M3 Validation Study and White Papers
Discussing Mental Health Screening in Primary Care + Behavioral Health
# Table of Contents

I. Multidimensional Screening for Mood & Anxiety Disorders in Primary Care Settings  
   A. Viewing Depression, Anxiety, and Bipolar Disorder in Primary Care Practices ........................................... 4  
   B. On the Elicitation of Bipolar Disorder Diagnoses in Primary Care ......................................................... 5  
   C. Alcohol and Substance Use in Primary Care ............................................................................................ 9  
   D. Pain Management ................................................................................................................................. 12

II. Alternatives for Specific Populations  
   A. The M3 Screen for Postpartum Mood & Anxiety Disorders vs. The Edinburgh Postnatal Depression Scale ................................................................. 17  
   B. Alternatives for Veterans ..................................................................................................................... 20  
   C. Mood and Anxiety Disorders in College Students .................................................................................. 22  
   D. Employee Assistance Programs ......................................................................................................... 25

III. M3 Validation Study & Related Discussions  
   A. Feasibility and Diagnostic Validity of the M-3 Checklist ........................................................................ 28  
   B. Specificity of Bipolar Spectrum Conditions in the Comorbidity of Mood and Substance Use Disorders ........................................................ 38  
   C. Screening for Depression in Adults (USPSTF) ..................................................................................... 44

© M3 Information 2016. The contents of this booklet may not be reproduced, in part or whole, without the express written consent of M3 Information.
I. Multidimensional Screening for Mood and Anxiety Disorders in Primary Care Practices
A. Viewing Depression, Anxiety and Bipolar Disorder in Primary Care Practices

Improving efficiency in healthcare delivery is the golden fleece of current (and also longstanding) reform efforts. As one important component of this goal, over the years dreams of a valid and effective screening program for patients suffering from mood and anxiety disorders have led to numerous initiatives among managed care providers. The logic, plain and simple, is that anxious and depressed patients consume a disproportionate amount of healthcare dollars, especially when their identification is missed or delayed. Effective and targeted treatment, so it goes, should work to keep costs down.

The problem is: despite considerable effort and investment, existing screens have largely proven that a different approach is needed. Clinicians must be pushed to use them, but still find them too time consuming or otherwise impractical. Several studies have indicated that even when screens do identify cases, often this has no impact on intervention or outcome.

What gives? The sense we have at M3 Information is that: 1) existing screens do not reflect the reality of mood and anxiety presentations in the field; 2) nor do they promote the investment of the patients in the process, who, upon being confronted with waiting room forms, find paper-and-pencil diagnosis off-putting and confusing; and 3) clinicians find it daunting to synthesize what data they do collect, to assess compliance and progress, and to keep patients on course with the treatment.

Existing screens focus on only a single diagnosis. In the real world, patients often present with comorbid mood and anxiety symptoms, subsyndromal illness, co-morbid substance abuse, and of course, comorbid physical symptoms, including chronic pain, GI complaints, etc. Looked at piecemeal, these overlapping symptoms may not reach criteria for, say, a depressive disorder; but, taken as a whole these cases typically represent patients absolutely in need of mental health treatment. More in keeping with the direction of the DSM-V development, the M3 is less “diagnosis-centric” and more dimensional in its approach. While it does provide individual risk assessments for depression, anxiety, PTSD and bipolar disorder, it also offers a single measure, the M3 Checklist Score, which reflects the patient’s overall need for treatment, and includes a measure of functional impairment and substance use patterns. We find that, by being less obsessed with the proper identification of depression versus anxiety, we increase the overall case identification rate. Also, by taking care to screen for bipolar disorder – something no other anxiety or depression screen does – we help to avoid one very troubling and expensive source of iatrogenic morbidity; i.e., antidepressant-induced mania and mixed mood states.
B. On the Elicitation of Bipolar Disorder Diagnoses in Primary Care

M3 is committed to improving the accurate diagnosis of Bipolar Disorder (BD) and to minimizing its misdiagnosis, as the current trend in non- and mistreatment entails tremendous but avoidable morbidity and mortality, wasteful healthcare costs, and lost productivity. Yet, in our ongoing effort we have occasionally confronted the concern that eliciting BD in primary care (PC) is fraught, and that without appropriate guidelines and expertise in place, the risks of identification may outweigh any potential benefits.

Evidence-based guidance on this matter is available and speaks in favor of screening for BD (see below, p2). Yet, there have been no prospective large-scale study comparing diagnosis as usual over against diagnosis based on a validated screen for BD. Concerns that PC doctors will be overwhelmed by the treatment demands, will mistreat such cases, etc., are put forth purely on hypothetical grounds or based on anecdotal evidence, and typically without bearing in mind the alternative; i.e., the status quo. How do these same individuals fare when left improperly diagnosed and, furthermore, how do they consume healthcare services? Rather than simply conjure a hypothetical bipolar patient newly diagnosed by his PCP, we must compare such an individual moving through the system along two divergent paths. A proper analysis of this issue needs to consider the following:

1) **Regarding misdiagnosis:** At present, which carries with it the greater risk of mistreatment and its associated costs: the cohort of bipolar patients misdiagnosed as unipolar depression/anxiety OR the group correctly identified and managed by their PCP? *We believe the available evidence strongly demonstrates that misdiagnosis leads to mistreatment and hence to marked increases in morbidity, mortality and disability. Meanwhile, there is no scientific evidence that boosting the identification of this condition in primary care would result in higher rates of such negative outcomes.*

2) **Regarding non-diagnosis:** How many entirely new cases of mental illness would be uncovered with an effective bipolar screen; i.e., those BD patients who currently receive no psychiatric diagnosis whatsoever? To reflect accurately the natural history of this condition and to properly distinguish true non-diagnosis from misdiagnosis, a 12-month incidence rate should be utilized. A patient hypomanic in January may not be found, and misdiagnosed, with unipolar depression until September. *We believe these patients are more often misdiagnosed than undiagnosed, but the undiagnosed will ultimately consume an unnecessary surplus of health care services and express high rates of medical, psychiatric and substance abuse complications.*
3) **Regarding overall healthcare utilization:** How do the misdiagnosed and the undiagnosed BD patients compare to the properly identified BD cohort in terms of overall healthcare consumption, employment status, marital status, legal history, and substance abuse history? *We interpret the studies quoted below to support the view that recognizing BD at the earliest possible time saves lives, reduces suffering and sustains home and work life functioning.*

The following, from just one of many reviews on this topic, are excerpted from Keck, Kessler & Ross (2008)¹:

> For every behavioral health care dollar spent on outpatient care for bipolar disorder, $1.80 was spent on inpatient care, suggesting that strategies to prevent acute episodes could decrease the financial burden of the illness. *When the diagnosis of bipolar disorder is missed or the presentation is misdiagnosed and inappropriately treated as unipolar depression, as frequently occurs . . . this can lead to even higher healthcare costs.*²,³

Birnbaum et al.⁴ compared treatment patterns and costs for patients with recognized and unrecognized bipolar disorder with those of depressed patients without bipolar disorder, using claims data from seven large national employers for 585,584 individuals younger than 65 years of age…*The patients with unrecognized bipolar disorder incurred significantly greater mean monthly medical ($1179) and indirect ($570) costs in the 12 months after initiation of antidepressant treatment compared with those with recognized bipolar disorder ($801 and $514).* These investigators concluded that accurate recognition of bipolar disease was associated with lower medical costs as well as lower indirect costs due to work loss.

Shi et al.⁵ compared hospital use, suicide risk, and healthcare costs of patients with recognized and unrecognized bipolar disorder with those of patients with non-bipolar mood disorders, using data from 25,460 adults in the California Medicaid program with a new episode of antidepressant therapy. *They found that patients with unrecognized bipolar disorder were more likely to attempt suicide and be hospitalized than those with recognized bipolar disorder and non-bipolar mood disorders.* Patients with bipolar disorder also had significantly higher total costs (almost $1,000 per person per year more in direct healthcare costs) than those with non-bipolar illness in the first year post-treatment; likewise, healthcare costs for patients with recognized bipolar disorder were also nearly $700 less per person per year than for those with unrecognized bipolar disorder, primarily due to lower costs for ambulatory care and hospital services.
McCombs et al. examined California Medicaid data from patients who had started new courses of antidepressant therapy for whom at least 6 years of post-treatment data were available. They found that growth in costs for patients with unrecognized bipolar disorder over 6 years was 171%, greatly exceeding the increase in costs associated with recognized bipolar disorder (82%) and depression (95%). They also found that costs increased by $91/month for each month the diagnosis of bipolar disorder was delayed . . . These researchers concluded that early diagnosis of bipolar disorder may significantly reduce healthcare costs.

Studies have also found that bipolar disorder can have severe and often enduring negative effects on occupational functioning, resulting in significant indirect costs through lost productivity. Wyatt and Henter estimated that the economic costs of bipolar disorder from a societal perspective were $45 billion a year in the United States, with the economic losses due to work impairment accounting for the largest proportion (nearly $18 billion annually) of this total. This estimate translates into annual workplace costs of over $125,000 for a company with 1,000 employees.

A primary care study found that employed patients with bipolar disorder missed seven times as much work as other patients.

A recent study by Guo et al. found that treatment of comorbid conditions (including substance use disorders, cerebrovascular disease, ischemic heart disease, and hypertension) accounted for 70% of treatment costs for patients with bipolar disorder. Because of the risks associated with treating bipolar disorder with antidepressant monotherapy, it is important that primary care physicians be educated about how to screen for and treat bipolar disorder.
REFERENCES


C. Alcohol and Substance Abuse in Primary Care

Substance Abuse Disorders (SUDs), which include alcohol and substance abuse and dependence, create tremendous burdens on US health and healthcare costs and impede economic growth through lost worker productivity. In an effort to manage the SUDs – and the associated mental, physical and economic fallout – policy-makers have naturally looked toward screening within primary care as a rational starting point. Unfortunately, there are some pitfalls to this approach. With the typical in-office screening, physicians have come to expect that many patients lack the insight and honesty to ensure full disclosure. Then there is the impracticality of actually counting drinks and the frequency of use, and placing this information in the context of a patient’s culture, stage of life, and lifestyle.

A fresh start is in order, one that takes hold of the problem in a new way. Alcohol and substance abuse often begin with individuals trying to manage either incipient or burgeoning psychiatric symptoms. A strong association between mood and anxiety disorders and SUDs has long been established in the literature. What is perhaps less well appreciated is the particularly strong link between Bipolar Disorders and SUD. Merikangas et al.\textsuperscript{1} demonstrated in their 20-year prospective study, reported in 2008, that Bipolar Spectrum conditions have the strongest association with alcohol and benzodiazepine abuse and dependence, while unipolar depression was associated strongly with only benzodiazepine dependence. Compared to those without a mood disorder, a diagnosis of Bipolar Disorder type II was found to bring with it a 9-fold increase in alcohol abuse and 21-fold increase in alcohol dependence.

Swendsen et al.\textsuperscript{2}, in their 10-year prospective study reported in 2010, extended these findings to include a survey of anxiety as well as mood disorders. Again, a particularly strong link was found between bipolar disorders and SUD. But, a range of anxiety conditions were also found to have higher associations with SUD than did unipolar depression.

From here, the plot thickens. As one might imagine, over time alcohol and substance abuse worsen the very symptoms the users are trying to ameliorate. While numerous studies have documented the pernicious effects of substance abuse on mood disorders, two recent reports stand out. In their 2008 report, Kemp et al.\textsuperscript{3} found that SUD was implicated in a worsening course of bipolar disorder and with its resistance to treatment. And in a 2009 report, Jaffee et al.\textsuperscript{4} found that, among bipolar patients, the duration of alcohol and substance use was linearly correlated with an increased risk of an ensuing depressive episode.
How do we get a handle on these complex problems? In order to find meaningful misuse patterns, rather than counting ounces of vodka and puffs of marijuana, better to focus on when such agents are being used to manage mood and anxiety symptoms, from the mildest to the most severe. The M3 was designed to help clinicians find a wide range of mood and anxiety disorders, and not merely depression, while it also asks patients whether they have resorted to alcohol or non-prescribed drugs to manage some of their mood or anxiety symptoms. The symptoms of stress and distress are the context within which querying for alcohol and drug use can best determine when patterns of use are problematic.

Furthermore, the M3 allows clinician’s to track patients over time, each time asking about alcohol and drugs along with its survey of mood and anxiety symptoms. Being able to recognize the slow development of bipolar symptoms, discoverable through this tracking, is an invaluable tool in our effort to manage the SUDs. One version of this very narrative was depicted in a 2008 report of Oronsky & Martin. In it they found evidence that, among patients suffering from chronic pain with an unrecognized diagnosis of bipolar disorder, the widespread practice of prescribing antidepressants as a treatment for pain actually increased the risk of opiate addiction. They write:

“In our experience, psychological dependence on narcotics diminishes with appropriate treatment for bipolar disorder. If an antidepressant has already been prescribed, the patient should be closely evaluated for a worsening of psychiatric and/or pain symptoms and discontinuing the antidepressant medication in this case may be warranted.”

The onset or worsening of bipolar symptoms over time, within the context of ongoing antidepressant treatment, is one of the many trends the M3 was engineered to track. As these authors suggest, being able to catch a worsening of psychiatric symptoms at the earliest possible time helps to prevent an iatrogenic mishap: making patients worse despite a well-intentioned treatment. Giving antidepressants to an undiagnosed bipolar patient is one such potential mishap. But, at a more general level, it is important to appreciate how symptoms of overactivation, be they bipolar or anxiety-related, place patients at risk for alcohol and substance misuse. An effective primary care mental health screen must be able to measure all of the dimensions of mood and anxiety disorders (not merely depression) along with usage patterns of alcohol and non-prescribed drugs. This is precisely what the M3 was designed to do.
REFERENCES


5. Oronskey BY, Martin M. Do antidepressants contribute to prescription narcotic medicine dependence and addiction? *Int J Fam Pract* 2008; 6(2)
D. Pain Management

Chronic pain syndromes place a tremendous burden on our healthcare system, with a recent NIH-sponsored study estimating the cost at $635 billion a year in the US (*Institute of Medicine, Consensus Report, 2011*). Clearly, improving the management of chronic pain within primary care, and medicine generally, is of utmost importance. In this critical effort, the **M3** screen and monitor represent a revolutionary advance over existing instruments. Based on numerous studies and reviews in the literature (*e.g.; Bair, Robinson, Katon, Kroenke, 2003*), the relationship between chronic pain and depression is well known and now well-established. Consequently, any instrument that increases the accurate identification of depression in patients will be a boon to those patients and their doctors. What is perhaps less well known is the very strong relationship between anxiety disorders and chronic pain. With the **M3**’s multi-dimensional approach to mood and anxiety diagnosis and its high correlation with the SF-12, it provides an excellent way to find, assess and monitor the benefits of treatment in terms of symptom burden, social and work functioning, and quality of life globally.

Here are a few less well known facts concerning anxiety disorders and chronic pain:

- Among patients with chronic pain and comorbid anxiety disorders 77% experienced their anxiety symptoms before the onset of their pain syndrome (Knaster, Karlsson, Estlander, Kalso, 2012).
- Patients with Generalized Anxiety Disorder (GAD) suffer from more lifetime pain interference compared to non-anxious patients, and also have disproportionally high medical health care costs. (Olfson & Gameroff, 2007)
- "The co-occurrence of medically unexplained pain and generalized anxiety was associated with poorer negative outcomes in terms of quality of life, disability, and health care utilization." (Beedso, Hoyer, Jacobi, et al. 2009)
- Patients with arthritis had a 12-month prevalence rate for GAD of 5.6% compared to 2.7% of people without arthritis; those with migraine a prevalence of 9.1% compared to 2.5% without migraine; and those with back pain 6.2% compared to 2.5% without back pain. (McWilliams, Goodwin & Cox, 2007)
- Some studies have found rates of Post-Traumatic Stress Disorder (PTSD) to exceed 25% among chronic pain patients when the pain is related to accident or injury. (Jaspers, 1998)
• "The added morbidity of depression and anxiety with chronic pain is strongly associated with more severe pain, greater disability, and poorer health-related quality of life." (Bair, Wu, Damush, et al. 2008).

• "...treating a GAD-only patient averaged $2647 per year. However, GAD patients with comorbid depression-only, comorbid pain-only, and comorbid depression and pain had increased costs of 60%, 200% and 400% over those GAD-only patients, respectively." (Zhu, Zhao, Ye, et al. 2009).

The following chart was adapted from Beesdo, Jacobi, Hoyer, et al. 2010.

---

Prevalence and comorbidity of lifetime pain (not mutually exclusive groups) and 12-month anxiety & depressive disorders (N=4,181) EPS=explained pain symptoms; UPS=unexplained pain symptoms; PD=pain disorder.

The M3 also allows clinicians to track patients' progress over time, each time asking about alcohol and drugs along with its broadly tuned survey of mood and anxiety symptoms. The M3's capacity to track and help recognize the development of bipolar symptoms and substance misuse is an invaluable tool not available in other mono-diagnostic rating scales. One cautionary tale related to substance use and bipolar comorbidity was reported in a 2008 by Oronsky & Martin.
They found evidence that among patients suffering from chronic pain with an unrecognized diagnosis of bipolar disorder, the widespread practice of prescribing antidepressants as a treatment for pain actually increased the risk of opiate addiction. They write:

“In our experience, psychological dependence on narcotics diminishes with appropriate treatment for bipolar disorder. If an antidepressant has already been prescribed, the patient should be closely evaluated for a worsening of psychiatric and/or pain symptoms and discontinuing the antidepressant medication in this case may be warranted.”

The problem depicted here is not a rare event. In another recent study (Wilde, Gota & Muzina, 2010) one quarter of the fibromyalgia patients referred to a tertiary care center for rheumatology screened positive for bipolar disorder. In their conclusion they echo the analysis of Oronsky and Martin above. Identifying a worsening of psychiatric symptoms at the earliest possible time helps to prevent a costly iatrogenic mishap: making patients worse by providing the wrong treatment. Giving antidepressants to an undiagnosed bipolar patient is one such potential mishap. The onset or worsening of bipolar symptoms over time, within the context of ongoing antidepressant treatment, is one of the many trends the M3 was engineered to track.

But, at a more general level, it is important to appreciate how symptoms of overactivation, be they bipolar or anxiety-related, place patients at risk for increased pain, as well as alcohol and substance misuse. An effective primary care mental health screen must be able to measure all of the dimensions of mood and anxiety disorders (not merely depression) along with usage patterns of alcohol and non-prescribed drugs. This is precisely what the M3 was designed to do.
REFERENCES


Committee on Advancing Pain Research, Care, and Education. Relieving pain in America: a blueprint for transforming prevention, care, education, and research. Institute of Medicine Brief Report, June 2011


McWilliams LA, Goodwin RD, Cox BJ. Depression and anxiety associated with three pain conditions: results from a nationally representative sample. Pain 111, 2004, 179-192


Oronsky BY, Martin M. Do antidepressants contribute to prescription narcotic medicine dependence and addiction? Int J Fam Pract 2008; 6(2)


II. Alternatives for Specific Populations
A. The M3 Screen for Postpartum Mood & Anxiety Disorders vs. The Edinburgh Postnatal Depression Scale

The Edinburgh Postnatal Depression Scale (EPDS) was introduced to the public in 1987 when Cox, Holden and Sagovsky published their report on only 84 subjects in the British Journal of Psychiatry. The scale was validated using the Goldberg Standard Psychiatric Interview as its gold standard. It has gone on to be investigated and revalidated in a wide range of settings and languages, but typically with only tens of patients per study, but nonetheless has held up as a standard screen for Postpartum Depression ever since.

But, there are many weaknesses in the 10-item EPDS, weaknesses that are addressed and corrected by the M3, a 27-item, multi-diagnostic screen validated by comparison with the Mini International Neuropsychiatry Interview in a 650-subject University of North Carolina study (Gaynes et al, Ann Fam Med 8-2; 2010). Below is a list of the most critical problems.

The EPDS provides 7 questions in its survey of depressive symptoms and 3 questions targeting anxiety, with only one of these three aimed at panic anxiety. But postpartum mood and anxiety episodes are phenotypically much more complex than this simple instrument can capture.

1) Bernstein et al. (Depression and Anxiety, 25-1; 2008) found that, among postpartum depressed women, sad mood was less prominent while psychomotor restlessness and poor concentration were more significant markers. They conclude: “These differences between postpartum and other depressives suggest the need to include agitation/restlessness and impaired concentration/decision-making among screening questions for postpartum depression.” The M3 queries: “Q2 – I can’t concentrate or focus”; and “Q8 – I feel tense, anxious, or can’t sit still”; but there are no comparable questions in the EPDS.

2) Altshuler et al (J of Clin Psychiatry 59-2; 1998) describe how anxiety syndromes other than generalized anxiety can manifest in the postpartum period; these include panic disorder and obsessive-compulsive disorder. While the EPDS offers one question concerning panic, and none that target OCD, the M3 provides two panic disorder questions and three targeting OCD. Wisner et al (JAMA Psych 70-5; 2013), in their study of 10,000 mothers who screened positive on the EPDS 83% had comorbid anxiety disorders.
3) In that same study of 10,000 women, Wisner et al found that among those who tested positive on the EPDS... "a striking 22.6% had bipolar disorder." Meanwhile, Sharma et al (Am J Psychiatry 166-11; 2009) comments that: “…54% of 56 outpatients... with the referral diagnosis of postpartum depression were rediagnosed as having a lifetime diagnosis of bipolar disorder.” Meanwhile, Ghaemi et al (Psychopathology 37-5; 2004) found that postpartum depression was one of “the five most powerful predictors of bipolar disorder.” Bipolar symptoms are not touched upon by the EPDS, but these symptoms are now understood to be essential to any screen designed to uncover acute depression. Fiedorowicz et al (Am J Psychiatry 168-1; 2011) point out that, among depressed patients, subsyndromal symptoms of hypomania often indicate an eventual progression to a bipolar diagnosis. Missing these symptoms is a source of misdiagnosis with profound implications. Afain, Sharma et al conclude: “… evidence suggests that bipolar II depression arising in the postpartum period is often misdiagnosed as unipolar major depressive disorder. The consequences of this misdiagnosis can be particularly serious because of delayed initiation of appropriate treatment and the inappropriate prescription of antidepressants.”

The M3 has not been studied specifically in the postpartum period, but the 10 items comprising the EPDS do not refer in any way to stressors specific to the postpartum. It is essentially a screen for non-specific depression and anxiety. It is quite evident that, from among the 27- items in the M3, those questions that correspond to EPDS items would perform every bit as effectively. But, at the same time, the M3 surveys symptoms “invisible” to the EPDS. Broadly tuned to elicit symptoms of bipolar and unipolar depression, as well as a range of anxiety disorders, the M3 offers a dimensional perspective impossible for the EPDS to provide. Furthermore, because of its dimensional approach—combined with its single number total score that reflects overall impairment—the M3 can help to find subsyndromal and mixed mood and anxiety conditions.

The US Preventative Services Task Force (USPSTF), in its newly released Recommendation Statement, entitled Screening for Depression in Adults (JAMA 315-4; 2016), prescribes “screening for depression in the general adult population, including pregnant and postpartum women.” But the recommendations go on to state that “All positive screening results should lead to additional assessment that considers … comorbid psychological problems (e.g. anxiety, panic attacks, or substance abuse), alternate diagnoses, and medical conditions.” How much more efficient—as compared to stacking screen upon screen—is a single questionnaire that probes for the full range of symptomatology in the postpartum?
The answer here is quite evidently that a single multi-dimensional screen represents a great advance over the mono-diagnostic approach that the EPDS has to offer*.

It is time to modernize behavioral health screening to reflect our understanding of the variability and complexity of mental illness. This is especially so for the postpartum period, where mood and anxiety hybrid syndromes and subsyndromes that can worsen over time, are so very common. The M3 provides the necessary flexibility and scope, as well as the sensitivity and specificity on par with the comparable mono-diagnostic screens.

*A relevant USPSTF study concerning screening for depression, especially in postpartum women can, can be found on page 44
B. Alternatives for Veterans

The Archives of General Psychiatry, in its November 2010 issue, published a report by Ilgen, Bohnert, Ignacio, et al, surveying diagnoses among the 7700 suicides committed by US Veterans over the six year span ending in 2006. Their findings were significant, but perhaps most noteworthy is the lack of action prompted by these findings.

Among men, a bipolar disorder diagnosis carried the greatest risk of suicide, accounting for 9% of all suicides, and 19% of all suicides with a reportable diagnosis. The authors conclude: "This makes bipolar disorder particularly appropriate for targeted intervention efforts . . . " What the authors do not emphasize in their commentary is the likelihood that their method, a retrospective chart review of veterans who made medical visits to the VHA, actually under-reported the incidence of bipolar-related suicide. Many psychiatrically ill individuals do not seek medical care, and this is more likely among male veterans. So, with scant medical records to base their findings upon, the lack of a diagnosis (or any kind) among half of suicides represents a vast dark matter of psychiatric illness. The likelihood is high that among these fatalities without an associated diagnosis, there was a substantial number of additional bipolar cases.

In support of this inference is the study's finding that psychiatric diagnoses appeared to play a considerably stronger role in risk for suicide among female veterans, who had more than twice the incidence of any diagnosis as compared to their male counterparts. But this otherwise puzzling imbalance almost certainly reflects the greater willingness of women to seek medical attention. The more medical interventions made, the more diagnoses available upon retrospective analysis.

There are other reasons to suspect that bipolar disorder was an under-reported factor in the risk for suicide. Studies have consistently found highly significant rates of misdiagnosed bipolar patients, with 20% to 40% of unipolar depression diagnoses later reclassified as bipolar (Ghaemi et al - J Affect Disord 52, 1999; Ghaemi et al - J Clin Psychiatry 61, 2000; Hirschfeld et al - J Am Board Fam Pract 18, 2005). And in their careful review of the literature Keck, Kessler, and Ross (J Psychiatric Pract 14-S2, 2008) reported: "Misdiagnosis of bipolar disorder can lead to inappropriate treatment, worsening symptoms, and increased hospitalization and emergency room visits." And, entirely in agreement with Ilgen et al, they state: "...untreated bipolar illness is associated with a high risk of suicide, which has been reported to range from 15% to 19%. . . suicidal behaviors were more prevalent among patients with bipolar disorder than unipolar depression . . . "
Another important finding in the Ilgen et al article was that anxiety disorders other than PTSD among veterans accounted for 31% of all diagnoses associated with suicide. In fact, the risk of death by suicide among these non-PTSD anxiety patients was slightly higher than the risk associated with PTSD itself.

What has not occurred since the release of this data is any change in the VHA's and DOD's method of screening for mood and anxiety disorders. Primary care patients in the VHA receive routine screens for depression, PTSD, and alcohol abuse. Yet, no care is taken to screen out the bipolar patients among those found to be acutely depressed, they provide no baseline bipolar screen, and there is no adequate screen for anxiety disorders other than PTSD. With non-PTSD anxiety disorders and bipolar disorder combined accounting for 50% of the suicides with assigned diagnoses, can the VHA afford to ignore these diagnoses in their routine screening procedures? This question becomes even more pointed when one considers that an alternative screening system, a simple 27-item self-reported questionnaire called the M3 Checklist, has been validated and is available for unipolar and bipolar mood disorders and a range of anxiety disorders, in addition to PTSD.

Sincerely,

M3 Collaborative

Robert Post, M.D.

Gerald Hurowitz,, M.D.

Michael Byer
C. Mood and Anxiety Disorders in College Students

Freshman year of college is a wonderful time: a time for independence, for exposure to bigger ideas and to new social adventures. But, numerous studies have demonstrated that the late teens and early twenties are also the age at which many mood and anxiety disorders first develop.\(^1,2,3\) The great annual migration of 18 year olds delivers them to a land of newfound independence, but with this comes the challenges of having to manage a personal schedule and mini-household, to balance an increased workload, and to find one's way socially in entirely new surroundings. While many freshman get a big thrill out of these developments and flourish during this time, even for those who eventually adapt it can be quite stressful. The stress of separating from one’s family and being asked to quickly fit in and get to work, coming at a vulnerable time in terms of brain development, carries with it a particular risk for mood or anxiety symptoms. More than has been widely appreciated, this is a time when vigilance about mental illness is of critical importance.

Naturally, we need to find and help kids who are struggling through this important transition to adulthood, but screening the college-aged for mood and anxiety disorders is particularly critical because identifying problems late in the game often leads to illness that becomes harder to treat.\(^4\) Numerous studies have shown that delays in the identification of depression, anxiety and bipolar disorder complicate treatment and lead to more suffering in the long term, including early death.\(^4,5,6\) Untreated depression, insomnia, irritability and anxiety all bring down academic performance\(^7\), but also lead to a greater likelihood of alcohol and substance abuse, setting up a vicious cycle because misuse of recreational drugs will worsen the mood and anxiety symptoms the student is trying to manage in this way.\(^8,9\)

It is important to focus not merely on depression as some mental health screens do. The anxiety disorders are actually more prevalent than, and may also co-exist with depression; and, they carry with them a greater risk of inciting substance and alcohol abuse.\(^8,10\) Meanwhile, a diagnosis of bipolar disorder is missed in up to 30% of individuals\(^11,12\) diagnosed with depression. The consequences of this misidentification are especially pernicious: bipolar patients who receive antidepressants will often grow sicker, are more prone to suicide, and ultimately become more difficult to treat.\(^4\) Moreover, bipolar patients have the greatest risk of falling into drug and alcohol dependence.\(^8,10\)
The M3 is a screening and monitoring system for mood and anxiety symptoms, as well as alcohol and substance misuse. A computer interfacing, self-rated 27-item questionnaire that can be completed in less than 3 minutes, the M3 indicates the risk of depression, anxiety, PTSD and bipolar disorders. It is accessible through numerous channels that your typical 18 year old can appreciate: on the web at www.whatsmym3.com, at doctors’ offices and the student health clinic at www.M3Clinician.com, and on any smart phone via the app WhatsmyM3. This accessibility means that students may screen themselves, but also track and follow their response to treatment over time, whether that involves psychotherapy or medication or both.

It is vital that we improve our mental health services to our young people in college. Now there is a new way to provide that kind of support. By helping colleges and clinicians find and manage cases of mental illness, it also supports their effort to keep these kids happy, active, engaged and mentally activated.
REFERENCES


D. Employee Assistance Programs: One Solution to Improve Workplace Performance and Lower Health Care Costs

**M3** is an evidenced based solution to integrate the Behavioral and Physical Health sides of Employee Wellness Initiatives. **M3 Clinician** facilitates wellness programs achieving this by engaging workers in the “wellness” setting to let them know their mental health may be placing them at increased risk for physical health problems and connecting them with their doctor. **M3** is reviewed by the American Academy of Family Physicians and in line with programs at and American Hospital Insurance Plans and CMS.

**M3** is a collaborative of former NIMH researchers, business people and clinicians who are developing game changing but practical solutions to improve mental health care benefiting the individual, the clinician and the employer. Specifically this solution focuses on providing a feasible solution to improve people’s mental health and reduce avoidable medical costs. As you will see research from both employer assistance programs and the physical health side both share that improving mental health is the most effective and most cost efficient way to improve well being, improve productivity and to lower healthcare costs.

### Reduce Absenteeism, Boost Productivity and Improve Work Quality

More than 90 percent of employees agree that their mental health and personal problems spill over into their professional lives, and have a direct impact on their job performance. Mental health conditions are actually the second leading cause of absenteeism. Some interesting statistics:

- Untreated and mistreated mental illness costs the United States $150 billion in lost productivity each year, and U.S. businesses foot up to $44 billion of this bill.
- Workplace stress causes about 1 million employees to miss work each day.
- Three out of four employees who seek care for workplace issues or mental health problems see substantial improvement in work performance after treatment.
- According to the RAND Corporation, depression results in more “bed” days than many other medical ailments, including ulcers, diabetes, high blood pressure and arthritis.

### Lower Medical Costs

The relationship between physical and mental health is undeniable. People who have untreated mental health issues use ore general health services than those who seek mental health care when they need it.
This translates to dramatic and unnecessary increases in an organization’s health care bill.

A few alarming facts:

- People with high rates of medical service use have four times the prevalence of depression and anxiety disorders.

- Effective treatment of mental illnesses is associated with improved outcomes for chronic physical disorders.

- 43% of all adults suffer adverse health effects from stress, and stress is linked to the six leading causes of death: heart disease, cancer, lung ailments, accidents, cirrhosis of the liver and suicide. In fact, chronic stress may double the risk of heart attack. Both depression and chronic stress can weaken the immune system and make people vulnerable to a host of illnesses.

- Researchers estimate that 50 to 80 percent of all medical illnesses reported to physicians have a strong emotional or stress-related component.

**M3** is able to implement a process to facilitate whole person care. **M3** is the only patient rated screen that is sensitive to a wide range of diseases and provides longitudinal monitoring to help drive outcomes. The **M3 Clinician** portal facilitates clinicians providing better and more engaged care. This revolutionary approach recognizes real cases and does so in a feasible way that is easy to implement, affordable and engages the patient. The real benefit is to the person, who is a healthier individual and a more productive worker and a patient who does not unduly tax the healthcare system.
III. M3 Validation Study and Related Discussions
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

ABSTRACT

PURPOSE Mood and anxiety disorders are the most common psychiatric conditions seen in primary care, yet they remain underdetected 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.

INTRODUCTION

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

Bradley N. Gaynes, MD, MPH
Joanne DeVeaugh-Geiss, MA, LPA
Sam Weir, MD
Hongbin Gu, PhD
Cora MacPherson, PhD
Herbert C. Schulberg, PhD, MSHyg
Larry Culpepper, MD, MPH
David R. Rubinow, MD

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

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

ANNALS OF FAMILY MEDICINE • WWW.ANNFAMMED.ORG • VOL. 8, NO. 2 • MARCH/APRIL 2010
prescriptions for mood and anxiety disorders in the United States, underrecognition and inadequate treatment of these disorders in primary care practices remain substantial concerns. Tools to improve the identification and management of these disorders are therefore being developed to address this problem.

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), and they do not help to differentiate unipolar from bipolar spectrum illness. The latter has a 3.9% prevalence in community settings, but it can be as high as 9.8% in primary care. 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. Indeed, bipolar patients seen in primary care appear to be at a particularly increased risk of inappropriate treatment.

Also of concern is that depression-screening instruments often do not address anxiety syndromes (28% prevalence), 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, but it is also quite common in the general primary care population, where it is often underrecognized. 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 screens 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 Interview (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.annfammed.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, served as the reference standard to evaluate the performance of the M-3. The MINI is widely used, well accepted, and validated in general medical settings. It has good concordance with other diagnostic measures and can be administered by telephone when in-person interviews are not feasible or practical. 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 patients 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 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 index and the (0, 1) method by Holmes. 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. For depression, anxiety, and PTSD, the Youden index and (0, 1) method 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
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 depressive 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 1. Baseline Characteristics of the Study Sample**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Diagnosis by MINI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Disorder n=423</td>
</tr>
<tr>
<td></td>
<td>(65.4%)</td>
</tr>
<tr>
<td>Mean age (SD) [range], y</td>
<td>46.4 (16.6)</td>
</tr>
<tr>
<td></td>
<td>[18-92]</td>
</tr>
<tr>
<td>Women, %</td>
<td>69.5</td>
</tr>
<tr>
<td>Race, %</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>67.4</td>
</tr>
<tr>
<td>Black</td>
<td>27.4</td>
</tr>
<tr>
<td>Other</td>
<td>5.2</td>
</tr>
<tr>
<td>Marital status, %</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>53.8</td>
</tr>
<tr>
<td>Single</td>
<td>26.2</td>
</tr>
<tr>
<td>Divorced</td>
<td>9.6</td>
</tr>
<tr>
<td>Separated</td>
<td>1.4</td>
</tr>
<tr>
<td>Widowed</td>
<td>4.2</td>
</tr>
<tr>
<td>Living with partner</td>
<td>4.8</td>
</tr>
<tr>
<td>Education, %</td>
<td></td>
</tr>
<tr>
<td>Without high school</td>
<td>2.9</td>
</tr>
<tr>
<td>Diploma or GED</td>
<td>30.6</td>
</tr>
<tr>
<td>High school diploma/GED</td>
<td>13.5</td>
</tr>
<tr>
<td>Associate/technical degree</td>
<td>29.4</td>
</tr>
<tr>
<td>College diploma</td>
<td>23.6</td>
</tr>
<tr>
<td>Graduate degree</td>
<td></td>
</tr>
<tr>
<td>Household’s gross income, %</td>
<td></td>
</tr>
<tr>
<td>≤$14,999</td>
<td>10.9</td>
</tr>
<tr>
<td>$15,000-39,999</td>
<td>24.6</td>
</tr>
<tr>
<td>$40,000-59,999</td>
<td>11.8</td>
</tr>
<tr>
<td>≥$60,000</td>
<td>40.4</td>
</tr>
<tr>
<td>Not know/refused</td>
<td>12.3</td>
</tr>
<tr>
<td>Employment status, %</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>13.8</td>
</tr>
<tr>
<td>Employed</td>
<td>69.3</td>
</tr>
<tr>
<td>Retired, not working</td>
<td>17.0</td>
</tr>
</tbody>
</table>

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

* 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.
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 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 who did, 65% talked about depression, 9% about bipolar illness, 28% about anxiety, and 6% about PTSD.

---

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

<table>
<thead>
<tr>
<th>Test Result</th>
<th>Depresion (n = 142)</th>
<th>Bipolar (n = 60)</th>
<th>Anxiety (n = 182)</th>
<th>PTSD (n = 41)</th>
<th>Any Diagnosis (n = 224)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-3 subscore cutoff (≥)</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>Any positive screen</td>
</tr>
<tr>
<td>Question No.</td>
<td>1-7</td>
<td>20-23</td>
<td>8-19</td>
<td>13-16</td>
<td>1-23</td>
</tr>
<tr>
<td>Sensitivity (95% CI)</td>
<td>0.84 (0.77-0.89)</td>
<td>0.88 (0.77-0.95)</td>
<td>0.82 (0.75-0.87)</td>
<td>0.88 (0.74-0.96)</td>
<td>0.83 (0.77-0.88)</td>
</tr>
<tr>
<td>Specificity (95% CI)</td>
<td>0.80 (0.76-0.83)</td>
<td>0.70 (0.66-0.74)</td>
<td>0.78 (0.74-0.81)</td>
<td>0.76 (0.73-0.80)</td>
<td>0.76 (0.72-0.80)</td>
</tr>
<tr>
<td>Positive LR (95% CI)</td>
<td>4.19 (3.47-5.06)</td>
<td>2.94 (2.53-3.44)</td>
<td>3.65 (3.05-4.39)</td>
<td>3.69 (3.08-4.44)</td>
<td>3.48 (2.90-4.16)</td>
</tr>
<tr>
<td>Negative LR (95% CI)</td>
<td>0.20 (0.14-0.29)</td>
<td>0.17 (0.08-0.33)</td>
<td>0.23 (0.17-0.32)</td>
<td>0.16 (0.07-0.36)</td>
<td>0.22 (0.17-0.30)</td>
</tr>
<tr>
<td>Positive M-3 screen, %</td>
<td>34</td>
<td>35</td>
<td>39</td>
<td>28</td>
<td>44</td>
</tr>
<tr>
<td>Diagnosed by MINI, %</td>
<td>22</td>
<td>9</td>
<td>28</td>
<td>6</td>
<td>35</td>
</tr>
<tr>
<td>Positive predictive value (95% CI) [n/n]</td>
<td>0.54 (0.47-0.61)</td>
<td>0.23 (0.18-0.29)</td>
<td>0.59 (0.53-0.65)</td>
<td>0.20 (0.15-0.27)</td>
<td>0.65 (0.59-0.70)</td>
</tr>
<tr>
<td>Negative predictive value (95% CI) [n/n]</td>
<td>0.95 (0.92-0.96)</td>
<td>0.98 (0.96-0.99)</td>
<td>0.92 (0.88-0.94)</td>
<td>0.99 (0.97-1.0)</td>
<td>0.89 (0.86-0.92)</td>
</tr>
</tbody>
</table>

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

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

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

* Number with a negative screen and who do not have that diagnosis by MINI total number with a negative screen by M-3.
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 Questionnaire (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 scale (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.

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, PTSD, and bipolar disorder 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, is consistent with the rate obtained in other studies of primary care settings, including those conducted in low-income primary care settings. 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.

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 instruments and is greater than some, but not all, 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, and 965 reported with the GAD-7 scale).

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 disorder) or a dimensional approach (mild-severe) toward diagnoses. 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 instruments 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

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 DeVeaux-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. Rubino has received research support from the National Institute of Mental Health.

References


19. Lecrubier Y. Is depression under-recognised and undertreated? In Arch Gen Psychiatry. 2007;64(9):1039.


34. Amsterdam JD, Wang CH, Shwartz 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. 2006;163(2):232-239.


36. Erratum in: Arch Gen Psychiatry. 2007;64(9):1039. 
Specificity of Bipolar Spectrum Conditions in the Comorbidity of Mood and Substance Use Disorders

Results From the Zurich Cohort Study

Kathleen R. Merikangas, PhD; Richard Herrell, PhD; Joel Swendsen, PhD; Wulf Rossler, MD, MSc; Vladeta Ajdacic-Gross, PhD; Jules Angst, MD

Context: Although an association between mood disorders and substance use disorders has been well established, there is a lack of long-term prospective data on the order of onset and subtypes of mood disorders associated with specific substances and their progression. 

Objective: To estimate the respective risks posed by subtypes of mood disorders or bipolar spectrum conditions for the subsequent development of substance use disorders.

Design: Six waves of direct diagnostic interviews were administered to a sample of young adults during a 20-year period. Mood disorders and syndromes assessed at each interview were used to predict the cumulative incidences of substance use disorders at subsequent interview waves.

Participants: We followed up 591 individuals (292 men and 299 women) who were selected at study enrollment from a representative sample of young adults in Zurich, Switzerland.

Main Outcome Measures: Structured Diagnostic Interview for Psychopathologic and Somatic Syndromes, a semistructured clinical interview that collected data on the spectrum of expression of mood disorders and substance use disorders for DSM-III-R and DSM-IV.

Results: Individuals having manic symptoms were at significantly greater risk for the later onset of alcohol abuse/dependence, cannabis use and abuse/dependence, and benzodiazepine use and abuse/dependence. Bipolar II disorder predicted both alcohol abuse/dependence and benzodiazepine use and abuse/dependence. In contrast, major depression was predictive only of later benzodiazepine abuse/dependence. 

Conclusions: In comparison with major depression, bipolar II disorder was associated with the development of alcohol and benzodiazepine use and disorders. There was less specificity of manic symptoms that tended to predict all levels of the substances investigated herein. The different patterns of association between mood disorders and substance use trajectories have important implications for prevention and provide lacking information about underlying mechanisms.

Main Outcome Measures: Structured Diagnostic Interview for Psychopathologic and Somatic Syndromes, a semistructured clinical interview that collected data on the spectrum of expression of mood disorders and substance use disorders for DSM-III-R and DSM-IV.
demonstrated unidirectional patterns either for the order of onset of these disorders \(^9,10,23,24\) or for treatment response and syndrome severity changes as a function of comorbid disorder type. \(^23,25,26\) This absence of uniformity has been attributed to the likelihood that multiple mechanisms of association are simultaneously active. \(^20\) An alternative explanation is that current diagnostic boundaries do not sufficiently differentiate between profiles of comorbidity risk and, specifically, that a more careful separation of major depression from bipolar spectrum conditions may improve the prediction of secondary SUDs. \(^27\)

Unlike the patterns observed for major depression, relatively consistent evidence has been found for causal associations underlying bipolar disorder and alcohol abuse/dependence. Ambulatory monitoring investigations have shown that while increases in depressed mood have little predictive value in explaining daily variation in alcohol use, diverse aroused states such as nervousness or happiness are highly significant predictors of such consumption. \(^28\) Mania has also been shown to predict the later onset of alcohol dependence, \(^9\) and recent epidemiological data suggest that the time to onset of the secondary syndrome is shorter in individuals with this form of comorbidity if the primary condition is bipolar disorder. \(^29\) A direct causal model may, therefore, account for a large percentage of comorbidity between bipolar and alcohol use disorders, and potentially for major depression cases that fulfill broader criteria for bipolar spectrum conditions. This possibility is of particular importance in that a substantial percentage of individuals with lifetime major depression manifest bipolar disorder. \(^30,35\)

Bipolar spectrum conditions are frequent in the general population \(^36\) and remain associated with substance use disorders at levels below diagnostic thresholds. \(^37\) However, the precise nature of these associations remains uncertain and may reflect several different underlying mechanisms. Aroused affect and expansive temperament characteristics, for example, may be largely specific to bipolar spectrum conditions while increasing the risk of abuse/dependence for a wide range of substances. This notion is consistent both with elevated risk behaviors and novelty seeking in patients with bipolar disorder \(^38,39\) and the pharmacologic properties of a given substance. This latter possibility is supported by observations that bipolar disorder may predict the onset of some forms of substance dependence \(^40\) while having a fully opposite temporal relationship with others. \(^41\) Discriminat-

ing between these alternatives, therefore, requires the prospective study of multiple classes of psychoactive substances in association with both major depression and bipolar spectrum conditions.

Despite considerable research on this topic, findings from epidemiological investigations are characterized by near-complete reliance on retrospective estimates of disorder onset, and only a portion of studies have examined mood disorders in relation to multiple forms of SUD. The present prospective study was undertaken in an attempt to address these issues using the community-based sample of the Zurich Cohort Study. Young adults were interviewed directly in their homes 6 times during a 20-year period. Mood and SUDs, as well as subthreshold manifestations of these diagnostic categories, were assessed at each interview wave. The primary objectives were to examine the degree to which major depression and bipolar spectrum conditions may increase the risk of later onset of various forms of SUD and to assess the degree to which the observed patterns of comorbidity correspond to specific models of association.

**STUDY POPULATION**

The Zurich Cohort Study is composed of a subset of 4547 individuals (2201 males and 2346 females) representative of the canton of Zurich in Switzerland who were screened using the Symptom Checklist-90 (SCL-90)–Revised. \(^42\) The screening occurred in 1978, when male participants were aged 19 years and female participants were aged 20 years. To enhance the probability of cases, a 2-stage sampling procedure was used to identify 591 subjects (292 men and 299 women), two-thirds of whom scored at the 85th percentile or higher on the SCL-90 (high scorers) and one-third of whom were a random sample of individuals who scored below the 85th percentile. \(^43,44\)

All subjects who met inclusion and exclusion criteria provided written informed consent or assent as appropriate, after receiving a complete description of the study and having an opportunity to ask questions. Diagnostic interviews were conducted in 1979, 1981, 1986, 1988, 1993, and 1999. Of the initial sample, 91% participated in at least 2 interviews; 82% in at least 3 interviews; 74% in at least 4 interviews; and 63% in at least 5 interviews. Sixty-nine percent of the original sample remained in the cohort during the 20-year study; 424 and 367 individuals participated in the 10- and 20-year interview waves, respectively (Table 1). Those who dropped out did not differ significantly in baseline measures of demographic characteristics or status at study enrollment, and no differences were observed in dropout rates for high and low

**MEASURES**

The diagnostic instrument used in the Zurich study was the SCL-90. \(^45\) Data were weighted to yield estimates of the diagnostic prevalence of bipolar spectrum conditions in the Zurich Cohort Study population.
mantic Syndromes (SPIKE), a semistructured instrument that was developed for epidemiological studies. Interviewers for the Zurich study were professional psychologists or psychiatrists with extensive experience in psychopathologic diagnosis and treatment. Initial training of interviewers involved group sessions to establish interrater reliability. After achieving an acceptable level of agreement, pairs of interviewers conducted sessions to establish pairwise reliability. Reliability was also maintained through periodic group sessions over the duration of the study. The interrater reliability of the SPIKE was K = 0.89 and 0.91 for the symptoms of depression and anxiety (generalized anxiety disorder) and K = 0.90 for the corresponding syndromal diagnoses. The validity of the SPIKE has also been assessed by comparing physician ratings and medical records from local inpatient and outpatient settings with SPIKE diagnoses established by independent clinicians. High and moderate specificity levels were found for both threshold and subthreshold depression (0.90), anxiety (0.83), and mania (0.67).

Information was collected on childhood characteristics, treatment history, psychiatric and somatic syndromes, and use and abuse of various substances. Symptoms, duration and frequency, subjective degree of distress, treatment, social consequences, history, age at onset, and family history were assessed for each syndrome. The core phenomenological probe for all of the syndromes was asked for each of the interim years between interviews in order to cover the entire assessment period. Screening probes were based solely on the major phenomenological features of each syndrome and were administered for each diagnostic category. Positive endorsement of the screening probe was followed by queries about specific symptoms and then by questions about symptom duration, frequency, and severity; treatment history; and impairment in work, social, and leisure activities. With this approach, diagnostic criteria for subthreshold syndromes and multiple systems of nosology could be applied. The classification of psychiatric disorders was based on the DSM-IV criteria for major depressive disorder, bipolar II (BP-II) disorder, and substance abuse/dependence. We also included subthreshold bipolar conditions that required either a reduced duration (<4 days) or number of symptoms (<4 symptoms) with regard to the complete criteria for each disorder.

ESTIMATION OF CUMULATIVE INCIDENCE

For analyses of cumulative incidence, mood disorders were classified according to a 4-level hierarchical variable, as follows: no mood disorder, major depressive disorder, manic symptoms, and BP-II disorder. With time, the highest level reached by each person was carried forward. For example, those who fulfilled criteria for manic/hypomanic symptoms at wave 3 and major depression at wave 4 were classified as having manic/hypomanic symptoms at wave 4. The frequency of respondents without a history of alcohol use was so small that the non-use and use groups were combined for the initial level of alcohol exposure in trajectory analyses. Similarly, because of the small overall number of cases of cannabis and benzodiazepine abuse/dependence, these disorder categories were combined for each substance. Substance use disorders were, therefore, classified as follows: alcohol nonuse, abuse or dependence; cannabis non-use, use, or abuse/dependence; and benzodiazepine nonuse, use, or abuse/dependence. At each time point, the dependent variable (substance use or abuse/dependence) was set to the highest value reached by the subject during each wave up to and including that wave. The principal independent variables (mood syndromes) represent the highest diagnostic level reached by the subject during each wave. The generalized ordinal logistic model used here (see the "Statistical Analysis" section) estimates odds ratios (ORs) for the association over these data points for each individual. Participants who met criteria for substance dependence at the initial interview were excluded from the analysis, including 2 individuals with alcohol dependence, 5 with benzodiazepine abuse/dependence, and 21 with cannabis abuse/dependence.

STATISTICAL ANALYSIS

A generalized ordinal logistic model was used to estimate the prospective effect of mood disorders on the development of subsequent SUDs. Unlike the proportional odds, this method estimates an equation for each cut point in the outcome measure, allowing the effect of the predictor variable to vary across the levels of the ordinal measure. Thus, separate measures of effect (ORs) can be estimated for each level, such as abuse/dependence, and these cut points can be understood as diagnostic thresholds. The OR for the alcohol abuse threshold, therefore, estimates the effect on abuse/dependence compared with nonuse or use; the effect for the dependence threshold estimates the effect on crossing the dependence threshold compared with nonuse, use, or abuse. The predictors were entered as indicator variables, setting no disorder as the referent category. The final models given herein include time (number of years in the study) and sex. Missing values were replaced with the last observed level for each variable that would bias the results negatively if subjects reached a higher level in a wave for which they were not assessed. Sampling weights were incorporated to account for the design, and robust variance estimates were calculated to account for the repeated measures. Analyses were conducted using commercially available software: Stata (StataCorp LP, College Station, Texas). The unweighted counts of participants meeting criteria for each of the mood disorder categories were as follows: no disorder (n = 315), unipolar depression (n = 87), manic symptoms (n = 148), and BP-II disorder (n = 41). The percentages of individuals in the cohort affected by mood and SUDs are given in Table 2. The cumulative incidence of major depressive disorder by 1993 was 9.7% of the total sample, and while no individual met criteria for bipolar I disorder by that year, 4.4% met criteria for BP-II and 23.5% had manic symptoms. Of the total sample, 17.9% met criteria for alcohol abuse/dependence by the last interview wave in 1999, 8% met criteria for cannabis abuse/dependence, and 3.4% met criteria for benzodiazepine abuse/dependence. The cumulative incidence of cannabis use never surpassed 18% of the cohort and remained relatively stable since the 1986 interview. In contrast, a small but steady progression in benzodiazepine use was observed at each wave, and most of the cohort (83.2%) had used benzodiazepines by the 1999 interview.

MOOD DISORDERS

AS PREDICTORS OF SUD COMORBIDITY

Table 3 gives the generalized ordinal logistic model used to estimate the effect of mood disorders on subsequent substance use disorders. These associations were not significant for major depression and all categories of SUDs...
with the exception of benzodiazepine abuse/dependence (OR, 13.2; 95% confidence interval [CI], 2.56-67.67). In contrast, all of the substances examined manic symptoms were highly consistent predictors of use or abuse/dependence. The magnitude of risk for later alcohol dependence (OR, 4.44; 95% CI, 1.55-12.73) was nearly twice that of alcohol abuse (OR, 2.41; 95% CI, 1.22-4.76), although the overall number of persons with manic symptoms who developed later alcohol abuse was greater than the number with alcohol dependence. The development of later cannabis abuse/dependence (OR, 4.82; 95% CI, 2.01-11.60) and benzodiazepine abuse/dependence (OR, 11.54; 95% CI, 4.80-27.78) was also significantly predicted by manic symptoms, as was general use of both substances. The role of DSM-IV BP-II in predicting later alcohol abuse (OR, 9.11; 95% CI, 2.66-31.21) and dependence (OR, 21.05; 95% CI, 6.57-67.47), as well as benzodiazepine use (OR, 6.86; 95% CI, 1.95-24.19) and abuse/dependence (OR, 14.10; 95% CI, 2.65-75.09) in contrast to manic symptoms, this category of mood disorder did not predict earlier development of cannabis abuse or dependence.

### Table 2. Cumulative Incidence of Mood Disorders and Substance Use Disorders

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mood</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major depressive disorder</td>
<td>2.5</td>
<td>4.9</td>
<td>6.7</td>
<td>9.3</td>
<td>10.3</td>
</tr>
<tr>
<td>Manic symptoms</td>
<td>1.8</td>
<td>17.0</td>
<td>22.4</td>
<td>23.5</td>
<td>24.2</td>
</tr>
<tr>
<td>Bipolar II disorder</td>
<td>1.2</td>
<td>3.1</td>
<td>4.2</td>
<td>4.4</td>
<td>4.5</td>
</tr>
<tr>
<td><strong>Substance Use</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abuse</td>
<td>10.6</td>
<td>10.6</td>
<td>8.4</td>
<td>9.2</td>
<td>9.5</td>
</tr>
<tr>
<td>Dependence</td>
<td>0.2</td>
<td>2.0</td>
<td>6.4</td>
<td>8.7</td>
<td></td>
</tr>
<tr>
<td>Cannabis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use</td>
<td>18.1</td>
<td>17.4</td>
<td>17.0</td>
<td>17.9</td>
<td>18.0</td>
</tr>
<tr>
<td>Abuse/dependence</td>
<td>6.4</td>
<td>7.4</td>
<td>7.9</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>Benzodiazepines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use</td>
<td>74.3</td>
<td>76.2</td>
<td>81.7</td>
<td>83.2</td>
<td>83.5</td>
</tr>
<tr>
<td>Abuse/dependence</td>
<td>2.5</td>
<td>3.5</td>
<td>3.6</td>
<td>3.6</td>
<td></td>
</tr>
</tbody>
</table>

*Values are given as weighted percentage.

### Table 3. Mood Disorders as Predictors of Alcohol, Cannabis, and Benzodiazepine Use, Abuse, and Dependence

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Abuse</th>
<th>Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major depressive disorder</td>
<td>1.8 (0.6-2.9)</td>
<td>2.2 (0.7-7.2)</td>
</tr>
<tr>
<td>Manic symptoms</td>
<td>2.4 (1.2-4.8)</td>
<td>4.4 (1.6-12.7)</td>
</tr>
<tr>
<td>Bipolar II disorder</td>
<td>9.1 (2.7-31.2)</td>
<td>21.1 (6.6-67.5)</td>
</tr>
<tr>
<td>No. of years</td>
<td>1.0 (1.0-1.1)</td>
<td>1.1 (1.1-1.2)</td>
</tr>
<tr>
<td>Male sex</td>
<td>6.3 (3.0-13.4)</td>
<td>12.5 (5.1-30.5)</td>
</tr>
<tr>
<td>Cannabis Use</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major depressive disorder</td>
<td>1.5 (0.7-3.6)</td>
<td>2.3 (0.7-6.9)</td>
</tr>
<tr>
<td>Manic symptoms</td>
<td>2.2 (1.2-4.1)</td>
<td>4.8 (2.0-11.6)</td>
</tr>
<tr>
<td>Bipolar II disorder</td>
<td>2.1 (0.8-6.0)</td>
<td>0.8 (0.2-3.7)</td>
</tr>
<tr>
<td>No. of years</td>
<td>1.0 (1.0-1.0)</td>
<td>1.0 (1.0-1.0)</td>
</tr>
<tr>
<td>Male sex</td>
<td>2.4 (1.4-4.3)</td>
<td>3.1 (1.0-9.4)</td>
</tr>
<tr>
<td>Benzodiazepines</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major depressive disorder</td>
<td>2.7 (0.7-10.2)</td>
<td>13.2 (2.6-67.7)</td>
</tr>
<tr>
<td>Manic symptoms</td>
<td>3.6 (1.3-10.0)</td>
<td>11.9 (2.7-52.8)</td>
</tr>
<tr>
<td>Bipolar II disorder</td>
<td>11.9 (2.7-52.8)</td>
<td>14.1 (2.7-75.0)</td>
</tr>
<tr>
<td>No. of years</td>
<td>1.0 (1.0-1.1)</td>
<td>1.0 (1.0-1.1)</td>
</tr>
<tr>
<td>Male sex</td>
<td>0.3 (0.1-0.5)</td>
<td>0.7 (0.2-2.6)</td>
</tr>
</tbody>
</table>

*Values are given as odds ratio (95% confidence interval).

### Comment

Past epidemiological research on the comorbidity of mood disorders and SUDs has been based almost exclusively on cross-sectional methods or retrospective estimates of disorder onset,1-6,9-12,14 and few studies of any type have attempted to differentiate subthreshold bipolar conditions from major depression. Using longitudinal data and mutually exclusive categories of mood disorder, the present findings confirm the results of retrospective clinical and community studies of the strong association between bipolar disorder and the development of alcohol use disorders.1-6,9-12,14 Moreover, the presence of manic symptoms was found to constitute a comprehensive risk factor for the future development of all categories of substance use or abuse/dependence investigated in this study. In comparison with major depression, the widespread association between manic symptoms and the full trajectory of all 3 substance categories provides new evidence for the specificity of bipolar disorder in the development of SUDs and further suggests that manic symptoms, even when below current diagnostic thresholds for hypomania, may be a powerful risk factor for these disorders. The patterns of association observed when carefully separating bipolar spectrum conditions from major depression also differed in important ways from previous population-based surveys. In particular, major depression was associated with an increased risk only of benzodiazepine abuse/dependence during the 2 decades covered by the study. The lack of association between major depression and later alcohol use disorders is in contrast to prospective epidemiological research that did not control for the influence of bipolar spectrum conditions and52 may help clarify the inconsistent findings reported for this form of comorbidity.20 However, this consistency with previous observations in community samples of the lack of association between major depression and later cannabis abuse/dependence as many investigations have, therefore, focused on causal explanations for these forms of comor-
bidity. Among the potential causal mechanisms to be considered, the risk for several forms of mood disorder posed by bipolar diagnoses and manic symptoms indicates that self-medication is potentially less appropriate than more global explanatory models of association. The notion of self-medication assumes a contrast between the effects induced by a given substance and the nature of psychiatric symptoms experienced by the patient, therefore encouraging more targeted use of specific drug classes. Psychoeducation and other treatment approaches may, therefore, benefit from alternative models of substance use prevention in this population. In particular, these approaches may focus on reducing the use of diverse substances on the basis of the broad risks posed by temperamental and personality characteristics in persons with this disorder, certain of which also aggregate in substance-dependent samples.

This same causal model implicating behavioral, temperamental, or personality risk factors does not seem applicable to major depression, given its limited prospective association with most SUD categories. However, when compared with mutually exclusive groups of individuals with BP-II disorder or manic symptoms, major depression was independently associated with the later development of benzodiazepine abuse/dependence. In understanding this association, it is important to note that the percentage of individuals prescribed benzodiazepines is often greater than 50% in samples with mood disorders and that a greater percentage of individuals prescribed benzodiazepines subsequently develop abuse compared with those who acquire this substance through other means. The significant association of both bipolar disorder and major depression with benzodiazepine abuse/dependence, therefore, may be attributable in part to the lack of association in BP-II disorder with benzodiazepine abuse/dependence, as observed in the present findings. In contrast, as symptoms of BP-II disorder, the BP-I category is more closely related to impulsive acts for the worship of mood-producing substances and, the moderate sample sizes may have prevented the detection of smaller but clinically pertinent effects. The cohort was assessed through age 40 years, but the prospective nature of analyses limited the last assessment of mood disorders to age 35 years. While this is past the age of greatest risk for initial onset of most forms of mood disorder, it is possible that individuals fulfilling criteria, notably for BP-I, will be identified in future assessments. For these reasons, the findings may not be applicable to older cohorts and are limited to the clinical disorders and conditions described in the “Methods” section. The present method also does not permit the examination of specific factors such as impulsivity or other personality traits separately from bipolar spectrum conditions. The strategy for weighting of coefficients and the use of robust variance estimation yielded results that are representative of the reference population, but power remains limited for detecting the influence of sex and other individual differences. The analyses were also designed to examine the risk posed by mood disorders for the later onset of SUDs, with temporally primary manic symptoms and BP-II disorder, but bidirectional associations remain possible and should be examined in subsequent investigations. However, the emergence of mood syndromes after the onset of SUDs poses important questions about the nature of these clinical entities.

In terms of clinical implications, the prospective links observed between bipolar spectrum conditions and subsequent alcohol use disorders support recent interventions that incorporate both disorders in developing treatment strategies. The present results further indicate that such treatment programs should be expanded to include a diversity of substance classes. Insofar as prevention of SUDs, consideration of BP-II disorder and manic symptoms is clearly merited, and early intervention in the form of psychoeducation and regular screening of substance use practices may be particularly beneficial in individuals with subthreshold manic symptoms. Although Angst et al demonstrated that 16% of cohort participants with SUDs and 55% of those with mood disorders received treatment, the effect of services on the transitional to middle adulthood. In particular, these approaches may focus on reducing the use of diverse substances on the basis of the broad risks posed by temperamental and personality characteristics in persons with this disorder.

LIMITATIONS AND IMPLICATIONS

The significant association of both bipolar disorder and major depression with benzodiazepine abuse/dependence, therefore, may be attributable in part to the lack of association in BP-II disorder with benzodiazepine abuse/dependence, as observed in the present findings. In contrast, as symptoms of BP-II disorder, the BP-I category is more closely related to impulsive acts for the worship of mood-producing substances and, the moderate sample sizes may have prevented the detection of smaller but clinically pertinent effects. The cohort was assessed through age 40 years, but the prospective nature of analyses limited the last assessment of mood disorders to age 35 years. While this is past the age of greatest risk for initial onset of most forms of mood disorder, it is possible that individuals fulfilling criteria, notably for BP-I, will be identified in future assessments. For these reasons, the findings may not be applicable to older cohorts and are limited to the clinical disorders and conditions described in the “Methods” section. The present method also does not permit the examination of specific factors such as impulsivity or other personality traits separately from bipolar spectrum conditions. The strategy for weighting of coefficients and the use of robust variance estimation yielded results that are representative of the reference population, but power remains limited for detecting the influence of sex and other individual differences. The analyses were also designed to examine the risk posed by mood disorders for the later onset of SUDs, with temporally primary manic symptoms and BP-II disorder, but bidirectional associations remain possible and should be examined in subsequent investigations. However, the emergence of mood syndromes after the onset of SUDs poses important questions about the nature of these clinical entities.

In terms of clinical implications, the prospective links observed between bipolar spectrum conditions and subsequent alcohol use disorders support recent interventions that incorporate both disorders in developing treatment strategies. The present results further indicate that such treatment programs should be expanded to include a diversity of substance classes. Insofar as prevention of SUDs, consideration of BP-II disorder and manic symptoms is clearly merited, and early intervention in the form of psychoeducation and regular screening of substance use practices may be particularly beneficial in individuals with subthreshold manic symptoms. Although Angst et al demonstrated that 16% of cohort participants with SUDs and 55% of those with mood disorders received treatment, the effect of services on the transitional to middle adulthood. In particular, these approaches may focus on reducing the use of diverse substances on the basis of the broad risks posed by temperamental and personality characteristics in persons with this disorder.

LIMITATIONS AND IMPLICATIONS

The significant association of both bipolar disorder and major depression with benzodiazepine abuse/dependence, therefore, may be attributable in part to the lack of association in BP-II disorder with benzodiazepine abuse/dependence, as observed in the present findings. In contrast, as symptoms of BP-II disorder, the BP-I category is more closely related to impulsive acts for the worship of mood-producing substances and, the moderate sample sizes may have prevented the detection of smaller but clinically pertinent effects. The cohort was assessed through age 40 years, but the prospective nature of analyses limited the last assessment of mood disorders to age 35 years. While this is past the age of greatest risk for initial onset of most forms of mood disorder, it is possible that individuals fulfilling criteria, notably for BP-I, will be identified in future assessments. For these reasons, the findings may not be applicable to older cohorts and are limited to the clinical disorders and conditions described in the “Methods” section. The present method also does not permit the examination of specific factors such as impulsivity or other personality traits separately from bipolar spectrum conditions. The strategy for weighting of coefficients and the use of robust variance estimation yielded results that are representative of the reference population, but power remains limited for detecting the influence of sex and other individual differences. The analyses were also designed to examine the risk posed by mood disorders for the later onset of SUDs, with temporally primary manic symptoms and BP-II disorder, but bidirectional associations remain possible and should be examined in subsequent investigations. However, the emergence of mood syndromes after the onset of SUDs poses important questions about the nature of these clinical entities.

In terms of clinical implications, the prospective links observed between bipolar spectrum conditions and subsequent alcohol use disorders support recent interventions that incorporate both disorders in developing treatment strategies. The present results further indicate that such treatment programs should be expanded to include a diversity of substance classes. Insofar as prevention of SUDs, consideration of BP-II disorder and manic symptoms is clearly merited, and early intervention in the form of psychoeducation and regular screening of substance use practices may be particularly beneficial in individuals with subthreshold manic symptoms. Although Angst et al demonstrated that 16% of cohort participants with SUDs and 55% of those with mood disorders received treatment, the effect of services on the transitional to middle adulthood.

Funding/Support: This study was supported by grant 3200-050881.97/1 from the Swiss National Science Foundation.

REFERENCES


The US Preventive Services Task Force (USPSTF) makes recommendations about the effectiveness of specific preventive care services for patients without related signs or symptoms. It bases its recommendations on the evidence of both the benefits and harms of the service and an assessment of the balance. The USPSTF does not consider the costs of providing a service in this assessment. The USPSTF recognizes that clinical decisions involve more considerations than evidence alone. Clinicians should understand the evidence but individualize decision making to the specific patient or situation. Similarly, the USPSTF notes that policy and coverage decisions involve considerations in addition to the evidence of clinical benefits and harms.

Summary of Recommendation and Evidence

The USPSTF recommends screening for depression in the general adult population, including pregnant and postpartum women. Screening should be implemented with adequate systems in place to ensure accurate diagnosis, effective treatment, and appropriate follow-up. (B recommendation) (Figure 1)

Rationale

Importance
Depression is among the leading causes of disability in persons 15 years and older. It affects individuals, families, businesses, and society and is common in patients seeking care in the primary care setting. Depression is also common in postpartum and pregnant women and affects not only the woman but her child as well.

Detection
The USPSTF found convincing evidence that screening improves the accurate identification of adult patients with depression in primary care settings, including pregnant and postpartum women.

Benefits of Early Detection and Intervention and Treatment
The USPSTF found adequate evidence that programs combining depression screening with adequate support systems in place improve clinical outcomes (ie, reduction or remission of depression symptoms) in adults, including pregnant and postpartum women.

Harms of Early Detection and Intervention and Treatment
The USPSTF found convincing evidence that treatment of adults and older adults with depression identified through screening in primary care settings with antidepressants, psychotherapy, or both decreases clinical morbidity.

The USPSTF also found adequate evidence that treatment with cognitive behavioral therapy (CBT) improves clinical outcomes in pregnant and postpartum women with depression.

Harms of Early Detection and Intervention and Treatment
The USPSTF found adequate evidence that the magnitude of harms of screening for depression in adults is small to none.

The USPSTF found adequate evidence that the magnitude of harms of treatment with CBT in postpartum and pregnant women is small to none.
The USPSTF found that second-generation antidepressants (mostly selective serotonin reuptake inhibitors [SSRIs]) are associated with some harms, such as an increase in suicidal behaviors in adults aged 18 to 29 years and an increased risk of upper gastrointestinal bleeding in adults older than 70 years, with risk increasing with age; however, the magnitude of these risks is, on average, small. The USPSTF found evidence of potential serious fetal harms from pharmacologic treatment of depression in pregnant women, but the likelihood of these serious harms is low. Therefore, the USPSTF concludes that the overall magnitude of harms is small to moderate.

**USPSTF Assessment**
The USPSTF concludes with at least moderate certainty that there is a moderate net benefit to screening for depression in adults, including older adults, who receive care in clinical practices that have adequate systems in place to ensure accurate diagnosis, effective treatment, and appropriate follow-up after screening (Figure 1). The USPSTF also concludes with at least moderate certainty that there is a moderate net benefit to screening for depression in pregnant and postpartum women who receive care in clinical practices that have CBT or other evidence-based counseling available after screening.

**Clinical Considerations**
**Patient Population Under Consideration**
This recommendation applies to adults 18 years and older (Figure 2). It does not apply to children and adolescents, who are addressed in a separate USPSTF recommendation statement (available at http://www.uspreventive servicestaskforce.org).
Assessment of Risk

The USPSTF recommends screening in all adults regardless of risk factors. However, a number of factors are associated with an increased risk of depression. Among general adult populations, prevalence rates vary by sex, age, race/ethnicity, education, marital status, geographic location, and employment status. Women, young and middle-aged adults, and nonwhite persons have higher rates of depression than their counterparts, as do persons who are undereducated, previously married, or unemployed. Persons with chronic illnesses, other mental health disorders, or a family history of psychiatric disorders are also at increased risk.

Risk factors in older adults include disability and poor health status related to medical illness, complicated grief, chronic sleep disturbance, loneliness, and history of depression. Risk factors during pregnancy and postpartum include poor self-esteem, child-care stress, prenatal anxiety, life stress, decreased social support, single/unpartnered relationship status, history of depression, difficult infant temperament, previous postpartum depression, lower socioeconomic status, and unintended pregnancy.

Screening Tests

Commonly used depression screening instruments include the Patient Health Questionnaire in various forms and the Hospital Anxiety and Depression Scales in adults, the Geriatric Depression Scale in older adults, and the Edinburgh Postnatal Depression Scale in postpartum and pregnant women. Positive screening results should lead to additional assessment that considers severity of depression and comorbid psychological problems, alternate diagnoses, and medical conditions.

Screening Interval

The optimal timing and interval for screening for depression is not known. A pragmatic approach might include screening all adults who have not been screened previously and using clinical judgment in consideration of risk factors, comorbid conditions, and life events to determine if additional screening of high-risk patients is warranted.

Treatment and Interventions

Effective treatment of depression in adults generally includes antidepressants or specific psychotherapy approaches, alone or in combination. Given the potential harms to the fetus and newborn child from certain pharmacologic agents, clinicians are encouraged to consider evidence-based counseling interventions when managing depression in pregnant or breastfeeding women.

Balance of Benefits and Harms

The net benefit of screening for depression in the general adult population is moderate.

Other Relevant USPSTF Recommendations

The USPSTF has made recommendations on screening for depression in children and adolescents and screening for suicide risk in adolescents, adults, and older adults. These recommendations are available on the USPSTF website (http://www.uspreventiveservicestaskforce.org).

Risk factors for depression during pregnancy and postpartum include poor self-esteem, child-care stress, prenatal anxiety, life stress, decreased social support, single/unpartnered relationship status, history of depression, difficult infant temperament, previous postpartum depression, lower socioeconomic status, and unintended pregnancy.

Commonly used depression screening instruments include the Patient Health Questionnaire (PHQ) in various forms and the Hospital Anxiety and Depression Scales in adults, the Geriatric Depression Scale in older adults, and the Edinburgh Postnatal Depression Scale (EPDS) in postpartum and pregnant women. Positive screening results should lead to additional assessment that considers severity of depression and comorbid psychological problems (eg, anxiety, panic attacks, or substance abuse), alternate diagnoses, and medical conditions.

Screening Timing and Interval

There is little evidence regarding the optimal timing for screening. The optimum interval for screening for depression is also unknown; more evidence for all populations is needed to identify.
ideal screening intervals. A pragmatic approach in the absence of data might include screening all adults who have not been screened previously and using clinical judgment in consideration of risk factors, comorbid conditions, and life events to determine if additional screening of high-risk patients is warranted.

Treatment
Effective treatment of depression in adults generally includes antidepressants or specific psychotherapy approaches (eg, CBT or brief psychosocial counseling), alone or in combination. Given the potential harms to the fetus and newborn child from certain pharmacologic agents, clinicians are encouraged to consider CBT or other evidence-based counseling interventions when managing depression in pregnant or breastfeeding women.

Other Approaches to Prevention
The Community Preventive Services Task Force, which makes evidence-based recommendations on preventive services for community populations, recommends collaborative care for the management of depressive disorders as part of a multicomponent, health care system–level intervention that uses case managers to link primary care providers, patients, and mental health specialists. More information about the Community Preventive Services Task Force and its recommendations on depression interventions is available on its website (http://www.thecommunityguide.org).

Useful Resources
The USPSTF has made recommendations on screening for depression in children and adolescents and screening for suicide risk in adolescents, adults, and older adults (available at http://www.uspreventiveservicestaskforce.org).

The Substance Abuse and Mental Health Services Administration maintains a national registry of evidence-based programs and practices for substance abuse and mental health interventions (http://nrepp.samhsa.gov/) that may be helpful for clinicians looking for models of how to implement depression screening.

Other Considerations
Implementation
The USPSTF recommends that screening be implemented with adequate systems in place. “Adequate systems in place” refers to having systems and clinical staff to ensure that patients are screened and, if they screen positive, are appropriately diagnosed and treated with evidence-based care or referred to a setting that can provide the necessary care. These essential functions can be provided through a wide range of different arrangements of clinician types and settings. In the available evidence, the lowest effective level of support consisted of a designated nurse who advised resident physicians of positive screening results and provided a protocol that facilitated referral to evidence-based behavioral treatment. At the highest level, support included screening; staff and clinician training (1- or 2-day workshops); clinician manuals; monthly training lectures; academic detailing; materials for clinicians, staff, and patients; an initial visit with a nurse specialist for assessment, education, and discussion of patient preferences and goals; a visit with a trained nurse specialist for follow-up assessment and ongoing support for medication adherence; a visit with a trained therapist for CBT; and a reduced copayment for patients referred for psychotherapy.

Multidisciplinary team–based primary care that includes self-management support and care coordination has been shown to be effective in management of depression.

Other Components of Primary Care
These components of primary care are detailed in recommendations from the Community Preventive Services Task Force. It recommends collaborative care for the treatment of major depression in adults 18 years and older on the basis of strong evidence of effectiveness in improving short-term treatment outcomes. As defined, collaborative care and disease management of depressive disorders include a systematic, multicomponent, and team-based approach that “strengthens and supports self-care, while assuring that effective medical, preventive, and health maintenance interventions take place” to improve the quality and outcome of patient care.

Costs
The economic burden of depression is substantial for individuals as well as society. Costs to an individual may include emotional suffering, reduced quality of personal relationships, possible adverse effects from treatment, cost of mental health and medical visits and medications, time away from work and lost wages, and cost of transportation. Costs to society may include loss of life, reduced productivity (because of both diminished capacity while at work and absenteeism from work), and increased costs of mental health and medical care.

Research Needs and Gaps
Gaps in the evidence on screening for depression in older adults in primary care include a lack of information from large-scale randomized controlled trials (RCTs) in settings that are applicable to the US population. More research is needed on the accuracy of screening tools in languages other than English and Spanish and to identify the timing and optimal screening interval in all populations. Data are lacking on both the accuracy of screening and the benefits and harms of treatment in pregnant women, as well as for the balance of benefits and harms of treatment with antidepressants in postpartum women. Finally, research is needed to assess barriers to establishing adequate systems of care and how these barriers can be addressed.

Discussion
Burden of Disease
Major depressive disorder (MDD) is a common and significant health care problem. It is the leading cause of disability among adults in high-income countries and is associated with increased mortality due to suicide and impaired ability to manage other health issues. Depression has a major effect on quality of life for the patient and affects family members, especially children. Depression also imposes a significant economic burden through direct and indirect costs. In the United States, an estimated $22.8 billion was spent on depression treatment in 2009, and lost productivity cost an additional estimated $23 billion in 2011.
Scope of Review

The USPSTF commissioned a systematic evidence review to update its 2009 recommendation, which focused on the direct evidence of the benefits and harms of screening for depression in adult populations, including older adults and pregnant and postpartum women. The USPSTF also reviewed the evidence on the accuracy of screening depression instruments and the benefits and harms of screening in pregnancy.

Accuracy of Screening Tests

General Adult Population and Older Adults

The accuracy of screening tests in the general adult population was established in the 2002 and 2009 USPSTF reviews and found to be convincing.

Pregnant and Postpartum Women

Twenty-three studies (n = 5398), including 8 studies of the English-language version, compared the accuracy of the EPDS with a diagnostic interview. Sensitivity of the English-language EPDS with a cutoff score of 13 ranged from 0.67 (95% CI, 0.18-0.96) to 1.00 (95% CI, 0.67-1.00), and specificity for detecting MDD was consistently at least 0.90. In the 2 trials conducted in the United States, including a recent study in low-income African American women, sensitivity for detecting MDD ranged from 0.78 to 0.81. This suggests that the average sensitivity of the EPDS with a cutoff score of 13 in the United States is approximately 0.80, and the positive predictive value for detecting MDD would be 47% to 64% in a population with a 10% prevalence of MDD. The Spanish-language version also showed acceptable performance characteristics. No studies of screening in pregnant and postpartum women with the 9-item PHQ or other versions met inclusion criteria.

Effectiveness of Screening and Treatment

General Adult Population and Older Adults

Nine good- or fair-quality trials addressed screening in general adults (5 trials; n = 2924) and older adults (4 trials; n = 890). Seven studies were conducted in the United States, and 2 (in older adults) were conducted in the Netherlands. Most studies were published in the 1990s and early 2000s; only 1 (in older adults) of the 9 trials was published since the previous systematic review. One study in general adults directly compared screening with usual care case-finding, while the other studies screened all patients for depression, enrolled only those screening positive, and returned results of screening to clinicians in the intervention group only. Studies included a range of additional treatment components along with providing screening result feedback to clinicians.

Improvements in remission, response rates, or both in the general adult population ranged from 17% to 87%. Other outcomes were sparsely reported. The effect of screening on remission, response rates, or both in the trials of older adults was minimal. However, both of the trials in older adults that showed a paradoxical effect were conducted in the Netherlands, and the trial with the worst outcomes had a number of features that may have affected its reliability, including external referrals for depression treatment, very low uptake of treatment (19%), and high mortality and morbidity in the intervention group, suggesting that the control and intervention groups may have been different at baseline.

The 2009 USPSTF recommendation concluded that the evidence was sufficient to establish the benefits of treatment of depression in general adult populations, including older adults. A systematic review of intention-to-treat trials comparing 3 groups of adult patients who received antidepressants, psychotherapy, or a control condition reported a 46% remission rate with antidepressants and a 48% remission rate with psychotherapy after 10 to 16 weeks. Two systematic reviews concluded that antidepressants were effective in treating depression in older adults. In 1 review, older adults who received antidepressants were twice as likely to have remission from major or minor depression as older adults who received placebo (odds ratio [OR], 2.03 [95% CI, 1.67-2.46]). The other review indicated that among community-dwelling older adults, 36% of those who received antidepressants were in remission at the end of the study compared with 21% of those who received placebo (OR, 2.13 [95% CI, 1.61-2.86]). In addition, 2 good-quality systematic reviews on the efficacy of psychotherapy in older adults found that older adults who received psychotherapy were more than twice as likely to have remission as those who received no treatment (OR, 2.47 [95% CI, 1.76-3.47] vs 2.63 [95% CI, 1.96-3.53]).

Pregnant and Postpartum Women

The USPSTF identified 6 fair- or good-quality trials (n = 11 869) (5 in postpartum women and 1 in pregnant women) that assessed the effect of screening for depression in pregnant and postpartum women. Trial participants were identified through primary care settings using the EPDS (cutoff scores varied) and included women with and without depression. None of the trials simply compared usual care with screening plus usual care. Two trials assessed minimal additional intervention beyond screening or feedback of screening results in postpartum and pregnant women. Women, 2 trials assessed the effects of screening plus provider supports in postpartum women, and 2 trials assessed feedback of screening results plus adjunctive counseling by home health visitors in postpartum women. Studies varied by geographic location (United States, northern Europe, United Kingdom, and Hong Kong), length of follow-up (11 weeks to 16 months), and baseline depression rates (10% to 28%).

Despite the variation in trial design and population, results were reasonably consistent across the range of designs. Trials in postpartum women showed 28% to 59% reductions in risk of depression at follow-up compared with usual care. The reported effect was smaller (18%) and did not reach significance in the trial of pregnant women but was in the same direction. The 4 studies that reported remission or response rates reported significant improvements in both postpartum and pregnant women. The most applicable trial (US trial of screening plus provider supports) found that 45% of intervention participants reported a 5-point or greater reduction in 9-item PHQ score (an improvement considered to be clinically important) compared with 35% of usual care participants (OR, 1.74 [95% CI, 1.05-2.86]; adjusted for depression history, marital status, income, education, age, and degree of parenting stress).
Eighteen trials examined the benefits of treatment interventions in women who screened positive for depression in primary care or community settings. Fifteen trials were in postpartum women (usually 6-12 weeks postpartum) and 3 trials were in pregnant women, but all reported outcomes during the postpartum period. Only 1 small, short-term trial of screen-detected depression in postpartum women included antidepressants as an intervention. The most commonly studied approach was CBT or related interventions that included CBT components. All 10 trials of CBT or CBT-related interventions, including the 2 trials in pregnant women, showed an increased likelihood of remission with treatment in the short term (<7.8 months). The magnitude of effect in pregnant women was similar to that in postpartum women. Pooling results that used only the longest follow-up period within 1 year showed a 35% increase in the likelihood of remission with CBT (DerSimionian and Laird pooled relative risk, 1.34 [95% CI, 1.19-1.50]; K = 10; I² = 7.9%) compared with usual care. The other 8 non-CBT studies examined a diverse range of interventions but did not provide sufficient evidence to draw conclusions for any one approach. There was also insufficient evidence to assess differences in effectiveness for patient subgroups.

Potential Harms of Screening and Treatment

General Adult Population and Older Adults

One trial in general adults reported no adverse events attributable to screening in a subset of participants with newly identified depression; none of the other effectiveness trials in general adults reported on harms. One trial in older adults reported para- doxical effects from screening, as previously discussed. No additional studies addressing harms of screening were identified in the review.

The 2009 USPSTF review found 7 studies that compared suicide-related events in adults who received SSRIs and other second-generation antidepressants vs placebo. No studies reported a significant increase in completed suicide rates in adults who received antidepressants compared with those who received placebo, although completed suicides were rare and, as a result, the power to detect a significant difference was limited. For adults older than 65 years, antidepressant use seemed to be protective against suicidal behavior (OR, 0.06 [95% CI, 0.01-0.58]). In addition, the 2009 USPSTF review identified 1 fair-quality study on bleeding risk in older adults who received SSRIs. Although patients 16 years and older were at increased risk of upper gastrointestinal bleeding during SSRI use, the risk increased significantly with age, from 4.1 hospitalizations per 1000 adults aged 65 to 70 years to 12.3 hospitalizations per 1000 adults aged 80 to 89 years. The odds of upper gastrointestinal bleeding in adults aged 40 to 79 years who were taking SSRIs (adjusted OR, 3.0 [95% CI, 2.1-4.4]) were much higher when they were also taking a nonsteroidal anti-inflammatory drug (adjusted OR, 15.6 [95% CI, 6.6-36.6]).

Pregnant and Postpartum Women

Direct and indirect evidence support moderate certainty that screening for depression in general adults is of moderate net benefit. The evidence for older adults is less clear, because the trials that assessed the direct effect of screening found no benefit and possibly even harm. However, given the strength of the indirect evidence (the accuracy of screening in older adults and the effectiveness of treatment in older adults), the inclusion of adults older than 65 years in the studies of all adults, and the weakness of the direct evidence on screening in older adults, the USPSTF concludes that the weight of evidence still favors a net benefit. However, more research on optimal screening approaches in older adults is imperative.

Estimate of Magnitude of Net Benefit

General Adult Population and Older Adults

The evidence from 5 RCTs, in addition to indirect evidence reviewed for the 2009 recommendation, supports moderate certainty that screening for depression in general adults is of moderate net benefit. The evidence for older adults is less clear, because the trials that assessed the direct effect of screening found no benefit and possibly even harm. However, given the strength of the indirect evidence (the accuracy of screening in older adults and the effectiveness of treatment in older adults), the inclusion of adults older than 65 years in the studies of all adults, and the weakness of the direct evidence on screening in older adults, the USPSTF concludes that the weight of evidence still favors a net benefit. However, more research on optimal screening approaches in older adults is imperative.

Pregnant and Postpartum Women

Direct and indirect evidence support moderate certainty that screening for depression in pregnant and postpartum women is of moderate net benefit. Six RCTs with varying degrees of additional support found direct benefit of screening, 23 studies confirmed the accuracy of the EPDS for identifying MDD, and 10 RCTs found benefit of treatment with CBT. Although most of the evidence (except for evidence on harms of SSRIs) is in postpartum women, the direction and magnitude of effect in pregnant women was consistent with the outcomes for postpartum women and for adults in general. It is important to note that the evidence on treatment benefit is primarily for nonpharmacologic interventions (ie, CBT), there is evidence of a small risk of harm to fetal health with SSRI use in pregnant women, and there is a lack of evidence on harms of SSRI use in postpartum women. Therefore, it is important that a range of treatment options are available for pregnant and postpartum women with...
depression who are identified through screening and that treatment choices are made through shared decision making.

Response to Public Comment
A draft version of this recommendation statement was posted for public comment on the USPSTF website from July 28, 2015, to August 24, 2015. A number of comments requested a more detailed definition of what constitutes an “adequate system” for screening. The USPSTF revised the implementation section to clarify that a range of staff types, organizational arrangements, and settings can be used to support the goals of depression screening and provided a link to the Substance Abuse and Mental Health Services Administration registry of evidence-based mental health interventions as a resource. Comments suggested that access to depression screening and management resources would be useful. The USPSTF has now provided links to evidence-based depression screening and management toolkits for primary care settings. There were several requests to clarify the potential harms of SSRIs; in response, the USPSTF added information to the Discussion section. Finally, many concerns were expressed about barriers to effectively implementing screening within adequate systems of care; the USPSTF noted these as a research need.

Objective of Previous USPSTF Recommendation
In 2009, the USPSTF recommended screening all adults when staff-assisted depression care supports are in place and selective screening based on professional judgment and patient preferences when such support is not available. In recognition that such support is now much more widely available and accepted as part of mental health care, the current recommendation statement has omitted the recommendation regarding selective screening, as it no longer represents current clinical practice. The current statement also specifically recommends screening for depression in pregnant and postpartum women, subpopulations that were not specifically reviewed for the 2009 recommendation.

Recommendations of Others
The American Academy of Family Physicians recommends screening for depression in the general adult population, including pregnant and postpartum women. Screening should be implemented with adequate systems in place to ensure accurate diagnosis, effective treatment, and appropriate follow-up. The American Academy of Pediatrics recommends that pediatricians screen mothers for postpartum depression at the infant’s 1-, 2-, and 4-month visits. The American College of Preventive Medicine recommends that primary care clinicians screen all adults for depression and that all primary care clinicians should have systems in place, either within the primary care setting itself or through collaborations with mental health professionals, to ensure the accurate diagnosis and treatment of this condition. The American College of Obstetricians and Gynecologists recommends that clinicians screen patients at least once during the perinatal period for depression and anxiety symptoms. Screening must be coupled with appropriate follow-up and treatment when indicated (practices should be prepared to initiate medical therapy, refer patients to appropriate care, or both), and systems should be in place to ensure follow-up for diagnosis and treatment. The Canadian Task Force on Preventive Health Care does not recommend routinely screening for depression in adults who are at average risk of depression or in subgroups of the population who may be at increased risk of depression. The Institute for Clinical Systems Improvement recommends that clinicians use a standardized instrument to screen for depression if it is suspected based on risk factors or presentation. The Community Preventive Services Task Force recommends collaborative care for the management of depressive disorders based on strong evidence of effectiveness in improving depression symptoms, adherence to treatment, response to treatment, and remission and recovery from depression. This collaboration is designed to improve the routine screening and diagnosis of depressive disorders, as well as the management of diagnosed depression.
Role of the Funder/Sponsor: AHRQ staff assisted in the following: development and review of the research plan; commission of the systematic evidence review from an Evidence-based Practice Center, coordination of expert review and public comment of the draft evidence report and draft recommendation statement, and the writing and preparation of the final recommendation statement and its submission for publication. AHRQ staff had no role in the approval of the final recommendation statement or the decision to submit for publication.

Disclaimer: Recommendations made by the USPSTF are independent of the US government. They should not be construed as an official position of the AHRQ or the US Department of Health and Human Services.

Additional Contributions: We thank Elisabeth Kato, MD, MRP, of AHRQ, who contributed to the writing of the manuscript, and Lisa Nicollera, MA, of AHRQ, who assisted with coordination and editing.

REFERENCES