

Table S1. Search terms applied to electronic databases †

Database	Search terms
The African Journal Archive	depression OR distress
African Journals Online	(depressive OR depression OR affective OR mood OR postpartum OR postnatal) AND (diagnosis OR sensitivity OR specificity OR validation OR screening OR psychometric OR "factor analysis" OR "factor structure" OR reliability OR validity OR consistency)
Cumulative Index to Nursing and Allied Health Literature	(MH "affective disorders" OR MH "Beck Depression Inventory, revised edition" OR MH "Center for Epidemiological Studies Depression scale" OR MH "Edinburgh Postnatal Depression Scale" OR MH "Geriatric Depression Scale" OR MH "Hamilton Rating Scale for Depression" OR MH "Self-Rating Depression Scale" OR MH "Bech-Rafaelson Melancholia Scale" OR AB depressive OR AB depression OR AB "psychological distress" OR AB idiom) AND (MH Africa OR MH refugees OR AB Africa) AND (MH diagnosis OR MH psychometrics OR MH "factor analysis" OR MH "reliability and validity" OR MH "rasch analysis" OR AB "construct validity" OR AB "convergent validity" OR AB "divergent validity" OR AB "discriminant validity" OR AB "content validity" OR AB "face validity" OR AB "criterion validity" OR AB "concurrent validity" OR AB "predictive validity" OR AB validity OR AB reliability OR AB consistency)
Embase	(depression:cl OR "distress syndrome":cl OR "Bech-Rafaelsen Melancholia Scale":cl OR "Beck Depression Inventory":cl OR "Beck Hopelessness Scale":cl OR "Brief Psychiatric Rating Scale":cl OR "depression inventory":cl OR "Diagnostic Interview Schedule":cl OR "Edinburgh Postnatal Depression Scale":cl OR "General Health Questionnaire":cl OR "Geriatric Depression Scale":cl OR "Hamilton scale":cl OR "Montgomery Asberg Depression Rating Scale":cl OR "Psychiatric Symptom Index":cl OR "Self-rating Depression Scale":cl OR "Symptom Checklist 90":cl OR depressive:ab OR depression:ab OR "psychological distress":ab OR idiom:ab) AND (Africa:cl OR refugee:cl OR Africa:ab) AND ("psychiatric diagnostic":cl OR diagnosis:cl OR "sensitivity and specificity":cl OR validity:cl OR reliability:cl OR "mass screening":cl OR "factorial analysis":cl OR "Rasch analysis":cl OR psychometry:cl OR screening:ab OR psychometric:ab OR "factor analysis":ab OR "factor structure":ab OR "Rasch analysis":ab OR "latent trait analysis":ab OR "construct validity":ab OR "convergent validity":ab OR "divergent validity":ab OR "discriminant validity":ab OR "content validity":ab OR "face validity":ab OR "criterion validity":ab OR "concurrent validity":ab OR "predictive validity":ab OR "validity":ab OR "reliability":ab OR "consistency":ab)

Supplemental Digital Content to accompany: [authors]. Reliability and validity of depression assessment among persons with HIV in sub-Saharan Africa: systematic review and meta-analysis. *J Acquir Immune Defic Syndr*

Medical Literature Analysis and Retrieval System Online	(“depressive disorder”[MeSH Terms] OR “depression”[MeSH Terms] OR “affective symptoms”[MeSH Terms] OR “mood disorders”[MeSH Terms] OR “depression, postpartum”[MeSH Terms] OR “stress, psychological”[MeSH Terms] OR “depressive”[TIAB] OR “depression”[TIAB] OR “psychological distress”[TIAB] OR “idiom”[TIAB]) AND (“Africa”[MeSH Terms] OR “refugees”[MeSH Terms] OR “Africa”[TIAB]) AND (“diagnosis”[MeSH Terms] OR “sensitivity and specificity”[MeSH Terms] OR “reproducibility of results”[MeSH Terms] OR “validation studies as topic”[MeSH Terms] OR “validation studies”[Publication Type] OR “screening”[TIAB] OR “psychometric”[TIAB] OR “factor analysis”[TIAB] OR “factor structure”[TIAB] OR “Rasch analysis”[TIAB] OR “latent trait analysis”[TIAB] OR “construct validity”[TIAB] OR “convergent validity”[TIAB] OR “divergent validity”[TIAB] OR “discriminant validity”[TIAB] OR “content validity”[TIAB] OR “face validity”[TIAB] OR “criterion validity”[TIAB] OR “concurrent validity”[TIAB] OR “predictive validity”[TIAB] OR “validity”[TIAB] OR “reliability”[TIAB] OR “consistency”[TIAB])
PsycINFO	(DE “affective disorders” OR DE “stress” OR AB “depressive” OR AB “depression” OR AB “psychological distress” OR AB “idiom”) AND (DE “african cultural groups” OR DE “refugees” or AB “Africa”) AND (DE “test validity” OR DE “statistical validity” OR DE “factor analysis” OR DE “screening” OR DE “psychometrics” OR AB “psychometric” OR AB “factor analysis” OR AB “factor structure” OR AB “Rasch analysis” OR AB “latent trait analysis” OR AB “construct validity” OR AB “convergent validity” OR AB “divergent validity” OR AB “discriminant validity” OR AB “content validity” OR AB “face validity” OR AB “criterion validity” OR AB “concurrent validity” OR AB “predictive validity” OR AB “validity” OR AB “reliability” OR AB “consistency”)
World Health Organization African Index Medicus	depression OR distress

† All database searches were completed on January 27, 2012, with the exception of searches conducted using the African Journal Archive, African Journals Online, and the World Health Organization African Index Medicus (which were completed May 30, 2012). The Medical Literature Analysis and Retrieval System Online search was updated on December 25, 2013.

Table S2. Characteristics of studies on the reliability and/or validity of depression assessment among persons with HIV in sub-Saharan Africa

Reference	Study design	Instrument	Evidence for reliability and/or validity
Akena et al. (110)	A consecutive sample of 94 HIV-positive adults (18+ years) attending an HIV primary care facility in Kampala, Uganda were administered a 7-item depression scale where binary response options for normal/healthy states were depicted by visual representations of each state. Patients previously diagnosed with major depressive disorder were excluded. Clinicians conducted a diagnostic interview using the MINI to establish the reference criterion diagnosis of depression.	Visual scale	A visual depression scale score ≥ 10 had 0.75 sensitivity and 0.71 specificity for detecting depression (AUC=0.82).
Akena et al. (98)	A consecutive sample of 368 HIV-positive adults (18+ years) attending an HIV primary care facility in Kampala, Uganda were administered the PHQ-9, K10, and CES-D. Psychiatrists blinded to the screening results evaluated all patients using the MINI to establish the reference criterion of major depressive disorder.	CES-D K6/K10 PHQ	For detection of major depressive disorder, the screening instruments performed as follows: CES-D ≥ 18 had 0.88 sensitivity and 0.81 specificity (AUC=0.94); K6 ≥ 13 had 0.77 sensitivity and 0.67 specificity (AUC=0.81); K10 ≥ 23 had 0.83 sensitivity and 0.72 specificity (AUC=0.82); PHQ-2 ≥ 3 had 0.83 sensitivity and 0.70 specificity (AUC=0.82); and PHQ-9 ≥ 10 had 0.91 sensitivity and 0.81 specificity (AUC=0.96).

Supplemental Digital Content to accompany: [authors]. Reliability and validity of depression assessment among persons with HIV in sub-Saharan Africa: systematic review and meta-analysis. *J Acquir Immune Defic Syndr*

Chishinga et al. (99)	A consecutive sample of 649 adults (16+ years) attending one of 16 different TB/HIV clinics throughout Zambia were administered the CES-D. Two hundred and five participants were HIV-positive, 66 had not been tested, and 123 were HIV-negative. Mental health clinical assistants blinded to the screening results administered the MINI to establish the reference criterion of major depressive disorder.	CES-D	The CES-D had an internal consistency of 0.84. Factor analysis suggested that a four-factor model (depressed affect, somatic symptoms, positive affect, and interpersonal difficulties) did not fit the data well. CES-D ≥ 22 had 0.73 sensitivity and approximately 0.70 specificity (by visual inspection of the receiver-operating characteristic curve) for detecting major depressive disorder (AUC=0.78). Reliability and validity data were not disaggregated by HIV serostatus.
Epino et al. (105)	Six hundred and ten HIV-positive, treatment-naïve adults initiating HIV treatment at one of nine health clinics in rural Rwanda were administered the HSCL.	HSCL	The HSCL had an internal consistency of 0.87. Principal components analysis supported the hypothesis of a single factor that explained 37% of the variance. The HSCL was negatively correlated with the MOS-HIV mental health summary score ($\rho = -0.47$, $P < 0.001$).
Kagee et al. (96)	A convenience sample of 185 HIV-positive adults receiving treatment were recruited from two public health clinics in South Africa and administered the 21-item BDI-II.	BDI-II	The BDI-II had an internal consistency of 0.90. Exploratory factor analysis yielded a three-factor solution: cognitive, affective, and somatic.
Kaharuza et al. (100)	A consecutive sample of 1,017 HIV-positive, treatment-eligible/ treatment-naïve adults enrolled in a randomized trial of treatment monitoring in rural Uganda were administered the CES-D at baseline prior to eligibility determination.	CES-D	The CES-D had an overall internal consistency of 0.90. Factor analysis revealed four underlying factors: depressed affect (internal consistency, 0.81), positive affect (0.73), somatic symptoms (0.79), and interpersonal difficulties (0.87). The CES-D had statistically significant associations with sex, lack of income, and low CD4+ T-lymphocyte cell count (all $P < 0.01$).

Supplemental Digital Content to accompany: [authors]. Reliability and validity of depression assessment among persons with HIV in sub-Saharan Africa: systematic review and meta-analysis. *J Acquir Immune Defic Syndr*

Myer et al. (101); Spies et al. (106)	A convenience sample of 465 HIV-positive adults (18-65 years) receiving care at three primary level HIV facilities in urban South Africa who had a Mini-Mental State Exam score ≥ 24 were administered the CES-D and K10. (The study by Spies et al. further limited the sample to participants with an HIV Dementia Scale score ≥ 9 .) Research nurses blinded to the CES-D and K10 scores administered the MINI in either Afrikaans or Xhosa to establish the reference criterion diagnosis of major depressive disorder.	CES-D K10	The K10 had an internal consistency of 0.87. The CES-D ≥ 20 had 0.79 sensitivity and 0.61 specificity in detecting a current major depressive episode (AUC=0.76). Among Afrikaans-speaking participants (N=118), the CES-D had 0.86 sensitivity and 0.76 specificity, while among Xhosa-speaking participants (N=309) the CES-D had 0.73 sensitivity and 0.56 specificity. K10 ≥ 28 had 0.67 sensitivity and 0.77 specificity for detecting a current major depressive episode (AUC=0.77).
Nakimuli-Mpungu et al. (109)	Cultural adaptation of the SRQ was facilitated through in-depth interviews held with a convenience sample of 9 mental health professionals and 11 laypersons, and through a focus group of 10 HIV-positive outpatients. A psychiatric nurse administered the Luganda version of the 20-item SRQ to a consecutive sample of 200 HIV-positive persons attending an outpatient HIV clinic. The first 20 participants were re-interviewed on the same day, and the kappa statistic was calculated to assess test-retest reliability. Psychiatric clinical officers administered the MINI to establish the reference criterion diagnosis of current major depression.	SRQ	Five items were revised to increase comprehension and relevance. One item was replaced to increase consistency with the Baganda expression of depressed mood. Two items were revised in recognizance of cultural considerations that would affect participant responses. The SRQ had an internal consistency of 0.84. The kappa statistic was 0.48 for test-retest reliability and 0.63 for inter-rater reliability. Principal component analysis revealed two factors (depression, anxiety) that accounted for 87% of the total variance. SRQ ≥ 6 had 0.84 sensitivity and 0.93 specificity for detecting current major depression (AUC=0.95).

Supplemental Digital Content to accompany: [authors]. Reliability and validity of depression assessment among persons with HIV in sub-Saharan Africa: systematic review and meta-analysis. *J Acquir Immune Defic Syndr*

Pence et al. (108)	A consecutive sample of 400 HIV-positive adults (18-55 years) on HIV antiretroviral therapy and attending an outpatient HIV treatment center in northwest Cameroon were administered the PHQ-9 by research staff. A health care professional blinded to the PHQ-9 scores administered the CIDI to establish the reference criterion diagnosis of past-month major depressive episode.	PHQ	PHQ-9 ≥ 10 had 0.27 sensitivity and 0.94 specificity for detecting past-month major depressive episode.
Reda (103)	A convenience sample of 302 HIV-positive adults on HIV antiretroviral therapy in Ethiopia were administered the Amharic version of the 14-item HADS by two trained nurses. The intra-class correlation coefficient was used to estimate test-retest reliability in a convenience sample of 144 participants who were re-interviewed two weeks after the initial interview.	HADS	The HADS had an internal consistency of 0.78, and the depression subscale had an internal consistency of 0.87. The intra-class correlation coefficient was 0.84 for the HADS and 0.86 for the depression subscale. Factor analysis revealed a single factor explaining 38% of the variation.

Supplemental Digital Content to accompany: [authors]. Reliability and validity of depression assessment among persons with HIV in sub-Saharan Africa: systematic review and meta-analysis. *J Acquir Immune Defic Syndr*

Shacham (97); Monahan et al. (107); Shacham et al. (112)	A convenience sample of 347 HIV-positive adults (18-61 years) attending a hospital-based psychosocial support group in western Kenya self-administered the English versions of the 53-item BSI and PHQ-9, with clarifications provided in Swahili by trained research assistants if needed. Of these, 24 participants participated in 4 focus group interviews. Intra-class correlation coefficients were used to assess test-retest reliability in a convenience sample of 187 participants who were re-interviewed 1 week later. The intra-class correlation coefficient was used to estimate test-retest reliability using a subsample of every other participant interviewed one week after the initial interview. The PHQ-9 was used to determine the reference criterion diagnosis of major depression.	BSI PHQ	Focus group interviews revealed some difficulties in understanding the Likert-type response options for the BSI and PHQ-9; and specific items in the BSI and PHQ-9 contained unfamiliar terminology, were subject to cultural gaps in understanding, or were interpreted through the lens of HIV. Internal consistency was 0.95 for the BSI total score, 0.78 for the 6-item depression subscale, and 0.78 for the PHQ-9. The intra-class correlation within participants was 0.73 for the BSI total score, 0.58 for the depression subscale, and 0.59 for the PHQ-9. Factor analysis performed for the BSI revealed a single factor that accounted for 30% of the total variance; for the PHQ-9, factor analysis revealed a single factor that accounted for 37% of the total variance. The PHQ-9 was correlated with the BSI total score (Pearson's $r=0.82$, P-value not reported), the BSI depression subscale ($r=0.64$), and general health rating (greatest scores observed among those with minimal depression and the lowest scores observed among those with severe depression, $P<0.001$). An 11-item perceived HIV stigma scale was correlated with the PHQ-9 ($r=0.30$), the BSI total score ($r=0.33$), and the BSI depression subscale ($r=0.40$). PHQ-2 ≥ 3 had 0.91 sensitivity and 0.77 specificity for detecting major depression (AUC=0.91), and 0.85 sensitivity and 0.95 specificity for detecting any depressive disorder (AUC=0.97).
Singh et al. (102)	A sample of 20 HIV-positive, treatment-naïve inpatients at a hospital in South Africa were administered the CES-D in either English or isiZulu by a counselor. A psychiatrist blinded to the CES-D scores conducted clinical interviews with each patient to establish the reference criterion diagnosis of major depression.	CES-D	CES-D ≥ 16 had 0.91 sensitivity and 0.44 specificity for detecting major depression.

Supplemental Digital Content to accompany: [authors]. Reliability and validity of depression assessment among persons with HIV in sub-Saharan Africa: systematic review and meta-analysis. *J Acquir Immune Defic Syndr*

Wouters et al. (104); Pappin et al. (111)	A consecutive sample of 716 HIV-positive adults (18+ years) enrolled in a public sector HIV treatment program in the Free State, South Africa were administered the Sesotho version of the 14-item HADS.	HADS	Internal consistency of the depression subscale was 0.63. Confirmatory factor analysis showed that the original two-factor model (depression, anxiety) demonstrated good fit but had a marginally poorer fit compared to a three-factor model (depression, anxiety, and negative affect). Depression symptom severity was correlated with HIV stigma ($P<0.01$) and widowhood ($P<0.05$), and inversely correlated with support group membership ($P<0.05$).
---	--	------	--

AUC = area under the receiver-operating characteristic curve; BSI = Brief Symptom Inventory; CES-D = Center for Epidemiologic Studies Depression Scale; CIDI = Composite International Diagnostic Interview; HADS = Hospital Anxiety and Depression Scale; HSCL = Hopkins Symptom Checklist; K6/K10 = Kessler non-specific psychological distress scale; MINI = Mini International Neuropsychiatric Interview; MOS-HIV = Medical Outcomes Study-HIV Health Survey; PHQ = Patient Health Questionnaire; SCID = Structured Clinical Interview for the DSM-IV; SF-36 = Medical Outcomes Study Short Form 36-Item Health Survey; SRQ = Self-Reporting Questionnaire

Table S3. Quality appraisal for studies on the criterion-related validity of depression assessment among persons with HIV in sub-Saharan Africa

Reference	Risk of bias					Applicability concerns			
	Patient selection	Index test	Criterion standard	Flow and timing	Overall risk	Patient selection	Index test	Criterion standard	Overall concerns
Akena et al. (78)	Unclear	At risk	Low	Low	At risk	Unclear	Low	Low	At risk
Akena et al. (65)	Low	At risk	Low	Low	At risk	Low	Low	Low	Low
Chishinga et al. (66)	Low	At risk	Low	Low	At risk	Low	Low	Low	Low
Shacham (64)	At risk	Low	At risk	Low	At risk	At risk	Low	Low	At risk
Myer et al. (68)	Low	Low	Low	Low	At risk	Low	Low	Low	Low
Nakimuli-Mpungu et al. (77)	Low	At risk	Low	Low	At risk	Low	Low	Low	Low
Pence et al. (76)	Low	Low	Low	Low	Low	Low	Low	Low	Low
Singh et al. (69)	Unclear	Low	Low	Unclear	At risk	Low	Low	Low	Low
Spies et al. (74)	Low	At risk	Low	Low	At risk	Low	Low	Low	Low