



VATAP SHORT REPORT: *OUTCOMES MEASUREMENT
IN MAJOR DEPRESSION
(NUMBER 3 IN A SERIES: OUTCOMES MEASUREMENT
IN VHA MENTAL HEALTH SERVICES)
SEPTEMBER 2004*

Item	Yes	Partly	No
Preliminary			
1. Appropriate contact details for further information?	√		
2. Authors identified?	√		
3. Statement regarding conflict of interest?			√
4. Statement on whether report externally reviewed?	√		
5. Short summary in non-technical language?	√		
Why?			
6. Reference to the question that is addressed and context of the assessment?	√		
7. Scope of the assessment specified?	√		
8. Description of the health technology?	√		
How?			
9. Details on sources of information?	√		
10. Information on selection of material for assessment?	√		
11. Information on basis for interpretation of selected data?	√		
What?			
12. Results of assessment clearly presented?	√		
13. Interpretation of the assessment results included?	√		
What Then?			
14. Findings of the assessment discussed?	√		
15. Medico-legal implications considered?	√		
16. Conclusions from assessment clearly stated?	√		
17. Suggestions for further actions?		√	

The VA Technology Assessment Program uses this checklist[®], developed by the International Network of Agencies for Health Technology Assessment (INAHTA), as a quality assurance guide to foster consistency and transparency in the health technology assessment (HTA) process. VATAP will add this checklist[®] to its reports produced since 2002.

This summary form is intended as an aid for those who want to record the extent to which a HTA report meets the 17 questions presented in the checklist. It is NOT intended as a scorecard to rate the standard of HTA reports – reports may be valid and useful without meeting all of the criteria that have been listed.



VA TECHNOLOGY ASSESSMENT PROGRAM

SHORT REPORT

Outcomes Measurement in Major Depression (Number 3 in a Series: Outcomes Measurement in VHA Mental Health Services)

Number 8 One in a series of reports on Outcomes Measurement for Serious Mental Illness September, 2004

Executive Summary

- Major depressive illness and depressive symptoms not classifiable as major depression are common in the general population (Figure 1) and in VHA; both decrease sufferers' quality of life and productivity, while increasing mortality, costs, and morbidity. Older medical patients in VHA with the highest quartile depressive symptoms have 50% greater medical costs than those with lowest quartile symptoms.
- Commonly associated with a variety of physical complaints, as many as 50% of cases of major depression present first to primary care providers and a large percentage are not diagnosed or effectively treated.
- While this report focuses on outcomes measurement for major depressive disorder, depressive symptoms occur in individuals with other diagnoses, such as bipolar disorder and dysthymia. The diagnostic challenge for all types of depressive illness is to distinguish pathology from depressed states secondary to life stress, alcohol or drugs, medical illness, or uncomplicated bereavement.
- Domains important to the assessment of outcomes for depressive disorders include: symptom severity, remission and relapse, suicidality, social functioning, disability days, mortality, and general health status.
- Process measures such as treatment adherence to guidelines have been used as quality indicators for depression care, although neither process nor outcomes measures alone fully capture quality (see "Overview" to this series).
- VHA's Office of Quality and Performance filed the measure, "percent of patients screened for depression" (for fiscal year '02), a process measure not to be equated with outcome, with the National Quality Measures Clearinghouse, while the monumental *Guide to Treatments that Work* (Craighead, 1998) tabulates trial outcomes in terms of symptom severity scales, such as the Beck Depression Inventory or Hamilton Rating Scale for Depression.
- Given the multiple domains altered by depressive illness, a multidimensional instrument (i.e., one that addresses more than a single domain such as symptom severity) is widely cited as important, as is evaluating quality of life.
- VATAP applied methods, including selection criteria, as detailed in the Overview, to standardized instruments developed for depression.
- Ten instruments specific to depression are indexed under that heading in VATAP's primary resource for this series, the *APA Handbook of Psychiatric Measures* (2000); all but one of these evaluate symptom

• VA TAP (11T) Office of Patient Care Services • Room D4-142 • 150 S. Huntington Avenue, Boston, MA 02130 •

• Tel: 617-278-4469 • Fax: 617-264-6587 •

• E-mail: VATAP@med.va.gov • Web: <http://www.va.gov/vatap> • <http://vaww.va.gov/vatap> •

severity only (i.e., do not address patient functioning), so VATAP also applied VHA selection criteria to generic scales that the literature indicates are used in the disorder, or to those (either generic or depression-specific) identified through other sources.

- Five additional generic instruments useful in depression and for which information comparable to that in the *APA Handbook* could be readily obtained were so identified along with one depression-specific instrument.
- Contrary to judgments by other technology assessment agencies (see Overview), that quality of life (QoL) instruments are not suitable for routine use, VHA investigators have reported encouraging findings with QoL instruments in depression, and the Medical Outcomes Study (MOS) relied on the SF-36, which covers many of the same domains as QoL scales. Accordingly, those instruments are included in this report.
- One depression-specific instrument, the Depression Outcomes Module (DOM) meets all criteria. However, VHA clinical leaders would need to judge feasibility for use of its five component scales with correspondingly complex analyses and follow-up procedures in a range of clinical settings representative of VHA.
- Two severity of depressive symptoms scales, Zung Self Rating Depression Scale (Zung SDS), and Hamilton Rating Scale for Depression (HAM-D) meet all criteria except assessment of functioning.
- One scale of functional disability, Sheehan Disability Scale (SDS) was used for major depressive disorder in the National Comorbidity Survey; it meets four of VHA's eight criteria.
- One QoL instrument [the Quality of Well-Being Scale, self-administered (QWB SA)] meets all criteria, as do two other generic health status scales, Short Form 36-item

health survey (SF-36) and the Global Assessment of Function (GAF); the latter is already used within VHA.

- QWB-SA and SF-36 are self-administered by patients and may present validity concerns in a severely ill patient population, but VHA investigators have used QWB-SA in depressed patients, and SF-36 was central to the depression component of the large Medical Outcomes Study.
- As VATAP found in other sections of this series, VHA clinical leaders and managers thus have a variety of constructs and instruments from among which to choose those most suitable to their outcomes measurement needs and preferences for depression care.
- New depression instruments, such as the Quality of Life in Depression Scale, are under development and will be monitored by VATAP.
- The Patient Health Questionnaire, 9-item (PHQ-9), a newer depression severity scale, is brief and includes one functional status item. Clinical use of the copyrighted instrument is free, but it lacks an electronic form and documentation of sensitivity to change is incomplete. Otherwise, it may be a strong contender for depression outcomes monitoring in VHA.

Background

While this report focuses on outcomes measurement for major depressive disorder, depressive symptoms occur in individuals with other diagnoses: bipolar disorder involves depressive as well as manic episodes, and dysthymia is characterized by depressive symptoms of generally lesser severity and a more chronic course.

“A major depressive disorder (MDD) is the most commonly diagnosed psychiatric disorder among adults, with U.S lifetime prevalence rates

of 20-25% for women and 9-12% for men; point prevalence rates are about 6% and 3% for women and men, respectively"... "These prevalence rates and gender differences are relatively constant across the adult life span. Depression engenders not only extraordinary personal and family suffering, but also significant societal burdens, such as an increased use of social and medical services, enormous financial costs for treatment, and lost productivity due to absenteeism from work..." (Craighead, 1998).

Definitions

"Depression does not always present as sadness. It can also be characterized by somatic symptoms or at times physical symptoms with no clear-cut organic basis. Similarly, not all sadness is depression. A definitive diagnosis of depression requires satisfying criteria listed in the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. To make the diagnosis of major depression or dysthymia, use SIGECAPS, a mnemonic system which is a concise version of the DSM-IV criteria. Both disorders require active treatment." (Brigham and Women's Hospital, 2001).

Major depression: depressed mood or interest plus four SIGECAPS (below) for two or more weeks.

Dysthymia: depressed mood or interest plus three SIGECAPS most days for two or more years:

- S**leep increase/decrease
- I**nterest in formerly compelling or pleasurable activities diminished
- G**uilt, low self esteem
- E**nergy poor
- C**oncentration poor
- A**ppetite increase/decrease
- P**sycomotor agitation or retardation
- S**uicidal ideation

(VA/DoD, 2000; Brigham and Women's Hospital, 2001).

Glick (2001) provides an exceptionally concise definition: "**Major depression is defined as having at least five symptoms during the same**

two-week period, with marked change in function. At least one of the symptoms must be depressed mood or loss of interest or pleasure. In addition, changes in at least three of the following must be observed: weight, sleep, activity level, energy, ability to think or concentrate, or suicidal ideation. Subtypes being investigated include depression associated with psychotic symptoms, atypical features, or seasonal patterns ... A major diagnostic task is differentiating major depression from depressed states secondary to life stress, alcohol or drugs, medical illness, or bereavement..."

Pyne (1997) reports: "Depression is currently diagnosed according to the presence of signs, symptoms, and functional impairment (American Psychiatric Association [APA] 1994). Criteria for its signs and symptoms are well established and incorporated into numerous diagnostic instruments. Functional impairment criteria, however, are less clear..."

Doris (1999) further clarifies: "Depressive symptoms, such as unhappiness and disappointment, are common. They affect up to a third of the population. When symptoms become qualitatively different, or interfere with normal function, they are considered to be pathological. The clinical syndrome of illness is known as depressive disorder, clinical depression, or major depression. A subtype of depression, characterized by loss of pleasure in almost all activities, loss of reactivity to usually pleasurable stimuli, a distinct quality of depressed mood, with symptoms worse in the morning, early morning wakening, marked psychomotor slowing or agitation, significant loss of appetite or weight loss and excessive or inappropriate guilt is known as melancholia..."

Treatment

Geddes (2002) summarizes systematic review results for depression treatment:

Beneficial:

- Antidepressant drugs are effective in acute treatment of all grades of depressive disorders, in all common treatment settings,

and in people with or without co-existent physical illness;

- There is no evidence of a clinically significant difference in the benefits of different antidepressant drugs, although the drugs do vary in adverse effects;
- Collaborative working between primary care physicians and psychiatrists, case management, and telephone support or patient education improves the effectiveness of drug treatment.

Likely to be beneficial:

- Problem solving therapy and St. John’s Wort in mild to moderate depression;
- Combined drug and psychological treatment in severe depression.

Unknown effectiveness:

- Care pathways and non-directive counseling;
- Psychological treatments in severe depression;
- Exercise, bibliotherapy, and befriending.

Outcomes

Geddes (2002) identifies a range of outcomes: depressive symptoms; social functioning; occupational functioning; quality of life; admission to hospital; rates of self harm; relapse of depressive symptoms; and rates of adverse events. Sapin (2004) reiterates the impact of major depression on patient functioning and quality of life, using SF-36, Quality of Life in

Depression Scale, and the EuroQoL EQ-5D in a study of depressed patients in primary care.

VHA’s Office of Quality and Performance (2002) filed the measure: “percent of patients screened for depression”, a process measure not to be equated with outcome, with the National Quality Measures Clearinghouse, while Nathan and Gorman’s monumental *Guide to Treatments that Work* tabulates trial outcomes in terms of symptom severity scales, specifically the Beck Depression Inventory (BDI) or Hamilton Rating Scale for Depression (HAM-D) (Craighead, 1998).

Burden of depression: Epidemiology, disability, and human suffering

Figure 1 shows incidence and prevalence of several forms of depression, as summarized by Doris (1999), with VHA prevalence for 2002.

Geddes (2002) reports major depression prevalence as 5-10% of people seen in primary care settings, while two to three times that number may have depressive symptoms without meeting criteria for major depression. According to these authors, the data make depressive disorders the fourth most important cause of disability worldwide, with the expectation that it will advance to second position by 2020. Young (2001) used a national sample to confirm prevalence of depressive disorder in primary care, and also found that most adults with a probable anxiety or depressive disorder did not receive appropriate care in 1997-98.

Figure 1: Epidemiology of depressive disorders in the general population

Adapted from Doris (1999); VHA data from Blow (2002)

Measure	Depressive symptoms	Major depression	Dysthymia
Prevalence	Up to 33% of population	17%	3.6%
	12% in VHA healthcare facilities (inpatient and outpatient) in 2002		
Incidence		Female: 2% Male: 1%	
Age at onset		Female: 23 Male: 26	

In a study using the 36-item short form health survey (SF-36) for outcomes measurement, Wells (1989) and Hays (1995) found that patients with either current depressive disorder or depressive symptoms have comparable or worse physical, social, and role functioning, worse perceived general health, and greater bodily pain than did patients with eight major chronic medical conditions. These studies also found that depression and chronic medical conditions had unique and additive effects on patient functioning. Pyne (1997b), using the Quality of Well Being scale (QWB), found comparable results in VHA.

Kazis (1998; 2004) used SF-36 to document substantially worse health status in VHA outpatients (enrollees in the Veterans Health Study) compared to non-VHA populations; the same authors found younger veterans to be sicker, with lower subjective quality of life. These authors attributed lower QoL of younger veterans to mental health differences: screening scores for depression in younger veterans were positive in 51% versus 33% and 16% in groups of older veterans. However, older populations are not exempt from depression: Hickie (2000) confirmed that major depression is a serious health problem in older persons, occurring in 1%-3% of the elderly population, and resulting in significant disability, cognitive impairment, suicide, medical illness and increased mortality.

The burden of depression to VHA and its patients is further indicated in additional published research reports:

- Foster (1999) screened 574 male veterans in VHA primary care for depression: 13% were positive and 33% of those showed evidence of a major depressive episode.
- Hankin (1999) reported that 31% of 2,160 Boston-area male VHA outpatients screened positively for depression. These authors used screening test characteristics and study results to calculate an overall VHA outpatient population prevalence for depression of 29%.
- Charbonneau (2003) used VHA administrative databases to identify 12,678

patients from 14 VHA hospitals in the Northeast during two fiscal years only ('98 and '99) who were eligible for depression care profiling; this sample represented 8.5% of the patient population at the study sites for that time period.

- Blow (2002) reports key findings from the National Registry for Depression:
 - The majority of veterans (54%) in the registry had an additional psychiatric diagnosis; 40% had one additional diagnosis and 20% had two or more.
 - Post-traumatic stress disorder and substance abuse were the most commonly diagnosed co-morbid psychiatric conditions.
 - 88% of depressed veterans had a co-morbid medical condition, most commonly hypertension and arthritis.
- Finally, patients discharged from the military with serious mental disorders are eligible for VHA aftercare: Mojtabai (2003) identified 2,086 patients who had been treated for major depression in the military between 1993 and 1996 for an analysis of factors predicting use of VHA care after discharge from the military.

Who better to describe the human cost of depression-associated suffering than a writer? William Styron (1990) describes his own experience:

“Depression is a disorder of mood, so mysteriously painful and elusive in the way it becomes known to the self—to the mediating intellect—as to verge close to being beyond description. It thus remains nearly incomprehensible to those who have not experienced it in its extreme mode, although the gloom, “the blues” which people go through occasionally and associate with the general hassle of everyday existence is of such prevalence that they do give many individuals a hint of the illness in its catastrophic form. But at the time of which I write I had descended far past those familiar, manageable doldrums...

“...my brain had begun to endure its familiar siege: panic and dislocation, and a sense that

my thought processes were being engulfed by a toxic and unnamable tide that obliterated any enjoyable response to the living world. That is to say more specifically that instead of pleasure—certainly instead of the pleasure I should be having in this sumptuous showcase of bright genius (at the Picasso Museum in Paris)—I was feeling in my mind a sensation close to, but indescribably different from, actual pain...the pain persisted during my museum tour and reached a crescendo when, back at the hotel, I fell onto the bed and lay gazing at the ceiling, nearly immobilized and in a trance of supreme discomfort...a condition of helpless stupor in which cognition was replaced by that 'positive and active anguish' (as described by William James)..."

Assessment Methods

Please see the "Methods" section of the Overview report for additional detail. To recapitulate briefly:

VATAP's customary approach to assessment is the qualitative systematic review, which is usually tailored to evaluating literature on the efficacy of health care interventions.

For the current project, VATAP's charge was to identify the most appropriate among the available standardized outcome instruments for use by VHA mental health services. While the aim remained a methodologically transparent and systematic product, the size and scope of the literature encouraged VATAP to draw heavily on existing compendia rather than on original research reports, with reference to the latter where needed or helpful. The *Handbook of Psychiatric Measures* [American Psychiatric Association (APA), 2000] was particularly central to compiling this series of reports.

The series thus attempts to capitalize on existing compendia and reviews of standardized mental health care outcome measures while providing another level of synthesis. It can be considered a secondary review of other resources, supplemented from the primary research

literature as needed, and framed in the context of VHA mental health services. VATAP used the material in Overview Appendix B to generate selection criteria for VHA outcome instruments. The same criteria used in the Overview for global instruments of symptoms, disability, and functioning will be applied to depression instruments here:

Criteria for VHA outcomes instruments

1. Original purpose congruent with intended VHA use for quality of care tracking and reporting, documenting effective treatment of veteran patients;
2. Multidimensional;
3. Acceptable reliability and validity;
4. Sensitivity to change;
5. Feasible for routine use;
6. Electronic data entry and analysis;
7. Readily interpretable by non-professionals;
8. Free or obtainable to VHA at minimal cost.

In addition to the methods used for the Overview, VATAP scanned articles identified through its literature database searches for those describing or analyzing instruments used in depression. This approach allowed VATAP to identify additional instruments to those indexed under "Depression" in the *APA Handbook of Psychiatric Measures* (2000), all of which addressed symptom severity to the exclusion of functioning.

Search strategies

Comprehensive searches of the psychological and biomedical databases, MEDLINE[®], HEALTHStar[®], PSYCInfo[®], Current Contents[®], EMBASE[®], The Cochrane Library[®], and the extensive local monograph collections of McLean Psychiatric Hospital (Bedford, Massachusetts), and the Countway Library of Medicine at Harvard University were carried out. The monographic literature contributed several highly useful books on mental health instruments and outcomes evaluation.

Bibliographic search strategy terms included many exploded MeSH[®] subject headings for major depression. These terms and their synonyms combined with terms and free text

words describing treatment outcome, outcome measures, outcome assessment, treatment efficacy, instruments and surveys, severity measurement, and outcome evaluation yielded substantial results. Over 1,400 references, including end references, ranging from 1976 to 2001 were retrieved.

Quality of Life (QoL) measures are sometimes used in depression (Pyne, 1997a; 1997b; 2001). However, length and complexity generally render them unsuitable for routine clinical use on a large scale.

Results

Application of VHA criteria

The American Psychiatric Association *Handbook* indexes ten standardized measures under “depression” (abstracted details in Table 1; summarized in Table 2); all except the Depression Outcomes Module (DOM) address symptom severity, while DOM includes both symptoms and functioning.

VATAP identified five additional scales from other sources, including one quality of life measure (QWB) and its self-administered version (QWB-SA) that are not specific to depression but have been used for the disorder, including by VHA investigators.

One of these additional scales, the Short Form-36 item (SF-36), was used in the Medical Outcomes Study (Sherbourne, 1997; Wells, 1989), a pivotal study with parallels to VHA’s present mental health care outcome measurement effort (Tarlov, 1989) and that has been cited as a benchmark by VHA investigators (Pyne, 1997). Glick (1991) used the GAS, a predecessor of GAF, in an international study of depression treatment outcomes.

One scale of functional disability, Sheehan Disability Scale (SDS) was used for major depressive disorder in the National Comorbidity Survey Replication (Kessler, 2003); it meets four of VHA’s eight criteria.

Work by VHA investigators provides further guidance on instruments useful in a depression setting:

- Pyne (2001) reports results for the Quality of Well Being (QWB) in depressed VHA inpatients. These authors compared four depression instruments to find variables most predictive of acute treatment. QWB subscales for physical and social activity added to an initial prediction model based on admission BDI score significantly improved treatment response prediction accuracy to 86%.
- Rogers (2004) and Kazis (2004) used the SF-36 in a veteran population; the former to compare health status of VHA patients to non-VHA ambulatory patients in the Medical Outcomes Study, and the latter to improve response choices on two role functioning scales.

Using the model established in the Overview, which is based on data elements provided in the *APA Handbook*, abstracted detail on instruments that are potentially useful as depression outcome measurements is presented in Table 1, and summarized in Table 2. Instruments are included in this report’s tables and discussion to the extent that comparable information to that in the *APA Handbook* could be obtained from other sources.

Summary And Discussion

VATAP identified fifteen instruments (ten depression-specific, five generic) with potential for depression treatment outcomes measurement. Eight of ten (all but DOM and PHQ-9) depression-specific instruments assess symptom severity only, while DOM meets all criteria. However, DOM comprises five separate instruments and could commit VHA to a more cumbersome measurement system than that for which it might be fully prepared. A single measure may be preferable. PHQ-9 meets six of eight criteria.

Kristiansson (1996) discusses the relative merits and shortcomings of self- (quicker and less resource-intensive, but more prone to distorted reporting) versus clinician- administered scales in depression. Demyttenaere and de Fruyt (2003) reiterate the same concerns.

Geddes (2002) reported that trials often use continuous scales, such as the Hamilton Depression Rating Scale (HAM-D) or Beck Depression Inventory (BDI) to measure depressive symptoms. Clinician reports and global self-reports are also used in trials. HAM-D, however, is not ideal for older patients as it includes a number of somatic items that may be positive in older people who are not depressed; further, several authors (Mulder, 2003; Kristiansson, 1996) report other reservations about the HAM-D content.

While other health care systems have decided against QoL measures for mental health care outcomes (see the Overview to this series), Evenson (1998) makes persuasive arguments for using QoL measures: it makes sense to measure comfort rather than cure; complex programs require complex outcome measures; QoL measures place consumer interest foremost; finally, QoL constitutes good politics.

Two QoL instruments met all VHA criteria for depression outcomes measurement, should VHA decide in favor of one of these. As noted above, Pyne (2001) documents an additional use for QWB subscales: adding admission QWB subscale scores for physical and social activity to an existing “benchmark” model improved acute treatment response prediction.

The need for consensus on preferred approaches for VHA is further supported in that one of the high ranking global instruments discussed in the “Overview”, the Global Assessment of Function (GAF), is a barely visible presence in the depression literature.

Möller (2000) suggests a multi-method approach that combines observer- and self- assessment of multiple domains: “*In the clinical setting,*

depressed patients are assessed with a variety of scales. The use of a carefully selected battery of scales can give the clinician a clear assessment of a patient’s disease symptoms and a measure of improvement during and after treatment, along with the patient’s assessment of his own well being and social functioning. Frequently, the choice of rating scales is rather arbitrary and may be from familiarity rather than being based on clinical or scientific need.”

References

- Bland RC. Epidemiology of affective disorders: a review. *Canadian Journal of Psychiatry*, 1997; 42: 367-377.
- Blow FC, Owen RE, Valenstein M, Austin K, Khanuja K, McCarthy JF. (Serious Mental Illness Treatment Research and Evaluation Center (SMITREC)). Specialty care for veterans with depression in the VHA: 2002 National Registry for Depression (NARDEP) Report. 2002.
- Burvill PW. Recent progress in the epidemiology of major depression. *Epidemiologic Reviews*, 1995; 17: 21-31.
- Charbonneau A, Rosen AK, Ash AS, Owen RR, Kader B, Spiro A, 3rd, et al. Measuring the quality of depression care in a large integrated health system. *Medical Care*, 2003; 41: 669-680.
- Craighead WC, LW; Ilardi, SS. Psychosocial treatments for major depressive disorder. In: Nathan PE, Gorman JM, Ed. *A Guide to Treatments That Work*. New York: Oxford University Press, 1998.
- Demyttenaere K, De Fruyt J. Getting what you ask for: on the selectivity of depression rating scales. *Psychotherapy and Psychosomatics*, 2003; 72: 61-70.
- Doris A, Ebmeier K, Shajahan P. Depressive illness. *Lancet*, 1999; 354: 1369-1375.
- Druss BG, Rohrbaugh RM, Rosenheck RA. Depressive symptoms and health costs in older medical patients. *American Journal of Psychiatry*, 1999; 156: 477-479.
- Evenson RC, Vieweg BW. Using a quality of life measure to investigate outcome in outpatient

treatment of severely impaired psychiatric clients. *Comprehensive Psychiatry*, 1998; 39: 57-62.

Fortney J, Rost K, Zhang M, Pyne J. The relationship between quality and outcomes in routine depression care. *Psychiatric Services*, 2001; 52: 56-62.

Foster MA, Ragsdale K, Dunne B, Jones E, Ihnen GH, Lentz C, et al. Detection and treatment of depression in a VA primary care clinic. *Psychiatric Services*, 1999; 50: 1494-1495.

Garratt AM, Ruta DA, Abdalla MI, Buckingham JK, Russell IT. The SF36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? *British Medical Journal*, 1993; 306: 1440-1444.

Geddes J, Butler R. Depressive disorders. *Clinical Evidence*, 2002; 951-973.

Glick ID, Burti L, Suzuki K, Sacks M. Effectiveness in psychiatric care. I. A cross-national study of the process of treatment and outcomes of major depressive disorder. *Journal of Nervous and Mental Disease*, 1991; 179: 55-63.

Groessler E. *Personal Communication with: Flynn K. Personal Communication on: QWB and QWB-SA. Format: Email; September- October, 2003.*

Handbook of Psychiatric Measures / Task Force for the Handbook of Psychiatric Measures. Washington, DC: American Psychiatric Association, 2000.

Hankin CS, Spiro A, 3rd, Miller DR, Kazis L. Mental disorders and mental health treatment among U.S. Department of Veterans Affairs outpatients: the Veterans Health Study. *American Journal of Psychiatry*, 1999; 156: 1924-1930.

Hays RD, Wells KB, Sherbourne CD, Rogers W, Spritzer K. Functioning and well-being outcomes of patients with depression compared with chronic general medical illnesses. *Archives of General Psychiatry*, 1995; 52: 11-19.

Hermann RC, Leff HS, Palmer RH, Yang D, Teller T, Provost S, et al. Quality measures for mental health care: results from a national inventory. *Medical Care Research and Review*, 2000; 57 Suppl 2: 136-154.

Hickie I, Burke D, Tobin M, Mutch C. The impact of the organisation of mental health services on the quality of assessment provided to older patients with

depression. *Australian and New Zealand Journal of Psychiatry*, 2000; 34: 748-754.

Hirschfeld RM. Major depression, dysthymia and depressive personality disorder. *British Journal of Psychiatry Supplement*, 1994; 23-30.

Horvitz-Lennon M, Normand SL, Frank RG, Goldman HH. "Usual care" for major depression in the 1990s: characteristics and expert-estimated outcomes. *American Journal of Psychiatry*, 2003; 160: 720-726.

Kazis LE, Miller DR, Clark J, Skinner K, Lee A, Rogers W, et al. Health-related quality of life in patients served by the Department of Veterans Affairs: results from the Veterans Health Study. *Archives of Internal Medicine*, 1998; 158: 626-632.

Kazis LE, Miller DR, Clark JA, Skinner KM, Lee A, Ren XS, et al. Improving the response choices on the veterans SF-36 health survey role functioning scales: results from the Veterans Health Study. *Journal of Ambulatory Care Management*, 2004; 27: 263-280.

Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *Journal of the American Medical Association*, 2003; 289: 3095-3105.

Kristiansson E, McDowell I. Depression. In: McDowell I, Newell C, Eds. *Measuring health: a guide to rating scales and questionnaires.* New York: Oxford University Press, 1996.

Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Medical Care*, 2003; 41: 1284-1292.

Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatric Annals*, 2002; 32: 509-515.

Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 2001; 16: 606-613.

McKenna SP, Doward LC, Kohlmann T, Mercier C, Niero M, Paes M, et al. International development of the Quality of Life in Depression Scale (QLDS). *Journal of Affective Disorders*, 2001; 63: 189-199.

Mojtabai R, Rosenheck RA, Wyatt RJ, Susser ES. Use of VA aftercare following military discharge among patients with serious mental disorders. *Psychiatric Services*, 2003; 54: 383-388.

Möller HJ. Rating depressed patients: observer- vs self-assessment. *European Psychiatry*, 2000; 15: 160-172.

Montano CB. Primary care issues related to the treatment of depression in elderly patients. *Journal of Clinical Psychiatry*, 1999; 60 Suppl 20: 45-51.

Mulder RT, Joyce PR, Frampton C. Relationships among measures of treatment outcome in depressed patients. *Journal of Affective Disorders*, 2003; 76: 127-135.

Office of Quality and Performance. Accessed: March 8, 2002. VHA performance measurement system: technical manual. FY2002. http://vaww.oqp.med.va.gov/oqp_services/performance_measurement/tech_man.asp.

Owen RR, Cannon D, Thrush CR. Mental Health QUERI Initiative: expert ratings of criteria to assess performance for major depressive disorder and schizophrenia. *American Journal of Medical Quality*, 2003; 18: 15-20.

Pyne JM, Bullock D, Kaplan RM, Smith TL, Gillin JC, Golshan S, et al. Health-related quality-of-life measure enhances acute treatment response prediction in depressed inpatients. *Journal of Clinical Psychiatry*, 2001; 62: 261-268.

Pyne JM, Patterson TL, Kaplan RM, Gillin JC, Koch WL, Grant I. Assessment of the quality of life of patients with major depression. *Psychiatric Services*, 1997b; 48: 224-230.

Pyne JM, Patterson TL, Kaplan RM, Ho S, Gillin JC, Golshan S, et al. Preliminary longitudinal assessment of quality of life in patients with major depression. *Psychopharmacology Bulletin*, 1997a; 33: 23-29.

Rogers WH, Kazis LE, Miller DR, Skinner KM, Clark JA, Spiro A, 3rd, et al. Comparing the health status of VA and non-VA ambulatory patients: the veterans' health and medical outcomes studies. *Journal of Ambulatory Care Management*, 2004; 27: 249-262.

Rosenberg R. Outcome measures of antidepressive therapy. *Acta Psychiatrica Scandinavica Supplement*, 2000; 402: 41-44.

Rosenheck R, Van Stone W. Evaluating VA outcomes using the Global Assessment of Functioning (GAF). Unpublished Manuscript. July 16, 2001.

Rost K, Smith GR, Burnam MA, Burns BJ. Measuring the outcomes of care for mental health problems. The case of depressive disorders. *Medical Care*, 1992; 30: MS266-273.

Rost K, Williams C, Wherry J, Smith GR, Jr. The process and outcomes of care for major depression in rural family practice settings. *Journal of Rural Health*, 1995; 11: 114-121.

Sapin C, Fantino B, Nowicki ML, Kind P. Usefulness of EQ-5D in assessing health status in primary care patients with major depressive disorder. *Health and Quality of Life Outcomes*, 2004; 2: 20.

Sherbourne CD, Wells KB, Sturm R. Measuring health outcomes for depression. *Evaluation and the Health Professions*, 1997; 20: 47-64.

Smith GR, Mosley CL, Booth BM. Measuring health care quality: major depressive disorder. AHCPR discussion paper. 1996. Report Number AHCPR Pub. No. 96-NO23.

Styron W. Darkness Visible: A Memoir of Madness. New York: Random House, 1990: 84 Pages.

Tarlov AR, Ware JE, Jr., Greenfield S, Nelson EC, Perrin E, Zubkoff M. The Medical Outcomes Study. An application of methods for monitoring the results of medical care. *Journal of the American Medical Association*, 1989; 262: 925-930.

Wells KB, Burnam MA, Rogers W, Hays R, Camp P. The course of depression in adult outpatients. Results from the Medical Outcomes Study. *Archives of General Psychiatry*, 1992; 49: 788-794.

Wells KB, Stewart A, Hays RD, Burnam MA, Rogers W, Daniels M, et al. The functioning and well-being of depressed patients. Results from the Medical Outcomes Study. *Journal of the American Medical Association*, 1989; 262: 914-919.

Young AS, Klap R, Sherbourne CD, Wells KB. The quality of care for depressive and anxiety disorders in the United States. *Archives of General Psychiatry*, 2001; 58: 55-61.

APPENDIX

Criteria for VA use of standardized measures of mental health care outcomes

1. Original purpose congruent with intended VA use for quality of care tracking and reporting, performance measure for effective treatment of veteran patients
2. Multidimensional, i.e., measuring both symptoms and function
3. Acceptable reliability and validity
4. Sensitive to change
5. Feasible for routine use
6. Electronic data entry, analysis
7. Reasonable cost for VHA to obtain
8. Readily interpretable by non-professionals

Notes:

- Unless otherwise noted, the information in this table was obtained from The American Psychiatric Association's *Handbook of Psychiatric Measures* (2000);
- ? indicates lack of information relative to that criterion in *APA Handbook* or alternate source;
- Darker shaded rows represent measures meeting all criteria; lighter shaded rows, those meeting all but one criterion.

Abbreviations in appendix

APA, American Psychiatric Association [*Handbook of Psychiatric Measures* (2000)]
BDI, Beck Depression Inventory
CES-D, Center for Epidemiologic Studies Depression Scale
D-ARK, Depression-Arkansas Scale
DIS, Diagnostic Interview Schedule
DOM, Depression Outcomes Module
GAF, Global Assessment of Function
*GDS, Geriatric Depression Scale
GHQ, General Health Questionnaire
HADS, Hospital Anxiety and Depression Scale
HAM-D, Hamilton Rating Scale for Depression
IDS, Inventory of Depression Symptomatology
IQOLA, International Quality of Life Assessment

MADRS, Montgomery-Asberg Depression Rating Scale
MHI, Mental Health Inventory
MDD, Major Depressive Disorder
MOS, Medical Outcomes Study
PHQ-9, Patient Health Questionnaire, 9-item
SDS, Sheehan Disability Scale
SSI, Somatic Symptom Inventory
QWB, Quality of Well-Being
QWB-SA, Quality of Well Being, Self-Administered
ROC, receiver operating characteristic
RS, Raskin Scale (Three-Area Severity of Depression Scale)
SF-36, Short Form Health Survey
Zung SDS, Zung Self-rating Depression Scale

Table 1. Measures evaluation matrix: Abstracted details for instruments for depression

Name	Purpose	Symptoms + function	Reliability	Validity	Change	Time	Acceptability	Who?	Electronic	Cost	Interpretation
<i>Depression-specific instruments</i>											
BDI	To measure the behavioral manifestations of depression (depressive symptom severity) in adolescents and adults. Information in AP Handbook reflects 1978 revision rather than 1996 revision.	Symptoms only.	Cronbach's alpha (internal consistency): .76-.95 (high). Consistency ratings by diagnostic group: Single major depressive episode, alpha = .8; recurrent major depressive episode, alpha = .86; dysthymic disorder, alpha = .79; alcohol abuse or dependence, alpha = .90; heroin abuse or dependence, alpha = .88. Test-retest reliability good for retesting at 1-6 hours and with time frame of "right now".	High but not complete correlation with other standard measures of depressive symptom severity. Discriminative ability in psychiatric samples is poor: scores cannot be used to determine diagnosis.	Yes	5-10 minutes as self-report questionnaire; 15 minutes as interview; longer in severely obsessive patients.	Likely to be good: brief and simple, can be self-administered. Widely used in treatment studies to show symptom change over time.	Self- or interviewer-administered.	Yes	Copy-righted, manual + 25 record forms, \$75; additional forms, \$40/25.*	Summed score of item sets. Subscale scores for a cognitive-affective factor and a somatic-performance factor.
CES-D	To measure symptoms of depression in community populations: health correlates of depressive symptoms, changes in severity of symptoms over time.	Symptoms only	Internal consistency high across a variety of populations (Cronbach's alpha generally .85 in community samples, .90 in psychiatric samples). Split-half reliability is also high (.77-.92). Test-retest reliability over 2-8 weeks is moderate (0.51-0.67), which is desirable for symptoms expected to show change over time.	Correlation with Symptom Checklist-90 high in outpatient samples with depression, alcoholism, drug addiction, or schizophrenia. Correlations with HAM-D variable (.49 for patients with acute depression, .85 for schizophrenia). Fair agreement with short form of GDS. Moderate and variable correlations with Raskin Scale. <i>Measure is not specific to depression</i> , studies do not support use in	Yes	5 minutes	Used in "many studies" as a screen for depressive illness.	Self-report	Not noted by APA	Public domain	Sum of item scores with cut-off to identify depressive illness

• VA TAP (11T) Office of Patient Care Services • Room D4-142 • 150 S. Huntington Avenue, Boston, MA 02130 •

• Tel: 617-278-4469 • Fax: 617-264-6587 •

• E-mail: VATAP@med.va.gov • Web: <http://www.va.gov/vatap> • <http://vaww.va.gov/vatap> •

VA Technology Assessment Program Short Report: Outcomes Measurement in Major Depression

Name	Purpose	Symptoms + function	Reliability	Validity	Change	Time	Acceptability	Who?	Electronic	Cost	Interpretation
				undiagnosed populations without a follow-up interview.							
HAM-D	To measure severity of depressive symptoms in patients with primary depressive illness: estimate severity before treatment; gauge effect of treatment; detect relapse or recurrence.	Symptoms only.	Varies with conditions, but generally acceptable: Cronbach's alpha = .48-.85.	Correlations with global measures of depressive severity = .65-.9. Correlation with clinician rated instruments = .8-.9. Validity not high in all populations: depressive symptoms in older patients may be over-rated because HAM-D relies on somatic symptoms. Validity as outcome or symptom severity measure has been questioned (Mulder, 2003)	Yes. Loss of insight, obsessive symptoms, agitation, and hypochondriasis may be less sensitive to changes in overall severity than other items.	15-20 minutes	"most commonly used observer rated depressive symptom rating scale"	Clinicians. Trained non-clinicians in research	Yes	Public domain.	Total score with thresholds for severity in categories. Rosenberg (2000): summing item scores in HAM-D is not a valid measure of depression severity.
HADS	To screen for presence of mood disorder in medically ill patients in community or hospital settings, with depression subscale oriented to symptoms of anhedonia rather than sadness	Symptoms only.	Depression subscale: high internal consistency (Cronbach's alpha, 0.90); item-total correlations, 0.39-0.60; test-retest in healthy respondents, 0.90	Correlation between depression subscale and MADRS, 0.7. In psychiatric populations without medical illness, sensitivity is high (cut-point of 8), but specificity is inadequate.	"may be useful"	"few minutes"	Short, easy to use screen for depressive disorder in non-psychiatric populations.	Self-report. If depression is identified, other scales are needed.	Not noted by APA	Copyrighted	Use not indicated in non-medically ill psychiatric populations.
IDS Clinician-administered (IDS-C) and self report (IDS-SR) versions; 28 and 30 item forms for both	To measure signs and symptoms of depression in both inpatients and outpatients. Wider range in total score than other instruments, so can be used in less severely ill populations	Symptoms only	IDS-C: Cronbach's alpha = .92-.94; joint reliability = .94 IDS-SR: Cronbach's alpha = .93-.94 in mixed populations, .77 in symptomatic patients.	Highly correlated (.88-.94) with other depression rating scales; can be used to classify endogenous and non-endogenous depression.	Yes	IDS-SR, 15-20 minutes; IDS-C, 30 minutes	Likely to be good, brief and can be self-administered; Wider range for total score than other instruments allows use in less severely ill populations. However, lack of wide use provides	Self report or clinician administered.	Not noted by APA.	Copyrighted, but no fee for use.	Total score for overall severity

• VA TAP (11T) Office of Patient Care Services • Room D4-142 • 150 S. Huntington Avenue, Boston, MA 02130 •
• Tel: 617-278-4469 • Fax: 617-264-6587 •

• E-mail: VATAP@med.va.gov • Web: <http://www.va.gov/vatap> • <http://vaww.va.gov/vatap> •

VA Technology Assessment Program Short Report: Outcomes Measurement in Major Depression

Name	Purpose	Symptoms + function	Reliability	Validity	Change	Time	Acceptability	Who?	Electronic	Cost	Interpretation
versions.							more limited framework of reference for clinicians and researchers.				
MADRS	To measure overall severity of depressive symptoms on patients with major depression	Symptoms only	Comparable to other observer-rated depression scales.		Yes, one of developers' goals.	15 minutes	Probably good. Brief (10-item checklist) and easy to administer		No	Public domain	Summary score
Raskin Scale	To provide an empirical estimate of severity of depressive symptoms to identify individuals sufficiently depressed to warrant treatment. "However, no studies have been done to determine whether those who are identified by the scale are more likely to benefit from treatment than those who are not."	Experience, behavioral indicators, and secondary signs of depression	Measures of internal consistency have not been reported. Intra-class reliability = .88	Sensitivity to change in clinical state over time demonstrated in many clinical trials.		"few minutes", but must be determined on basis of sufficient interaction with patient for accurate rating.	"offers only a global assessment, and no clear documentation of changes in specific symptoms"... "brief and has high face validity, but only limited psychometric data available"	Mental health professionals	Not noted	No copyright restrictions	Total depression severity score (3-15)
PHQ-9	Self-administered version of PRIME-MD, an instrument to assist primary care clinicians in making criteria-based diagnoses of DSM-IV disorders commonly encountered in medical patients: mood, anxiety, somatiform, alcohol, and eating (Kroenke, 2002). Establishes provisional depressive diagnoses and grades symptom severity.	Yes, one item assesses global functioning.	Excellent internal consistency, Cronbach's alpha = .86-.89. Excellent test-re-test reliability, 0.84 (Kroenke, 2001)	Positive predictive value for major depression in sample with 7% prevalence: 31-51%. "Substantial" association between increasing score and likelihood of major depression; area under curve in ROC analysis: 0.95 (Kroenke, 2002).	Under investigation (Kroenke, 2002)	Brief, 9-item instrument, self-administered. Clinician review takes < 3 minutes in 85% of cases (Kroenke, 2001; 1999)	Probably good: Brief, 9-item instrument with good construct and criterion validity. Can be a dual purpose instrument to establish diagnosis and grade severity (Kroenke, 2001)	Self administered, then clinician reviews responses	No	Copyrighted, but clinical use is free.	Scores from 0-27: higher score = greater severity
Zung SDS	Self-administered measure of depression severity: all-inclusive re symptoms, but short, simple, and quantitative	Symptoms only.	Cronbach's alpha = .79 (satisfactory).	Significant correlations with: MMPI-D, HAM-D	Yes, but less so than other scales	5-30 minutes, depending on patient pathology		No training requirements (self-administered)	Not noted	Free, but permission to use needed	Total score and 4 sub-scales

• VA TAP (11T) Office of Patient Care Services • Room D4-142 • 150 S. Huntington Avenue, Boston, MA 02130 •
 • Tel: 617-278-4469 • Fax: 617-264-6587 •

• E-mail: VATAP@med.va.gov • Web: <http://www.va.gov/vatap> • <http://vaww.va.gov/vatap> •

VA Technology Assessment Program Short Report: Outcomes Measurement in Major Depression

Name	Purpose	Symptoms + function	Reliability	Validity	Change	Time	Acceptability	Who?	Electronic	Cost	Interpretation
DOM	To assess the process of care, patient characteristics, and outcomes of care for patients with major depressive disorder in primary care settings; primarily designed to be used as part of an outcomes management system in which data are aggregated at the provider or system level to monitor and improve patient outcomes. Five components:	Yes, symptoms (by means of D-ARK scale) and functioning (by means of SF-36 questions)	D-ARK: Cronbach's alpha = 0.90-0.92; Medical record review: 100% agreement between two research assistants. Patient follow-up assessment: good agreement (78-100%) one week later in reports of health service utilization.	D-ARK: high agreement with diagnoses of major depression for DSM Axis I disorders (kappa = .7-.89).	"too much measurement error to examine change over time in an individual patient, it can be used to examine between-group differences in severity of depression" (Rost, 1995)		Rost (1992), indicates that DOM is feasible in specialty settings	No specific training needed, patient components self-administered.	Internet	Copyrighted, but available (\$35) for unlimited use in research or clinical care if no cost to patients.	Outcomes adjusted for case mix can be compared for groups of patients, clinicians, or health care systems.
Five components of DOM:											
Patient screener for MDD	To identify patients likely to have MDD and who should complete full DOM					3 items		clinician			
Patient baseline assessment	80 items, completed by patient at initial visit					25 minutes					
Clinician baseline assessment	20 items addressing diagnostic and exclusion criteria for major depression					3 minutes					
Patient follow-up assessment: 83 items	83 items, performed every 4-6 months until patient is in remission					25 minutes					
	Medical record review form					10 minutes by hand, less if electronic medical record.					
Generic instruments											
MHI	To assess level of mental health among psychiatrically healthy samples, to provide screening for general clinical practice to detect symptoms of psychopathology; and to measure the affective	Yes	38-item version: Cronbach's alpha, .83-.91; test-retest reliability, .56-.64; correlations among 5 subscales, .34-.75 5-item version: item-scale correlations, .65-.80.	Statistically significant advantages of MHI over GHQ in detecting: any DIS disorder; over SSI in detecting an affective disorder MHI-5 is equal to MHI-18 and GHQ-30 in detecting any DIS	Not noted by APA	"few minute" for 5-item version, < 10 minutes for 18-item version, 10-20 minutes for 38-item version	Valid with younger and older populations, but not for children or psychiatric populations. Used in MOS population, [Course of	Self-administered	Not noted	RAND usually gives permission to reproduce and use with proper attribution.	Highest score = least favorable health. Lowest score = most favorable health.

• VA TAP (11T) Office of Patient Care Services • Room D4-142 • 150 S. Huntington Avenue, Boston, MA 02130 •

• Tel: 617-278-4469 • Fax: 617-264-6587 •

• E-mail: VATAP@med.va.gov • Web: <http://www.va.gov/vatap> • <http://vaww.va.gov/vatap> •

VA Technology Assessment Program Short Report: Outcomes Measurement in Major Depression

Name	Purpose	Symptoms + function	Reliability	Validity	Change	Time	Acceptability	Who?	Electronic	Cost	Interpretation
	components of psychological distress and well-being. Intended as a measure that is brief, easy to administer and score, easily interpreted. Focuses on depression, anxiety, other affective states. 5-18- and 38-item versions, plus alternate form to assess general mental health repeatedly in group-level studies.			disorder. MHI-5 is inferior for the full range of affective disorders. Moderate correlations with SF-36 for: physical functioning, physical role functioning, bodily pain, general health, vitality, social functioning, and emotional role functioning.			Depression sub-study (Wells, 1992)].				
GAF	Can be made to fit	Yes	Fair to excellent	Good concurrent validity, but depends on info available to guide rating	More sensitive than other measures	Once info obtained, 1-2 minutes	Already used in VHA	Clinician, with structured interview, guidelines, users' guide	Yes (within VA)	Public domain	Summary score of functioning
QWB	Generic measure of health-related quality of life (Groessler, 2002)	Yes, symptoms and functioning	Test-retest good, other reliability measures not applicable due to structure of measure and multi-factorial nature of HRQOL concept.	Well-validated in many disease populations over 25 years, used by VHA investigators for depression	Yes	12-15 minutes	Probably good re use in large NIH studies	Trained interviewer	Yes	Free; Manuals \$79	4 domain scores (3 functioning + 1 symptom) weighted by preference to create total expression of well-being
QWB-SA	Generic measure of health-related quality of life (Groessler, 2002)	Yes, symptoms and functioning	?	Newer, but validated in large studies	Yes	10 minutes	More useful for clinicians than QWB, better coverage of mental health	Self-administered	Yes Internet version pending	Free for non-profit use	4 domain scores (3 functioning + 1 symptom) weighted by preference to create total expression of well-being
SDS	Composite of 3 self-rated items designed to measure extent to which panic, anxiety or depressive symptoms impair major sectors (work, social and family life) in the patient's life;	Function only	Inter-item correlations fairly high (.70-.79); internal consistency high (Cronbach's alpha =0.89)	Construct validity "substantial", but 3 functioning items are highly inter-related, so may not be useful as a comprehensive global measure	Yes	1-2 minutes	Probably good, brief and self- or clinician-administered; "may usefully supplement diagnostic assessment".	Self report, administered by clinician, or rated by both independently	Not noted by APA	Copy- righted	3 items summed into single measure of global functional impairment (0-30)

• VA TAP (11T) Office of Patient Care Services • Room D4-142 • 150 S. Huntington Avenue, Boston, MA 02130 •

• Tel: 617-278-4469 • Fax: 617-264-6587 •

• E-mail: VATAP@med.va.gov • Web: <http://www.va.gov/vatap> • <http://vaww.va.gov/vatap> •

VA Technology Assessment Program Short Report: Outcomes Measurement in Major Depression

Name	Purpose	Symptoms + function	Reliability	Validity	Change	Time	Acceptability	Who?	Electronic	Cost	Interpretation
	intended to supplement symptom assessments. Intended for use in treatment outcome studies.						"can be administered and scored without interrupting flow of routine clinical care."				
SF-36	Generic measure of Perceived health status.	Yes, subscales include physical functioning, role functioning, bodily pain, general health, vitality, social functioning, emotional role functioning, mental health.	Estimates of internal consistency (alpha coefficients): 0.62-0.94, with majority ≥ 0.80 ; Test-retest coefficients: 0.43-0.90 for 6-month interval, 0.60-0.81 for 2-week interval.	Correlates moderately well with other measures (e.g., Sickness Impact Profile and Duke Health Profile); Levels of medical and psychiatric severity correlate as expected with clinical outcome criteria (e.g., burden of care). Useful in predicting clinically and socially relevant outcomes (ability to work, health services utilization).	Yes	10-12 minutes	Assesses health concepts that represent basic human values relevant to functional status and well-being; thus can serve an important function in everyday practice. Used in MOS and IQOLA Project. Garrat (1993) demonstrated good acceptability of UK version in Scotland.	Self-administered by patients > 14 years.s	Not noted by APA	Copyrighted, but permission to use routinely granted royalty-free; Handbook \$53.	Scale scores range 0 – 100, with higher scores indicating better health. Extensive normative data from US general population and various subgroups are available.

* <http://marketplace.psychcorp.com/PsychCorp.com/Cultures/en-US/Products>.

Table 2. Summary: instruments specific to depression, global instruments used in depression studies, and quality of life instruments used in depression research by VA investigators

Notes: Darker shaded rows represent measures meeting all criteria, lighter shaded rows, those meeting all but one criterion.

Measure	Criteria							
	Purpose	> 1 dimension	Reliability, validity	Change	Feasible	Electronic	Interpretation	Cost
Depression-specific scales								
BDI	X		X	X	X	X	X	
CES-D	X		X	X	X	?	X	X
HADS			X	?	X		?	
HAM-D	X		?	X	X	X	X	X
IDS	X		X	X	X	?	X	X
MADRS	X		X	X	X		X	X
PHQ-9		X	X	X	?		X	X
Raskin			X	X	X	?	X	X
Zung SDS	X		X	X	X	?	X	X
DOM	X	X	X	(between-groups)	?	X	X	X
Generic scales								
GAF	X		X	X	X	X	X	X
QWB	X	X	X	X	?		X	X
SDS			X	X	X		X	
QWB-SA	X	X	X	X	X	X	X	X
SF-36	X	X	X	X	X	?	X	X

Authors

Author: Karen Flynn, DDS, MS
Contributors: Elizabeth Adams, RRT, MPH
Elaine Alligood, MLS

The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs. The contributions of the following reviewers are gratefully acknowledged. Participation as a reviewer does not imply endorsement.

Reviewers

Valerie A. Lawrence, M.D.
Physician Advisor, VA Technology Assessment Program
Audie L. Murphy VA Medical Center
Associate Professor, Dept of Medicine
University of Texas Health Science Center
San Antonio, Texas

Richard Owen, MD
Director, HSR&D Center for Mental Health Care and Outcomes Research
Central Arkansas Veterans Healthcare System
John L. McClellan Memorial Veterans Hospital
Little Rock, Arkansas

