CJASN ePress. Published on November 17, 2020 as doi: 10.2215/CJN.05540420

Article

Depression Screening Tools for Patients with Kidney Failure

A Systematic Review

Karli Kondo[1,2] *Jennifer R. Antick*[3,4] *Chelsea K. Ayers*[1] *Devan Kansagara*[1,5,6] *and Pavan Chopra*[5]

Abstract

Background and objectives Patients with kidney failure experience depression at rates higher than the general population. Despite the Centers for Medicare and Medicaid Services' ESRD Quality Incentive Program requirements for routine depression screening for patients with kidney failure, no clear guidance exists.

Design, setting, participants, & measurements For this systematic review, we searched MEDLINE, PsycINFO, and other databases from inception to June 2020. Two investigators screened all abstracts and full text. We included studies assessing patients with kidney failure and compared a tool to a clinical interview or another validated tool (*e.g.*, Beck Depression Inventory II). We abstracted data related to sensitivity and specificity, positive and negative predictive value, and the area under the curve. We evaluated the risk of bias using the Quality Assessment of Diagnostic Accuracy Studies 2.

Results A total of 16 studies evaluated the performance characteristics of depression assessment tools for patients with kidney failure. The Beck Depression Inventory II was by far the best studied. A wide range of thresholds were reported. Shorter tools in the public domain such as the Patient Health Questionnaire 9 and Geriatric Depression Scale 15 (adults over 60) performed well but were not well studied. Short tools such as the Beck Depression Inventory–Fast Screen may be a good option for an initial screen.

Conclusions There is limited research evaluating the diagnostic accuracy of most screening tools for depression in patients with kidney failure, and existing studies may not be generalizable to US populations. Studies suffer from limitations related to methodology quality and/or reporting. Future research should target widely used, free tools such as the Patient Health Questionnaire 2 and the Patient Health Questionnaire 9.

Clinical Trial registry name and registration number: Systematic Review Registration: PROSPERO CRD42020140227.

CJASN 15: ●●●–●●●, 2020. doi: https://doi.org/10.2215/CJN.05540420

[1]Evidence Synthesis Program, Veterans Affairs Portland Health Care System, Portland, Oregon
[2]Research Integrity Office, Oregon Health & Science University, Portland, Oregon
[3]School of Graduate Psychology, Pacific University, Hillsboro, Oregon
[4]Legacy Good Samaritan Medical Center, Portland, Oregon
[5]Department of Medicine, Oregon Health & Science University, Portland, Oregon
[6]Center to Improve Veteran Involvement in Care, Veterans Affairs Portland Health Care System, Portland, Oregon

Correspondence: Dr. Karli K. Kondo, Department of Veterans Affairs, Evidence Synthesis Program, VA Portland Health Care System, Mail Code R&D 71, 3710 South West US Veterans Hospital Road, Portland, OR 97239-2999. Email: kondo@ohsu.edu

Introduction

The incidence and prevalence of kidney failure in the United States have increased steadily over the past 4 decades (1). Patients with kidney failure experience major depressive disorder at three to six times the rate of the general US population, depending on the method of assessment (2,3). Comorbid depression is associated with treatment nonadherence, poorer quality of life, worse sleep, more frequent emergency department visits, hospitalizations, suicide, and all-cause mortality (4–7).

The Centers for Medicare and Medicaid Services requires routine depression screening for patients with kidney failure as part of their ESRD Quality Incentive Program (ESRD-QIP) (8). However, there is no system-wide screening protocol, leading to wide variation in the way depression is assessed. Established evidence and guidelines suggest that screening for depression in the general population is both accurate and can improve outcomes (9). However, screening may lead to false positives and concomitant iatrogenic harm from unnecessary pharmacotherapy or resource-heavy psychotherapy. Screening may also lead to false negatives in which depression goes untreated. Patients with kidney failure differ from the general population both because they experience higher rates of comorbid depression, and because they often have symptoms related to their underlying diagnoses and treatments that mimic the somatic symptoms of depression.

Given the wide variation in depression screening options and lack of a gold standard assessment tool for patients with kidney failure, a clear understanding of the validity of the available screening tools is vital. The purpose of this review is to identify depression screening tools (and/or thresholds) appropriate for patients with kidney failure, and to better understand the effect of depression screening in this population.

Materials and Methods

This is part of a larger systematic review commissioned by the Veterans Health Administration that examined both screening and the effectiveness of interventions for patients with kidney failure and comorbid depression (10). The protocol, which follows PRISMA guidelines (11), was registered to PROSPERO before study initiation (CRD42020140227).

Data Sources and Searches

We searched Ovid MEDLINE, PsycINFO, Elsevier EMBASE, and Ovid EBM Reviews Cochrane Database of Systematic Reviews (Database of Abstracts of Reviews of Effects, Health Technology Assessment Database, Cochrane CENTRAL, *etc*.) from database inception through June 2020. We reviewed the bibliographies of relevant articles and contacted experts to identify additional studies. Search strategies were developed in consultation with a research librarian, and were peer reviewed by a second research librarian using the instrument for Peer Review of Search Strategies (12; Supplemental Material). All studies identified were completed before the onset of the COVID-19 pandemic.

Study Selection

Studies were eligible if they: (*1*) assessed depression in patients with kidney failure or stage 5 CKD; (*2*) compared an index (examined) tool to a "gold standard" clinical interview or another well-validated tool (*e.g*., Beck Depression Inventory II [BDI-II] [13]); (*3*) were published in English; and (*4*) examined tools based on criteria from the Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV) or higher (Supplemental Table 1). Studies were independently reviewed by at least two reviewers. Discordant results were resolved through consensus or a third reviewer.

Data Abstraction and Quality Assessment

From each study, we abstracted details related to sample size, setting, population, inclusion and exclusion criteria, administration and timing of depression screening, the index and reference standard (comparison), and findings. Data were abstracted by one investigator and confirmed by a second. Two reviewers independently assessed study risk of bias using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2). The QUADAS-2 is a commonly used 17-item tool for assessing the risk of bias across four domains: (*1*) flow and timing, (*2*) application of the reference standard, (*3*) application of the index test, and (*4*) patient selection (14). See Supplemental Table 2 for a list of all items. Disagreements were resolved by consensus or a third reviewer.

Data Synthesis

We qualitatively synthesized findings, organized them in tables, and present forest plots of the summary measures (*e.g*., sensitivity, specificity). The data did not allow for quantitative pooling of results.

Results

We reviewed 8050 titles and abstracts and 189 full text studies. A total of 16 studies were included (Figure 1). Nine studies examined the performance of the BDI-II (13). Other tools include the Cognitive Depression Index (CDI) (15), the Center for Epidemiologic Studies–Depression Scale (CES-D) (16), the Hospital Anxiety and Depression Scale–Depressive Subscale (HADS-D) (17), the Geriatric Depression Scale 15 (GDS-15) (18,19), the Hamilton Depression Rating Scale (20), the Patient Health Questionnaire 9 (PHQ-9) (21), and others. We identified only one study of a depression tool specifically targeting patients on maintenance dialysis (Depression Inventory–Maintenance Hemodialysis [DI-MHD]) (22). Supplemental Table 3 and Table 1 provide study characteristics and Table 2 provides a brief description of the included tools and the gold standard interviews used as reference standards.

There were five US studies (24,33,35,37,46). Other studies were located in Australia (25), Canada (27), China (22), Italy (29), The Netherlands (30,36), Norway (34), Saudi Arabia (23), Turkey (28), and the United Kingdom (26,30).

Most studies included only patients undergoing hemodialysis. Only four studies also included participants undergoing peritoneal dialysis (32,34,35,37). Across studies reporting time on dialysis, the minimum (mean) months was 8.5 (interquartile range, 4–22) (34) and the maximum was 72.2 (SD=11.7) (28).

Of the 16 studies, 11 compared tools with a gold standard clinical interview (*e.g*., Structured Clinical Interview for DSM-IV [38]), and five used other established, validated assessment measures (*e.g*., BDI-II [13]) for comparison.

Seven studies examined thresholds for major depressive disorder (23–26,30,33,37), one of which also categorized less severe depression (23). The remaining nine studies did not describe differences between major depressive disorder, less severe depressive disorders (*e.g*., dysthymia, pervasive depressive disorder), and subclinical symptoms (Supplemental Table 3) (22,27–29,31,32,34–36).

The 16 studies were relatively similar in quality, with the risk of bias largely unclear for patient selection, the index test, and the reference standard. Figure 2 summarizes the number of studies rated as low, unclear, and high risk of bias across the four QUADAS-2 domains (Supplemental Table 2 reports individual study ratings).

Screening Tools Compared with a Gold Standard Clinical Interview

Included studies compared nine tools to clinical interviews, across a range of thresholds. Figure 3 illustrates the performance characteristics by tool and threshold. Supplemental Table 4 and Table 1 provide detail.

Beck Depression Inventory II. Five studies examined the accuracy of the BDI-II in diagnosing major depressive disorder compared with a gold standard clinical interview (24–26,30,37). Sample sizes ranged from 40 (26) to 96 (24). Two were conducted in the United States (24,37). One was a small, multicenter study ($n=62$) that reported an optimal BDI-II cutoff of ≥16. Sensitivity was 0.91 and specificity was 0.86, with an area under the curve (AUC) of 0.94 (37). The second was a multicenter study of adults 65 and older

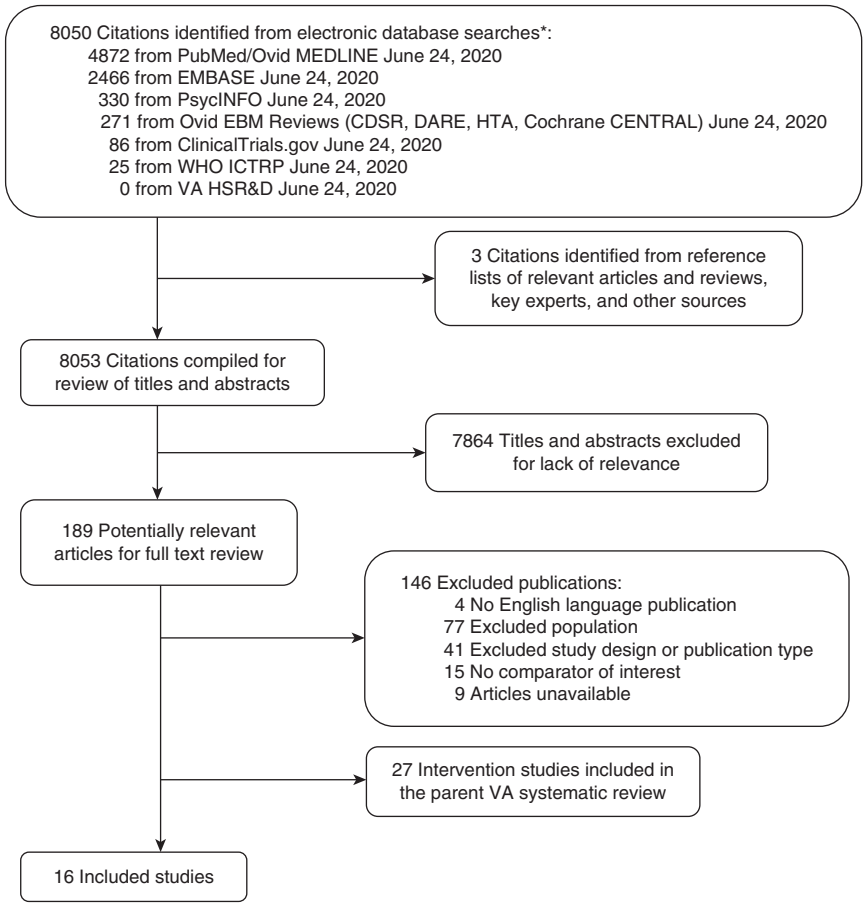

Figure 1. | **Literature flow chart.** *After deduplication. CDSR, Cochrane Database of Systematic Reviews; DARE, Database of Abstracts of Reviews of Effects; EBM, Evidence-based Medicine; HSR&D, Health Services Research and Development; HTA, Health Technology Assessment Database; ICTRP, International Clinical Trials Registry Platform; VA, Veterans Affairs; WHO, World Health Organization.

(n=96). At a cutoff of ≥10, sensitivity was 0.68, specificity was 0.77, and reported AUC was 0.73 (24).

One study examined a range of thresholds (30). The optimal threshold was ≥15, with a reported AUC of 0.93. Another study reported a much lower AUC (24). This study's population was limited to older adults, and age differences may have contributed to the difference in performance (24).

Among studies screening for depressive symptoms and disorders ranging from subclinical to major depressive disorder (22,31,32,34), sample sizes ranged from 43 (26) to 319 (22). Only one study (n=98) was conducted in the United States (31). At a threshold of ≥14, sensitivity was 0.62, specificity was 0.81, and AUC was 0.77 (31). The largest study (n=319), conducted in China, compared the BDI-II (≥19) to the Structured Clinical Interview for DSM-IV as part of a development and validation study for a depression screen designed specifically for patients undergoing maintenance hemodialysis (22). Sensitivity, specificity, positive predictive value, negative predictive value, and AUC were 0.83, 0.86, 0.63, 0.94, and 0.84, respectively.

Cognitive Depression Index. Four studies compared the CDI to a gold standard clinical interview (25,26,31,34), of which only two screened specifically for major depressive disorder (25,26). One study (n=45) examined a threshold of ≥11, with a sensitivity, specificity, and AUC of 0.79, 0.81, and 0.94, respectively (25). The second study identified an optimal threshold of ≥10. Sensitivity, specificity, and AUC were 0.78, 0.81, and 0.94, respectively (26).

The two studies screening for the range of depressive symptoms and diagnoses examined thresholds of ≥8 (n=98) (31) and ≥11 (n=109) (34). Sensitivity was 0.50 (31) and 0.82 (34), specificity was 0.83 (31) and 0.93 (34), and AUC was 0.76 (31) and 0.89 (34).

Of note, two studies (26,34) compared the BDI-II and the CDI to a clinical interview and both concluded that the BDI-II performed better.

Center for Epidemiologic Studies–Depression Scale. A multisite study (n=98) (31) compared the CES-D (≥18) to the Structured Clinical Interview for DSM-IV for depressive disorders and subclinical symptoms. Sensitivity, specificity, and AUC were 0.69, 0.83, and 0.89, respectively.

Depression Inventory–Maintenance Hemodialysis. A single validation study (n=319) conducted in China compared the Structured Clinical Interview for DSM-IV to both the BDI-II and the DI-MHD and found that at a cutoff of ≥25, the DI-MHD performed better than the BDI-II. Sensitivity, specificity, and AUC were 0.95, 0.93, and 0.94, respectively (22).

Table 1. Study characteristics

Author (Ref.)	*N*	Population	Index Test(s)	Reference Standard	% With Major Depressive Disorder Diagnosis[a]	
					Index Test (cutoff)	Reference Standard
Alsuwaida and Alwahhabi (23)	26	HD	SRQ (Arabic version)	Clinical interview	NR	15.4%
Balogun *et al.* (24)	96	Dialysis (HD/PD: NR)	BDI-II, GDS-15	Clinical interview	BDI-II (≥10): 37.1%, GDS-15 (≥5): 32.3% (subset *n*=62)	30.6% (subset *n*=62)
Bautovich *et al.* (25)	45	HD	BDI-II, CDI	Clinical interview	BDI-II/CDI: NR	13.3%
Chilcot *et al.*(26)	40	HD	BDI-II, CDI	MINI	BDI-II (≥16) on dialysis: 32.5%, off dialysis: 30%, CDI (≥10) on and off dialysis: 32.5%	22.5% (off dialysis)
Collister *et al.* (27)[b]	50	HD	Single question from the ESAS	HADS	ESAS: NR	HADS (≥7): 54%
Gencoz *et al.* (28)[b]	45	HD	Ham-D	SCID-I (Turkish Translation)	Ham-D: NR	4% MDD (18% other depressive disorders)
Giordano *et al.* (29)[b]	31	HD	GDS-15	BDI-II	GDS-15 (≥6): 58%	BDI-II (≥14): 61%
Grant *et al.* (30)[c]	57	HD	BDI-II	Clinical interview (based on ICD-10 diagnosis)	BDI-II (≥10): 56.1%, BDI-II (≥15): 31.6%	12.3%
Hedayati *et al.* (31)[b]	98	HD	BDI-II, CDI, CES-D, Feinstein Scale	SCID-I	BDI-II (≥14): 30.6%, CES-D (≥18): 30.6%	17.3% MDD (26.5% any depressive disorders)
Loosman *et al.* (32)[b]	62	HD and PD	BDI-II, HADS	MINI	BDI-II, HADS: NR; 33.9%	
Neitzer, 2012 (33)	134	HD	BDI-FS	BDI-II	BDI-FS (≥4): 30.1%	BDI-II (≥16): 28.7%
Preljevic *et al.* (34)[b]	109	HD and PD	BDI-II, CDI, HADS-D	SCID-I	BDI-II (≥16): 20.8%, CDI (≥11): NR, HADS-D (≥8): 20.1%	14.7% MDD (22% any depressive disorders)
Troidle *et al.* (35)[b]	97	CPD and HD	Two items from the KDQOL SF-36	BDI-II	KDQOL SF-36: NR	BDI-II (≥11): NR
van den Beukel *et al.* (36)[b]	133	HD: 72%	MHI5 of the SF-36	BDI-II/CDI (Dutch Translation)	MHI5 (≤70): 39%	BDI-II (≥16): 23% CDI (≥10): 23%
Watnick *et al.* (37)	62	HD and PD	BDI-II, PHQ-9	SCID-I	BDI-II, PHQ-9: NR	19.4%
Wang *et al.* (22)[b]	319	HD	BDI-II, DI-MHD (Chinese language)	SCID-I	BDI-II (≥19): 20.7%, DI-MHD (≥25): 20%	21.9%

HD, hemodialysis; SRQ, Self-Reporting Questionnaire; NR, not reported; PD, peritoneal dialysis; BDI-II, Beck Depression Inventory II; GDS-15, Geriatric Depression Scale 15; CDI, Cognitive Depression Index; MINI, Mini International Neuropsychiatric Interview; ESAS, Edmonton Symptom Assessment System; HADS: Hospital Anxiety and Depression Scale; Ham-D, Hamilton Depression Rating Scale; SCID-I, Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders IV; MDD, major depressive disorder; ICD-10, 10th revision of the International Statistical Classification of Diseases and Related Health Problems; CES-D, Center for Epidemiologic Studies–Depression Scale; BDI-FS, Beck Depression Inventory–Fast Screen; HADS-D, Hospital Anxiety and Depression Scale–Depressive Subscale; CPD, chronic pulmonary disease; KDQOL SF-36, Kidney Disease Quality of Life Short Form 36; MHI5, Mental Health Inventory 5; PHQ-9, Patient Health Questionnaire 9; DI-MHD, Depression Inventory–Maintenance Hemodialysis.
[a]The reported prevalence of major depressive disorder identified by the index test and the reference standard.
[b]Screened for depressive symptoms or milder forms of depression in addition to major depressive disorder.
[c]Included cutoff values for both major depressive disorder and for milder forms of depression and subclinical symptoms.

Table 2. Characteristics of included screening tools and gold standard semi-structured diagnostic interviews examined

Name	Abbrev.	Number of Items	Description	Time to Complete/Score
Gold standard semi-structured diagnostic interviews				
Mini International Neuropsychiatric Interview	MINI	NA	A short, structured diagnostic interview, developed jointly by psychiatrists and clinicians in the United States and Europe, for DSM-IV and ICD-10 psychiatric disorders.	Administration: 15 min
Structured Clinical Interview for DSM-IV Axis I Disorders (38)	SCID-I	NA	A semi-structured interview administered by a trained mental health professional familiar with DSM-IV Axis I diagnostic criteria.	Administration: 1–2 h
Screening tools				
Beck Depression Inventory II (13,39)	BDI-II	21	Widely used, validated self-report tool designed to assess depression severity in adolescents and adults. Most widely studied instrument in the kidney failure population. Closely mirrors DSM-IV criteria for major depressive disorder, and includes questions related to cognitive, affective, and somatic symptoms.	Respondent: 5–10 min Administrator: 5 min
Beck Depression Inventory II Fast Screen (40)	BDI-FS	7	Brief version of the BDI-II that excludes somatic symptoms. Designed to screen for major depressive disorder in medical patients.	Respondent: <5 min Administrator: <3 min, can be staff
Cognitive Depression Index (41)	CDI	15	A subset of the BDI-II (first 15 items), eliminating items related to somatic symptoms. It was developed for use in patients with CKD, with the goal of reducing the likelihood of the overdiagnosis of depression.	Respondent: 7–8 min Administrator: 5 min, These are estimates based on the BDI-II from which it is derived.
Center for Epidemiologic Studies–Depression Scale (16,39)	CES-D	20	Widely used, revised in 2004, evaluates depressive symptoms across four factors: depressive affect, well-being, somatic symptoms, and interpersonal relations.	Respondent: <10 min Administrator: <10 min (can be scored during administration)
Depression Inventory–Maintenance Hemodialysis (22)	DI-MHD		Developed specifically for patients with kidney failure.	Respondent: 5 min Administrator: no information yet available.
Edmonton Symptom Assessment System (27)	ESAS[a]	1	Single item that asks patients to rate depression from 0 (not depressed) to 10 (worst depression) (1).	Respondent: minimal Administrator: minimal
Hamilton Depression Rating Scale (20,42)	Ham-D	17	Assesses the frequency and intensity of depressive symptoms. Developed in 1960, and last revised in 1967 (20).	Administrator: 15–20 min
Hospital Anxiety and Depression Scale–Depression Subscale (17)	HADS-D	7	Subscale of the 14-item HADS that includes ratings of physical, cognitive, and affective symptoms of depression (17).	Respondent: ≤5 min (for entire HADS including anxiety items) Administrator: 1–2 min
Geriatric Depression Scale-15 (18,39)	GDS-15	15	Shortened version of the original 30-item GDS, assesses depressive symptoms in older adults, developed in 1982 (18,19).	Respondent: 2–5 min Administrator: 2 min
Kidney Disease Quality of Life Short Form - 36 (43)	KDQOL SF-36	1	The KDQOL SF-36 is a self-report measure developed for patients with kidney disease on dialysis. Six-options ranging from "all of the time" to "none of the time." The included study tests the use of two single questions. (1) "Have you felt so down in the dumps that nothing could cheer you up?" (2) "Have you felt downhearted and blue?"	Respondent: minimal Administrator: minimal
Mental Health Inventory 5 (44)	MHI5	5	Also known as the mental health subscale of the SF-36. Six option Likert scale ranging from "all of the time" to "none of the time."	Respondent: <5 min Administrator: minimal
Patient Health Questionnaire-9 (21,39)	PHQ-9	9	Developed in 2001. Screen of depression and severity. Widely used in the United States and internationally (21).	Respondent: <3 min Administrator: minimal

Table 2. (Continued)

Name	Abbrev.	Number of Items	Description	Time to Complete/Score
Self-Reporting Questionnaire (45)	SRQ	20	Developed by the WHO to screen for a range of mental health disorders (45).	Respondent: "a few min" Administrator: 1 min, yes/no additive

MINI, Mini International Neuropsychiatric Interview; NA, not applicable; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders IV; ICD-10, 10th revision of the International Statistical Classification of Diseases and Related Health Problems; SCID-I, Structured Clinical Interview for DSM-IV; BDI-II, Beck Depression Inventory–II; BDI-FS, Beck Depression Inventory–Fast Screen; CDI, Cognitive Depression Index; CES-D, Center for Epidemiologic Studies–Depression Scale; DI-MHD, Depression Inventory–Maintenance Hemodialysis; ESAS, Edmonton Symptom Assessment System; Ham-D, Hamilton Depression Rating Scale; HADS-D, Hospital Anxiety and Depression Scale–Depressive Subscale; GDS-15, Geriatric Depression Scale-15; KDQOL SF-36, Kidney Disease Quality of Life Short Form 36; SF-36, 36-item Short-Form Health Survey Questionnaire; MHI5, Mental Health Inventory 5; PHQ-9, Patient Health Questionnaire 9; SRQ, Self-Reporting Questionnaire; WHO, World Health Organization.
[a]The ESAS is a ten-item scale. However, only one item applies to depression.

Hamilton Depression Rating Scale A single study (n=45) conducted in Turkey compared the Hamilton Depression Rating Scale (≥10) to the Structured Clinical Interview for DSM-IV and screened for the range of depressive symptoms and disorders. Reported sensitivity was 1.00, specificity was 0.80, and AUC was 0.85 (28).

Hospital Anxiety and Depression Scale–Depression Subscale. Two studies examined the performance characteristics of the HADS-D (32,34). Both studies screened for depressive disorders and subclinical symptoms. One study (n=62; ≥6) reported sensitivity, specificity, and AUC values of 0.91, 0.76, and 0.89, respectively (32). The other (n=109; ≥8) reported sensitivity, specificity, and AUC values of 0.73, 0.87, and 0.91, respectively. Of note, this study also examined the BDI-II (≥16), and concluded it performed better than the HADS-D (34).

Geriatric Depression Scale 15 (adults aged 60+). A single study (n=96) compared the GDS-15 (≥5) to a gold standard interview for major depressive disorder. Sensitivity was 0.62, specificity was 0.82, and AUC was 0.81 (24).

Patient Health Questionnaire 9. A small multisite study (n=62) compared the PHQ-9 (≥10) to the Structured Clinical Interview for DSM-IV for major depressive disorder. Sensitivity and specificity were both 0.92, and AUC was 0.94 (37).

Self-Reporting Questionnaire. A single small study (n=26) conducted in Saudi Arabia compared the Self-Reporting Questionnaire (≥13) to Structured Clinical Interview for DSM-IV for major depressive disorder. Sensitivity, specificity, and AUC were 1.00, 0.82, and 0.96, respectively (23).

Screening Tools Compared with Other Tools

Five studies compared other, generally short, tools to established, validated tools *(i.e.*, BDI-II, HADS-D; see Figure 4, Supplemental Table 5, Table 1) (27,29,33,35,36). One study screened for major depressive disorder specifically, and evaluated the BDI–Fast Screen (40). It had high sensitivity and specificity compared with the BDI-II (≥16) (33). The GDS-15 (≥6) screened for a range of depression diagnoses and symptoms and appeared to perform well (29).

Timing of Screening

A small (n=43) multisite UK study in outpatient hemodialysis units compared depression screening (BDI-II, CDI) completed on and off dialysis (26). Findings indicated a high level of agreement among patients who were depressed. However, patients who were not depressed had higher mean overall BDI-II (9.6 [6.2] versus 7.3 [5.7], P=0.007) and somatic symptom item scores (4.4 [2.5] versus 3.3 [2.1], P=0.01) on assessments completed while undergoing dialysis (Supplemental Table 3).

Discussion

We identified 16 studies examining the performance characteristics of depression screening tools in patients

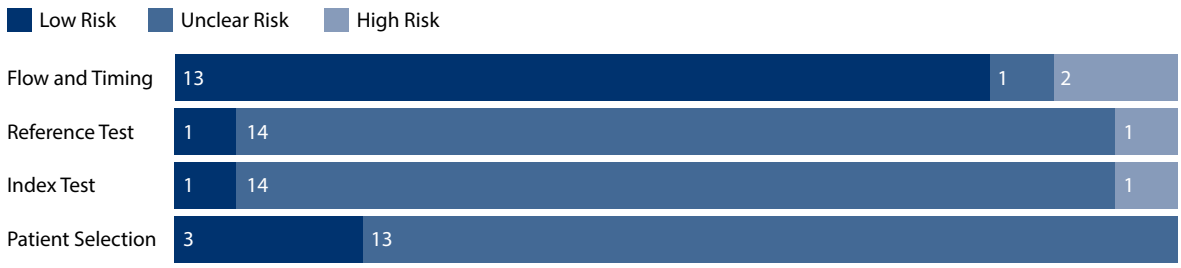

Figure 2. | **QUADAS-2 risk of bias summary.** Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2), independently assessed by two investigators.

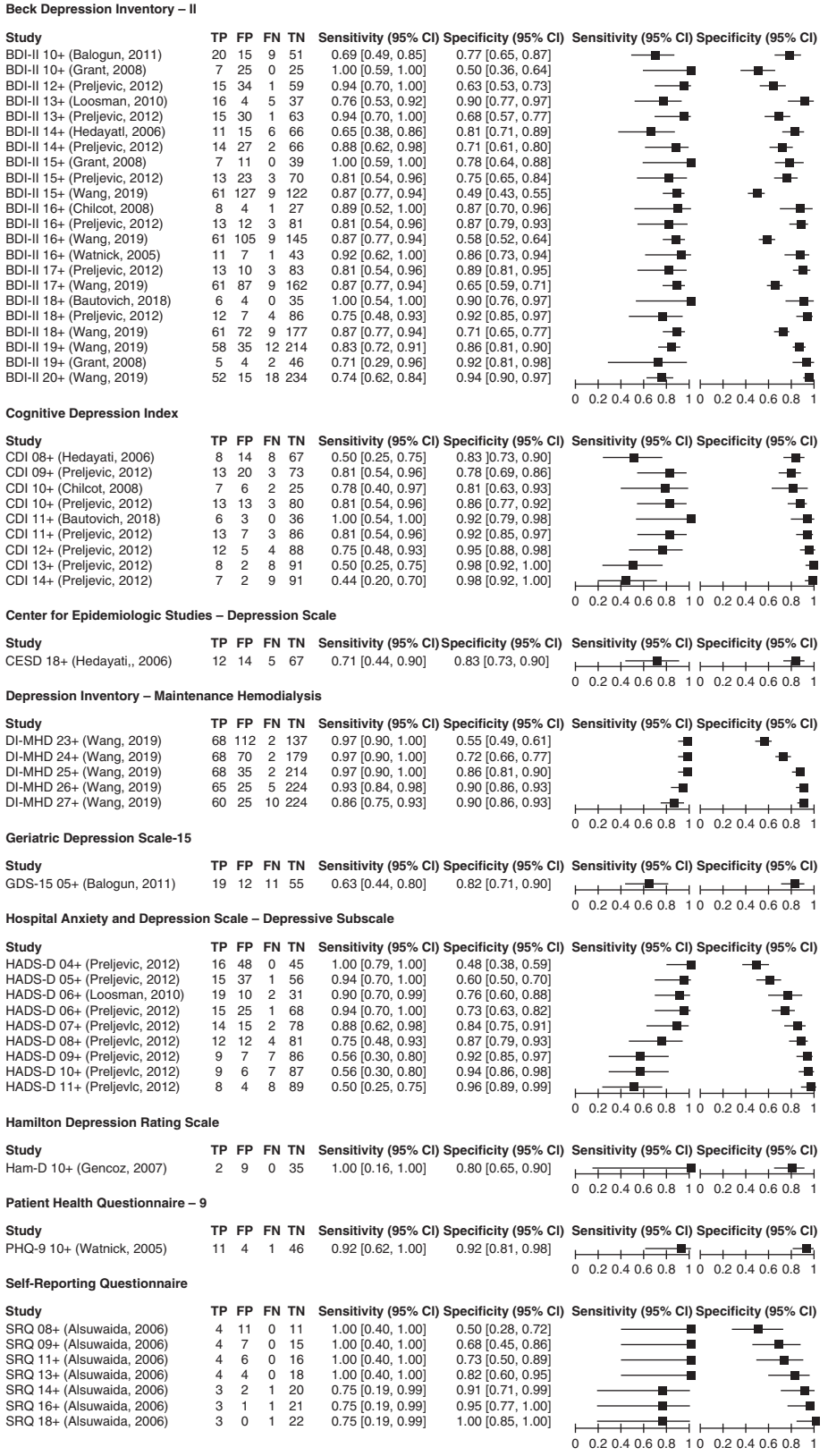

Figure 3. | **Performance characteristics of tools compared with a gold standard clinical interview.** BDI-II, Beck Depression Inventory–II; CES-D, Center for Epidemiologic Studies–Depression Scale; CI, confidence interval; DI-HMD, Depression Inventory–Maintenance Hemodialysis; FN, false negative; FP, false positive; GDS-15, Geriatric Depression Scale-15; HADS-D, Hospital Anxiety and Depression Scale–Depressive Subscale; Ham-D, Hamilton Depression Rating Scale; SRQ, Self-Reporting Questionnaire; TN, true negative; TP, true positive.

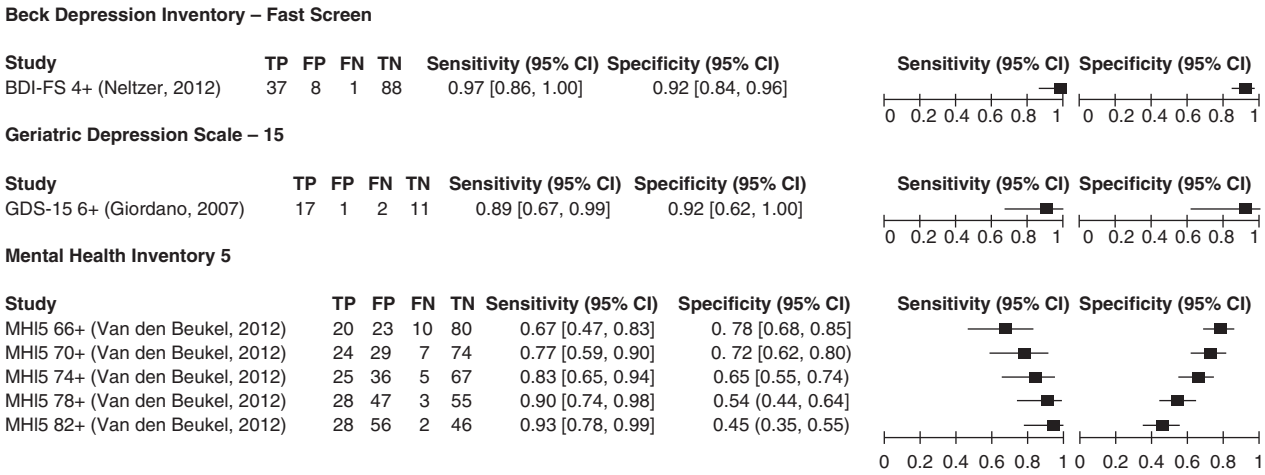

Figure 4. | **Performance characteristics of tools compared with a gold standard clinical interview.** BDI-FS, Beck Depression Inventory–Fast Screen; ESAS, Edmonton Symptom Assessment System; MHI5, Mental Health Inventory 5.

with kidney failure, and found depression can be accurately diagnosed. By far, the strongest body of evidence addressed the use of longer screening tools and those that require a per-administration cost (*e.g.*, BDI-II), which are not common in medical settings. We found promising evidence that shorter instruments in the public domain such as the PHQ-9 and GDS-15 (for older adults) performed well, but only one study compared each of these instruments to a clinical interview (24,37).

Across studies, sample sizes were small, and studies examined a wide (and inconsistent) range of thresholds. In addition, methodologic details, particularly related to the selection of patients and the conduct and/or interpretation of both the index and reference tests, were generally poorly reported. We identified few studies conducted in the United States, or countries with similar health systems, raising concerns about the generalizability of findings. Except for the BDI-II, the evidence base is quite limited due to few studies examining each tool. There was heterogeneity in how depression was operationalized across studies. Half of the studies evaluated the performance characteristics associated with thresholds intended to screen for major depressive disorder, whereas the other half defined depression broadly, some including subclinical depressive symptoms.

Figure 5 illustrates the effect of sensitivity and specificity across different population-based depression rates. We used data from a US study comparing the PHQ-9 to a gold standard interview (37) that screened specifically for major depressive disorder. At a threshold of ≥10, both sensitivity and specificity were 0.92. Holding these constant, we compared positive and negative predictive values across reported major depressive disorder prevalence rates for (*1*) general US populations (7.1%) (2); (*2*) US patients with kidney failure, diagnosed using a gold standard interview (22.8%) (3); and (*3*) US patients with kidney failure, diagnosed using a screening tool (39.3%) (3). Across populations, the negative predictive values, or accuracy of eliminating depression, are generally high, and false negatives are unlikely. However, the positive predictive values, or accuracy of correctly diagnosing depression, range from 0.47 to 0.88, suggesting in this example that for populations with a lower prevalence of depression, the potential for false positives may be high (Figure 5). Providers should keep these factors in mind when using the results of depression screening tools to guide treatment decisions.

Among the studies evaluating the BDI-II as a tool to identify major depressive disorder, the threshold that best optimized the balance between sensitivity and specificity for patients with kidney failure was ≥16. In fact, in some studies, the BDI-II performed reasonably well when compared with a gold standard clinical interview. The caveats

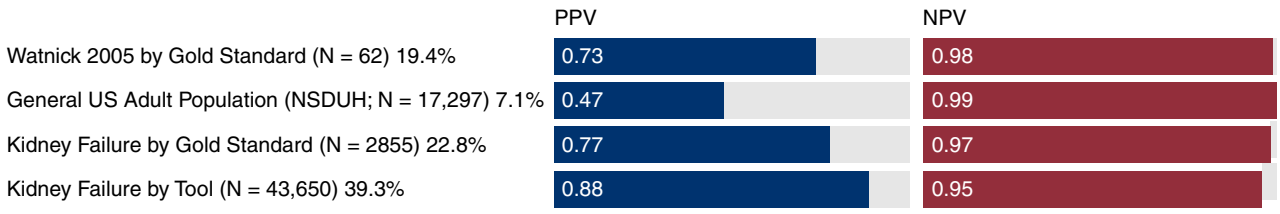

Figure 5. | **PHQ-9 ≥10 positive and negative predictive values: Three US subpopulations.** Both sensitivity and specificity are held constant at 0.92 based on findings from Watnick's 2005 study comparing the PHQ-9 to a gold standard clinical interview (37). Prevalence data for the general US population comes from the 2017 National Survey on Drug Use and Health (2), and prevalence for patients with kidney disease is reported in a meta-analysis of 249 unique populations (3). NPV, negative predictive value; PHQ-9, Patient Health Questionnaire 9; PPV, positive predictive value.

are the heterogeneity in how tools were administered, and that very few studies contributed data for the same thresholds. Interestingly, two studies found that, compared with a clinical interview, the BDI-II performed better than the CDI, a subset of the BDI without items related to somatic symptoms.

Shorter screening tools compared with established, validated tools performed well overall. Because the ESRD-QIP requires a follow-up after an initial positive screen, these short tools may be appropriate for an initial depression screen of all patients with kidney failure. In particular, the BDI–Fast Screen performed well when compared with the BDI-II. Of note, we identified no studies evaluating the PHQ-2, arguably the most commonly used short screen for depression in US medical settings.

One study examined differences in performance based on the timing of screening and found that participants who were not depressed reported significantly more somatic symptoms when they were screened during dialysis sessions versus off dialysis. Not only were scores on somatic items significantly higher, but BDI-II scores were significantly higher as well. This has implications for dialysis units working to streamline processes, as it illustrates the potential for overdiagnosis and overtreatment.

There are several important limitations; notably, small sample sizes and few studies examining specific tool thresholds. Many studies were conducted outside of the United States and examined participants and health systems that differ from US populations and settings. In addition, the lack of methodology detail reported in many of the studies contributed to uncertainty about study processes and poor or unclear quality ratings. The definition of depression varied widely, which hampered our ability to synthesize the body of research for each tool. Future studies should use standardized language and diagnostic criteria (*e.g.*, DSM-5) (47).

As described above, future research examining the diagnostic accuracy of depression tools in US populations is needed. In addition, although the PHQ-9 is used widely in medical settings, current research in kidney failure populations is extremely limited. Similarly, despite wide use of the PHQ-2 as an initial screen, no studies were identified. Research evaluating performance characteristics of both tools in this population are warranted. There are a handful of studies supporting the use of the BDI-II as a screening tool for major depressive disorder in this population. However, the BDI-II requires a per-use fee, is more commonly used in research and mental health than in medical settings, was developed to align to the out-of-date DSM-IV, and may be less informative for screening and assessment due to its reliance on somatic symptoms. Future research should target free tools (*e.g.*, CES-D, PHQ-9, PHQ-2) that are widely used in US medical settings.

Short, population-targeted tools may be appropriate as an initial screen for depression in dialysis settings, and we identified several high-performing tools that used the BDI-II as a reference standard. However, more research is needed to validate existing findings. Given the ESRD-QIP requirement of both an initial and follow-up screen (if warranted), future research should evaluate both quick screens and those that are more comprehensive. Finally, the DI-MHD appears to be the only screening tool for depression designed specifically for patients with kidney failure. It performed well in a large sample in China. Additional research validating the DI-MHD generally and in English-speaking patients has the potential to affect screening practices in this population.

We identified no studies examining the effect of screening on health outcomes, and only one study that examined differences in implementation. It compared differences in both overall BDI-II scores and somatic item scores when completed on versus off dialysis and touches on only one of many important depression screening implementation issues (*e.g.*, timing, location, administration). Implementation is an important issue, as patients' responses may differ based on setting and timing, due not only to the experience of somatic symptoms, but also the perception of privacy and other factors. Future implementation research may help to better identify depression in patients with kidney failure and minimize overtreatment. We identified no studies examining potential harm associated with the lack of a standardized depression screening tool (*e.g.*, false positive or negative depression screens contributing to over- or undertreatment). Harm research is especially important in patients with chronic conditions such as kidney failure, given that the physical symptoms of chronic illnesses and their treatment can mimic the somatic symptoms of depression. Also important, but missing, is evidence of potential demographic and clinical differences. Research in these areas will help decision makers to implement screening processes that are not only evidence based, but also the best fit for patient populations.

Our findings have implications for the selection and implementation of depression screening in patients with kidney failure, and highlight the moderate positive predictive values in this population. Clinicians should be prepared to validate positive screens before making treatment decisions that may be burdensome or introduce the possibility of harm.

There is limited research evaluating the diagnostic accuracy of most screening tools for depression in patients with kidney failure, and existing studies may not be generalizable to US populations. Studies suffer from limitations related to methodological quality and/or reporting. Future research should target widely used, free tools such as the PHQ-2 and the PHQ-9.

Disclosures

All authors have nothing to disclose.

Funding

This research was funded by the US Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Health Services Research & Development.

Acknowledgments

The authors wish to thank Ms. Robin Paynter for developing the search strategy and running electronic searches.

The findings and conclusions in this document are those of the authors who are responsible for its contents, the findings and conclusions do not necessarily represent the views of the Department of Veterans Affairs or the US government.

Ms. Chelsea K. Ayers, Dr. Devan Kansagara, Dr. Jennifer R. Antick, Dr. Karli Kondo, and Dr. Pavan Chopra were responsible for the

research idea and study design; Dr. Jennifer R. Antick and Dr. Karli Kondo were responsible for study selection and data extraction; Dr. Jennifer R. Antick and Dr. Karli Kondo were responsible for data analysis and interpretation; Ms. Chelsea K. Ayers, Dr. Devan Kansagara, Dr. Jennifer R. Antick, Dr. Karli Kondo, and Dr. Pavan Chopra were responsible for evaluation of study quality and strength of evidence; and Dr. Devan Kansagara, Dr. Jennifer R. Antick, and Dr. Pavan Chopra were responsible for supervision and mentorship. Each author contributed important intellectual content during manuscript drafting or revision, accepts personal accountability for the author's own contributions, and agrees to ensure that questions pertaining to the accuracy or integrity of any portion of the work are appropriately investigated and resolved.

Supplemental Material

This article contains the following supplemental material online at http://cjasn.asnjournals.org/lookup/suppl/doi:10.2215/CJN.05540420/-/DCSupplemental.

Supplemental Material. Search strategies (parent VA systematic review) and supplemental references.

Supplemental Table 1. PICOTS by key question.

Supplemental Table 2. QUADAS-2 risk of bias assessment.

Supplemental Table 3. Characteristics of studies examining the diagnostic accuracy of depression screening tools in patients with kidney failure.

Supplemental Table 4. Findings of studies examining the diagnostic accuracy of depression screening tools in patients with kidney failure compared with a gold standard diagnostic interview.

Supplemental Table 5. Studies comparing a depression tool to another validated depression tool.

References

1. United States Renal Data System (USRDS): 2018 USRDS Annual Data Report. Available at: https://www.usrds.org/media/2283/2018_volume_2_esrd_in_the_us.pdf. Accessed April 20, 2019
2. National Institute of Mental Health: Major Depression, 2017. Available at: https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHDetailedTabs2017/NSDUHDetailedTabs2017.htm#tab8-56A. Accessed October 3, 2020
3. Palmer S, Vecchio M, Craig JC, Tonelli M, Johnson DW, Nicolucci A, Pellegrini F, Saglimbene V, Logroscino G, Fishbane S, Strippoli GF: Prevalence of depression in chronic kidney disease: Systematic review and meta-analysis of observational studies. *Kidney Int* 84: 179–191, 2013
4. Xing L, Chen R, Diao Y, Qian J, You C, Jiang X: Do psychological interventions reduce depression in hemodialysis patients?: A meta-analysis of randomized controlled trials following PRISMA. *Medicine (Baltimore)* 95: e4675, 2016
5. Farrokhi F, Abedi N, Beyene J, Kurdyak P, Jassal SV: Association between depression and mortality in patients receiving long-term dialysis: A systematic review and meta-analysis. *Am J Kidney Dis* 63: 623–635, 2014
6. Weisbord SD, Mor MK, Sevick MA, Shields AM, Rollman BL, Palevsky PM, Arnold RM, Green JA, Fine MJ: Associations of depressive symptoms and pain with dialysis adherence, health resource utilization, and mortality in patients receiving chronic hemodialysis. *Clin J Am Soc Nephrol* 9: 1594–1602, 2014
7. Kurella M, Kimmel PL, Young BS, Chertow GM: Suicide in the United States end-stage renal disease program. *J Am Soc Nephrol* 16: 774–781, 2005
8. Centers for Medicare & Medicaid Services: ESRD Quality Incentive Program, 2018. Available at: https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/esrdqip/. Accessed April 5, 2018
9. Siu AL, Bibbins-Domingo K, Grossman DC, Baumann LC, Davidson KW, Ebell M, García FA, Gillman M, Herzstein J, Kemper AR, Krist AH, Kurth AE, Owens DK, Phillips WR, Phipps MG, Pignone MP; US Preventive Services Task Force (USPSTF): Screening for depression in adults: US preventive services task force recommendation statement. *JAMA* 315: 380–387, 2016
10. Kondo K, Ayers CK, Chopra P, Antick J, Kansagara D: End stage renal disease and depression: A systematic review, 2019. Washington, DC, Evidence Synthesis Program, Health Services Research and Development Service, Office of Research and Development, Department of Veterans Affairs. VA ESP Project #05-225
11. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group: Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med* 6: e1000097, 2009 10.1371/journal.pmed.1000097
12. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C: PRESS peer review of electronic search strategies: 2015 guideline statement. *J Clin Epidemiol* 75: 40–46, 2016
13. Beck AT, Steer RA, Brown GK: *BDI-II: Beck Depression Inventory Manual*, 2nd Ed., San Antonio, TX, Psychological Corporation, 1996
14. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM; QUADAS-2 Group: QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 155: 529–536, 2011
15. Kimmel PL, Weihs K, Peterson RA: Survival in hemodialysis patients: The role of depression. *J Am Soc Nephrol* 4: 12–27, 1993
16. The Center for Innovative Public Health Research: CESD-R: The Center for Epidemiologic Studies Depression Scale Revised. Available at: https://cesd-r.com/about-cesdr/. Accessed November 1, 2019
17. Snaith RP, Zigmond AS: The hospital anxiety and depression scale. *Br Med J (Clin Res Ed)* 292: 344, 1986
18. Sheikh JI, Yesavage JA: Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clin Gerontol* 5: 165–173, 1986
19. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, Leirer VO: Development and validation of a geriatric depression screening scale: A preliminary report. *J Psychiatr Res* 17: 37–49, 1982-1983
20. Hamilton M: Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 6: 278–296, 1967
21. Kroenke K, Spitzer RL, Williams JB: The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med* 16: 606–613, 2001
22. Wang YY, Zhang WW, Feng L, Gao D, Liu C, Zhong L, Ren JW, Wu YZ, Huang L, Fu LL, He YN: Development and preliminary validation of a depression assessment tool for maintenance hemodialysis patients. *Ther Apher Dial* 23: 49–58, 2019
23. Alsuwaida A, Alwahhabi F: The diagnostic utility of Self-Reporting Questionnaire (SRQ) as a screening tool for major depression in hemodialysis patients. *Saudi J Kidney Dis Transpl* 17: 503–510, 2006
24. Balogun RA, Turgut F, Balogun SA, Holroyd S, Abdel-Rahman EM: Screening for depression in elderly hemodialysis patients. *Nephron Clin Pract* 118: c72–c77, 2011
25. Bautovich A, Katz I, Loo CK, Harvey SB: Beck Depression Inventory as a screening tool for depression in chronic haemodialysis patients. *Australas Psychiatry* 26: 281–284, 2018
26. Chilcot J, Wellsted D, Farrington K: Screening for depression while patients dialyse: An evaluation. *Nephrol Dial Transplant* 23: 2653–2659, 2008
27. Collister D, Rodrigues JC, Mazzetti A, Salisbury K, Morosin L, Rabbat C, Brimble KS, Walsh M: Single questions for the screening of anxiety and depression in hemodialysis. *Can J Kidney Health Dis* 6: 1–7, 2019 10.1177/2054358118825441
28. Gençöz F, Gençöz T, Soykan A: Psychometric properties of the Hamilton Depression Rating Scale and other physician-rated psychiatric scales for the assessment of depression in ESRD patients undergoing hemodialysis in Turkey. *Psychol Health Med* 12: 450–459, 2007
29. Giordano M, Tirelli P, Ciarambino T, Gambardella A, Ferrara N, Signoriello G, Paolisso G, Varricchio M: Screening of depressive symptoms in young-old hemodialysis patients: Relationship between Beck Depression Inventory and 15-item Geriatric Depression Scale. *Nephron Clin Pract* 106: c187–c192, 2007
30. Grant D, Almond MK, Newnham A, Roberts P, Hutchings A: The Beck Depression Inventory requires modification in scoring

before use in a haemodialysis population in the UK. *Nephron Clin Pract* 110: c33–c38, 2008
31. Hedayati SS:Accuracy of depression evaluation in hemodialysis patients. *Nat Clin Pract Nephrol* 2: 409–410, 2006
32. Loosman WL, Siegert CE, Korzec A, Honig A: Validity of the hospital anxiety and depression scale and the Beck depression inventory for use in end-stage renal disease patients. *Br J Clin Psychol* 49: 507–516, 2010
33. Neitzer A, Sun S, Doss S, Moran J, Schiller B: Beck Depression Inventory-Fast Screen (BDI-FS): An efficient tool for depression screening in patients with end-stage renal disease. *Hemodial Int* 16: 207–213, 2012
34. Preljevic VT, Østhus TB, Sandvik L, Opjordsmoen S, Nordhus IH, Os I, Dammen T: Screening for anxiety and depression in dialysis patients: Comparison of the hospital anxiety and depression scale and the beck depression inventory. *J Psychosom Res* 73: 139–144, 2012
35. Troidle L, Wuerth D, Finkelstein S, Kliger A, Finkelstein F: The BDI and the SF36: Which tool to use to screen for depression? *Adv Perit Dial* 19: 159–162, 2003
36. van den Beukel TO, Siegert CE, van Dijk S, Ter Wee PM, Dekker FW, Honig A: Comparison of the SF-36 five-item mental health inventory and beck depression inventory for the screening of depressive symptoms in chronic dialysis patients. *Nephrol Dial Transplant* 27: 4453–4457, 2012
37. Watnick S, Wang PL, Demadura T, Ganzini L: Validation of 2 depression screening tools in dialysis patients. *Am J Kidney Dis* 46: 919–924, 2005
38. First M: *User's Guide for the Structured Clinical Interview for DSM-IV Axis I Disorders SCID-I: Clinician Version*, Washington, DC, American Psychiatric Press, 1997
39. Smarr KL, Keefer AL: Measures of depression and depressive symptoms: Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire-9 (PHQ-9). *Arthritis Care Res (Hoboken)* 63[Suppl 11]: S454–S466, 2011
40. Beck AT, Steer RA, Brown GK: BDI - FastScreen for medical patients. Available at: https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Personality-%26-Biopsychosocial/Brief/BDI-FastScreen-for-Medical-Patients/p/100000173.html?tab=product-details. Accessed August 6, 2020.
41. Sacks CR, Peterson RA, Kimmel PL: Perception of illness and depression in chronic renal disease. *Am J Kidney Dis* 15: 31–39, 1990
42. Psychiatry~ Behavioral Health Learning Network: Hamilton Depression Rating Scale (HAM-D). Available at: https://www.psychcongress.com/hamilton-depression-rating-scale-ham-d#:~:text=It%20generally%20takes%2015%2D20,are%20scored%20from%200%2D2.&text=Since%20its%20development%20in%201960%20by%20Dr. Accessed July 22, 2020
43. RAND Health Care: Kidney Disease Quality of Life Instrument (KDQOL). Available at: https://www.rand.org/health-care/surveys_tools/kdqol.html. Accessed July 22, 2020
44. National Multiple Sclerosis Society: Mental Health Inventory (MHI). Available at: https://www.nationalmssociety.org/For-Professionals/Researchers/Resources-for-Researchers/Clinical-Study-Measures/Mental-Health-Inventory-(MHI). Accessed July 22, 2020
45. World Health Organization: *A User's Guide to the Self Reporting Questionnaire (SRQ)*, Geneva, Division of Mental Health World Health Organization, 1994
46. Hedayati SS, Bosworth HB, Kuchibhatla M, Kimmel PL, Szczech LA: The predictive value of self-report scales compared with physician diagnosis of depression in hemodialysis patients. *Kidney Int* 69: 1662–1668, 2006
47. American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, 5th Ed., Arlington, VA, American Psychiatric Association, 2013

Received: April 20, 2020 **Accepted:** October 19, 2020

Published online ahead of print. Publication date available at www.cjasn.org.

See related Patient Voice, "Depression: A Side Effect of CKD," on pages xxx–xxx and editorial, "Screening for Depression in People with Kidney Failure," on pages xxx–xxx, respectively.