GAD-7, but some – those failing to reach the threshold for either diagnostic group – will be missed by using these instruments singly. One of the advantages of the M-3, which derives from its broad spectrum, is that patients with subclinical symptomatology in one diagnostic area are still likely to be correctly identified as someone in need of medical attention.

The false positives generated by the M-3 may also be clinically significant. Over half of the patients falsely diagnosed with Bipolar Disorder were found to have other MINI confirmed illness. And among the patients falsely assigned a PTSD diagnosis by the M-3, 70 percent received other psychiatric diagnoses by the MINI. More severely ill generally, we recommend psychiatric referral for the bipolar and PTSD cases discovered by the M-3. With such a high rate

known prevalence rate. The interim data from the UNC Study found 35% of the patients had a mood or anxiety diagnosis by MINI. Depression was found to have a prevalence rate roughly twice that found in some other studies; notably, Kessler et al, 2005. However, the Kessler study involved a nationwide household sampling, whereas the current study's subjects were primary care clinic patients; and medically ill patients are at generally greater risk for depression.

Table 3: Summary of Disorders Covered by each Instrument

	PHQ -9	GAD -7	CAP S	MDQ	M-3
Depression	X				X
Anxiety		X			X
Bipolar				X	X
PTSD			X		X
# of Questions	9	7	30	15	27

of alternative diagnoses among the false positives, such cases certainly do warrant a careful assessment by a specialist. The ability of the M-3 to identify such cases is valuable at the point of primary care contact.

The M-3 Covers a Broader Spectrum of Disorders: Each of the existing instruments focuses on a single disorder. Where these instruments would require patients to fill out several surveys (including a number of redundant questions) the single-paged M-3 checklist provides a succinct and comprehensive alternative. The spectrum of mood and anxiety disorders and the instruments' screening utility is summarized in table three.

The M-3 Impact: Why Use the M-3?: Another way to look at this is as follows: Of 100 patients who come into a primary physician's practice for care, for each of the instruments, how many mood and anxiety disorders will be missed? This question requires that we compare the instruments based on their ability to uncover cases at a

Table 4: Summary of Prevalence Rates by Study

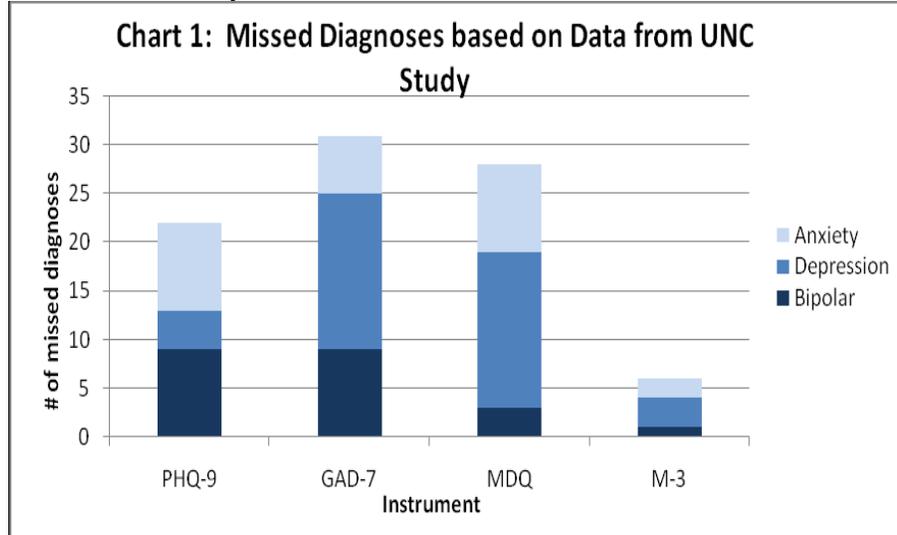
MINI Diagnosis	Prevalence in UNC Study	Prevalence in Literature Review
Depression (alone & „expanded“ w/anxiety)	16%	8%
Anxiety w/o depression	9%	10%
Bipolar	9%	10%
Any Disorder	35%	28%

2. Kroenke, 2007 Anxiety disorders in primary care
5. Das, 2005 Screening for Bipolar Disorder in a Primary Care Practice
6. Depression guideline panel. Depression in primary care: Volume 1, 1993

A practice that used the PHQ-9 alone, with a sensitivity of 0.75, would presumably identify 12 of the 16 patients out of 100 with depression. However, the PHQ-9 would miss the 9 patients diagnosed with anxiety without depression and less than half of the 9 bipolar disorder patients (only those who were currently depressed). Missing this significant number of anxiety and bipolar patients entirely is obviously troubling, but misidentifying bipolar cases as unipolar depressed patients is arguably worse, because treating bipolar depressed patients with antidepressant monotherapy is strictly contraindicated. Similarly, the GAD-7 alone would capture 7 from among the 9 out of 100 patients with anxiety disorder, and another few cases from among the expanded depression group, but would fail to identify roughly 20 patients with bipolar disorder and unipolar depression (Chart 1).

**Expected Number of Cases Missed by Instrument
Using Prevalence from UNC Study
(per 100 patients seen in practice)**

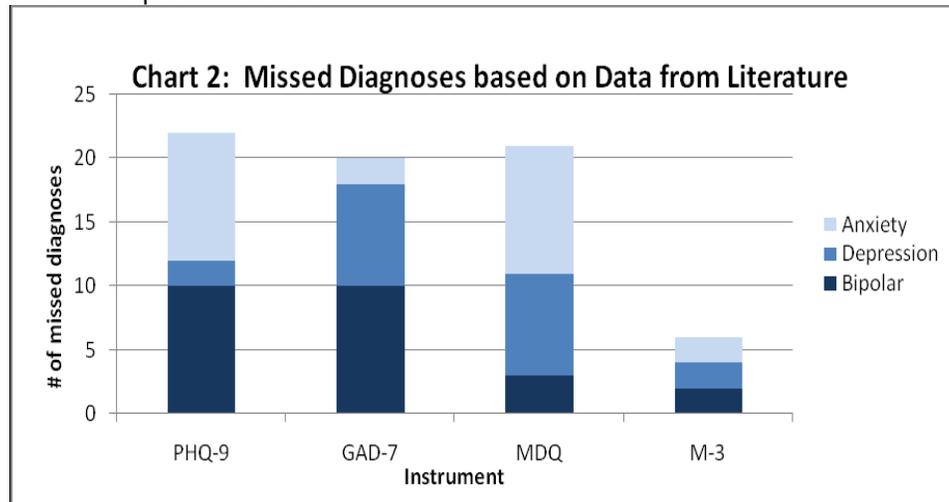
Chart 1: UNC Study



Using prevalence rates available in the literature, we find a very similar outcome (Chart 2).

**Expected Number of Potential Patients Missed by Instrument
Using Prevalence from Literature
(per 100 patients seen in practice)**

Chart 2: Popular Literature



- 2. Kroenke, 2007 Anxiety disorders in primary care
- 5. Das, 2005 Screening for Bipolar Disorder in a Primary Care Practice
- 6. Depression guideline panel. Depression in primary care: Volume 1, 1993

The M-3 is Quick and Easy: In the current study physicians and patients completed feasibility questionnaires after using the M-3 checklist in the office visit. Eighty percent of the primary care physicians read the checklist responses in 30 seconds or less and none reported that the checklist was too complicated. Eighty percent of the physicians felt the M-3 was helpful in reviewing their patients' emotional health.

Among patients who received a diagnosis by the MINI, 75% said that the M-3 helped them talk to their doctor about their mood or feelings.

Gateway Approach to Scoring the M-3: Numerous approaches were explored in seeking the best way to score the M-3. Full scoring, weighted scoring, different combinations of questions, different weights for each answer, counting the numbers of 3 and 4 responses, and multiplying lifestyle scores (Q.24-27) by diagnosis subset scores are just some of the avenues that were tried. In the end, the „Gateway Method“ was chosen, because it provides the best balance of robust sensitivity and specificity (S&S) while permitting a quick, visually intuitive method for scoring by hand.

The Gateway approach has two steps:

1. First, responses to the suicide question (Q. 5) and the „lifestyle“ questions (Q.24-27) are tallied. If the patient answered > 0 on the suicide question or > 1 on the sum of the lifestyle questions, the clinician is prompted to score the full M-3 checklist (i.e., the patient goes through the gateway). Otherwise, the scoring is finished and the patient receives no diagnoses by the M-3. This prevents 56% of the forms from having to be fully scored. The lifestyle questions are scored (0,1,2,3,4) → (0,0,1,2,2).
2. Once through the gateway, scores are calculated for each of the three diagnoses: depression, anxiety, and bipolar disorder. All questions are scored

(0,1,2,3,4) → (0,0,1,2,2) with the exception of the suicide question, which is scored slightly more sensitively with (0,1,2,3,4) → (0,1,1,2,2). Patients who score at or higher than the cut-off in a diagnostic category are assigned that diagnosis. Patients who receive a diagnosis of anxiety are then further examined for PTSD.

Methods for Choosing Cut-points: There are several methods for choosing the cut point for a screening tool. A comparison by Kelly and colleagues (2008) found the Youden Index and the (0,1) Method to be the two most effective techniques. The Youden Index is a commonly used measure of diagnostic effectiveness that optimizes and gives equal weight to sensitivity and specificity. The (0,1) method also puts equal weight on sensitivity and specificity, but focuses on minimizing the distance between the ROC curve and the point (0,1). An ROC curve which passes through the point (0,1) would indicate a perfect diagnostic tool. At an interim analysis, we used both of these methods on the first 525 patients enrolled in the study. For depression, anxiety, and PTSD, the two methods agreed on the optimal cut point. For bipolar disorder, the two methods chose different cut-points and the midpoint was used for the M-3 (rather than 1 or 3 we chose 2). To validate these cut points, the remaining 122 patients were run the M-3 using the method developed on the first cohort. For each diagnosis, the results were similar or better than the S&S found in the original cohort.

