Available online at www.sciencedirect.com
ScienceDirect
General Hospital Psychiatry 32 (2010) 345–359
General Hospital Psychiatry

Psychiatry and Primary Care

Recent epidemiologic studies have found that most patients with mental illness are seen exclusively in primary care medicine. These patients often present with medically unexplained somatic symptoms and utilize at least twice as many health care visits as controls. There has been an exponential growth in studies in this interface between primary care and psychiatry in the last 10 years. This special section, edited by Jürgen Unutzer, M.D., will publish informative research articles that address primary care-psychiatric issues.

The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review

Kurt Kroenke, M.D.[a,*], Robert L. Spitzer, M.D.[b], Janet B.W. Williams, D.S.W.[b], Bernd Löwe, M.D., Ph.D.[c]

[a]*From Regenstrief Institute, Inc. and the Department of Medicine, Indiana University, Indianapolis, IN 46202, USA*
[b]*Biometrics Research Department, New York State Psychiatric Institute and Department of Psychiatry, Columbia University, New York, NY 10032, USA*
[c]*Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Hamburg-Eppendorf and Schön Klinik, Hamburg-Eilbek, Germany*

Received 4 October 2009; accepted 9 March 2010

Abstract

Background: Depression, anxiety and somatization are the most common mental disorders in primary care as well as medical specialty populations; each is present in at least 5–10% of patients and frequently comorbid with one another. An efficient means for measuring and monitoring all three conditions would be desirable.
Methods: Evidence regarding the psychometric and pragmatic characteristics of the Patient Health Questionnaire (PHQ)-9 depression, generalized anxiety disorder (GAD)-7 anxiety and PHQ-15 somatic symptom scales are synthesized from two sources: (1) four multisite cross-sectional studies (three conducted in primary care and one in obstetric-gynecology practices) comprising 9740 patients, and (2) key studies from the literature that have studied these scales.
Results: The PHQ-9 and its abbreviated eight-item (PHQ-8) and two-item (PHQ-2) versions have good sensitivity and specificity for detecting depressive disorders. Likewise, the GAD-7 and its abbreviated two-item (GAD-2) version have good operating characteristics for detecting generalized anxiety, panic, social anxiety and post-traumatic stress disorder. The optimal cutpoint is ≥10 on the parent scales (PHQ-9 and GAD-7) and ≥3 on the ultra-brief versions (PHQ-2 and GAD-2). The PHQ-15 is equal or superior to other brief measures for assessing somatic symptoms and screening for somatoform disorders. Cutpoints of 5, 10 and 15 represent mild, moderate and severe symptom levels on all three scales. Sensitivity to change is well-established for the PHQ-9 and emerging albeit not yet definitive for the GAD-7 and PHQ-15.
Conclusions: The PHQ-9, GAD-7 and PHQ-15 are brief well-validated measures for detecting and monitoring depression, anxiety and somatization.

1. Introduction

The Primary Care Evaluation of Mental Disorders (PRIME-MD) was an instrument developed and validated in the early 1990s to efficiently diagnose five of the most common types of mental disorders presenting in medical populations: depressive, anxiety, somatoform, alcohol and eating disorders [1]. Patients first completed a one-page 27-item screener and, for those disorders for which they screened positive, were asked additional questions by the clinician using a structured interview guide. The latter was modelled after lengthier structured psychiatric interviews which were useful in research but impractical in clinical practice [2–4]. The PRIME-MD proved to have good operating characteristics and seemed reasonably efficient: it took an average of 5.6 min of clinician time to administer the PRIME-MD to patients without a mental disorder diagnosis and 11.4 min to patients with a diagnosis.

* Corresponding author. Tel.: +1 317 630 7447; fax: +1 317 630 6611.
E-mail address: kkroenke@regenstrief.org (K. Kroenke).

0163-8343/$ – see front matter © 2010 Published by Elsevier Inc.
doi:10.1016/j.genhosppsych.2010.03.006

This modest investment of time was still a barrier to use given the competing demands in primary care where visits typically average 15 min or less and patients may have multiple acute and chronic medical disorders, preventive medicine needs and documentation requirements. Therefore, in two large studies enrolling 6000 patients (3000 from general internal medicine and family practice clinics and 3000 from obstetrics-gynecology clinics), a self-administered version of the PRIME-MD called the Patient Health Questionnaire (PHQ) was developed and validated [5,6]. In the past decade, the PHQ in general and the PHQ-9 depression scale in particular have gained increasing use in both research and practice.

Given the popularity of the PHQ-9 for assessing and monitoring depression severity, a new seven-item anxiety scale using a response set similar to the PHQ-9 was initially developed to diagnose generalized anxiety disorder (GAD) (hence its name, the GAD-7) and validated in 2740 primary care patients [7]. Though originally developed to diagnose generalized anxiety disorder, the GAD-7 also proved to have good sensitivity and specificity as a screener for panic, social anxiety and post-traumatic stress disorder [8]. Finally, the PHQ-15 was derived from the original PHQ studies and is increasingly used to assess somatic symptom severity and the potential presence of somatization and somatoform disorders [9].

Each PHQ module can be used alone (e.g., the PHQ-9 if depression is the condition of interest), together with other modules or as part of the full PHQ. Although the PHQ was originally developed to detect five disorders, the depression, anxiety and somatoform modules (in that order) have turned out to be the most popular. Also, most primary care patients with depressive or anxiety disorders present with somatic complaints and co-occurrence of somatic, anxiety and depressive symptoms [the Somatic-Anxiety-Depressive (*SAD*) triad] is exceptionally common [10–14].

Thus, our article focuses on the PHQ-9 depression, GAD-7 anxiety, and PHQ-15 somatic symptom scales, drawing on data from both the original studies and other subsequent studies in the literature. We call this composite measure the Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales (PHQ-SADS) (Appendix A). This review is particularly timely because of the large amount of clinical research published on the PHQ scales over the past decade, the frequent overlap of depressive, anxiety, and somatic symptoms, and the increasing emphasis as *DSM-V* develops to conduct dimensional as well as categorical assessments [15].

2. Literature search

We searched MEDLINE from 1999 through September 2009 using the following search terms: *PRIME-MD*, *Patient Health Questionnaire*, *PHQ-9*, *PHQ-8*, *PHQ-2*, *GAD-7*, *PHQ-15*. A total of 561 publications were identified, and the abstracts were reviewed (full bibliography is available upon request from the authors). Studies that gathered original data or that synthesized data from multiple studies as either a meta-analysis or a systematic review were retrieved and assessed for inclusion in our narrative review. Studies that enrolled small samples (typically <100 patients) or that focused on the full PHQ or on the alcohol or eating modules only were excluded. Since our review focuses in particular on the operating characteristics and utility of the PHQ depressive, anxiety and somatic symptom scales, we also cite only the best and representative studies when there are numerous papers on a secondary topic (use of the PHQ scales in patients with comorbid disorders or special populations, testing in another language or country, dissemination of the scales, etc.). Because of the breath of our topic, this article is a narrative review of the three scales informed by a comprehensive literature search rather than an exhaustive literature synthesis and meta-analysis.

3. Overview of key validation studies

Together, the four original validation studies represented nearly 10,000 patients. Participants in the three primary care studies had a mean age of 46–55 years old; 60–69% were women, 8–30% were African-American and 4–9% were Hispanic [1,5,7]. Participants in the obstetrics-gynecology study had a mean age of 31; 100% were women, 15% were African-American and 39% were Hispanic [6]. Table 1 summarizes key psychometric characteristics of the PHQ-9, GAD-7 and PHQ-15. In addition to the four original studies, there have been two meta-analyses of the PHQ-9 [16,17] and a large validation study of the PHQ-15 [18]. Selected findings from the pivotal studies in Table 1 along with other relevant articles are discussed in subsequent sections of this review.

4. Depressive symptoms

4.1. Diagnostic performance and psychometric characteristics

The PHQ-9 can be used either as a diagnostic algorithm to make a probable diagnosis of major depressive disorder (MDD) or as a continuous measure with scores ranging from 0 to 27 and cutpoints of 5, 10, 15 and 20 representing mild, moderate, moderately severe and severe levels of depressive symptoms. MDD should be considered in patients who endorse ≥5 of the 9 symptoms as present "more than half the days" (the 9th item counts if endorsed "several days") and one of the first two symptoms (depressed mood or loss of interest) is endorsed. Most patients are able to complete all items (<5% items were missing in PHQ studies). If a self-administered PHQ-9 is subsequently reviewed in-person with the respondent, missing items may be probed, whereas a practical and diagnostically conservative approach to items left blank on

Table 1
Psychosometric characteristics of PHQ-9, GAD-7 and PHQ-15 derived from key studies

Scale/Study	Sample size • DSM-IV disorder	Criterion Validity[a]		Reliability			Area under Curve
		Sensitivity	Specificity	Internal (Cronbach α)	Test-Retest	Self-rated *vs.* Interviewer	
PHQ-9	• All MDD						
Kroenke [19]	6000 patients	.88	.88	.86-.89	.84	.84	.95
Gilbody [16]	14 studies[b]	.80	.92	–	–	–	–
Wittkampf [17]	4 studies[a]	.77	.94	–	–	–	–
GAD-7	2740 patients			.92	.83	.83	–
Spitzer [7,8]							
	• GAD	.89	.82	–	–	–	.91
	• Panic	.74	.81	–	–	–	.85
	• Social anxiety	.72	.80	–	–	–	.83
	• PTSD	.66	.81	–	–	–	.83
PHQ-15							
Kroenke [9]	6000 patients	–	–	.80	–	–	
Ravesteijn [18]	906 patients[c]			.80	.83		
	• Somatoform	.78	.71		.60		.76

GAD=generalized anxiety disorder.

[a] For sensitivity and specificity, PHQ-9 studies used cutpoint ≥10 (except Gilbody who looked at both PHQ-9 diagnostic algorithm and cutpoint ≥10 and found no difference); GAD-7 study used cutpoint ≥10; and Ravesteijn used ≥3 severe symptoms (i.e., rated by patient as "bothered a lot") on the PHQ-15.

[b] Gilbody and Wittkampf papers were meta-analyses of PHQ-9 diagnostic studies.

[c] For Ravesteijn study, line 1 are the psychometrics using PHQ-15 as a continuous score, and line 2 are psychometrics using a cutpoint of ≥3 severe symptoms.

mailed or automated reports is to score blank items as 0. Another approach in some research studies has been to replace the missing value by the mean of the completed items if less than 20% of items are missing. This approach is less conservative but has a lower risk of missing persons with depression, anxiety or somatization.

Psychometric characteristics of the PHQ-9 are summarized in Table 1. Secondary analysis of data from the 6000 patients in the two original PHQ studies established the reliability and validity of the PHQ-9 depression scale [19,20]. A sufficient number of subsequent studies have evaluated the operating characteristics of the PHQ-9 to allow two diagnostic meta-analyses to be conducted [16,17]. In a systematic review of 16 case-finding measures for depression, Williams et al. concluded from 38 studies involving more than 32,000 primary care patients that the PHQ-9 was equal or superior to other depression measures [21]. This was confirmed in two subsequent studies comparing the PHQ-9 to other brief measures [22,23]. The PHQ-9 performance is also similar regardless of the mode of administration (e.g., patient self-report, interviewer-administered either in-person or by telephone, touch-screen computer) [24]. The PHQ-9 performs similarly across sex, age [25,26] and racial/ethnic groups [27,28].

Sensitivity to change — an essential characteristic of measures used to monitor response to treatment — has been repeatedly established for the PHQ-9 [26,29,30]. Although a small cross-sectional study in 49 depressed patients showed only a moderate correlation (r=.52) between the PHQ-9 and the commonly-used reference standard 17-item Hamilton Depression Rating Scale (HAM-D) [31], two randomized controlled treatment trials enrolling 388 patients and conducting longitudinal assessments showed comparable sensitivity to change of the PHQ-9 and HAM-D [32,33]. A five-point decline represents a clinically significant improvement [26]. A method for equating PHQ-9 severity scores to other depression measures has been reported [34]. Finally, a modified version of the PHQ-9 to assess a lifetime history of depression has also been developed [35].

4.2. Patients with medical comorbidity

The PHQ-9 has been used in clinical studies across a variety of medical conditions including neurological disorders [36,37], cardiovascular disease [38–42], HIV disease [43,44], diabetes [45,46], chronic kidney disease [47], cancer [48–50], rheumatological disorders [51,52], gastrointestinal disease [53] dermatological disorders [54], obstetrics-gynecology practices [6,55,56], ophthalmologic and otolaryngologic disorders [56,57], pain and other somatic symptoms [58–64], brain [65] and spinal cord injury [66] and a variety of chronic medical conditions [67,68]. In dissemination studies, the PHQ-9 has been favorably received by both primary care and mental health specialists [69,70].

4.3. Nursing home and cognitively impaired patients

The Minimum Data Set (MDS) is part of the US federally mandated process for clinical assessment of all residents in Centers for Medicare and Medicaid (CMS) certified nursing homes. To determine changes for the third revision (MDS 3.0), CMS sponsored a national validation study involving 3258 nursing home (NH) residents [71]. Most NH residents (86%) successfully completed the PHQ-9 interview, and NH staff were able to complete an observational version of the PHQ-9 for 424 (92%) of the 461 residents who did not complete the resident interview. There was excellent agreement between PHQ-9 results obtained independently

by NH facility nurses and researchers (kappa=0.968). Moreover, compared to the 15-item Geriatric Depression Scale, the PHQ-9 required less time to complete, correlated more strongly with the criterion standard psychiatric assessment (.83 vs. .71), and showed more internal consistency across varying levels of cognitive ability. Most nurses rated the PHQ-9 as superior to depression assessment as currently performed in MDS 2.0 (88% of nurses), felt the PHQ-9 results could inform care plans (84%) and provide new insights into resident mood (86%). Most (77%) also reported that they felt that all residents who gave answers understood them (only 6% disagreed). Thus, the PHQ-9 is the depression measure in MDS 3.0. Another recent study found the PHQ-2 superior to four other measures (including the Geriatric Depression Scale) for depression screening in nursing home and assisted living residents [72].

4.4. Adolescents

Johnson et al. developed the PHQ-A, a version of the PHQ modified for use in adolescent populations, and validated it in a sample of 403 primary care adolescent patients, 50 of whom had major depression as determined by an independent criterion standard psychiatric interview [73]. Compared to the PHQ-9, the PHQ-A depression module has more items (14 vs. 9), a yes–no response set (thus precluding a severity score) and much less cumulative validation data. However, the PHQ-A is one of the few depression measures specifically validated in adolescents [74]. The US Preventive Services Task Force recently recommended depression screening in adolescents (12–18 years of age) when systems are in place to ensure accurate diagnosis, psychotherapy (cognitive-behavioral or interpersonal) and follow-up . They concluded evidence is insufficient to warrant a recommendation to screen children (7–11 years of age). They advised that either the PHQ-A or the Beck Depression Inventory-Primary Care Version be used for screening. A different multidisciplinary expert consensus group produced the Guidelines for Adolescent Depression in Primary Care (GLAD-PC) report [75]. The GLAD-PC Toolkit suggests several different depression tools, one being a modified version of the PHQ-9 which preserves the four-item response set (and thus its continuous severity score) and added three additional items (two on suicidal ideation and one to screen for dysthymia). The traditional version of the PHQ-9 has recently proven sensitive to change in a small pilot study of adolescent depression [76]. Ideally, for studies that determine the operating characteristics of these different versions of the PHQ as well as alternative adolescent depression measures, the use of a criterion standard psychiatric interview would be desirable.

4.5. Pregnancy and postpartum period

Both pregnancy and the postpartum period are times of high-risk for the onset or recurrence of depression. The 10-item Edinburgh Postnatal Depression Scale (EPDS) is the most commonly-used depression scale for this population. Several studies have tested the PHQ-9 against a criterion standard psychiatric interview. In a longitudinal study of 506 women assessed at six time points over 9 months following delivery of their child, the PHQ-2 was found to be highly sensitive and the PHQ-9 was highly specific for identifying postpartum depression [77]. In head-to-head comparisons with the EPDS, the PHQ has been found to be more accurate (n=160 subjects) [78], comparable (n=415) [79], or slightly less accurate (n=123) [80]. The PHQ has recently been used in a depression screening initiative involving 1336 pregnant and postpartum women receiving obstetrical care at publicly funded health care clinics [81]. A recent evidence-based review of postnatal depression screening concluded that the optimal method to identify postnatal depression has not yet been identified [82].

4.6. Abbreviated versions: PHQ-2 and PHQ-8

The PHQ-2 consists of the first two items of the PHQ-9 — depressed mood and loss of interest (anhedonia) — of which at least one is required to establish a diagnosis of MDD or any *DSM-IV* depressive disorder. These two items are comparable to many longer case-finding measures for depression screening whether asked with a yes–no response set as in the original PRIME-MD [83] or with the four-response set of the PHQ-9 [44,84–87]. PHQ-2 scores can range from 0 to 6, and a cutpoint $\geq$3 suggests clinically significant depression which should prompt either completion of the full PHQ-9 or a clinical interview to assess for MDD. The PHQ-2 also appears to be responsive when monitoring depression outcomes [86]. In a study of 1523 psychiatric outpatients evaluated in the Methods to Improve Diagnostic Assessment and Services project, the PHQ-2 items were found to be the most discriminatory for major depression from a databank of items [88].

The PHQ-8 omits the ninth item that asks about thoughts of death or self-harm. For large epidemiological studies or other research where depression is a secondary measure, some investigators have used the PHQ-8 because depression is not the main outcome, and it is just not feasible to have a back-up system for the very rare "suicidal" response to the ninth item. Even in clinical trials of depressed patients, when conducted in primary care or other medical populations rather than psychiatric patients, most patients who endorse the ninth item are agreeing with the first part of the item (passive thoughts of "being better off dead") rather than the second part (active thoughts of "hurting yourself in some way"). The PHQ-8 has also been used in clinical or research settings where follow-up to positive responses to the ninth item of the PHQ-9 may be delayed (e.g., mailed or web-based screening).

The PHQ-8 has been shown to have comparable operating characteristics to the PHQ-9 in terms of diagnosing depressive disorders when using a *DSM-IV* based diagnostic algorithm [20,85]. Research indicates that deletion of the

ninth item has only a minor effect on scoring because it is, by far, the least frequently endorsed item on the PHQ-9 [20,28,89,90]. In 1004 primary care patients, the correlation between PHQ-9 scores and PHQ-8 scores was very high (r=.998); only three patients with a PHQ-9 score of 10 or higher had a PHQ-8 score of less than 10 [85]. In our two PHQ validation studies totaling 6000 patients, we have confirmed this (data not previously published) by finding very high correlations between the PHQ-9 and PHQ-8 (r=.997 in both samples) and a similar area under the curve (0.95) by receiver operating characteristic (ROC) analysis for the two measures in diagnosing major depression. Thus, identical cutpoints can be used on the PHQ-8 and PHQ-9 for both clinical and research purposes [91].

4.7. Alternative response sets

In some settings, the response set has been converted from the standard verbal options to number of days in the past 1 or 2 weeks. This has facilitated use of the PHQ in certain survey settings [91] as well as for telephone-based automated depression monitoring by interactive voice recording [49] Data on MDD prevalence rates and construct validity suggest that the number of days response set may be an acceptable option [91]. Since the standard response set has much stronger validation data, however, direct comparisons of the two response sets in the same population would be advisable before widespread adoption of the number of days alternative.

When respondents are asked to report the number of days they have been bothered by each PHQ-9 symptom in the past 2 weeks, the PHQ-9 scores for 0–1, 2–6, 7–11 and 12–14 days are 0, 1, 2 and 3, respectively [91]. The rationale is that 0–1 days is clearly less than "several days" and probably not clinically significant; 7–11 days is "more than half the days" and the training manual for the *Structured Clinical Interview for DSM-IV Axis I Disorders* instructs the interviewer to consider more than 11 days as the cutpoint if respondents insist on explicit guidance for what is meant by "nearly every day in the past 2 weeks" [92]. When a 1-week time frame is used, the PHQ-9 scores for 0–1 days, 2–3 days, 4–5 days and 6–7 days are 0, 1, 2 and 3, respectively [49].

4.8. Dissemination

The PHQ-9 has had a considerable amount of uptake both clinically and by researchers in less than a decade. Examples include its:

- Selection as the clinical tool used in the intervention arm of numerous randomized effectiveness trials in medical populations for monitoring and adjusting depression therapy [32,49,50,93–100]
- Use in its full or abbreviated forms in large federally-sponsored surveys and programs including the Behavioral Risk Factor Surveillance Survey [101]; the National Health and Nutrition Examination Survey [102]; the Medical Expenditure Panel Survey [103]; the National Epidemiologic Survey on Alcohol and Related Conditions [87]; the Medicare Health Support program [104], the Millennium Cohort Study [105] and population-based studies in Germany [106], Australia [107] and Canada [108].
- Adoption as a standard measure for depression screening by the Veterans Administration [109], Department of Defense [110], Kaiser [111] and several integrated health care systems [112], managed behavioural care organizations [113] and public health departments [114].
- Popularity as the most commonly used depression measure in the United Kingdom's National Health Service which requires use of a measure as part of depression treatment in primary care [115].
- Incorporation into toolkits or programs to improve depression care by the MacArthur Foundation [116] and the Robert Wood Johnson Foundation [117] as well as professional organizations like the American Heart Association [118].
- Consideration by the American Psychiatric Association as the dimensional depression measure in its *DSM-V* classification manual [119].
- Acceptability, utility and sustainability in psychiatry and primary care practice networks [69,70,120].

5. Anxiety symptoms

5.1. Original PHQ anxiety module

The original PHQ anxiety module focused on two diagnoses: *panic disorder* and *other anxiety disorder*. The 15-item panic module yielded a probable diagnosis of panic disorder for individuals who answered "yes" to the first four questions and endorsed ≥4 of 11 somatic symptoms during an anxiety attack. In addition to validation in the two original PHQ studies, further research has strengthened the evidence for the panic disorder section [89,121–125]. The other anxiety disorder section primarily included criteria for GAD but its yes-no response format did not permit calculation of a severity score that proved so useful with the PHQ-9. Thus, a study in more than 2700 primary care patients was conducted to develop and validate a 7-item anxiety measure (GAD-7) with a similar response set to the PHQ-9 in order to establish probable diagnoses of GAD as well as grade its severity [7].

5.2. Diagnostic performance and psychometric characteristics of GAD-7 for GAD

GAD-7 scores can range from 0 to 27, with 5, 10 and 15 representing mild, moderate and severe levels of anxiety symptoms [7,8]. Psychometric characteristics of the GAD-7 are summarized in Table 1. At a cutpoint of ≥10, both sensitivity and specificity exceeded .80, and sensitivity was

nearly maximized. Scale characteristics were not influenced by age, sex or race/ethnicity. However, the proportion of primary care patients who score at this level was quite high (23%). Thus, a cutpoint of ≥15 maximizes specificity and approximates a prevalence (9%) more in line with current epidemiologic estimates of GAD prevalence in primary care. However, sensitivity at this high cutpoint is low (48%).

Although the GAD-7 inquires about symptoms in the past 2 weeks, the *DSM-IV* diagnostic criteria for GAD specify at least a 6 month duration of symptoms. Nonetheless, the operating characteristics of the scale were good because most patients with high symptom scores had in fact chronic symptoms. Of the 433 patients with GAD-7 scores ≥10, 96% had symptoms ≥1 month and 67% had symptoms ≥6 months. Notably, the National Comorbidity Survey has shown that cases with episodes of GAD for 1–5 months do not differ greatly from those with episodes of ≥6 months in onset, persistence, impairment, co-morbidity, parental GAD or sociodemographic correlates [126]. Kessler et al. conclude that there is little basis for excluding people from a diagnosis of GAD based simply on duration of symptoms.

There was a strong association between increasing GAD-7 severity scores and worsening function on all six Medical Outcomes Study Short-Form 20-item General Health Survey (SF-20) health-related quality of life scales [7]. Convergent validity of the GAD-7 was good, as demonstrated by its correlations with two anxiety scales: the Beck Anxiety Inventory (r=.72) and the anxiety subscale of the SCL-90 (r=.74). Factor analysis that included the GAD-7 anxiety items and the PHQ-8 depression items confirmed two distinct dimensions, with all depression items having the highest factor loadings on one factor (0.58–0.75) and all anxiety items having the highest factor loadings on the second factor (0.69–0.81). Although sensitivity to change of the GAD-7 is suggested by the results of a large randomized depression trial [100] where anxiety was a secondary outcome, responsiveness will be better established in treatment trials or other types of longitudinal studies where anxiety is the primary outcome.

The validity of the GAD-7 for population-based epidemiologic studies is supported by results from a nationally representative face-to-face household survey of 5030 subjects conducted in Germany [127]. Factor analysis substantiated the 1-dimensional structure of the GAD-7 and its factorial invariance for gender and age. Internal consistency was identical across all subgroups (0.89). Approximately 5% of subjects had GAD-7 scores of 10 or greater and 1% had GAD-7 scores of 15 or greater.

5.3. Operating characteristics of GAD-7 for other anxiety disorders

The four most common anxiety disorders (excluding simple phobias which seldom present clinically) are generalized anxiety disorder, panic disorder, social anxiety disorder and posttraumatic stress disorder (PTSD) [128]. In the GAD-7 validation study, PTSD was present in 8.6% of the patients, GAD in 7.6%, panic disorder in 6.8%, social anxiety disorder in 6.2% and any anxiety disorder in 19.5%, all of which are in the prevalence range reported in previous primary care studies.

Although originally developed for GAD, the GAD-7 also proved to have good sensitivity and specificity as a screener for panic, social anxiety and post-traumatic stress disorder [8]. Not surprisingly, the area under the curve by ROC analysis is greatest (.91) for GAD, but also quite good for panic disorder(.85), social anxiety disorder (.83) and PTSD (.83). Each of the four anxiety disorders was strongly and similarly associated with impaired functioning on all six SF-20 scales and with self-reported disability days. In fact, the number of anxiety disorders rather than the specific type of disorder was the factor most strongly associated with impairment.

5.4. Abbreviated versions: GAD-2

The GAD-2 consists of the first two items of the GAD-7 which in turn correspond to the two core diagnostic criteria for GAD. Scores on the GAD-2 range from 0 to 6, and a cutpoint ≥3 denotes a screening cutpoint for clinically significant anxiety which should prompt completion of the full GAD-7 and a clinical interview to determine the type of anxiety disorder and whether treatment and/or referral is warranted [8]. Notably, the area under the curve by ROC analysis for the GAD-2 was similar to that for the GAD-7 for three of the four anxiety disorders and only slightly lower (.80 vs. .83) for PTSD.

5.5. Dissemination

Compared to the PHQ-9, uptake of the GAD-7 is in a more nascent phase given its later publication (2006 vs. 1999). Some experts are recommending its routine use [129], and new research studies using the GAD-7 are starting to emerge [49,100,130,131]. Unlike the PHQ-9 which can serve as both a diagnostic and severity measure, the GAD-7 is principally a measure of anxiety severity; the likelihood of an anxiety disorder increases with higher GAD-7 scores, but a clinical interview is required to confirm the presence and type of disorder.

6. Somatic symptoms

6.1. Three limitations in the classification of somatoform disorders

Three findings from recent research regarding the classification of somatoform disorders is relevant to the PHQ-15 [132,133]. First, fewer chronic symptoms are probably needed than has traditionally been required for the diagnosis of somatization disorder which captures less than 10–20% of the patients with chronic and disabling somatization in primary care. Second, focusing primarily on

current rather than lifetime symptom counts may be desirable due to the greater reliability and comparable validity of current counts. Third, the requirement that symptoms be "medically unexplained" is problematic since confidently attributing the etiology of pain, fatigue, dizziness, gastrointestinal complaints and numerous other common somatic symptoms to specific medical or psychological conditions that frequently coexist in the same patient is often difficult. Even disease-specific physical symptoms in patients with medical disorders may be explained as much by comorbid depression or anxiety as by the severity of the medical disorder [134]. By allowing lower symptom thresholds, focusing on current symptoms and not requiring adjudication of symptom etiology, the PHQ-15 addresses these three issues [135].

6.2. Diagnostic performance and psychometric characteristics of the PHQ-15

The PHQ-15 includes 15 symptoms that account for more than 90% of symptoms seen in primary care (exclusive of upper respiratory symptoms such as cough, nasal symptoms, sore throat, ear ache, etc.). [9]. The PHQ-15 asks patients to rate how much they have been bothered by each symptom during the past month on a 0 ("not at all") to 2 ("bothered a lot") scale. Thus, the total score ranges from 0 to 30, with cutpoints of 5, 10 and 15 representing thresholds for mild, moderate and severe somatic symptom severity, respectively. In the original study, the majority (88%) of patients who endorsed ≥3 medically unexplained symptoms at the level of "bothered a lot" and who had at least a several year history of poorly explained symptoms had a somatoform diagnosis [136,137]. In a recent study of the PHQ-15 in 906 primary care patients, a cutpoint of ≥3 severe (i.e., "bothered a lot") symptoms during the past 4 weeks had a sensitivity of 78% and a specificity of 71% for a *DSM-IV* somatoform diagnosis [18]. These somewhat lower operating characteristics may be because the PHQ-15, unlike *DSM-IV*, asks only about current (rather than lifetime) symptoms and does not adjudicate whether or not a symptom is medically unexplained. Indeed, both *DSM-IV* criteria (lifetime recall and symptom attribution) have suboptimal reliability and validity in diagnosing somatoform disorders [132,133].

In the original PHQ studies of 6000 unselected primary care patients, higher PHQ-15 scores were strongly associated with worsening function on all six SF-20 scales as well as increased disability days and health care utilization [9]. In a smaller study of 172 primary care patients with at least moderate somatization enrolled in a clinical trial, Cronbach's alpha was 0.79 and there was a moderately strong correlation (r=0.52) between PHQ-15 scores and medically unexplained symptom counts elicited by an independent structured psychiatric interview [138]. There is some support for the PHQ-15's sensitivity to change in clinical trials or longitudinal studies, both as a primary [139–141] and a secondary [142,143] outcome measure.

There are several factors arguing for the PHQ-15 as an excellent measure of somatic symptom burden and potential somatization [135,144]. First, approximately 10% of primary care and obstetric-gynecology outpatients have a score of 15 or greater, a prevalence consistent with other studies of clinically significant somatization. Second, increasing scores on the PHQ-15 are strongly associated with functional impairment, disability, and health care use. Third, items on the PHQ-15 overlap better with other validated somatization screeners than any other two screeners do with one another. Fourth, it is an excellent measure for identifying high-utilizing somatizing patients in health care systems [145,146]. For example, patients with somatization as identified by the PHQ-15 had approximately twice the outpatient and inpatient medical care utilization and twice the annual medical care costs of nonsomatizing patients. Adjusting the findings for the presence of psychiatric and medical comorbidity had relatively little effect on this association. Fifth, total self-reported PHQ somatic symptom counts have been shown to be highly associated with clinician-rated somatoform disorder symptom counts [138,147].

6.3. Dissemination

Uptake of the PHQ-15 to date has largely been in the area of research, in part due to lower clinician interest in chronic symptoms and somatoform disorders as well as skepticism about effective treatments [148,149]. However, recent reviews have identified effective pharmacological and behavioral interventions [150–152]. Besides the studies already cited, examples of other major studies where the PHQ-15 has been a key measure include:

- An international study of mental disorders in primary care conducted in 15 countries that used the earlier yes-no PRIME-MD version of the PHQ-15 [11];
- A prospective study of physical symptoms in predicting hospitalization and mortality in 3498 elderly patients [153];
- A study of somatization in 10,507 consecutive patients attending 340 Australian general practices [154];
- A study of somatization in a representative sample (n=2510) of the German general population [155];
- A 20-year cohort study of more than 100,000 military personnel [156]
- A 10-year follow-up study of 30,000 veterans from the Gulf War era [157];
- A study of traumatic brain injury in 2525 US soldiers returning from Iraq [158];
- A primary care study of the impact of somatization on disability [159];

The PHQ-15 has also been used as a secondary measure in a variety of smaller studies examining mental disorders or physical symptom syndromes.

7. SAD triad

7.1. Overlap and additive effects

The comorbidity of somatic, anxiety and depressive symptoms (the "SAD" triad) is well-established [12,13]. Three recent large epidemiologic studies in primary care confirm that "pure" forms are much less common than overlapping syndromes; most patients reporting high levels of one symptom type also report high levels of one or both of the other types of symptoms [14,154,160]. Also, somatic, anxiety and depressive symptoms have independent, additive and differential effects on multiple domains of health-related quality of life, functional status, disability and health care use [9,14,161,162].

7.2. Incremental impairment

Table 2 summarizes data on health-related quality of life (HRQL) and symptom severity from the original studies [7,9,19]. When moving across the ordinal categories of minimal to mild to moderate to severe scores, there is a large and progressive decrement across multiple domains of HRQL for all three types of symptoms. For simplicity, only mean scores are displayed (standard deviations are provided in the original papers); however, most pair-wise comparisons of mean SF-20 scores between adjacent symptom severity levels within each scale are highly significant ($P<.001$). Figs. 1 and 2 show a similar relationship between increasing severity level and self-reported symptom-related difficulty and disability days for all three scales.

7.3. Similar cutpoints

Scores of 0–4, 5–9, 10–14, and ≥15 represent minimal, mild, moderate, and severe levels of symptom burden, respectively, on the PHQ-9, GAD-7, and PHQ-15. This

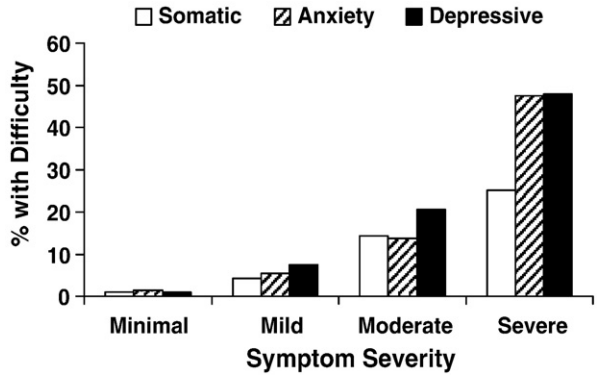

Fig. 1. Proportion of patients with minimal, mild, moderate, and severe levels of symptoms who respond with either "very difficult" or "extremely difficult" to the last question of the PHQ: "How difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?" Somatic symptoms are measured by the PHQ-15, anxiety symptoms by the GAD-7 and depressive symptoms by the PHQ-9. On all three scales, minimal symptoms are represented by a score of 0 to 4, mild symptoms by a score of 5–9, moderate symptoms by a score of 10–15 and severe symptoms by a score ≥15. Data is derived from the three original validation papers [7,9,19].

results in a simple 5-10-15 cutpoint rule-of-thumb for all three scales. Also, a score ≥10 is the most commonly-recommended cutpoint for "clinically significant" symptoms on all three scales. Finally, a score ≥3 is the suggested cutpoint for the PHQ-2 and GAD-2.

However, inflexible adherence to a single cutpoint should be discouraged. For one thing, the maximum scores are not identical: 21, 27, and 30 for the GAD-7, PHQ-9, and PHQ-15, respectively. More importantly, one might choose a different cutpoint depending upon the population being assessed [23,163](community vs. primary care vs. mental health setting) and the purpose of the

Table 2
Relationship between depressive, anxiety and somatic symptom severity and SF-20 functional status

Level of Symptom *Severity* PHQ score (% of Dep/Anx/Som)	*Mean SF-20 score*[a] By depression (Dep), anxiety (Anx), and somatic (Som) symptom severity level																	
	Mental			Social			Role			General			Pain			Physical		
	Dep	Anx	Som	Dep	Anx	Som	Dep	Anx	Som	Dep	Anx	Som	Dep	Anx	Som	Dep	Anx	Som
Minimal 0–4 (62/56/35)	81	82	82	92	91	93	87	84	91	72	68	78	69	71	79	84	84	88
Low 5–9 (23/24/35)	66	65	73	79	79	87	70	69	81	54	52	64	56	56	63	73	74	80
Medium 10–14 (8/12/20)	52	54	62	70	69	74	58	59	64	44	43	49	51	51	49	68	66	69
High ≥15 (7/8/9)	39	41	53	53	55	60	44	46	43	33	39	34	45	47	37	60	61	56

For simplicity, standard deviations are not displayed (but are provided in the original papers). However, most pair-wise comparisons of mean SF-20 scores between adjacent symptom severity levels within each scale were highly significant ($P<.001$).

[a] Each SF-20 scale is scored from 0 (worst) to 100 (best) health-related quality of life on that domain. SF-20 scores are adjusted for age, sex, education for all scales, and also number of physical disorders for Dep and Som.

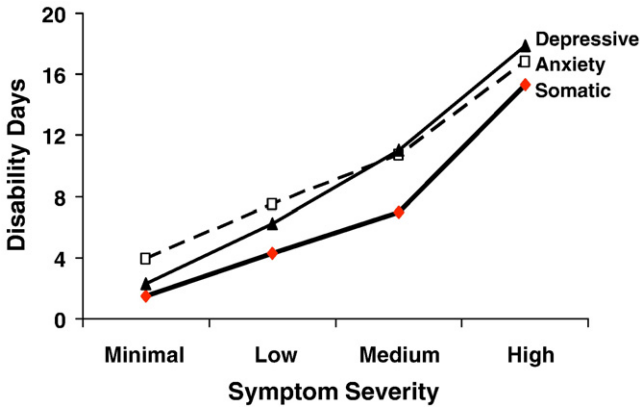

Fig. 2. Mean number of disability days in the past 3 months reported by patients with minimal, mild, moderate, and severe levels of somatic, anxiety, and depressive symptoms (as defined in Fig. 1). Disability days were determined by the patient's response to the question: "How many days altogether in the *past three months*, were you kept from your usual activities because you weren't feeling well?" Data is derived from the three original validation papers [7,9,19].

assessment (routine screening vs. evaluating suspected cases). Also, modest nuancing of the ≥10 cutpoint has been suggested by some investigators, such as ≥12 for the PHQ-9 [22,115,164] and ≥8 for the GAD-7 [8]. Indeed, these latter cutpoints for the PHQ-9 and GAD-7 are the ones being recommended as part of the United Kingdom's primary care Quality and Outcomes Framework (QOF) [115] and Improving Access to Psychological Therapies (IAPT) program [165], notable in light of the large amount of routine data these two initiatives will generate regarding the PHQ-9 and GAD-7.

7.4. Patient-rated difficulty item

The final question on the PHQ asks the patients to report "how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?" This single patient-rated difficulty item is not used in calculating any PHQ score or diagnosis but rather represents the patient's global impression of symptom-related impairment. It may be useful in decisions regarding initiation of or adjustments to treatment since it is strongly associated with both psychiatric symptom severity as well as multiple measures of impairment and HRQL. In the two large validation studies [5,6], the percentage of subjects with a PHQ diagnosis varied significantly ($P<.001$) based upon response to this single item, ranging from 16% of the subjects who responded "not difficult at all," to 35% who responded "somewhat difficult," to 73% who responded "very difficult," to 88% who responded "extremely difficult." This question was also associated with functional impairment: the mean correlation of this item with each of the 6 SF-20 scales was .36, with the highest being 0.51 for mental health.

7.5. Ultra-brief scales

Scales as short as 2–4 items have good operating characteristics for depression screening and appear to perform as well as longer case-finding measures [166]. Validated individually as abbreviated screeners for depression and anxiety, the PHQ-2 and GAD-2 have also been validated when combined as an ultra-brief screener — the PHQ-4 — for depression and anxiety in large clinical (n=2149) [167] and general population (n=5030) samples [168]. The construct and factorial validity of the two-item depression and two-item anxiety subscales was confirmed, as was the recommended cutpoint of ≥3 on each subscale. Depression and anxiety scores were shown to have independent effects on impairment yet also had synergistic value in identifying patients with single or dual disorders: of the 2013 primary care patients who completed the PHQ-4 12.1% had anxiety only (GAD-2 score ≥3), 4.4% had depression only (PHQ-2 score ≥3), and 8.5% had both anxiety and depression [167].

7.6. PHQ-SADS

The PHQ-SADS (Appendix A) includes the PHQ-15, GAD-7 and PHQ-9 scales, plus the first five items of the panic module. A positive response to the first panic question alone has a sensitivity and specificity of 93% and 78%, respectively, while each additional "yes" response to the remaining four items increases specificity with only a minimal decline in sensitivity [121]. Because the 11 somatic symptoms in the original PHQ only marginally improve the operating characteristics of the panic module, they are omitted from the PHQ-SADS for the sake of brevity. The final item on the PHQ-SADS is the respondent's global rating of symptom-related difficulty. As noted, scores of 5, 10 and 15 indicates increasing levels of severity on all three scales, and elevated scores on two or more scales suggests comorbidity. Responses to the single-item difficulty question can further guide treatment decisions.

7.7 Efficiency and acceptability

In the two PHQ validation studies totaling 6000 patients, the median time for physician review of the full PHQ after completion by a patient was less than 2 min [5,6]. Since the full PHQ assesses two additional disorders (substance and eating), the PHQ-SADS should take even less time. In these same two studies, 87–89% of physicians found the PHQ very or somewhat useful in management or treatment planning, 88–93% of patients felt very or somewhat comfortable answering the PHQ questions and 89–93% of patients believed the questions were very or somewhat helpful in getting their doctor to better understand or treat the problems that they were having.

8. Conclusion

While we have summarized the salient PHQ data, a comprehensive comparison of the PHQ with all alternative scales is beyond the scope of this review. Evidence-based literature syntheses of depression measures have recently been published [21,166]. Another limitation is the uncertain efficacy of screening for mental disorders on outcomes. For example, it appears that depression screening alone may not be sufficient to improve patient outcomes [169]. On the other hand, there is substantial evidence that when screening is part of a multi-component intervention, depression outcomes are improved [170,171]. That is why the US Services Preventive Services Task Force recommends routine screening for depression only if there are systems in place to deliver adequate treatment and follow-up [172]. A similar degree of evidence has not yet accumulated to justify routine screening for anxiety or somatoform disorders.

Nonetheless, all three conditions are sufficiently prevalent, disabling and costly in both primary care and specialty clinics to trigger case detection at a relatively low index of suspicion. Brief, self-administered scales are an efficient method for stratifying patients into screen-negative and screen-positive groups, thus allowing busy clinicians to prioritize their limited interview time for the smaller group of patients with high scores. Scales are also efficient tools for monitoring response to treatment; measurement-based care has been shown to enhance depression outcomes [173] and is likely to be beneficial for anxiety and somatization as well. Although the uptake of standardized scales into clinical practice has been slow [174,175], the potential benefits are reviewed elsewhere [175,176], including avoiding over- or undertreatment, monitoring response, adjusting treatment, assessing quality of care and standardizing communication among clinicians regarding disease severity.

Finally, scales are useful for identifying comorbidity, e.g., the somatizing depressed patient, or the patient with mixed anxiety-depression. This can be especially important in patients not responding to standard treatments, e.g., the depressed patient with persistent anxiety or pain who may need additional therapies. The PHQ depressive, anxiety and somatic symptom scales that have been developed, studied and clinically applied over the past decade constitute valid and efficient instruments for detecting, differentiating and monitoring the SAD triad. The PHQ-SADS (Appendix A) combines all three scales to provide a continuous severity measure of each of these three common and overlapping symptom domains. Such dimensional assessment is being advocated in the development of *DSM-V* to complement categorical diagnoses [15]. Finally, translations of the PHQ scales are currently available in more than 60 languages which facilitate their use in studying mental disorders and improving clinical outcomes on a global level.

Appendix A. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales (PHQ-SADS)

Patient Health Questionnaire (PHQ-SADS)

This questionnaire is an important part of providing you with the best health care possible. Your answers will help in understanding problems that you may have. Please answer every question to the best of your ability

A. During the *last 4 weeks*, how much have you been bothered by any of the following problems?	**Not bothered (0)**	**Bothered a little (1)**	**Bothered a lot (2)**
1. Stomach pain	☐	☐	☐
2. Back pain	☐	☐	☐
3. Pain in your arms, legs, or joints (knees, hips, etc.)	☐	☐	☐
4. Feeling tired or having little energy	☐	☐	☐
5. Trouble falling or staying asleep, or sleeping too much	☐	☐	☐
6. Menstrual cramps or other problems with your periods	☐	☐	☐
7. Pain or problems during sexual intercourse	☐	☐	☐
8. Headaches	☐	☐	☐
9. Chest pain	☐	☐	☐
10. Dizziness	☐	☐	☐
11. Fainting spells	☐	☐	☐
12. Feeling your heart pound or race	☐	☐	☐
13. Shortness of breath	☐	☐	☐
14. Constipation, loose bowels, or diarrhea	☐	☐	☐
15. Nausea, gas, or indigestion	☐	☐	☐

PHQ-15 Score ☐ = _____ + _____

B. Over the *last 2 weeks*, how often have you been bothered by any of the following problems?	**Not at all (0)**	**Several days (1)**	**More than half the days (2)**	**Nearly every day (3)**
Feeling nervous anxiety or on edge	☐	☐	☐	☐
Not being able to stop or control worrying	☐	☐	☐	☐
Worrying too much about different things	☐	☐	☐	☐
Trouble relaxing	☐	☐	☐	☐
Being so restless that it is hard to sit still	☐	☐	☐	☐
6. Becoming easily annoyed or irritable	☐	☐	☐	☐
7. Feeling afraid as if something awful might happen	☐	☐	☐	☐

GAD-7 Score ☐ = _____ +_____ + _____

C. Questions about anxiety attacks.	YES	NO
a. In the *last 4 weeks*, have you had an anxiety attack — suddenly feeling fear or panic?		
If you checked "NO", go to question E.	☐	☐
b. Has this ever happened before?	☐	☐
c. Do some of these attacks come *suddenly out of the blue* — that is, in situations where you don't expect to be nervous or uncomfortable?	☐	☐

d. Do these attacks bother you a lot or are you worried about having another attack?	☐	☐
e. During your last bad anxiety attack, did you have symptoms like shortness of breath, sweating, or your heart racing, pounding or skipping?	☐	☐

D. Over the *last 2 weeks*, how often have you been bothered by any of the following problems?	**Not at all (0)**	**Several days (1)**	**More than half the days (2)**	**Nearly every day (3)**
1. Little interest or pleasure in doing things	☐	☐	☐	☐
2. Feeling down, depressed, or hopeless	☐	☐	☐	☐
3. Trouble falling or staying asleep, or sleeping too much	☐	☐	☐	☐
4. Feeling tired or having little energy	☐	☐	☐	☐
5. Poor appetite or overeating	☐	☐	☐	☐
6. Feeling bad about yourself — or that you are a failure or have let yourself or your family down	☐	☐	☐	☐
7. Trouble concentrating on things, such as reading the newspaper or watching television	☐	☐	☐	☐
8. Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	☐	☐	☐	☐
9. Thoughts that you would be better off dead of or hurting yourself in some way	☐	☐	☐	☐

PHQ-9 Score ☐ = _____ + _____ + _____

E. If you checked off *any* problems on this questionnaire, how *difficult* have these problems made it for you to do your work, take care of things at home, or get along with other people?

Not difficult at all	Somewhat difficult	Very difficult	Extremely difficult
☐	☐	☐	☐

Developed by Robert L. Spitzer, Janet B.W. Williams, Kurt Kroenke, and colleagues, with an educational grant from Pfizer. The names PRIME-MD and PRIME-MD TODAY are trademarks of Pfizer.

References

[1] Spitzer RL, Williams JBW, Kroenke K, Linzer M, deGruy FV, Hahn SR, et al. Utility of new procedure for diagnosis mental-disorders in primary-care: the PRIME-MD 1000 Study. JAMA 1994;272:1749–56.

[2] Spitzer RL, Williams JBW, Gibbon M, First MB. The Structured Clinical Interview for *DSM-III-R* (SCID). Arch Gen Psychiatry 1992;49:624–9.

[3] Robins LN, Helzer JE, Croughan J, Ratcliff KS. National Institute of Mental Health Diagnostic Interview Schedule. Its history, characteristics, and validity. Arch Gen Psychiatry 1981;38:381–9.

[4] Wittchen HU. Reliability and validity studies of the WHO–Composite International Diagnostic Interview (CIDI): a critical review. J Psychiatr Res 1994;28:57–84.

[5] Spitzer RL, Kroenke K, Williams JBW. Validation and utility of a self-report version of PRIME-MD — the PHQ primary care study. JAMA 1999;282:1737–44.

[6] Spitzer RL, Williams JBW, Kroenke K, Hornyak R, McMurray J. Validity and utility of the PRIME-MD Patient Health Questionnaire in assessment of 3000 obstetric-gynecologic patients: The PRIME-MD Patient Health Questionnaire Obstetrics Gynecology Study. Amer J Obstet Gynecol 2000;183:759–69.

[7] Spitzer RL, Kroenke K, Williams JBW, Lowe B. A brief measure for assessing generalized anxiety disorder — the GAD-7. Arch Intern Med 2006;166:1092–7.

[8] Kroenke K, Spitzer RL, Williams JBW, Monahan PO, Lowe B. Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007;146:317–25.

[9] Kroenke K, Spitzer RL, Williams JBW. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med 2002;64:258–66.

[10] Kroenke K, Spitzer RL, Williams JB, Linzer M, Hahn SR, deGruy III FV, et al. Physical symptoms in primary care. Predictors of psychiatric disorders and functional impairment. Arch Fam Med 1994;3:774–9.

[11] Simon GE, Von Korff M, Piccinelli M, Fullerton C, Ormel J. An international study of the relation between somatic symptoms and depression. N Engl J Med 1999;341:1329–35.

[12] Kroenke K. Patients presenting with somatic complaints: epidemiology, psychiatric comorbidity and management. Int J Methods Psychiatr Res 2003;12:34–43.

[13] Kroenke K, Rosmalen JG. Symptoms, syndromes, and the value of psychiatric diagnostics in patients who have functional somatic disorders. Med Clin North Am 2006;90:603–26.

[14] Lowe B, Spitzer RL, Williams JBW, Mussell M, Schellberg D, Kroenke K. Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment. Gen Hosp Psychiatry 2008;30:191–9.

[15] Helzer JE, Kraemer HC, Krueger RF. The feasibility and need for dimensional psychiatric diagnoses. Psychol Med 2006;36:1671–80.

[16] Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J Gen Intern Med 2007;22:1596–602.

[17] Wittkampf KA, Naeije L, Schene AH, Huyser J, van Weert HC. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. Gen Hosp Psychiatry 2007;29:388–95.

[18] van Ravesteijn H, Wittkampf K, Lucassen P, van de Lisdonk E, van den Hoogen H, van Weert H, et al. Detecting somatoform disorders in primary care with the PHQ-15. Ann Fam Med 2009;7:232–8.

[19] Kroenke K, Spitzer RL, Williams JBW. The PHQ-9 — validity of a brief depression severity measure. J Gen Intern Med 2001;16:606–13.

[20] Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr Ann 2002;32:509–15.

[21] Williams Jr JW, Pignone M, Ramirez G, Perez SC. Identifying depression in primary care: a literature synthesis of case-finding instruments. Gen Hosp Psychiatry 2002;24:225–37.

[22] Lowe B, Spitzer RL, Grafe K, Kroenke K, Quenter A, Zipfel S, et al. Comparative validity of three screening questionnaires for *DSM-IV* depressive disorders and physicians' diagnoses. J Affect Disord 2004;78:131–40.

[23] Henkel V, Mergl R, Kohnen R, Allgaier AK, Moller HJ, Hegerl U. Use of brief depression screening tools in primary care: consideration of heterogeneity in performance in different patient groups. Gen Hosp Psychiatry 2004;26:190–8.

[24] Fann JR, Berry DL, Wolpin S, ustin-Seymour M, Bush N, Halpenny B, et al. Depression screening using the Patient Health Questionnaire-9 administered on a touch screen computer. Psychooncology 2009;18:14–22.

[25] Klapow J, Kroenke K, Horton T, Schmidt S, Spitzer R, Williams JBW. Psychological disorders and distress in older primary care

patients: a comparison of older and younger samples. Psychosom Med 2002;64:635–43.
[26] Lowe B, Unutzer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care 2004;42:1194–201.
[27] Huang FY, Chung H, Kroenke K, Spitzer RL. Racial and ethnic differences in the relationship between depression severity and functional status. Psychiatr Serv 2006;57:498–503.
[28] Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL. Using the patient health questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. J Gen Intern Med 2006;21:547–52.
[29] Lowe B, Kroenke K, Herzog W, Grafe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). J Affect Disord 2004;81:61–6.
[30] Lowe B, Schenkel I, Carney-Doebbeling C, Gobel C. Responsiveness of the PHQ-9 to psychopharmacological depression treatment. Psychosomatics 2006;47:62–7.
[31] Wittkampf K, van Ravesteijn H, Bass K, van de Hoogen H, Schene A, Bindels P, et al. The accuracy of Patient Health Questionnaire-9 in detecting depression and measuring depression severity in high-risk groups in primary care. Gen Hosp Psychiatry 2009;31:451–9.
[32] Williams LS, Kroenke K, Bakas T, Plue LD, Brizendine E, Tu WZ, et al. Care management of poststroke depression — a randomized, controlled trial. Stroke 2007;38:998-1003.
[33] Mannel M, Kuhn U, Schmidt U, Ploch M, Murck H. St. John's wort extract LI160 for the treatment of depression with atypical features — a double-blind, randomized, and placebo-controlled trial. J Psychiatr Res 2010; doi: 10.1016/j.jpsychires.2010.01.010. [in press].
[34] Rogers WH, Adler DA, Bungay KM, Wilson IB. Depression screening instruments made good severity measures in a cross-sectional analysis. J Clin Epidemiol 2005;58:370–7.
[35] Cannon DS, Tiffany ST, Coon H, Scholand MB, McMahon WM, Leppert MF. The PHQ-9 as a brief assessment of lifetime major depression. Psychological Assessment 2007;19:247–51.
[36] Callahan CM, Boustani MA, Unverzagt FW, Austrom MG, Damush TM, Perkins AJ, et al. Effectiveness of collaborative care for older adults with Alzheimer disease in primary care — a randomized controlled trial. JAMA 2006;295:2148–57.
[37] Williams LS, Jones WJ, Shen J, Robinson RL, Kroenke K. Outcomes of newly referred neurology outpatients with depression and pain. Neurology 2004;63:674–7.
[38] McManus D, Pipkin SS, Whooley MA. Screening for depression in patients with coronary heart disease (data from the heart and soul study). Am J Cardiology 2005;96:1076–81.
[39] Holzapfel N, Zugck C, Muller-Tasch T, Lowe B, Wild B, Schellberg D, et al. Routine screening for depression and quality of life in outpatients with congestive heart failure. Psychosomatics 2007;48:112–6.
[40] Stafford L, Berk M, Jackson HJ. Validity of the Hospital Anxiety and Depression Scale and Patient Health Questionnaire-9 to screen for depression in patients with coronary artery disease. Gen Hosp Psychiatry 2007;29:417–24.
[41] Whooley MA, de JP, Vittinghoff E, Otte C, Moos R, Carney RM, et al. Depressive symptoms, health behaviors, and risk of cardiovascular events in patients with coronary heart disease. JAMA 2008;300:2379–88.
[42] Thombs BD, de JP, Coyne JC, Whooley MA, Frasure-Smith N, Mitchell AJ, et al. Depression screening and patient outcomes in cardiovascular care: a systematic review. JAMA 2008;300:2161–71.
[43] Justice AC, McGinnis KA, Atkinson JH, Heaton RK, Young C, Sadek J, et al. Psychiatric and neurocognitive disorders among HIV-positive and negative veterans in care: Veterans Aging Cohort Five-Site Study. AIDS 2004;18:S49–59.
[44] Monahan PO, Shacham E, Reece M, Kroenke K, Ong'or WO, Omollo O, et al. Validity/reliability of PHQ-9 and PHQ-2 depression scales among adults living with HIV/AIDS in western Kenya. J Gen Intern Med 2009;24:189–97.
[45] Glasgow RE, Nutting PA, King DK, Nelson CC, Cutter G, Gaglio B, et al. A practical randomized trial to improve diabetes care. J Gen Intern Med 2004;19:1167–74.
[46] Katon WJ, Simon G, Russo J, Von Korff M, Lin EHB, Ludman E, et al. Quality of depression care in a population-based sample of patients with diabetes and major depression. Med Care 2004;42:1222–9.
[47] Drayer RA, Piraino B, Reynolds CF, Houck PR, Mazumdar S, Bernardini J, et al. Characteristics of depression in hemodialysis patients: symptoms, quality of life and mortality risk. Gen Hosp Psychiatry 2006;28:306–12.
[48] Dwight-Johnson M, Ell K, Lee PJ. Can collaborative care address the needs of low-income Latinas with comorbid depression and cancer? Results from a randomized pilot study. Psychosomatics 2005; 46:224–32.
[49] Kroenke K, Theobald D, Norton K, Sanders R, Schlundt S, McCalley S, et al. Indiana Cancer Pain and Depression (INCPAD) Trial: design of a telecare management intervention for cancer-related symptoms and baseline characteristics of enrolled participants. Gen Hosp Psychiatry 2009;31:240–53.
[50] Ell K, Xie B, Quon B, Quinn DI, Dwight-Johnson M, Lee PJ. Randomized controlled trial of collaborative care management of depression among low-income patients with cancer. J Clin Oncol 2008;26:4488–96.
[51] Lowe B, Willand L, Eich W, Zipfel S, Ho AD, Herzog W, et al. Psychiatric comorbidity and work disability in patients with inflammatory rheumatic diseases. Psychosom Med 2004;66:395–402.
[52] Rosemann T, Backenstrass M, Joest K, Rosemann A, Szecsenyi J, Laux G. Predictors of depression in a sample of 1,021 primary care patients with osteoarthritis. Arthritis Care Res 2007;57:415–22.
[53] Persoons P, Vermeire S, Demyttenaere K, Fischler B, Vandenberghe J, Van Oudenhove L, et al. The impact of major depressive disorder on the short- and long-term outcome of Crohn's disease treatment with infliximab. Alimentary Pharmacol Therapeutics 2005;22:101–10.
[54] Picardi A, Amerio P, Baliva G, Barbieri C, Teofoli P, Bolli S, et al. Recognition of depressive and anxiety disorders in dermatological outpatients. Acta Dermato-Venereologica 2004;84:213–7.
[55] Scholle SH, Haskett RF, Hanusa BH, Pincus HA, Kupfer DJ. Addressing depression in obstetrics/gynecology practice. Gen Hosp Psychiatry 2003;25:83–90.
[56] Leithner K, ssem-Hilger E, Fischer-Kern M, Loeffler-Stastka H, Sam C, Ponocny-Seliger E. Psychiatric morbidity in gynecological and otorhinolaryngological outpatients: a comparative study. Gen Hosp Psychiatry 2009;31:233–9.
[57] Lamoureux EL, Tee HW, Pesudovs K, Pallant JF, Keeffe JE, Rees G. Can clinicians use the PHQ-9 to assess depression in people with vision loss? Optometry Vision Sci 2009;86:139–45.
[58] Maizels M, Smitherman TA, Penzien DB. A review of screening tools for psychiatric comorbidity in headache patients. Headache 2006;46: S98–S109.
[59] Persoons P, Luyckx K, Desloovere C, Vandenberghe J, Fischler B. Anxiety and mood disorders in otorhinolaryngology outpatients presenting with dizziness: validation of the self-administered PRIME-MD Patient Health Questionnaire and epidemiology. Gen Hosp Psychiatry 2003;25:316–23.
[60] Tietjen GE, Brandes JL, Digre KB, Baggaley S, Martin V, Recober A, et al. High prevalence of somatic symptoms and depression in women with disabling chronic headache. Neurology 2007;68:134–40.
[61] Turner JA, Dworkin SF. Screening for psychosocial risk factors in patients with chronic orofacial pain — recent advances. J Am Dent Assoc 2004;135:1119–25.
[62] Gameroff MJ, Olfson M. Major depressive disorder, somatic pain, and health care costs in an urban primary care practice. J Clin Psychiatry 2006;67:1232–9.
[63] Kroenke K, Bair M, Damush T, Hoke S, Nicholas G, Kempf C, et al. Stepped Care for Affective Disorders and Musculoskeletal Pain

(SCAMP) study design and practical implications of an intervention for comorbid pain and depression. Gen Hosp Psychiatry 2007;29:506–17.
[64] Dobscha SK, Corson K, Perrin NA, Hanson GC, Leibowitz RQ, Doak MN, et al. Collaborative care for chronic pain in primary care: a clustered randomized trial. JAMA 2009;301:1242–52.
[65] Fann JR, Bombardier CH, Dikmer S, Esselman P, Warms CA, Pelzer E, et al. Validity of the patient health questionnaire-9 in assessing depression following traumatic brain injury. J Head Trauma Rehab 2005;20:501–11.
[66] Bombardier CH, Richards JS, Krause JS, Tulsky D, Tate DG. Symptoms of major depression in people with spinal cord injury: implications for screening. Arch Physical Med Rehab 2004;85:1749–56.
[67] Lowe B, Grafe K, Kroenke K, Zipfel S, Quenter A, Wild B, et al. Predictors of psychiatric comorbidity in medical outpatients. Psychosom Med 2003;65:764–70.
[68] Turvey CL, Willyard D, Hickman DH, Klein DM, Kukoyi O. Telehealth screen for depression in a chronic illness care management program. Telemed J E-Health 2007;13:51–6.
[69] Nease Jr DE, Nutting PA, Dickinson WP, Bonham AJ, Graham DG, Gallagher KM, et al. Inducing sustainable improvement in depression care in primary care practices. Jt Comm J Qual Patient Saf 2008;34:247–55.
[70] Duffy FF, Chung H, Trivedi M, Rae DS, Regier DA, Katzelnick DJ. Systematic use of patient-rated depression severity monitoring: Is it helpful and feasible in clinical psychiatry? Psychiatr Serv 2008;59:1148–54.
[71] Saliba D, Buchanan J. Development and validation of a revised nursing home assessment tool: MDS 3.0. Report for Centers for Medicare and Medicaid Services. HHS; 2009.
[72] Watson LC, Zimmerman S, Cohen LW, Dominik R. Practical depression screening in residential care/assisted living: five methods compared with gold standard diagnoses. Am J Geriatric Psychiatry 2009;17:556–64.
[73] JOHNSON JG, Harris ES, Spitzer RL, Williams JBW. The Patient Health Questionnaire for Adolescents: validation of an instrument for the assessment of mental disorders among adolescent primary care patients. J Adolesc Health 2002;30:196–204.
[74] Zuckerbrot RA, Jensen PS. Improving recognition of adolescent depression in primary care. Arch Pediatr Adolesc Med 2006;160:694–704.
[75] Zuckerbrot RA, Cheung AH, Jensen PS, Stein RE, Laraque D. Guidelines for Adolescent Depression in Primary Care (GLAD-PC): I. Identification, assessment, and initial management. Pediatrics 2007;120:e1299–312.
[76] Richardson L, McCauley E, Katon W. Collaborative care for adolescent depression: a pilot study. Gen Hosp Psychiatry 2009;31:36–45.
[77] Gjerdingen D, Crow S, McGovern P, Miner M, Center B. Postpartum depression screening at well-child visits: validity of a 2-question screen and the PHQ-9. Ann Fam Med 2009;7:63–70.
[78] Weobong B, Akpalu B, Doku V, Owusu-Agyei S, Hurt L, Kirkwood B, et al. The comparative validity of screening scales for postnatal common mental disorder in Kintampo, Ghana. J Affect Disord 2009;113:109–17.
[79] Bennett IM, Coco A, Coyne JC, Mitchell AJ, Nicholson J, Johnson E, et al. Efficiency of a two-item pre-screen to reduce the burden of depression screening in pregnancy and postpartum: an IMPLICIT network study. J Am Board Fam Med 2008;21:317–25.
[80] Hanusa BH, Scholle SH, Haskett RF, Spadaro K, Wisner KL. Screening for depression in the postpartum period: A comparison of three instruments. J Womens Health 2008;17:585–96.
[81] Yonkers KA, Smith MV, Lin HQ, Howell HB, Shao L, Rosenheck RA. Depression screening of perinatal women: an evaluation of the Healthy Start Depression Initiative. Psychiatr Serv 2009;60:322–8.
[82] Hewitt CE, Gilbody SM, Brealey S, Paulden M, Palmer S, Mann R, et al. Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis. Health Technol Assess 2009;13:1-230.
[83] Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression. Two questions are as good as many. J Gen Intern Med 1997;12:439–45.
[84] Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2 — Validity of a two-item depression screener. Med Care 2003;41:1284–92.
[85] Corson K, Gerrity MS, Dobscha SK. Screening for depression and suicidality in a VA primary care setting: 2 items are better than 1 item. Am J Managed Care 2004;10:839–45.
[86] Lowe B, Kroenke K, Grafe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). J Psychosom Res 2005;58:163–71.
[87] Li CY, Friedman B, Conwell Y, Fiscella K. Validity of the Patient Health Questionnaire 2 (PHQ-2) in identifying major depression in older people. J Am Geriatrics Soc 2007;55:596–602.
[88] Mitchell AJ, McGlinchey JB, Young D, Chelminski I, Zimmerman M. Accuracy of specific symptoms in the diagnosis of major depressive disorder in psychiatric out-patients: data from the MIDAS project. Psychol Med 2009;39:1107–16.
[89] Rief W, Nanke A, Klaiberg A, Braehler E. Base rates for panic and depression according to the Brief Patient Health Questionnaire: a population-based study. J Affect Disord 2004;82:271–6.
[90] Lee PW, Schulberg HC, Raue PJ, Kroenke K. Concordance between the PHQ-9 and the HSCL-20 in depressed primary care patients. J Affect Disord 2007;99:139–45.
[91] Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009;114:163–73.
[92] First MB, Spitzer RL, Gibbon M, Williams JB. Structured Clinical Interview for DSM-IV Axis I Disorders (SCID). Washington, DC: American Psychiatric Press Inc.; 1996.
[93] Unutzer J, Katon W, Callahan CM, Williams Jr JW, Hunkeler E, Harpole L, et al. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. JAMA 2002;288:2836–45.
[94] Dietrich AJ, Oxman TE, Williams Jr JW, Schulberg HC, Bruce ML, Lee PW, et al. Re-engineering systems for the treatment of depression in primary care: cluster randomised controlled trial. BMJ 2004;329:602–5.
[95] Dobscha SK, Corson K, Hickam DH, Perrin NA, Kraemer DF, Gerrity MS. Depression decision support in primary care — a cluster randomized trial. Ann Intern Med 2006;145:477–87.
[96] Fortney JC, Pyne JM, Edlund MJ, Williams DK, Robinson DE, Mittal D, et al. A randomized trial of telemedicine-based collaborative care for depression. J Gen Intern Med 2007;22:1086–93.
[97] Ell K, Quon B, Quinn DI, Dwight-Johnson M, Wells A, Lee PJ, et al. Improving treatment of depression among low-income patients with cancer: the design of the ADAPt-C study. Gen Hosp Psychiatry 2007;29:223–31.
[98] Strong V, Waters R, Hibberd C, Murray G, Wall L, Walker J, et al. Management of depression for people with cancer (SMaRT oncology 1): a randomised trial. Lancet 2008;372:40–8.
[99] Richards DA, Lovell K, Gilbody S, Gask L, Torgerson D, Barkham M, et al. Collaborative care for depression in UK primary care: a randomized controlled trial. Psychol Med 2008;38:279–87.
[100] Kroenke K, Bair MJ, Damush TM, Wu J, Hoke S, Sutherland J, et al. Optimized antidepressant therapy and pain self-management in primary care patients with depression and musculoskeletal pain: a randomized controlled trial. JAMA 2009;301:2099–110.
[101] Strine TW, Mokdad AH, Balluz LS, Gonzalez O, Crider R, Berry JT, et al. Depression and anxiety in the United States: findings from the 2006 Behavioral Risk Factor Surveillance System. Psychiatr Serv 2008;59:1383–90.
[102] Pratt LA, Brody DJ. Depression in the United States household population, 2005-2006. NCHS Data Brief 2008;7.

[103] Fleishman JA, Zuvekas SH. Global self-rated mental health: associations with other mental health measures and with role functioning. Med Care 2007;45:602–9.
[104] Taylor JK, Schoenbaurn M, Katon WJ, Pincus HA, Hogan DM, Unutzer J. Strategies for identifying and channeling patients for depression care management. Am J Managed Care 2008;14:497–504.
[105] Smith TC, Smith B, Jacobson IG, Corbeil TE, Ryan MAK. Reliability of standard health assessment instruments in a large, population-based cohort study. Ann Epidemiol 2007;17:525–32.
[106] Martin A, Rief W, Klaiberg A, Braehler E. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. Gen Hosp Psychiatry 2006;28:71–7.
[107] Pirkis J, Pfaff J, Williamson M, Tyson O, Stocks N, Goldney R, et al. The community prevalence of depression in older Australians. J Affective Disorders 2009;115:54–61.
[108] Patten SB, Schopflocher D. Longitudinal epidemiology of major depression as assessed by the Brief Patient Health Questionnaire (PHQ-9). Comprehensive Psychiatry 2009;50:26–33.
[109] Desai MM, Rosenheck RA, Craig TJ. Case-finding for depression among medical outpatients in the Veterans Health Administration. Med Care 2006;44:175–81.
[110] Engel CC, Oxman T, Yamamoto C, Gould D, Barry S, Stewart P, et al. RESPECT-Mil: feasibility of a systems-level collaborative care approach to depression and post-traumatic stress disorder in military primary care. Mil Med 2008;173:935–40.
[111] Arnow BA, Hunkeler EM, Blasey CM, Lee J, Constantino MJ, Fireman B, et al. Comorbid depression, chronic pain, and disability in primary care. Psychosom Med 2006;68:262–8.
[112] Katzelnick DJ, Von KM, Chung H, Provost LP, Wagner EH. Applying depression-specific change concepts in a collaborative breakthrough series. Jt Comm J Qual Patient Saf 2005;31:386–97.
[113] Bremer RW, Scholle SH, Keyser D, Houtsinger JV, Pincus HA. Pay for performance in behavioral health. Psychiatr Serv 2008;59:1419–29.
[114] Sederer LI, Silver L, McVeigh KH, Levy J. Integrating care for medical and mental illnesses. Prev Chronic Dis 2006;3:A33.
[115] Kendrick T, Dowrick C, McBride A, Howe A, Clarke P, Maisey S, et al. Management of depression in UK general practice in relation to scores on depression severity questionnaires: analysis of medical record data. BMJ 2009:338.
[116] Oxman TE, Dietrich AJ, Williams Jr JW, Kroenke K. A three-component model for reengineering systems for the treatment of depression in primary care. Psychosomatics 2002;43:441–50.
[117] Kilbourne AM, Schulberg HC, Post EP, Rollman BL, Belnap BH, Pincus HA. Translating evidence-based depression management services to community-based primary care practices. Milbank Quarterly 2004;82:631–59.
[118] Lichtman JH, Bigger Jr JT, Blumenthal JA, Frasure-Smith N, Kaufmann PG, Lesperance F, et al. Depression and coronary heart disease. recommendations for screening, referral, and treatment. Circulation 2008;118:1765–8.
[119] Andrews G, Anderson TM, Slade T, Sunderland M. Classification of anxiety and depressive disorders: problems and solutions. Depress Anxiety 2008;25:274–81.
[120] Lee PW, Dietrich AJ, Oxman TE, Williams Jr JW, Barry SL. Sustainable impact of a primary care depression intervention. J Am Board Fam Pract 2007;20:427–33.
[121] Lowe B, Grafe K, Zipfel S, Spitzer RL, Herrmann-Lingen C, Witte S, et al. Detecting panic disorder in medical and psychosomatic outpatients — comparative validation of the Hospital Anxiety and Depression Scale, the Patient Health Questionnaire, a screening question, and physicians' diagnosis. J Psychosom Res 2003;55: 515–9.
[122] Corapcioglu A, Ozer GU. Adaptation of revised Brief PHQ (Brief-PHQ-r) for diagnosis of depression, panic disorder and somatoform disorder in primary healthcare settings. Int J Psychiatry Clin Pract 2004;8:11–8.
[123] Means-Christensen AJ, Arnau RC, Tonidandel AM, Bramson R, Meagher MW. An efficient method of identifying major depression and panic disorder in primary care. J Behav Med 2005;28:565–72.
[124] Rollman BL, Belnap BH, Mazumdar S, Zhu F, Kroenke K, Schulberg HC, et al. Symptomatic severity of PRIME-MD diagnosed episodes of panic and generalized anxiety disorder in primary care. J Gen Intern Med 2005;20:623–8.
[125] Pilowsky DJ, Olfson M, Gameroff MJ, Wickramaratne P, Blanco C, Feder A, et al. Panic disorder and suicidal ideation in primary care. Depress Anxiety 2006;23:11–6.
[126] Kessler RC, Brandenburg N, Lane M, Roy-Byrne P, Stang PD, Stein DJ, et al. Rethinking the duration requirement for generalized anxiety disorder: evidence from the National Comorbidity Survey Replication. Psychol Med 2005;35:1073–82.
[127] Lowe B, Decker O, Muller S, Brahler E, Schellberg D, Herzog W, et al. Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. Med Care 2008;46:266–74.
[128] Issakidis C, Sanderson K, Corry J, Andrews G, Lapsley H. Modelling the population cost-effectiveness of current and evidence-based optimal treatment for anxiety disorders. Psychol Med 2004;34:19–35.
[129] Roy-Byrne P, Veitengruber JP, Bystritsky A, Edlund MJ, Sullivan G, Craske MG, et al. Brief intervention for anxiety in primary care patients. J Am Board Fam Med 2009;22:175–86.
[130] Bair MJ, Wu J, Damush TM, Sutherland JM, Kroenke K. Association of depression and anxiety alone and in combination with chronic musculoskeletal pain in primary care patients. Psychosom Med 2008;70:890–7.
[131] Sherbourne CD, Asch SM, Shugarman LR, Goebel JR, Lanto AB, Rubenstein LV, et al. Early identification of co-occurring pain, depression and anxiety. J Gen Intern Med 2009;24:620–5.
[132] Kroenke K, Sharpe M, Sykes R. Revising the classification of somatoform disorders: key questions and preliminary recommendations. Psychosomatics 2007;48:277–85.
[133] Dimsdale J, Creed F. The proposed diagnosis of somatic symptom disorders in *DSM-V* to replace somatoform disorders in *DSM-IV*–a preliminary report. J Psychosom Res 2009;66:473–6.
[134] Katon W, Lin EH, Kroenke K. The association of depression and anxiety with medical symptom burden in patients with chronic medical illness. Gen Hosp Psychiatry 2007;29:147–55.
[135] Kroenke K. Somatoform disorders and recent diagnostic controversies. Psychiatr Clin North Am 2007;30:593–619.
[136] Kroenke K, Spitzer RL, deGruy III FV, Hahn SR, Linzer M, Williams JB, et al. Multisomatoform disorder. An alternative to undifferentiated somatoform disorder for the somatizing patient in primary care. Arch Gen Psychiatry 1997;54:352–8.
[137] Kroenke K, Spitzer RL, deGruy FV, Swindle R. A symptom checklist to screen for somatoform disorders in primary care. Psychosomatics 1998;39:263–72.
[138] Interian A, Allen LA, Gara MA, Escobar JI, az-Martinez AM. Somatic complaints in primary care: Further examining the validity of the patient health questionnaire (PHQ-15). Psychosomatics 2006;47:392–8.
[139] Kroenke K, Messina III N, Benattia I, Graepel J, Musgnung J. Venlafaxine extended release in the short-term treatment of depressed and anxious primary care patients with multisomatoform disorder. J Clin Psychiatry 2006;67:72–80.
[140] Gerber MR, Wittenberg E, Ganz ML, Williams CM, McCloskey LA. Intimate partner violence exposure and change in women's physical symptoms over time. J Gen Intern Med 2008;23:64–9.
[141] Escobar JI, Gara MA, az-Martinez AM, Interian A, Warman M, Allen LA, et al. Effectiveness of a time-limited cognitive behavior therapy type intervention among primary care patients with medically unexplained symptoms. Ann Fam Med 2007;5:328–35.
[142] Kroenke K, West SL, Swindle R, Gilsenan A, Eckert GJ, Dolor R, et al. Similar effectiveness of paroxetine, fluoxetine, and sertraline in primary care: a randomized trial. JAMA 2001;286:2947–55.

[143] Reuber M, Burness C, Howlett S, Brazier J, Grunewald R. Tailored psychotherapy for patients with functional neurological symptoms: a pilot study. J Psychosom Res 2007;63:625–32.
[144] Kroenke K. Physical symptom disorder: a simpler diagnostic category for somatization-spectrum conditions. J Psychosom Res 2006;60:335–9.
[145] Barsky AJ, Orav EJ, Bates DW. Somatization increases medical utilization and costs independent of psychiatric and medical comorbidity. Arch Gen Psychiatry 2005;62:903–10.
[146] Barsky AJ, Orav EJ, Bates DW. Distinctive patterns of medical care utilization in patients who somatize. Med Care 2006;44:803–11.
[147] Rost KM, Dickinson WP, Dickinson LM, Smith RC. Multisomatoform disorder: agreement between patient and physician report of criterion symptom explanation. CNS Spectr 2006;11:383–8.
[148] Hahn SR. Physical symptoms and physician-experienced difficulty in the physician-patient relationship. Ann Intern Med 2001;134:897–904.
[149] Jackson JL, Kroenke K. The effect of unmet expectations among adults presenting with physical symptoms. Ann Intern Med 2001;134:889–97.
[150] Smith RC, Lein C, Collins C, Lyles JS, Given B, Dwamena FC, et al. Treating patients with medically unexplained symptoms in primary care. J Gen Intern Med 2003;18:478–89.
[151] Jackson JL, O'Malley PG, Kroenke K. Antidepressants and cognitive-behavioral therapy for symptom syndromes. CNS Spectr 2006;11:212–22.
[152] Kroenke K. Efficacy of treatment for somatoform disorders: a review of randomized controlled trials. Psychosom Med 2007;69:881–8.
[153] Sha MC, Callahan CM, Counsell SR, Westmoreland GR, Stump TE, Kroenke K. Physical symptoms as a predictor of health care use and mortality among older adults. Am J Med 2005;118:301–6.
[154] Clarke DM, Piterman L, Byrne CJ, Austin DW. Somatic symptoms, hypochondriasis and psychological distress: a study of somatisation in Australian general practice. Med J Aust 2008;189:560–4.
[155] Mewes R, Rief W, Brahler E, Martin A, Glaesmer H. Lower decision threshold for doctor visits as a predictor of health care use in somatoform disorders and in the general population. Gen Hosp Psychiatry 2008;30:349–55.
[156] Ryan MAK, Smith TC, Smith B, Amoroso P, Boyko EJ, Gray GC, et al. Millennium Cohort: enrollment begins a 21-year contribution to understanding the impact of military service. J Clin Epidemiol 2007;60:181–91.
[157] Kang HK, Li B, Mahan CM, Eisen SA, Engel CC. Health of US veterans of 1991 Gulf War: a follow-up survey in 10 years. J Occup Environ Med 2009;51:401–10.
[158] Hoge CW, McGurk D, Thomas JL, Cox AL, Engel CC, Castro CA. Mild traumatic brain injury in U.S. Soldiers returning from Iraq. N Engl J Med 2008;358:453–63.
[159] Harris AM, Orav EJ, Bates DW, Barsky AJ. Somatization increases disability independent of comorbidity. J Gen Intern Med 2009;24:155–61.
[160] Hanel G, Henningsen P, Herzog W, Sauer N, Schafert R, Szecsenyi J, et al. Depression, anxiety, and somatoform disorders: vague or distinct categories in primary care? Results from a large cross-sectional study. J Psychosom Res 2009;67:189–97.
[161] Ormel J, Vonkorff M, Ustun TB, Pini S, Korten A, Oldehinkel T. Common mental disorders and disability across cultures. Results from the WHO Collaborative Study on Psychological Problems in General Health Care. JAMA 1994;272:1741–8.
[162] Spitzer RL, Kroenke K, Linzer M, Hahn SR, Williams JBW, deGruy FV, et al. Health-related quality-of-life in primary-care patients with mental disorders — Results from the PRIME-MD 1000 Study. JAMA 1995;274:1511–7.
[163] Patel V, Araya R, Chowdhary N, King M, Kirkwood B, Nayak S, et al. Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires. Psychol Med 2008;38:221–8.
[164] Gilbody S, Richards D, Barkham M. Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ-9 and CORE-OM. Br J Gen Pract 2007;57:650–2.
[165] Clark DM, Layard R, Smithies R, Richards DA, Suckling R, Wright B. Improving access to psychological therapy: initial evaluation of two UK demonstration sites. Behav Res Ther 2009;47:910–20.
[166] Mitchell AJ, Coyne JC. Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. Br J Gen Pract 2007;57:144–51.
[167] Kroenke K, Spitzer RL, Williams JBW, Lowe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics 2009;50:613–21.
[168] Lowe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, et al. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. J Affect Disord 2009.
[169] Gilbody S, Sheldon T, House A. Screening and case-finding instruments for depression: a meta-analysis. CMAJ 2008;178:997–1003.
[170] Gilbody S, Bower P, Fletcher J, Richards D, Sutton AJ. Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. Arch Intern Med 2006;166:2314–21.
[171] Williams Jr JW, Gerrity M, Holsinger T, Dobscha S, Gaynes B, Dietrich A. Systematic review of multifaceted interventions to improve depression care. Gen Hosp Psychiatry 2007;29:91-116.
[172] U.S. Preventive Services Task Force. Screening for depression in adults: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2009;151:784–92.
[173] Gaynes BN, Rush AJ, Trivedi MH, Wisniewski SR, Balasubramani GK, McGrath PJ, et al. Primary versus specialty care outcomes for depressed outpatients managed with measurement-based care: results from STAR*D. J Gen Intern Med 2008;23:551–60.
[174] Gilbody S, House A, Sheldon T. Psychiatrists in the UK do not use outcomes measures. National survey. Br J Psychiatry 2002;180:101–3.
[175] Zimmerman M, McGlinchey JB, Chelminski I. An inadequate community standard of care: lack of measurement of outcome when treating depression in clinical practice. Primary Psychiatry 2008;15:67–75.
[176] Trivedi MH. Tools and strategies for ongoing assessment of depression: a measurement-based approach to remission. J Clin Psychiatry 2009;70(Suppl 6):26–31.