Using scales to monitor symptoms and treat depression (measurement based care)

AUTHOR: Mark Zimmerman, MD
SECTION EDITOR: Peter P Roy-Byrne, MD
DEPUTY EDITORS: Sara Swenson, MD, David Solomon, MD

All topics are updated as new evidence becomes available and our peer review process is complete.

Literature review current through: **Oct 2023.**
This topic last updated: **Feb 15, 2023.**

INTRODUCTION

It is necessary to evaluate outcome in order to determine the effectiveness of treatment. Clinicians treating hypertension do this by consistently measuring blood pressure, and treatment of diabetes always involves measuring serum glucose or hemoglobin A1C.

Psychiatrists treating depression can monitor progress by serially measuring severity of symptoms with a standardized scale. However, most psychiatrists do not. A survey of 500 psychiatrists in 2000 found that among the 340 who responded, 58 percent never used a scale to measure clinical change of depression and anxiety [1]. Another survey of 306 psychiatrists in 2006 to 2007 found that 29 percent never used scales and 32 percent did so rarely. More recently, a 2023 study of 108 clinicians (90 percent were psychiatrists) found that only 13 percent agreed or strongly agreed that they used a measurement based care approach at each visit to monitor treatment [2]. Among the psychiatrists who did not routinely monitor symptoms with a standardized scale, the primary reasons were lack of training and time. In the earlier studies, the psychiatrists did not believe that scales would be clinically helpful, but the majority of psychiatrists in the more recent surveys indicated that measuring symptoms would be helpful in monitoring outcomes and making treatment decisions [2,3].

Mental health clinicians typically assess progress of their depressed patients through unstructured interactions that yield unquantified judgments. Some clinicians ask only broad, global questions such as “How are you feeling?” or “How are you doing?” Many patients reply

with global responses such as “Okay” or “Fine.” However, these responses often do not accurately reflect the patient’s clinical status. As a result, it is increasingly recognized that incorporating standardized scales into clinical practice to measure depression may help clinicians evaluate the patient’s current status more accurately.

This topic reviews the use of depression rating scales in routine clinical practice. Initial treatment of depression and management of treatment resistant patients are discussed elsewhere. (See "Unipolar major depression in adults: Choosing initial treatment" and "Unipolar treatment-resistant depression in adults: Epidemiology, risk factors, assessment, and prognosis", section on 'Assessment and identification' and "Unipolar depression in adults: Choosing treatment for resistant depression".)

EVIDENCE OF EFFICACY

The evidence indicates that systematically monitoring depressive symptoms with a standardized scale can improve treatment outcomes. In some studies, monitoring was part of a program that included education or reinforcement of evidence-based treatment:

- A meta-analysis of nine controlled trials (seven randomized) evaluated measurement based care provided in specialist mental health settings for patients (n >4000) with different disorders, including depression, anxiety, personality disorder, and eating disorder, as well as personal concerns [4]. The trials compared treatment in which clinicians received feedback about symptoms assessed with standardized measures, with treatment that did not include feedback. Mental health outcomes within nine weeks of intake were better with treatment that included feedback, but the clinical benefit was small. In addition, outcomes 3 and 12 months after intake were comparable for treatment that included feedback and treatment that did not (five trials, 573 patients).

- A subsequent set of meta-analyses involved seven different randomized trials, which compared measurement based care with clinical judgement alone (standard care) in patients with depressive disorders treated with pharmacotherapy (n >2000) [5]. In the active treatment conditions, physicians reviewed routinely administered outcome measures to guide treatment decisions; the measures included the Patient Health Questionnaire – Nine Item (PHQ-9) (table 1), Hamilton Rating Scale for Depression (table 2), or Montgomery-Asberg Depression Rating Scale [6] (figure 1A-C). The trials were conducted in outpatient psychiatry or primary care clinics, and inpatient psychiatry units. Measurement based care was superior to standard care with regard to:

- Remission (53 versus 43 percent of patients)
- Depressive symptoms (clinically moderate benefit)
- Adherence to pharmacotherapy (odds ratio 1.7, 95% CI 1.2-2.3)

REASONS TO MEASURE OUTCOME

There are multiple theoretical reasons that support routinely using rating scales to assess outcome in depressed patients. In addition, measurement-based care can be implemented in such a way that it requires little additional time on the part of the clinician or support staff. Thus, potential benefits are large, with little cost. Measurement based care is consistent with multiple practice guidelines, including those from the American Psychiatric Association [7-9].

Identify nonresponders — Using a scale to monitor the patient's progress may help prompt clinicians to change treatment if the patient does not improve. It is our clinical experience that many patients remain on an antidepressant for several months without any benefit. Using a scale to quantify severity of depression may help identify nonresponders. However, there is no evidence showing that measurement-based care is superior to clinical interviews for detecting nonresponse in patients treated for unipolar major depression.

Studies indicate that patients who have not improved by at least 20 percent after two weeks of treatment are unlikely to improve after two to eight weeks of additional treatment [10,11]. As an example, a meta-analysis studied 6109 patients with unipolar major depression treated with mirtazapine or another antidepressant (41 controlled trials). Among patients who did not improve by at least 20 percent, 82 to 96 percent failed to achieve stable remission or a stable response (at least 50 percent improvement) after longer treatment with the same agent. However, no studies have established that changing treatment in nonresponders after two weeks results in improved outcome.

Detect residual symptoms — Measurement provides the clinician with information regarding the completeness of treatment success, and suboptimal outcomes prompt an intervention to further improve treatment outcome. As an example, reducing blood pressure from 165/105 to 145/95 is a positive yet incomplete response that warrants an adjustment in treatment.

Similarly, an incomplete response to treatment of depression should prompt the clinician to consider revising the treatment plan. Multiple studies have consistently found that residual symptoms were associated with an increased risk of relapse in depressed patients who otherwise responded to treatment [12-19]. However, there are no studies showing that

systematically using a scale improves the detection of residual symptoms. Furthermore, some patients with an incomplete response may not remit despite additional interventions [20].

Reduce dropout from treatment — Many depressed patients drop out of treatment within the first six months of initiating care [21,22]. The use of self-report scales to measure depression increases patients' active participation in their care, and this might facilitate participation in other therapeutic activities such as exercise. Patients who are more active in their treatment, and who believe that their clinicians better understand their clinical status, may in turn be more likely to continue treatment. A study of 24 depressed patients found that they overwhelmingly considered the use of scales helpful in managing their illness [23]. However, no studies have examined whether routine assessment with standardized scales reduces treatment dropout.

Help patients recognize improvement — Symptom assessment with a scale may help clinicians identify areas of improvement that otherwise may go unrecognized. As an example, a patient may remain dysphoric, pessimistic, unmotivated, self-deprecatory, and cognitively impaired despite two weeks of treatment and state at the beginning of the follow up visit that they are no better. Use of a scale may reveal however, that sleep, energy, and appetite are better. Identifying some areas of improvement may reduce the patient's hopelessness and therapeutic nihilism, thereby increasing treatment retention. In addition, recognizing mild levels of improvement (eg, 20 to 30 percent improvement over baseline) is important because it is a harbinger of future improvement [10,11].

Detect seasonal variation — It is easier to detect seasonal patterns of symptom fluctuation and diagnose seasonal affective disorder when looking at graphs of symptom scores. Thus, patients treated and followed longitudinally over a few years may benefit from the routine use of scales. (See "Seasonal affective disorder: Epidemiology, clinical features, assessment, and diagnosis".)

DESIRABLE FEATURES OF A DEPRESSION OUTCOME SCALE

Choosing a scale to monitor treatment depends upon the needs of the patient and clinician.

Patient perspective — Patients should find the measure user friendly with directions that are easy to follow. The questions should be understandable and relevant to the patient's problem. The scale should be brief, taking no more than two to three minutes to complete, so that upon routine administration at follow-up visits, patients are not inconvenienced by the need to arrive for their appointment 10 to 15 minutes early in order to complete the measure.

Clinician perspective — The instrument should have practical value for the clinician and improve the efficiency of conducting the clinical evaluation, and the clinician needs to trust the information that the measure provides. The scale should provide clinically useful information. In addition, clinicians and clinics should find that the instrument is user friendly; it should be easy to administer and score with minimal training. The outcome measure for depression must also have a sound basis in psychometrics to allay the concerns that many clinicians have about the validity of scales [23]. Thus, the instrument should demonstrate good reliability, validity, and sensitivity to change.

Desirable features — Bearing in mind the perspective of the patient and clinician, a useful depression scale has the following features [24]:

- Brevity
- Acceptable to patients and clinicians
- Sensitive to change
- Reliable (internal consistency and test-retest reliability)
- Convergent validity (high correlation with other measures of depression)
- Discriminant validity (low correlation with measures of other symptom domains such as phobias, compulsions, and mania)
- Indicator of symptom severity
- Indicator of remission status
- Assesses psychosocial function
- Assesses quality of life
- Assesses suicidal thoughts
- Easy to score
- Inexpensive to acquire and use
- Covers all of the diagnostic criteria for unipolar major depression in the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) [25]

POTENTIAL OBSTACLES IN MEASURING OUTCOME

There are several potential obstacles to systematically measuring outcome when treating depression in routine clinical practice. However, there are solutions to each obstacle.

Patient acceptability — If measurement is overly burdensome to patients, they may become dissatisfied and either drop out of treatment or seek care elsewhere. However, studies indicate that depressed outpatients readily accept measurement based care [23,26]:

- A study of 50 depressed outpatients who completed a self-report questionnaire found that the scale imposed very little or only a little burden in 98 percent of the patients [27]. In addition, 94 percent of the patients indicated a willingness to complete the scale at every visit in the future if their clinician believed that it was helpful.

- Two surveys of the potential feasibility of measurement based care asked patients if they were willing to spend five minutes completing a questionnaire before each appointment with their doctor [2,3]. Both studies found that more than 90 percent of the patients expressed a willingness to regularly complete brief questionnaires [2].

Patients may find briefer scales more acceptable for regular use than longer scales. A study compared a self-report scale consisting of 18 single items with a self-report scale consisting of 21 multiple choice questions, with four or five choices per question [27]. The 50 depressed patients who completed both scales were significantly more likely to view the shorter scale as less burdensome, and nearly three times as many patients preferred the shorter scale over the longer one.

Patients who are comfortable using the internet may prefer to complete scales online rather than at the clinician's office. One study found that both paper and electronic versions of a self-report scale were acceptable, but the electronic version was easier [26]. Other surveys of patients found that approximately 90 percent would be willing to use a mobile app to track their symptoms if their doctors recommended it [2,3].

If patients object to the use of any scale, it is reasonable to rely upon the clinical interview to assess symptoms or use a clinician-administered scale.

Clinician acceptability — An obstacle to using standardized scales in clinical practice is the perceived time burden of completing the scale. Clinicians may be reluctant to administer detailed instruments such as the Hamilton Rating Scale for Depression [28] (table 2) or Montgomery-Asberg Depression Rating Scale [6] (figure 1A-C) because of time pressures.

Self-report scales address this issue because they can improve the efficiency of conducting clinical interviews and need not interfere with the usual workflow. (See 'Self-report measurement' below and 'Self-report scales in the public domain' below.)

Cost — Some scales need to be purchased, require the clinician's time to administer the instrument, and may also require support staff time to copy, score, and file the instrument. However, there are self-report scales that are readily available, cost nothing to acquire, and require little in the way of support staff costs. (See 'Self-report scales in the public domain' below.)

SELF-REPORT MEASUREMENT

We suggest monitoring treatment of depression with self-report scales because they provide clinically meaningful information that correlates highly with clinician administered scales [26,29-33]. In addition, self-report questionnaires [34]:

- Do not require clinician time for administration
- Improve efficiency of clinical encounters
- May assess internal mental states more validly than clinician rating scales
- May help patients remember ongoing problems
- May capture sensitive information that the patient is reluctant to verbalize with the clinician
- Reduce or eliminate clinician bias to overestimate patient improvement
- May capture information that the clinician overlooks

Brief self-report scales can be readily completed at each follow up visit in the same way that blood pressure and weight are routinely assessed in primary care settings for patients being treated for hypertension and obesity [27,35].

However, there are disadvantages to self-report questionnaires. Patients may either minimize or over-report symptom severity (reporting bias), which in turn reduces validity of the scale. In addition, some patients cannot complete a scale due to illiteracy, physical impairment or debility, or impaired cognitive functioning.

COMMONLY USED SELF-REPORT DEPRESSION SCALES

There are many depression scales and a compendium of instruments is available [36]. Scales vary in length from a single item to more than 100. Some scales have been developed to

measure depression in specific populations such as patients who are postpartum, geriatric, or suffering from nonpsychiatric medical illnesses. Many scales have been used in clinical trials throughout the world and translated into different languages [34]. Several scales assessing the DSM-5 criteria have been developed and found to be valid and reliable measures of depression severity, although most scales were developed prior to DSM-5 and do not fully assess the diagnostic criteria for unipolar major depression.

There are several widely used, self-report depression scales that are in the public domain. (See 'Self-report scales in the public domain' below.) One commonly used self-administered scale that is not in the public domain is the Beck Depression Inventory-II (BDI-II) [37]. The scale has good internal consistency and item-scale correlations and is sensitive to change. However, the scale is copyrighted, users must pay a fee to purchase each copy of the scale administered, and unauthorized duplication for clinical use represents a violation of copyright. In addition, the scale is relatively long, and consists of 21 multiple choice items, with 4 statements per item. There is no evidence that the BDI-II is more valid or reliable than other depression scales, and thus it is difficult to justify its cost.

Self-report scales in the public domain — There are self-report scales that are copyrighted but in the public domain and thus available free of charge for unlimited clinical use. Each can be downloaded on the internet in user ready formats. Among the three commonly used scales that are in the public domain, there is no evidence that one is better than the others. However, the Patient Health Questionnaire – Nine Item (PHQ-9) has been studied and used most often (table 1) [38,39].

Patient Health Questionnaire - Nine Item — The Patient Health Questionnaire – Nine Item (PHQ-9) (table 1) is the standard among scales for monitoring symptoms of depression. The PHQ-9 consists of only nine items that correspond to the nine DSM-5 criteria for unipolar major depression, as well as an additional item assessing psychosocial impairment [38,39]. The PHQ-9 has been used in national and regional programs in the United States and United Kingdom that are intended to demonstrate the value of monitoring treatment of depression [35,40-42]. In addition, the PHQ-9 has been translated into many languages and is used internationally [34,43]. The scale is brief and should take less than two minutes to complete. The PHQ-9 has good test-retest reliability, internal consistency, and sensitivity to change in depression over time [44]. It has been extensively studied as a screening measure for major depression in primary care settings.

Compound symptom criteria are assessed with a single item. As an example, the PHQ-9 assesses insomnia and hypersomnia with a single item, and likewise, reduced or increased appetite. Thus, the nine-item format makes it easier to apply the DSM-5 diagnostic algorithm

for major depression, though at a cost of some information for the purpose of monitoring response to treatment.

The patient is instructed to rate each symptom item on a four-point Likert scale, indicating how often they have been bothered by the symptom over the past two weeks (0 = not at all, 1 = several days, 2 = more than half the days, 3 = nearly every day). Total scores on the PHQ-9 range from 0 to 27. Suggested cutoff scores for level of depression severity are as follows: a score of 0 to 4 indicates no depression, 5 to 9 mild depression, 10 to 14 moderate depression, 15 to 19 moderately severe depression, and a score of 20 to 27 indicates severe depression. However, several studies have found that the PHQ-9 overestimates depression severity compared with other self-report and clinician-rated measures of depression severity [45].

Clinically Useful Depression Outcome Scale — The Clinically Useful Depression Outcome Scale (CUDOS) contains 18 items assessing all of the DSM-5 criteria for unipolar major depression, as well as psychosocial impairment and quality of life (form 1) [24,46]. Compound DSM-5 symptom criteria referring to more than one construct (eg, insomnia or hypersomnia) are subdivided into their separate components. It usually takes less than two minutes to complete the scale. The CUDOS has good test-retest reliability, internal consistency, sensitivity to change, and can be used to screen for depression. The patient is instructed to rate the symptom items on a five point Likert scale indicating how well the item describes the patient during the past week (0 = not at all true/0 days; 1 = rarely true/one to two days; 2 = sometimes true/three to four days; 3 = usually true/five to six days; 4 = almost always true/every day). Total symptom scores on the scale range from 0 to 64, and empirically derived severity score ranges are 0 to 10 nondepressed, 11 to 20 minimal depression, 21 to 30 mild depression, 31 to 45 moderate depression, and 46 to 64 severe depression.

Quick Inventory of Depressive Symptomatology - Self Report 16 Item — The 16 items of the Quick Inventory of Depressive Symptomatology – Self Report 16 Item (QIDS-SR_{16}) (table 3) are multiple choice questions with four choices [29,47]. The items cover the symptoms of DSM-5 unipolar major depression, including single items that are used to assess indecisiveness and impaired concentration, guilt and worthlessness, and wishes for death and suicidal ideation. It usually takes 5 to 10 minutes to complete the scale. The QIDS-SR16 has good internal consistency, correlates significantly with clinician ratings of depression severity, and is sensitive to change. In scoring the QIDS-SR_{16}, the highest score is used of the four items that assess sleep disturbance (initial, middle, and terminal insomnia; and hypersomnia), the two items that assess psychomotor disturbance (agitation and retardation), and the four items that assess appetite and weight disturbance. Total scores on the scale range from 0 to 27, and scores of 0 to 5 indicate no depression, 6 to 10 mild depression, 11 to 15 moderate depression,

16 to 20 severe depression, and 21 to 27 indicate very severe depression. The QIDS-SR_{16} is available at the website.

WHEN AND WHERE TO MEASURE

Outpatients may complete a scale at every visit or at prefixed time points. We encourage having patients do so at each visit, reflecting the view that standardized, quantified assessment is an integral part of treatment. The availability of brief, reliable, and valid self-report scales to measure depression makes it is as feasible to assess depression as it is to measure medical vital signs [27,35].

A reasonable alternative is to have patients complete a scale at prefixed time points, eg, baseline and every three months. A practical problem with this approach is the need to keep track of when the follow-up assessments are to be conducted.

Patients have traditionally completed a scale while seated in the waiting room before the appointment. However, patients may prefer to complete the scale at their convenience in their home rather than arriving early or staying after their clinical appointment.

Scales can also be used to monitor patient progress daily in inpatient and partial hospital settings [48].

COMPUTERIZED AND PAPER FORMATS

Patients can complete scales using either a computer or paper and pencil. Clinicians should consider cost, missing data, and the results obtained with each format, when choosing between the two formats.

A web-based platform for administering outcome assessments can automatically score the instrument and reduce other administrative costs such as copying and handing out the instrument. In addition, a web-based system can generate and maintain a large database that can be used for developing quality improvement initiatives and reduce the cost of establishing such a database to evaluate treatment outcome. However, the use of a web-based system may increase the clinician's cost of acquiring and maintaining computer hardware and software.

A computer administered survey can prompt patients to ensure all questions are answered, and thus reduce missing data.

Studies comparing computer and paper scale administration have found that they are highly correlated and that patients find computer administration is easier and preferable [49-51]:

- A study of the Quick Inventory of Depressive Symptomatology-Self-Rated in 80 patients with major depression found that the correlation between electronic and paper responses was high (mean total score 15.3 versus 15.1; intraclass correlation 0.99) [26]. In addition, 49 percent of the patients found it easier to answer the questions with the computer, 22 percent found the paper version easier, and 29 percent found the two versions equivalent.

- A study of the Clinically Useful Depression Outcome Scale in 53 patients treated for depression found that the correlation between web based and paper responses was high (intraclass correlation 0.96), and that all of the patients preferred the web based version [52].

However, one cannot assume that paper and computer versions of the same test always produce equivalent results across all tests, because differences in format can influence results [53-57]. As an example, some computer administered tests present items one at a time on a screen, whereas paper tests present multiple items on a page. Presenting items singly or as a group of items can result in differential attention to the items and differences in item-scale correlations. Some computer administered tests do not allow a return to previous items, whereas paper tests allow prior responses to be changed. Patients might reconsider prior answers after answering subsequent questions. Some computer administered scales do not allow missing answers. The “candor hypothesis” suggests that respondents are more truthful when responding to computer administered tests. Each of these factors can influence the psychometric properties of a test and the cutoff value used to determine remission status, and it is necessary to demonstrate the equivalence between paper and electronic versions of any specific scale [53,54]. Although paper and electronic versions of the Patient Health Questionnaire – Nine Item (PHQ-9) have not been compared, we expect that such a comparison would show they are highly correlated, based upon studies of the Quick Inventory of Depressive Symptomatology Self-Rated, Clinically Useful Depression Outcome Scale, and other instruments [26,49-52].

ADVERSE SIDE EFFECTS SCALE

For those clinicians who already use a scale to monitor symptoms, it is possible to also use a self-report measure to assess medication side effects that otherwise may go undetected. Adverse side effects are one of the most common reasons that depressed patients discontinue their antidepressant [58,59].

A study compared side effects obtained with a self-report checklist with side effects that were documented in the charts of 300 outpatients treated for unipolar major depression with antidepressants by a private practice of six psychiatrists (who were blind to the purpose of the study) [60]. The checklist consisted of the Toronto Side Effects Scale (31 specific medication side effects) (table 4), which the investigators adapted for self-administration by the patients. Frequently occurring side effects were detected in significantly more patients by the self-report scale compared with the psychiatrists' evaluation (64 versus 26 percent of patients). The same was true for very or extremely bothersome side effects (52 versus 29 percent of patients). However, the study did not evaluate whether use of the scale improved medication adherence, and it is possible that using a self-report scale will elicit over-reporting in patients prone to somatic symptoms.

The Frequency, Intensity, and Burden of Side Effects Rating scale is a self-report instrument that is in the public domain and thus available free of charge for unlimited clinical use (figure 2) [61]. It consists of three global items that assess the frequency and intensity of side effects, and the degree that they interfere with functioning. Each item consists of a seven-point Likert scale. The items do not address specific side effects and thus clinicians need to follow up with specific questions. The Frequency, Intensity, and Burden of Side Effects Rating scale has good internal consistency. Test-retest reliability over two to three weeks was significant with correlations generally in the 0.4 to 0.5 range for each of the three items. Correlations with a more comprehensive side effects checklist were statistically significant, but of low magnitude (0.2 to 0.3).

SOCIETY GUIDELINE LINKS

Links to society and government-sponsored guidelines from selected countries and regions around the world are provided separately. (See "Society guideline links: Depressive disorders".)

SUMMARY AND RECOMMENDATIONS

- The evidence suggests that systematically using a rating scale to monitor symptoms can improve treatment outcomes. In addition, there are many theoretical reasons to use a scale. For patients with unipolar major depression, we suggest regularly monitoring depressive symptoms with a standardized scale. (See 'Evidence of efficacy' above and 'Reasons to measure outcome' above.)

- Clinicians should use self-report scales because they are readily accepted by patients and do not require the clinician's time to administer the instrument. In addition, the results from self-report scales are highly correlated with clinician administered instruments. (See 'Patient acceptability' above and 'Self-report measurement' above.)

- Several self-report depression scales are in the public domain and available free of charge for unlimited clinical use. The Patient Health Questionnaire – Nine Item (PHQ-9) has been studied and used most often (table 1). It is brief and easy to use, covers all of the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) diagnostic criteria for unipolar major depression, and has good psychometric properties. Reasonable alternatives are the Clinically Useful Depression Outcome Scale and the Quick Inventory of Depressive Symptomatology-Self Report 16 Item. (See 'Self-report scales in the public domain' above.)

- Outpatients can complete a scale either for every patient visit or at prefixed time points (eg, baseline and every three months). We encourage having patients do so at each visit, reflecting the view that standardized, quantified assessment is an integral part of treatment. One problem with completing the scale at prefixed time points is the need to keep track of when the follow-up assessments are to be conducted. Scales can be completed in the clinician's office or at home. (See 'When and where to measure' above.)

- Electronic (computer) versions of scales appear to provide the same information as paper versions and some patients find electronic scales easier and more convenient to use. (See 'Computerized and paper formats' above.)

Topic 14860 Version 33.0

→